

Patricia Esteban Sánchez

Biomarcadores metabólicos e
inmunológicos en enfermedades
infecciosas respiratorias:
Integración de ciencia de datos e
inteligencia artificial

Director/es

Arias Cabrero, Maykel Alberto
Galvez Buerba, Eva María

<http://zaguan.unizar.es/collection/Tesis>



Universidad de Zaragoza
Servicio de Publicaciones

ISSN 2254-7606



Tesis Doctoral

**BIOMARCADORES METABÓLICOS E
INMUNOLÓGICOS EN ENFERMEDADES
INFECCIOSAS RESPIRATORIAS: INTEGRACIÓN
DE CIENCIA DE DATOS E INTELIGENCIA
ARTIFICIAL**

Autor

Patricia Esteban Sánchez

Director/es

Arias Cabrero, Maykel Alberto
Galvez Buerba, Eva María

UNIVERSIDAD DE ZARAGOZA
Escuela de Doctorado

Programa de Doctorado en Bioquímica y Biología Molecular

2025

Tesis Doctoral

Biomarcadores metabólicos e inmunológicos en
enfermedades infecciosas respiratorias:
Integración de ciencia de datos e inteligencia
artificial

Autora

Patricia Esteban Sánchez

Director/es

Eva María Gálvez Buerba

Maykel Alberto Arias Cabrero

Facultad de ciencias

Programa de Doctorado en Bioquímica y Biología Molecular

2024



Universidad
Zaragoza



Instituto de Investigación
Sanitaria Aragón

ciber | **INFEC**



CSIC

CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

D.^a **Eva María Gálvez Buerba**, Doctora en Ciencias y Científica Titular en el Instituto de Carboquímica del Consejo Superior de Investigaciones Científicas (CSIC) y D. **Maykel Arias Cabrero**, Doctor en Ciencias, e Investigador Post Doctoral en el IIS Aragón.

CERTIFICAN:

Que la Tesis Doctoral titulada: “**Biomarcadores metabólicos e inmunológicos en enfermedades infecciosas respiratorias: Integración de ciencia de datos e inteligencia artificial**” ha sido realizada por la graduada en Farmacia y Bioinformática **Patricia Esteban Sánchez** en el Instituto de Carboquímica del CSIC y en el Instituto de investigación Sanitaria de Aragón (IISA) y que reúne, bajo nuestro juicio, las condiciones requeridas para que su autor opte al título de Doctor en Ciencias de la Universidad de Zaragoza.

Zaragoza, 19 de diciembre de 2024

“El conocimiento es el único tesoro que crece al compartirlo.”

Anónimo

Acc	Exactitud (del inglés, <i>Accuracy</i>)
AI	Aspergilosis invasiva
AIC	Criterio de información de Akaike (del inglés, <i>Akaike information criterion</i>)
AIF	Factor inductor de apoptosis
ANN	Redes neuronales artificiales (del inglés, <i>Artificial Neural Networks</i>)
ANOVA	Análisis de varianza (del inglés, <i>Analysis Of Variance</i>)
AUC	Área bajo la curva (del inglés, <i>Area Under the Curve</i>)
BDG	1,3-β D-glucano
BHT	Butilhidroxitolueno
BIC	Criterio de información bayesiano (del inglés, <i>Bayesian information criterion</i>)
BLAST	<i>Basic Local Alignment Search Tool</i>
bmGT	Bismetilgliotoxina
CART	Árboles de clasificación y regresión (del inglés, <i>Classification and Regression Trees</i>)
CBS	Centro de Biodiversidad Fúngica
CCL	Ligando 5 de quimioquina con motivo C-C
CEICA	Comité de Ética de la Investigación de la Comunidad Autónoma de Aragón
CIE	Clasificación Internacional de Enfermedades
CLSI	<i>Clinical and Laboratory Standards Institute</i>
CLX	Cloxacilina
CMI	Concentración Mínima Inhibitoria
CMV	Citomegalovirus
Corr	Correlación
cos2	Cuadrado del coseno
COV	Compuesto Orgánico Volátil
COVID-19	Enfermedad por coronavirus 2019 (del inglés, <i>COronaVirus Disease 2019</i>)
CPM	Concentración de Prevención de Mutaciones
CRISPR	Repeticiones palindrómicas cortas agrupadas regularmente espaciadas (del inglés, <i>Clustered Regularly Interspaced Short Palindromic Repeats</i>)
CRP	Proteína C reactiva (del inglés, <i>C-reactive protein</i>)
CTL	Linfocitos T citotóxicos
CXCL	Ligando 10 de quimioquina con motivo C-X-C
cyt c	Citocromo c
DALY	Años de vida ajustados por discapacidad (del inglés, <i>Disability Adjusted Life Years</i>)
DCM	Diclorometano
DO	Densidad óptica
ECMO	Oxigenación por membrana extracorpórea (del inglés, <i>Extracorporeal Membrane Oxygenation</i>)
ELISA	Ensayos por inmunoadsorción ligado a enzimas (del inglés, <i>Enzyme-Linked ImmunoSorbent Assay</i>)
EORTC/MSK	Grupo Colaborativo de Infecciones Fúngicas Invasoras de la Organización Europea para la Investigación y Tratamiento del Cáncer en conjunto con el Grupo de Estudio de Micosis del Instituto Nacional de Alergias y Enfermedades Infecciosas
EPI	Equipos de protección individual
EPOC	Enfermedad Pulmonar Obstructiva Crónica
E-sel	E-selectina
ETP	Epipolítiodioxopiperazina

Abreviaturas

EUCAST	<i>The European Committee on Antimicrobial Susceptibility Testing</i>
FN	Falso negativo
FOX	Cefoxitina
FP	Falso positivo
FUS	Ácido fusídico
GC-MS	Cromatografía de gases - espectrometría de masas (del inglés, <i>gas chromatography-mass spectrometry</i>)
GM	Galactomanano
GT	Gliotoxina
Gzm	Granzima
HEPES	Ácido 4-(2-hidroxietil)-1-piperazinaetanosulfónico
HPLC	Cromatografía líquida de alta eficacia (del inglés, <i>High Performance Liquid Chromatography</i>)
HPTLC	Cromatografía den capa fina de alto rendimiento (del inglés, <i>high-performance thin layer chromatography</i>)
IC	Índice de Combinación
IC 95 %	Intervalo de confianza del 95 %
ICAM-1	Molécula de adhesión intercelular 1 (del inglés, <i>Intercelular Adhesion Molecule 1</i>)
IDSA	Sociedad de enfermedades infecciosas de América (del inglés, <i>Infectious Diseases Society of America</i>)
IFI	Infección fúngica invasiva
IFN	Interferón
IHME	<i>Institute for Health Metrics and Evaluation</i>
IL	Interleuquina
IMC	Índice de Masa Corporal
IQR	Rango intercuartílico (del inglés, <i>interquartile range</i>)
ITS	Espaciador transcrito interno (ITS, del inglés, <i>Internal transcribed spacer</i>)
JTT	Jones-Taylor-Thornton
k	número de vecinos más cercanos
KNN	<i>k-Nearest Neighbors</i>
LB	Luria-Bertani
LBA	Lavado broncoalveolar
LIME	<i>Local Interpretable Model-agnostic Explanations</i>
L-Phe	L-Fenilalanina
LRI	Infecciones de las vías respiratorias inferiores (del inglés, <i>lower respiratory infections</i>)
L-Ser	L-Serina
LZD	Linezolid
MDG	Coeficiente de Gini de disminución media (del inglés, <i>Mean Decrease Gini</i>)
MET	Metanol
MH	Mueller-Hinton
MIBK	Metil isobutil cetona (del inglés, <i>methyl isobutyl ketone</i>)
ML	<i>Machine Learning</i>
NIH	Institutos Nacionales de Salud (del inglés, <i>National Institutes of Health</i>)
NIR	<i>Non Information Rate</i>
NK	Células sesinas naturales (del inglés, <i>Natural Killer</i>)
no-COVID-19	Otras infecciones respiratorias no COVID-19

NRPS	péptido sintasa no ribosomal
OMS	Organización Mundial de la Salud
PaO ₂ /FiO ₂	Relación de la presión parcial arterial de oxígeno a la fracción de oxígeno inspirado
PBS	Tampón fosfato salino
PCA	Análisis de componentes principales (del inglés, <i>Principal Component Analysis</i>)
PCR	Reacción en cadena de la polimerasa (del inglés, <i>Polymerase Chain Reaction</i>)
PCRASP	PCR para detectar <i>Aspergillus</i>
PKS	Policétido sintasa
POC	<i>Point-of-care</i>
RANTES	Quimiocina de regulación por activación expresada y secretada por los linfocitos T (del inglés, <i>regulated on activation, normal T-cell expressed and secreted</i>)
RBF	Función de base radial (del inglés, <i>radial basis function</i>)
RF	<i>Random Forest</i>
RL	Regresión logística
ROC	Característica Operativa del Receptor (del inglés, <i>Receiver Operating Characteristic</i>)
ROS	Especies Reactivas de Oxígeno (del inglés, <i>reactive oxygen species</i>)
rpm	Revoluciones por minuto
RPMI	<i>Roswell Park Memorial Institute</i>
RT-PCR	Reacción en cadena de la polimerasa con transcripción inversa
RT-qPCR	Reacción en cadena de la polimerasa cuantitativa con transcripción inversa
SAMS	<i>Staphylococcus aureus</i> sensible a meticilina
SARIV	<i>Staphylococcus aureus</i> con resistencia intermedia a vancomicina
SARM	<i>Staphylococcus aureus</i> resistente a meticilina
SARM-AC	<i>Staphylococcus aureus</i> resistente a meticilina adquiridas en la comunidad
SARS-CoV-2	Coronavirus del síndrome respiratorio agudo grave de tipo 2
SARV	<i>Staphylococcus aureus</i> resistente a vancomicina
SD	Desviación estándar
SDRA	Síndrome de Dificultad Respiratoria Aguda
SHAP	<i>Shapley Additive exPlanations</i>
SNS	Sistema Nacional de Salud
SPME	Microextracción en fase sólida (del inglés, <i>Solid Phase Microextraction</i>)
SpO ₂	Porcentaje de saturación de oxígeno de la hemoglobina en sangre arterial
STARD	Estándares para el informe de precisión diagnóstica (del inglés, <i>STAndards for the Reporting of Diagnostic accuracy</i>)
SVM	<i>Support Vector Machines</i>
TC	Tomografía computarizada
TF	Factor tisular (del inglés, <i>tissue factor</i>)
Th	Linfocitos T efectores
TM	Trombomodulina
TNF	Factor de Necrosis Tumoral (del inglés, <i>tumor necrosis factor</i>)
TSP	Trombospondina
UCI	Unidad de Cuidados Intensivos
UFC	Unidades formadoras de colonias
UV/Vis	Ultravioleta/Visible
VAN	Vancomicina

Abreviaturas

VCAM-1	Molécula de adhesión vascular 1 (del inglés, <i>vascular cell adhesion molecule-1</i>)
VIH	Virus de la Inmunodeficiencia Humana
VN	Verdadero negativo
VP	Verdadero positivo
VPN	Valor predictivo negativo
VPP	Valor predictivo positivo
vWF-A2	Factor de von Willebrand

Las enfermedades infecciosas respiratorias representan uno de los principales retos a los que se enfrentan los sistemas de salud a nivel mundial. La dificultad del manejo de graves enfermedades como la Aspergilosis Invasiva (AI) o el COVID-19 ponen de manifiesto la urgente necesidad del desarrollo de nuevas herramientas de diagnóstico y pronóstico. En este contexto, las nuevas técnicas de inteligencia artificial podrían aportar grandes beneficios.

La AI es una enfermedad grave, que afecta principalmente a pacientes inmunocomprometidos y supone una de las causas de muerte más comunes en este tipo de pacientes. Es producida por especies del género *Aspergillus*, mayoritariamente las pertenecientes a la sección *Fumigati*. Estas especies son capaces de generar multitud de metabolitos secundarios, entre las que destaca la gliotoxina (GT).

La GT posee un amplio espectro de actividades biológicas. Se decidió estudiar su efecto como antimicrobiano contra cepas de *Staphylococcus aureus* de interés clínico por su resistencia a múltiples fármacos. Nuestros resultados proporcionan observaciones novedosas y amplían los análisis anteriores al demostrar que la GT presenta actividad contra *S. aureus* resistente a meticilina y con sensibilidad reducida a vancomicina tanto *in vitro* como en un modelo *in vivo* desarrollado con *Caenorhabditis elegans*. Además, la GT muestra un efecto sinérgico con la vancomicina, lo que presenta una perspectiva prometedora.

Existe una gran controversia en relación a qué especies fúngicas son capaces de sintetizar GT. Mediante un estudio *in silico* analizamos la homología de secuencias entre varias proteínas de especies fúngicas similares a la proteína GliP de *A. fumigatus* (esencial en la biosíntesis de GT). El estudio *in vitro* corroboró la producción de GT y su derivado inactivo, la bismetilgliotoxina (bmGT) por especies de la sección *Fumigati* y por *Trichoderma virens*.

En cuanto a la bmGT, ha demostrado ser un biomarcador estable de AI, por lo que se estudió su utilidad como predictor de diagnóstico en una cohorte de pacientes hematológicos. La bmGT y la interleuquina 8 (IL-8) demuestran ser potenciales biomarcadores de AI probable cuando se combinan con los utilizados actualmente en clínica.

Por otro lado, la pandemia de COVID-19 ha acentuado la importancia de las infecciones respiratorias como una amenaza global. Combinamos la información de la composición de compuestos orgánicos volátiles expirados e inmunomoduladores séricos para estudiar el diagnóstico y pronóstico de pacientes con COVID-19 y otras infecciones respiratorias.

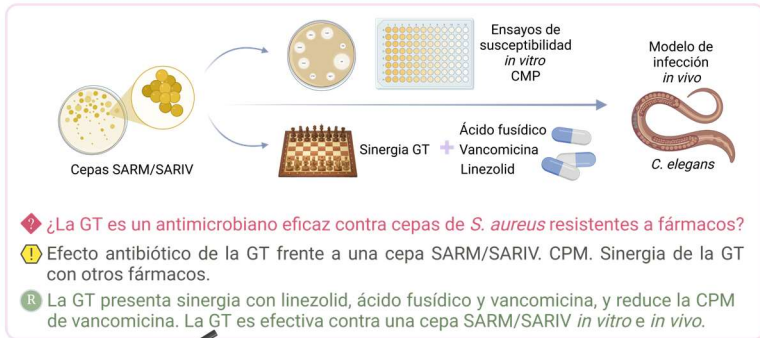
Resumen

Se muestran perfiles significativamente distintos de estos parámetros. Además, se han desarrollado modelos de predicción basados en *machine learning* que muestran la importancia de IL-6, IL-8, tolueno y trombomodulina en la predicción de gravedad de los pacientes COVID-19; y CXCL10 y metil isobutil cetona en otras infecciones respiratorias.

Nuestros resultados muestran que la integración del análisis de factores producidos por el patógeno y el hospedador durante una infección respiratoria puede mejorar el diagnóstico y el pronóstico de este tipo de enfermedades.

Resumen gráfico de la tesis doctoral

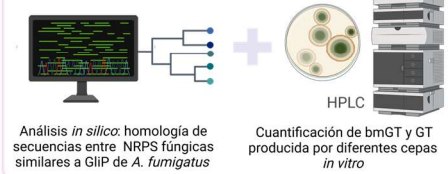
- ◆ Pregunta de investigación
- ⚠ ¿Qué nuevo conocimiento aportamos?
- Ⓡ Resultados obtenidos



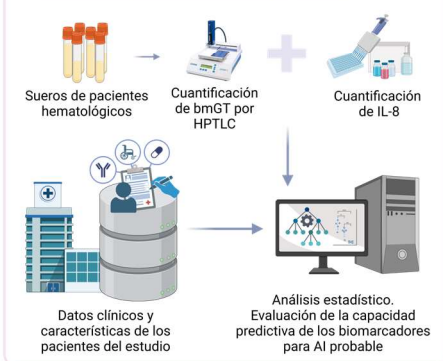
- ◆ ¿La GT es un antimicrobiano eficaz contra cepas de *S. aureus* resistentes a fármacos?
- ⚠ Efecto antibiótico de la GT frente a una cepa SARM/SARIV. CPM. Sinergia de la GT con otros fármacos.
- Ⓡ La GT presenta sinergia con linezolid, ácido fusídico y vancomicina, y reduce la CPM de vancomicina. La GT es efectiva contra una cepa SARM/SARIV *in vitro* e *in vivo*.



- ◆ ¿Qué cepas fúngicas producen GT y bmGT?
- ⚠ Nuevos datos sobre la producción de GT por diferentes cepas fúngicas.
- Ⓡ *A. fumigatus* B5233, *A. fumigatus* 1631562, *A. fischeri* CBS 420.96, *A. turcosus* CBS 140371, *A. pseudofischeri* CBS 404.67 y *T. virens* CBS 249.59 son capaces de producir GT y bmGT *in vitro*.



- ◆ La bmGT, ¿es un buen biomarcador de AI?
- ⚠ Estudio de la bmGT e IL-8 como biomarcadores individualmente y en conjunto con otros utilizados en clínica (GM, BDG) o en fase de estudio (CMV).
- Ⓡ bmGT e IL-8 demuestran ser potenciales biomarcadores de AI probable cuando se combinan con los utilizados en clínica.



- ◆ Perfil de COVs expirados en aire + inmunomarcadores en suero, ¿es útil para el diagnóstico y pronóstico de COVID-19 y no-COVID-19?
- ⚠ Diferencias en inmunomarcadores entre COVID-19 y no-COVID-19 y diferencias en COVs respecto a la gravedad. Perfil conjunto de inmunomarcadores y COVs y su valor pronóstico mediante el uso de ML.
- Ⓡ Perfiles significativamente distintos de los parámetros estudiados. Modelos de predicción basados en ML muestran la importancia de IL-6, IL-8, tolueno y TM en la predicción de gravedad en COVID-19; y CXCL10 y MIBK en no-COVID-19.

SARM: *Staphylococcus aureus* resistente a meticilina; SARIV: *Staphylococcus aureus* con resistencia intermedia a vancomicina; CPM: Concentración de Prevención de Mutaciones; NRPS: Péptido sintasa no ribosomal; HPLC: Cromatografía líquida de alta eficacia; GM: Galactomanano; BDG: 1,3-β D-glucano; CMV: Citomegalovirus; HPTLC: Cromatografía en capa fina de alto rendimiento; IL: interleuquina; AI: Aspergilosis invasiva; COVs: Compuestos Orgánicos Volátiles; ML: machine learning; CXCL10: ligando 10 de quimiocina con motivo C-X-C; no-COVID-19: pacientes con otras infecciones respiratorias que no son COVID-19; SHAP: Shapley Additive exPlanations; LIME: Local Interpretable Model-agnostic Explanations; RF: Random Forest; TM: trombomodulina; MIBK: metil isobutil cetona.

Abstract

Respiratory infectious diseases represent a major challenge for health systems worldwide. The complexity of managing severe conditions such as invasive aspergillosis (IA) or COVID-19 highlights the urgent need for the development of new diagnostic and prognostic tools. In this context, emerging artificial intelligence techniques have the potential to offer significant benefits.

IA is a severe disease primarily affecting immunocompromised patients and is one of the leading causes of mortality in this population. It is caused by species of the *Aspergillus* genus, predominantly those within the *Fumigati* section. These species produce a variety of secondary metabolites, with gliotoxin (GT) being particularly notable.

GT possesses a broad spectrum of biological activities. In this study, we investigated its antimicrobial effects against *Staphylococcus aureus* strains, which are clinically significant due to their multidrug resistance. Our findings offer novel insights and expand on previous studies by demonstrating that GT exhibits potent activity against methicillin-resistant *S. aureus* strains with reduced sensitivity to vancomycin. This antimicrobial effect was confirmed both *in vitro* and in an *in vivo* model using *Caenorhabditis elegans*. Furthermore, GT displayed a synergistic effect with vancomycin, presenting a promising perspective for future therapeutic applications.

The ability of different fungal species to synthesize gliotoxin (GT) remains a subject of considerable controversy. To address this, we conducted an *in silico* analysis to assess sequence homology between proteins from various fungal species and the GliP protein of *Aspergillus fumigatus*, which is essential for GT biosynthesis. Our *in vitro* studies further confirmed the production of GT and its inactive derivative, bismethylgliotoxin (bmGT), by species within the *Fumigati* section as well as by *Trichoderma virens*.

BmGT has been shown to serve as a stable biomarker for IA. In our study, we evaluated its diagnostic potential in a cohort of haematological patients. The results indicate that bmGT, when combined with IL-8 and existing clinical biomarkers, holds promise as a reliable predictor of probable IA, enhancing diagnostic precision in clinical practice.

On the other hand, the COVID-19 pandemic highlighted the importance of respiratory infections as a global threat. We combined composition data from exhaled volatile organic compounds and serum immunomodulators to study the diagnosis and prognosis of patients with COVID-19 and other respiratory infections. We identified significantly different profiles of these parameters. Additionally, we developed machine learning-based models that highlight the importance of IL-6, IL-8, toluene and thrombomodulin in

predicting the severity of COVID-19; and CXCL10 and methyl isobutyl ketone in other respiratory infections.

Our results demonstrate that integrating an analysis of factors produced by both the pathogen and the host during a respiratory infection can enhance the diagnosis and prognosis of these diseases.

Índice

1. INTRODUCCIÓN	1
1.1. Enfermedades infecciosas respiratorias	1
1.1.1. Descripción, epidemiología y agentes etiológicos.....	1
1.1.2. Diagnóstico de las enfermedades infecciosas respiratorias	4
1.2. Aspergilosis invasiva.....	7
1.2.1. Descripción y epidemiología de la aspergilosis invasiva	7
1.2.2. Diagnóstico de la aspergilosis invasiva	10
1.2.3. Agente etiológico de la aspergilosis invasiva: <i>Aspergillus</i> spp.	13
1.2.3.1. Taxonomía y distribución	14
1.2.3.2. Metabolismo secundario	16
1.2.3.2.1. Gliotoxina	18
1.2.3.2.1.1. Biosíntesis de la gliotoxina	18
1.2.3.2.1.2. Relevancia y aplicaciones de la gliotoxina.....	21
1.2.3.2.2. Bismetilgliotoxina.....	23
1.3. COVID-19	25
1.3.1. Descripción y epidemiología del COVID-19	25
1.3.2. Respuesta fisiológica a SARS-CoV-2.....	27
1.3.2.1. Inflamación	28
1.3.2.1.1. Inmunomoduladores solubles	29
1.3.2.2. Cambios en el metabolismo: Compuestos orgánicos volátiles expirados en aire	33
1.3.3. Diagnóstico de COVID-19.....	36
1.3.4. Pronóstico de COVID-19	40
1.4. <i>Machine learning</i> en diagnóstico y pronóstico clínico	43
1.4.1. Preparación de los datos	43
1.4.2. Algoritmos de <i>machine learning</i> para crear modelos de clasificación	45
1.4.2.1. Regresión logística	47
1.4.2.2. Árboles de decisión.....	48
1.4.2.3. <i>Support Vector Machines</i>	50
1.4.2.4. <i>K-Nearest Neighbors</i>	51
1.4.2.5. Redes neuronales artificiales	52
1.4.3. Evaluación de los modelos	53
1.4.4. Procedimientos para la mejora de los modelos de clasificación.....	56
1.4.4.1. Validación cruzada con ajuste de parámetros automatizado	56
1.4.4.2. <i>Ensemble learning</i>	57

1.4.4.2.1	Bagging	58
1.4.4.2.1.1	Random Forest	59
1.4.4.2.2	Boosting	62
1.4.5	Explicabilidad de los modelos de caja negra	63
1.4.5.1	Shapley Additive exPlanations	64
1.4.5.2	Local Interpretable Model-agnostic Explanations	65
2	CONTRIBUCIONES EXPERIMENTALES	71
2.1	CAPÍTULO 1 – Estudio de la actividad antibacteriana de la gliotoxina frente a <i>Staphylococcus aureus</i>	71
2.1.1	ANTECEDENTES	71
2.1.2	MATERIALES Y MÉTODOS	74
2.1.2.1	Cepas de <i>Staphylococcus aureus</i>	74
2.1.2.2	Ensayos de susceptibilidad antimicrobiana	74
2.1.2.3	Estudio de sinergia antimicrobiana	75
2.1.2.4	Determinación de la concentración de prevención de mutaciones 77	
2.1.2.5	Cultivo y sincronización de <i>Caenorhabditis elegans</i>	78
2.1.2.6	Cultivo e infección de larvas L4 de <i>Caenorhabditis elegans</i>	80
2.1.2.7	Ensayo de supervivencia de <i>Caenorhabditis elegans</i>	80
2.1.3	RESULTADOS	82
2.1.3.1	Susceptibilidad antimicrobiana de <i>Staphylococcus aureus</i> frente a gliotoxina	82
2.1.3.1.1	Efecto de la gliotoxina en el crecimiento de <i>Staphylococcus aureus</i>	82
2.1.3.1.2	Concentración mínima inhibitoria de gliotoxina y antibióticos frente a <i>Staphylococcus aureus</i>	83
2.1.3.2	Sinergia antimicrobiana de gliotoxina en combinación con fármacos antiestafilocócicos	83
2.1.3.3	Concentración de prevención de mutaciones de gliotoxina y vancomicina solas o en combinación frente a <i>Staphylococcus aureus</i>	86
2.1.3.4	Eficacia antibiótica <i>in vivo</i> de gliotoxina como monoterapia y en combinación con vancomicina frente a <i>Staphylococcus aureus</i> en un modelo de <i>Caenorhabditis elegans</i>	86
2.1.4	DISCUSIÓN	90
2.2	CAPÍTULO 2 – Estudio de la producción de gliotoxina y bismetilgliotoxina por diferentes especies fúngicas y su utilidad como biomarcador en aspergilosis invasiva	95
2.2.1	ANTECEDENTES	95
2.2.2	MATERIALES Y MÉTODOS	98

2.2.2.1	Análisis de la producción de gliotoxina y bismetilgliotoxina en diferentes especies fúngicas.....	98
2.2.2.1.1	Estudio bioinformático	98
2.2.2.1.1.1	Búsqueda de proteínas homólogas de Af293 GliP de <i>Aspergillus fumigatus</i> en otras especies.....	98
2.2.2.1.1.2	Análisis filogenético.....	98
2.2.2.1.2	Análisis y cuantificación de la producción de gliotoxina y bismetilgliotoxina por diferentes cepas fúngicas <i>in vitro</i>	99
2.2.2.1.2.1	Cepas fúngicas y condiciones de cultivo	99
2.2.2.1.2.2	Extracción química de gliotoxina y bismetilgliotoxina	101
2.2.2.1.2.3	Análisis y cuantificación mediante HPLC	101
2.2.2.2	Estudio de la combinación de bismetilgliotoxina con otros biomarcadores para el diagnóstico de aspergilosis invasiva en pacientes hematológicos.....	101
2.2.2.2.1	Pacientes y diseño del estudio	101
2.2.2.2.2	Detección y cuantificación de biomarcadores en suero de pacientes	102
2.2.2.2.3	Análisis de datos mediante técnicas estadísticas y <i>machine learning</i>	104
2.2.3	RESULTADOS	107
2.2.3.1	Producción de gliotoxina y bismetilgliotoxina en diferentes cepas fúngicas	107
2.2.3.1.1	Estudio bioinformático	107
2.2.3.1.1.1	Búsqueda de proteínas homólogas de GliP de <i>Aspergillus fumigatus</i> Af293 en otras especies.....	107
2.2.3.1.1.2	Árbol filogenético.....	110
2.2.3.1.2	Análisis y cuantificación de la producción de gliotoxina y bismetilgliotoxina por varias cepas fúngicas.....	111
2.2.3.2	Utilidad de la combinación de bismetilgliotoxina con otros biomarcadores para el diagnóstico de aspergilosis invasiva en pacientes hematológicos.....	112
2.2.4	DISCUSIÓN	126
2.3	CAPÍTULO 3 – Estudio de biomarcadores para el pronóstico y diagnóstico de COVID-19 y otras infecciones respiratorias.....	135
2.3.1	ANTECEDENTES	135
2.3.2	MATERIALES Y MÉTODOS	138
2.3.2.1	Participantes y diseño del estudio	138
2.3.2.2	Cuantificación de biomarcadores.....	138
2.3.2.3	Análisis estadístico y bioinformático	140

2.3.3	RESULTADOS	142
2.3.3.1	Diferencias entre sujetos con COVID-19, otras infecciones respiratorias no-COVID-19 y sanos	142
2.3.3.1.1	Comparación entre los niveles de compuestos orgánicos volátiles en aire expirado	142
2.3.3.2	Diferencias entre pacientes COVID-19 y no-COVID-19	145
2.3.3.2.1	Comparación entre los niveles de biomarcadores relacionados con la respuesta inmune	145
2.3.3.2.2	Análisis del perfil de correlación entre los parámetros en suero y los compuestos orgánicos volátiles en aire expirado	147
2.3.3.3	Compuestos orgánicos volátiles y parámetros en suero como biomarcadores predictivos de gravedad de la enfermedad	149
2.3.3.3.1	Potencial de los compuestos orgánicos volátiles y parámetros en suero para discriminar la gravedad de los pacientes	149
2.3.3.3.2	Desarrollo de modelos de predicción de gravedad para pacientes con COVID-19 y no-COVID-19	152
2.3.3.3.3	Explicación de los modelos <i>Random Forest</i> para la predicción de gravedad en pacientes con COVID-19 y no-COVID-19	158
2.3.4	DISCUSIÓN	168
3	CONCLUSIONES	175
4	CONCLUSIONS	177
5	ANEXOS	181
4.1	Anexo 1	181
4.2	Anexo 2	185
6	BIBLIOGRAFÍA	189

Introducción

1. INTRODUCCIÓN

1.1. Enfermedades infecciosas respiratorias

1.1.1. Descripción, epidemiología y agentes etiológicos

Las enfermedades infecciosas respiratorias representan uno de los desafíos más significativos para la medicina moderna y los sistemas de salud actuales (1). Estas patologías comprenden un amplio espectro de infecciones que afectan al tracto respiratorio. Su alta contagiosidad, junto con factores como la globalización, el cambio climático y la resistencia a los antimicrobianos, ha intensificado su impacto y dificultado su control y prevención.

Según la Clasificación Internacional de Enfermedades (CIE-11) (2), las infecciones del tracto respiratorio se agrupan en base a su sintomatología y la región anatómica afectada. Las infecciones agudas de las vías respiratorias superiores incluyen el resfriado común, la faringitis, la epiglotitis y la laringotraqueítis. En la mayoría de los casos, estas infecciones suelen ser transitorias, autolimitadas y bien toleradas. Sin embargo, las infecciones que también afectan las vías respiratorias inferiores (LRI, por sus siglas en inglés, *lower respiratory infections*) pueden ser graves y potencialmente mortales, como la bronquitis, bronquiolitis o neumonía (3). Estas enfermedades pueden ser nosocomiales o adquiridas en la comunidad. En términos de etiología microbiana, el patógeno causante suele ser viral, bacteriano, fúngico o parasitario (Figura 1).

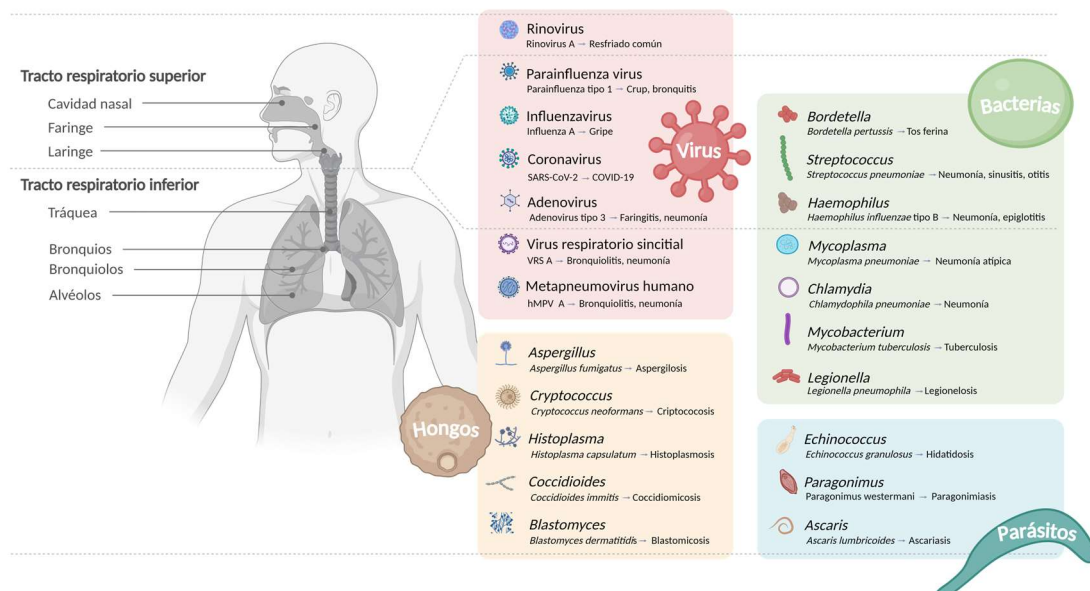


Figura 1. Etiología de las infecciones respiratorias. Principales tipos de microorganismos (virus, bacterias, hongos y parásitos) que pueden causar enfermedades infecciosas respiratorias. Se incluyen ejemplos representativos de cada categoría, así como las patologías específicas que causan (4, 5).

Introducción

La neumonía bacteriana representa un tercio de los casos de insuficiencia respiratoria aguda hipoxémica en pacientes inmunocomprometidos. Se recomienda la terapia temprana con antibióticos, pero disminuye las posibilidades de identificar los organismos causantes a aproximadamente el 50%. Los virus son la segunda causa más común de infecciones respiratorias graves en este tipo de pacientes. Cabe destacar que las pruebas positivas para virus en muestras respiratorias no necesariamente indican que el virus juegue un papel en la enfermedad aguda actual. Las infecciones fúngicas invasivas (causadas por *Aspergillus*, *Cryptococcus*, Mucorales y *Pneumocystis jirovecii*) representan alrededor del 15% de las infecciones respiratorias graves. Por otro lado, los parásitos pueden causar infecciones agudas graves, pero no es especialmente común en pacientes inmunocomprometidos (6).

Las infecciones respiratorias, especialmente las LRI como la neumonía, son una causa significativa de morbilidad y mortalidad a nivel mundial, especialmente en pacientes inmunocomprometidos (7). Gracias a los significativos avances en medicina, el número de muertes por infecciones respiratorias había disminuido notablemente a lo largo de los años hasta 2019 (Figura 2).

Tasa de mortalidad por infecciones respiratorias

Tasa de mortalidad estimada por infecciones respiratorias, incluidas la tuberculosis y COVID-19, por cada 100.000 personas.

Our World
in Data

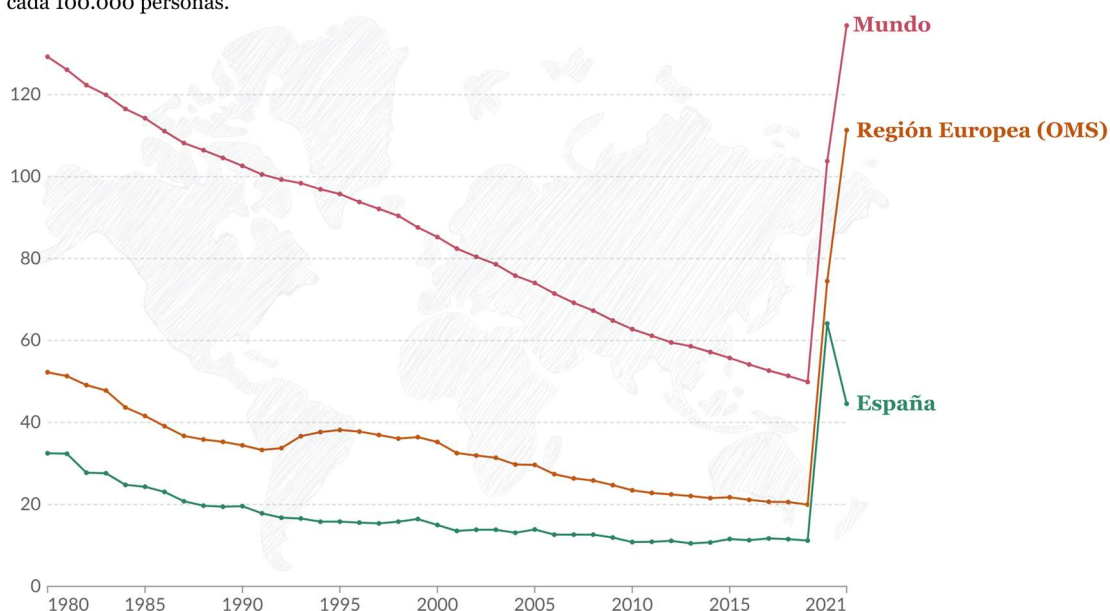


Figura 2. Tasa de mortalidad estimada por infecciones respiratorias incluyendo la tuberculosis y el COVID-19, en todo el mundo, Europa y España por cada 100.000 personas. Para permitir las comparaciones entre países y a lo largo del tiempo, esta métrica está estandarizada por edad. Fuente: Institute for Health Metrics and Evaluation (IHME), Global Burden of Disease (2024). Imagen adaptada de: OurWorldInData.org/causes-of-death | CC BY (8).

No obstante, el número absoluto de defunciones seguía siendo muy elevado, lo que conlleva un enorme impacto socioeconómico. Además, aunque se han realizado avances significativos en la comprensión de estas infecciones en las últimas décadas, el diagnóstico, pronóstico y tratamiento de las infecciones respiratorias continúa enfrentando grandes desafíos (7).

En 2019, la Organización Mundial de la Salud (OMS) clasificó a las LRI como las enfermedades infecciosas más mortales a nivel mundial, ocupando el cuarto lugar entre las diez principales causas de muerte (9). También fueron la cuarta causa a nivel mundial del aumento del número de años de vida ajustados por discapacidad (DALY, por sus siglas en inglés, *Disability Adjusted Life Years*), término que describe el número de años de vida saludable perdidos a causa de la enfermedad (10). Sin incluir la enfermedad por coronavirus 2019 (COVID-19, del inglés, *COronaVirus Disease 2019*), en 2021 se estimaron 344 millones de episodios de LRI y 2,18 millones de muertes a nivel global, lo que equivale a 27,7 muertes por cada 100.000 personas. Estas cifras incluyen una alta mortalidad en niños menores de 5 años, con 502.000 muertes en este grupo de edad (11).

El alcance y el impacto de la gravedad de las infecciones respiratorias se ha demostrado de manera dramática y catastrófica con la actual pandemia del COVID-19, enfermedad causada por el coronavirus del síndrome respiratorio agudo grave de tipo 2 (SARS-CoV-2) (12). El nuevo SARS-CoV-2 se informó por primera vez en diciembre de 2019 en la ciudad China de Wuhan, en la provincia de Hubei, y se extendió por multitud de países a finales de enero de 2020. La OMS declaró el brote como Emergencia de Salud Pública Mundial y la enfermedad como pandemia (13). A pesar del uso de diferentes precauciones, hubo un aumento diario de infecciones que se extendieron simultáneamente a más de 228 países en todo el mundo (14).

Por otro lado, la pandemia del COVID-19, el aumento de la resistencia a los fármacos antimicóticos y un mayor número de pacientes inmunocomprometidos han impulsado un reciente aumento mundial de infecciones fúngicas patógenas, incluidas las aspergilosis, candidiasis y mucormicosis. El 25 de octubre de 2022, la OMS publicó una lista de 19 hongos patógenos prioritarios identificados como las mayores amenazas para la salud pública, lo que representa la primera respuesta global para identificar y priorizar los patógenos fúngicos y su impacto en la salud pública mundial, así como para considerar las necesidades de investigación y desarrollo no cubiertas (15). La OMS ha agrupado los hongos patógenos prioritarios en aquellos de prioridad crítica (Figura 3), alta y media.

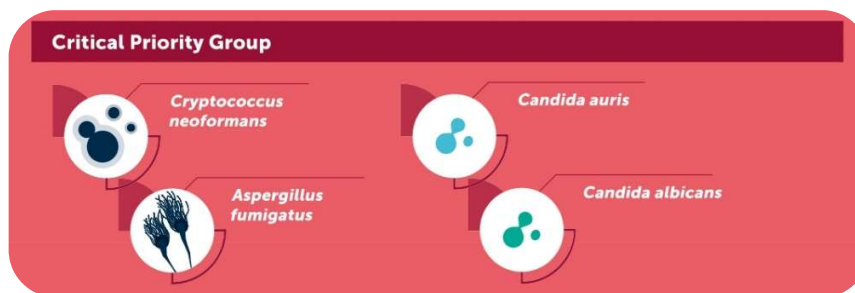


Figura 3. Lista de hongos patógenos fúngicos en el grupo de prioridad crítica según la Organización Mundial de la Salud (16).

Los hongos del grupo de prioridad crítica, como *Aspergillus*, son capaces de producir infecciones graves del tracto respiratorio, entre otras patologías. Sus esporas causan anualmente más de 3.000.000 casos de enfermedades crónicas y 300.000 casos de enfermedades invasivas en todo el mundo. Dependiendo del estado inmunológico del hospedador, la inhalación de estas esporas puede provocar un amplio espectro de enfermedades, tratándose en la mayoría de los casos de sinusitis, aspergilosis alérgica o enfermedad pulmonar crónica (17). Sin embargo, otros casos más graves corresponden a la aspergilosis invasiva (AI), que afecta principalmente a pacientes inmunocomprometidos (18, 19). Se trata de una infección fúngica invasiva (IFI), causada por dicho hongo oportunista, que logra diseminarse a órganos del hospedador más allá de los afectados inicialmente (20). La AI conlleva una elevada tasa de mortalidad (17, 18), la cual se ve aumentada sustancialmente si la infección es causada por cepas resistentes a los azoles (el fármaco oral de primera línea disponible contra las especies de *Aspergillus*) o si el diagnóstico se retrasa o se pasa por alto. La creciente resistencia a los tratamientos antimicóticos existentes se está convirtiendo en una preocupación importante; de hecho, la resistencia a los azoles ha aumentado un 40% desde 2006 (21).

En definitiva, la elevada prevalencia y morbimortalidad de las enfermedades infecciosas respiratorias pone de manifiesto la importancia del estudio de estas patologías, que es esencial para su prevención y control temprano, garantizando así mejores resultados en salud pública y la reducción de la carga global que éstas conllevan (22).

1.1.2. Diagnóstico de las enfermedades infecciosas respiratorias

El principal obstáculo al que se enfrentan los facultativos en el abordaje de las infecciones del tracto respiratorio es su diagnóstico tardío o erróneo (23). La detección temprana y el diagnóstico preciso de estas enfermedades son fundamentales para implementar tratamientos efectivos, reducir la propagación de infecciones y mejorar los

resultados de salud en los pacientes. En la actualidad, más del 50% de los pacientes con estas enfermedades no obtienen el diagnóstico de los agentes causales (24), debido a que su diagnóstico enfrenta varios desafíos. Entre ellos, la variabilidad de los síntomas (que a menudo se solapan con otras condiciones médicas), la falta de disponibilidad de muestras en condiciones óptimas, la gran variedad de microorganismos capaces de producir estas enfermedades, el tratamiento antimicrobiano administrado antes de la recolección de la muestra, y la influencia de la flora normal en el tracto respiratorio (24). Estos hechos resaltan la necesidad de desarrollar métodos diagnósticos rápidos y fiables que puedan ser utilizados en entornos clínicos.

La estrategia diagnóstica actual para las infecciones respiratorias consiste en una serie de evaluaciones clínicas, radiológicas y microbiológicas para centrarse en los agentes infecciosos más probables y decidir entre un enfoque invasivo o no invasivo (6). El uso de técnicas no invasivas es especialmente relevante en pacientes inmunocomprometidos o en estado crítico. Por lo tanto, los avances en técnicas no invasivas podrían permitir un diagnóstico más seguro y más temprano, mejorando así la supervivencia en pacientes con infecciones pulmonares graves (6). Los métodos actuales para diagnosticar la etiología de las LRI son limitados. Las imágenes radiológicas se utilizan tradicionalmente como referente, pero carecen de precisión (25). La obtención de muestras de esputo para cultivo microbiano plantea un desafío en la población pediátrica; la inducción de esputo puede ser desagradable y el lavado broncoalveolar (LBA) es un método invasivo. En los casos de pacientes inmunocomprometidos, la realización de una broncoscopia o un LBA conllevaría un riesgo significativamente elevado para la salud del paciente (6). Los métodos tradicionales, como el hemocultivo, carecen de sensibilidad, y, en general, el cultivo microbiano puede tardar varios días en producir resultados. Además de ser métodos lentos, requieren infraestructura y recursos especializados (26).

En los últimos años, el avance de la biotecnología y la medicina ha dado lugar a nuevas técnicas diagnósticas que ofrecen resultados más rápidos y precisos. Además, el desarrollo de tecnologías de diagnóstico *point-of-care* (POC) ha permitido realizar pruebas en el lugar de atención del paciente, facilitando un diagnóstico más inmediato y la implementación de medidas terapéuticas oportunas (26). Además, el progreso en el diagnóstico molecular, especialmente los enfoques novedosos como la metagenómica clínica y CRISPR (repeticiones palindrómicas cortas agrupadas regularmente espaciadas), no solo ha mejorado nuestra capacidad para la detección de patógenos del tracto respiratorio, sino que también ha traído nuevos desafíos (24).

Introducción

Recientemente, el número de nuevos biomarcadores ha aumentado notablemente en diferentes áreas, como la identificación de patógenos o la respuesta del hospedador (27). Actualmente se utilizan numerosos marcadores séricos en clínica, pero la eficacia del diagnóstico es menor cuando se usan solos o por separado. Una de las causas es que la forma de infección no se comprende completamente en la mayoría de las infecciones del tracto respiratorio. Por lo tanto, se sugiere utilizar combinaciones de diferentes biomarcadores para alcanzar una mayor precisión de diagnóstico (23).

Los beneficios de realizar un diagnóstico y pronóstico rápido de las infecciones respiratorias incluyen la mejora de la calidad de vida de los pacientes, evitando en muchos casos incluso su muerte, así como reducciones en la duración de su estancia tanto en emergencias como ingresado en el hospital y, por lo tanto, en los costes generales (28, 29). Un diagnóstico temprano y acertado permite un uso más adecuado de los medicamentos apropiados para la afección, evitando así efectos secundarios no deseados y tratando farmacológicamente a cada paciente de manera específica lo antes posible. Además, cabe destacar que las infecciones del tracto respiratorio son una de las razones más comunes para prescribir antibióticos, y el uso inadecuado de antibióticos está estrechamente relacionado con el desarrollo de resistencias (30, 31). Por lo tanto, el uso prudente de antibióticos es esencial para reducir la velocidad a la que avanzamos hacia un posible regreso a la era pre-antibiótica, en la que algunos pacientes padecen infecciones que no se pueden tratar con los medicamentos disponibles (28). Por otro lado, una identificación temprana de casos de infecciones respiratorias también permite la implementación eficaz de medidas de control de infecciones dentro de hospitales o centros de atención, resultando en un hecho crucial para reducir la propagación de casos secundarios a corto y largo plazo (29). Esta práctica cobra especial importancia en el contexto actual, marcado por la reciente amenaza global de la pandemia de SARS-CoV-2. La rápida detección y respuesta a las infecciones respiratorias puede significar la diferencia entre contener un brote o enfrentar una crisis de salud pública a gran escala.

Debido a su alta relevancia clínica y severidad, el diagnóstico y pronóstico de la aspergilosis invasiva y el COVID-19 serán aspectos clave de estudio en esta tesis doctoral.

1.2. Aspergilosis invasiva

1.2.1. Descripción y epidemiología de la aspergilosis invasiva

La AI es una enfermedad grave que afecta principalmente a pacientes inmunocomprometidos, cuya mortalidad varía del 70 % al 90 % y es una de las causas más comunes de muerte en este tipo de pacientes (32). Esta IFI oportunista es causada por hongos del género *Aspergillus* (33). *A. fumigatus*, concretamente las especies pertenecientes a la sección *Fumigati*, son las responsables de la mayoría de los casos (34).

Las enfermedades que derivan de la infección por *Aspergillus* generalmente afectan al aparato respiratorio, debido a que, en la mayoría de los casos, la entrada del hongo al hospedador es por inhalación de sus esporas (34, 35). La AI se produce cuando la infección se extiende rápidamente desde los pulmones al cerebro, el corazón, los riñones o la piel (36). Entre otros, los síntomas que provoca son fiebre, tos, dolor torácico y dificultad para respirar. El diagnóstico temprano y el tratamiento con antifúngicos son fundamentales para mejorar las posibilidades de recuperación.

En la década de 1990 se demostró que la incidencia de la AI se había cuatriplicado respecto a estudios epidemiológicos de dos décadas atrás. Los avances en el diagnóstico y la inclusión de los azoles en la estrategia farmacológica a finales de la década de 1980 significaron una mejoría considerable en la incidencia y la mortalidad asociada a la AI. Sin embargo, desde entonces se han realizado pocos progresos adicionales en este campo (34).

Los avances en medicina, en general, han permitido aumentar la esperanza de vida y, asimismo, la curación de enfermedades y patologías que antes parecían imposibles. Sin embargo, algunas técnicas y tratamientos han potenciado la aparición de casos de AI. Este hecho se debe al aumento de pacientes inmunodeprimidos, como consecuencia de los avances en trasplantes, quimioterapia y uso de anticuerpos monoclonales. A medida que mejora la supervivencia de los pacientes con cáncer y se desarrollan terapias innovadoras, un número cada vez mayor de pacientes en estado crítico están inmunocomprometidos (6). Por otro lado, este hecho también se basa en los riesgos potenciales de una mayor invasividad de los procedimientos quirúrgicos complejos en todos los campos de las especialidades médicas (37). Por ejemplo, la aspergilosis broncopulmonar alérgica derivada de procedimientos quirúrgicos es la forma más común de enfermedad debido a la presencia ubicua de los hongos del género *Aspergillus* (38).

Introducción

La prevalencia de la AI se correlaciona inversamente con la inmunocompetencia de los individuos (39). A continuación, se resumen los principales factores de riesgo para padecer esta enfermedad (Figura 4).



Figura 4. Factores de riesgo para padecer AI (40). UCI: Unidad de Cuidados Intensivos, VIH: Virus de la Inmunodeficiencia Humana.

El factor de riesgo principal para padecer AI es la neutropenia (número anormalmente bajo de neutrófilos), que es profunda y prolongada en pacientes que reciben quimioterapia intensiva como tratamiento de la leucemia aguda para un posterior trasplante de células madre (41). La mayor incidencia de AI se observa en pacientes con enfermedades hematológicas malignas, especialmente en los pacientes con trasplantes alogénicos de progenitores hematopoyéticos (42), con una mortalidad atribuible a la AI en estos pacientes que varía del 35 % al 46 % (43-46). Otro grupo de pacientes altamente susceptibles de padecer AI son los tratados en las Unidades de Cuidados Intensivos (UCI), donde los factores de riesgo predominantes son la enfermedad pulmonar obstructiva crónica (EPOC) y el tratamiento con corticoides (47). La prevalencia de esta enfermedad fúngica invasiva en pacientes no hemato-oncológicos se estima alrededor de un 7 % (48). De la misma manera, pueden estar en riesgo aquellos pacientes que presentan enfermedades autoinmunes o reciben tratamiento con diversos fármacos inmunosupresores (49). En la tabla 1 se muestran las incidencias registradas de AI en grupos de pacientes concretos.

Tabla 1. Frecuencias de AI en distintos grupos de pacientes inmunocomprometidos y aparentemente inmunocompetentes (50). EPOC: Enfermedad Pulmonar Obstructiva crónica; VIH: Virus de la Inmunodeficiencia Humana; TNF: Factor de necrosis tumoral (del inglés, *tumor necrosis factor*).

Diagnóstico de base agrupados por condición patológica	Incidencia (%)
Pacientes hemato-oncológicos	
Aplasia medular	9,8-11,5
Leucemia mieloide aguda	7,1-10,5
Leucemia linfática aguda	4,3
Otros cánceres hematológicos	0,35-2,4
Mieloma múltiple	0,2
Pacientes receptores de trasplante de precursores hematopoyéticos	
Alogénico	6,3-12
Autólogo	0,3-5,3
Pacientes receptores de trasplante de órganos sólidos	
Trasplante pulmonar	6
Trasplante cardíaco	5,2
Trasplante intestinal	2,2
Trasplante hepático	2
Trasplante de páncreas	1,1-2,9
Trasplante renal	0,7
Pacientes aparentemente inmunocompetentes	
Pacientes críticos con EPOC	0,36
Pacientes críticos	0,2
Pacientes inmunocomprometidos por otras causas	
Pacientes con infección por VIH	0,02-0,6
Pacientes receptores de anti-TNF	0,01

Como se ha mencionado, las personas con **neoplasias hematológicas** son el grupo más susceptible para sufrir IFI, tanto en la población adulta como pediátrica. Específicamente, la AI es la IFI más común en pacientes con neoplasias hematológicas agudas (51). Múltiples factores de riesgo están implicados en el desarrollo de AI en estos pacientes (52). Los principales factores que aumentan el riesgo de padecer AI en pacientes con este tipo de neoplasias se recogen en la tabla 2. Muchos de ellos están relacionados con las características del paciente.

Introducción

Tabla 2. Factores que aumentan el riesgo de padecer AI en pacientes con neoplasias hematológicas.

Factor de riesgo	Referencia
Neutrófilos < 500 por mm ³ y tiempo de neutropenia mayor a 10 días	(53)
Ser mayor de 65 años	(51, 54)
Polimorfismos genéticos asociados con la respuesta inmune	(51, 53, 55)
Tipo de enfermedad neoplásica y su progresión	(56)
Linfocitos T CD4 < 200 por mm ³	(57)
Infecciones previas	(54, 58-60)
Transfusiones de sangre repetidas previamente	(61)
Uso de corticosteroides	(62)
Regímenes intensivos de quimioterapia con radioterapia	(63)
Uso de nuevas estrategias terapéuticas, como agentes biológicos	(64, 65)

A pesar de la alta morbilidad y mortalidad, la falta de una comprensión profunda de la patogénesis de *A. fumigatus* y la respuesta del hospedador ha obstaculizado el desarrollo de nuevas estrategias diagnósticas y terapéuticas para el abordaje clínico de las infecciones fúngicas (21). Además de la dificultad del tratamiento (que se ve afectado por el número limitado de opciones terapéuticas, las interacciones farmacológicas, los efectos adversos y las resistencias a antifúngicos), uno de los principales obstáculos en el manejo de la AI es la alta complejidad del diagnóstico (66). Los avances recientes en la preparación de muestras, modelos de infección y técnicas de imagen aplicadas *in vivo* han abordado importantes lagunas en la investigación sobre infecciones fúngicas (21), lo que podría llevar a innovaciones significativas en el diagnóstico, pronóstico y tratamiento de la AI.

1.2.2. Diagnóstico de la aspergilosis invasiva

La AI persiste como una de las principales causas de morbilidad y mortalidad, sobre todo en pacientes hemato-oncológicos. A pesar de una mejora significativa en el arsenal antifúngico, la práctica hospitalaria habitual muestra que un tratamiento apropiado se administra demasiado tarde, principalmente debido al retraso en el diagnóstico (67). El diagnóstico temprano de la AI es uno de los pasos más críticos necesarios para tratar la infección de manera eficiente y reducir las altas tasas de mortalidad.

Los procedimientos de diagnóstico tradicionales, como cultivos o radiografías, ponen de manifiesto tanto la falta de sensibilidad como de especificidad (68). Y, por otro lado, los métodos de diagnóstico modernos no siempre son factibles para los pacientes de la UCI. Por ejemplo, las imágenes de tomografía computarizada (TC) requieren un equipo de rayos X especial que crea imágenes transversales del pulmón donde se visualiza la infección por el hongo. Además, otras técnicas novedosas no están bien validadas, como la detección de 1,3-β D-glucano (BDG), que es poco específico por ser un componente de la pared de la mayoría de las especies de hongos (68, 69).

Otro de los problemas graves es el criterio para detectar y diferenciar la AI de otras infecciones fúngicas. Dado que la prevalencia de infecciones por *Candida* spp. es mucho más elevada que las asociadas a *Aspergillus* spp., la mayoría de las herramientas de diagnóstico utilizadas (microscopía directa, cultivo de muestras, hemocultivo...) para ayudar a los médicos a diagnosticar las infecciones por hongos están diseñadas para detectar los hongos más comunes como los del género *Candida*.

La manera actual de realizar el diagnóstico de AI se recoge en los criterios diagnósticos revisados del Grupo Colaborativo de Infecciones Fúngicas Invasoras de la Organización Europea para la Investigación y Tratamiento del Cáncer en conjunto con el Grupo de Estudio de Micosis del Instituto Nacional de Alergias y Enfermedades Infecciosas (EORTC/MSIC) (70). La AI se clasifica de acuerdo con las definiciones de consenso revisadas en las categorías de enfermedad probada, probable o posible, dependiendo de factores del hospedador, criterios clínicos y criterios micológicos (67):

- 🍄 **Posible:** Factores del hospedador + criterios clínicos.
- 🍄 **Probable:** Factores del hospedador + criterios clínicos + micológicos.
- 🍄 **Probada:** Detección de *Aspergillus* en cultivos de cavidad estéril del paciente o sangre o estudio histológico que demuestre invasión fúngica.

El diagnóstico de la AI es complejo y, por lo tanto, requiere un enfoque multidimensional que utilice las características del paciente, los hallazgos clínicos y radiológicos y la evidencia micológica (67, 71, 72). Los factores del hospedador, criterios clínicos y criterios micológicos necesarios para el diagnóstico de los pacientes se muestran en la Figura 5. Para un diagnóstico definitivo de AI, se requiere una biopsia que muestre la invasión del tejido por hifas en combinación con la identificación de una especie de *Aspergillus* (73). Esta técnica es altamente invasiva, y no suele realizarse hoy en día en los hospitales españoles, por lo que habitualmente se da el diagnóstico “posible”

Introducción

o “probable”. El examen histopatológico, citopatológico o microscópico directo de una muestra obtenida mediante aspiración con aguja o biopsia en la que se observan hifas acompañadas de evidencia de daño tisular asociado, califica para una enfermedad invasiva por hongos, pero no para una AI en el caso de no haber identificado *Aspergillus* mediante cultivo o diagnóstico molecular (74).

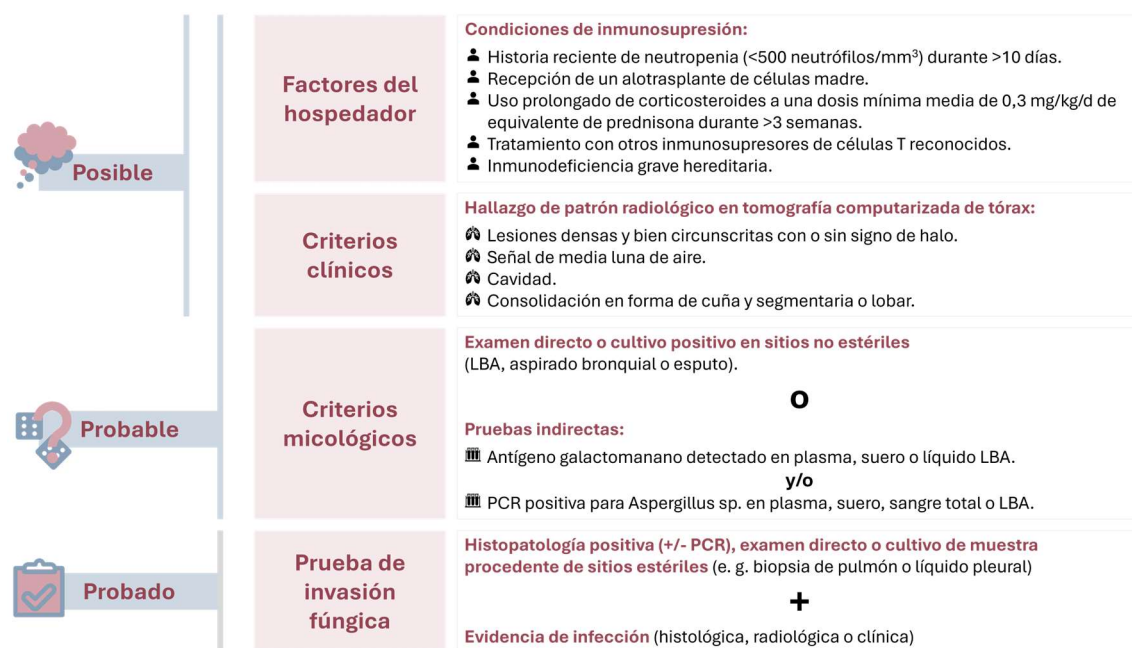


Figura 5. Representación esquemática de la clasificación diagnóstica de la aspergilosis invasiva (AI) según el Consorcio de Educación e Investigación del Grupo de Estudio de la Organización Europea para la Investigación y el Tratamiento del Cáncer y Micosis (70, 71). Una AI posible debe incluir al menos un criterio del hospedador y un criterio clínico. La AI probable debe incluir al menos un criterio del hospedador, un criterio clínico y un criterio micológico. La AI probada se define independientemente de la presencia o ausencia de criterios del hospedador/clínicos/micológicos (66). LBA: lavado broncoalveolar; PCR: reacción en cadena de la polimerasa, en este caso específica para especies de *Aspergillus*.

Un paciente es diagnosticado como AI “posible” cuando existe la presencia de alguno de los factores del hospedador, así como algún criterio clínico de los mostrados en la figura 5.

Para cambiar la clasificación de AI “posible” a AI “probable” es necesario un resultado positivo en las pruebas micológicas: Actualmente el criterio es la detección de dos pruebas de PCR positivas consecutivas para *Aspergillus* (PCRASP, factor de diagnóstico de nueva incorporación al cultivo de hongos convencional) (71) y/o la detección positiva del antígeno galactomanano (GM, componente de la pared celular de hongos del género *Aspergillus* que es liberado durante su crecimiento), ambos en suero, plasma o líquido de LBA. El umbral para la detección positiva del antígeno GM en suero se establece en un índice de densidad óptica $\geq 0,5$ (70).

Actualmente, no está claro cuál es el valor de la PCR positiva para *Aspergillus* en el líquido de LBA. También es objeto de controversia si, en pacientes con resultado positivo en la PCR para *Aspergillus* en LBA, se puede distinguir la colonización (cuando el microorganismo se encuentra entre la flora del paciente, pero no causa enfermedad) de la infección (cuando el microorganismo está presente en el organismo y está causando enfermedad) (75).

Una de las causas de la dificultad del diagnóstico de AI es la existencia de falsos positivos (FP) y falsos negativos (FN) en las pruebas micológicas. La detección de GM presenta FPs ante diversas situaciones: reacción cruzada con algunas bacterias, uso de piperacilina-tazobactam (76), uso de amoxicilina-ácido clavulánico, ingesta de determinados alimentos o aumento de la permeabilidad intestinal (77, 78). La detección de BDG también presenta FPs, ya que, como se ha comentado, es un componente de la pared de casi todas las especies de hongos. El resultado FP está vinculado al uso de hemodiálisis con celulosa, tratamientos con inmunoglobulinas, betalactámicos e infecciones por *Streptococcus pneumoniae* (79). Por otra parte, se producen FNs en infecciones por *Cryptococcus neoformans* y *Zygomycetes*, ya que estos últimos presentan valores indetectables de BDG (69). El diagnóstico correcto de estos FPs supondría evitar el tratamiento antimicótico de manera segura, minimizando el sobretratamiento, así como la toxicidad de los medicamentos y reduciendo los costes asociados.

El diagnóstico y el tratamiento de la AI en el enfermo crítico constituyen un reto para el médico intensivista. La ausencia de síntomas y signos clínicos específicos y la modesta rentabilidad de las pruebas diagnósticas hacen que sea necesario un alto grado de sospecha. El tratamiento debe iniciarse precozmente, pues el retraso en su inicio se asocia a mayor mortalidad (40).

Considerando la relevancia clínica de la AI y el riesgo elevado que presenta en pacientes inmunocomprometidos, es de suma importancia explorar diversas estrategias para optimizar su diagnóstico.

1.2.3. Agente etiológico de la aspergilosis invasiva: *Aspergillus* spp.

Aspergillus es un género de hongos que contiene más de 340 especies pertenecientes al filo de los ascomicetos (80). Se trata de un hongo filamentoso, formado por hifas hialinas septadas. Al final de cada hifa se encuentra una estructura llamada

Introducción

conidióforo (Figura 6), que da origen a las conidias (esporas asexuales) necesarias para la reproducción asexual, aunque ciertas especies también son capaces de reproducirse de forma sexual (81).



Figura 6. Conidias del hongo *Aspergillus fumigatus* agrupadas en forma de ramo en el conidióforo, estructura que genera las esporas asexuales y de la cual se pueden liberar con gran facilidad. Fotografías obtenidas de (82, 83).

Estos hongos son saprófitos. Son los principales responsables de la degradación de múltiples sustancias orgánicas de la naturaleza, descomponiendo restos vegetales y animales, dando lugar a sustancias más simples que pueden ser utilizadas por otros organismos (80, 84).

Varios de estos hongos se utilizan predominantemente para la fermentación de alimentos y la producción a gran escala de enzimas, ácidos orgánicos y compuestos bioactivos (80). Sin embargo, también son ampliamente conocidos patógenos oportunistas en animales y humanos.

Aspergillus se caracteriza por ser un hongo ubicuo. Está presente en casi cualquier sustrato (suelo, aire, alimentos, plantas, materia orgánica en descomposición...), pudiéndose encontrar normalmente en el polvo y los alimentos. Las esporas de *Aspergillus* viajan fácilmente como aerosoles por vía aérea, siguiendo la forma casi universal de transmisión. Esto se debe a su pequeño tamaño de spora (2-3 micras), su mayor termotolerancia con una capacidad de germinación entre 37 °C y 40 °C y su rápida replicación. Estas características optimizan su patogenicidad (54, 57), lo que explica que sea una especie que frecuentemente causa infecciones respiratorias (85).

1.2.3.1. Taxonomía y distribución

Aspergillus fue descrito originalmente en 1729 por el botánico y sacerdote Antonio Micheli. El género debe su nombre a la semejanza de su estructura asexual formadora de

esporas con el aspergillum, un instrumento para rociar agua bendita (84). Para consternación tanto de los sistemáticos aficionados como profesionales, las especies de *Aspergillus* son morfológicamente muy similares y difíciles de distinguir. Además, el género *Aspergillus* presenta una gran diversidad genómica y variabilidad en la naturaleza (84). Las especies del género *Aspergillus* se clasificaron originalmente de manera imprecisa, basándose en la estructura del conidióforo. Posteriormente, se han llevado a cabo clasificaciones más acertadas, en función de las diferencias o similitudes en la secuenciación del DNA ribosómico o de las regiones del espaciador transcrito interno (ITS, del inglés *Internal transcribed spacer*). Estos hallazgos han conducido a reordenamientos de las clasificaciones previas y se prevé que seguirán generando más cambios en el futuro (86).

La taxonomía del género *Aspergillus* se divide en 4 subgéneros (*Aspergillus*, *Circumdati*, *Fumigati* y *Nidulantes*) y 19 secciones, entre las que se destacan las secciones *Fumigati*, *Flavi*, *Terrei*, *Nigri*, y *Nidulantes*; cuyos organismos de referencia son, respectivamente, *A. fumigatus*, *A. flavus*, *A. terreus*, *A. niger* y *A. nidulans* (80, 87).

La sección ***Fumigati*** comprende especies como *A. fumigatus*, *A. fischeri* y especies crípticas como *A. lentulus*, *A. udagawae* y *A. novofumigatus* (80, 86, 87). Esta sección es la más relevante en clínica debido a que sus especies son las más patógenas para el ser humano. Las especies crípticas son menos infecciosas, principalmente debido a una menor velocidad de crecimiento y a una menor termotolerancia respecto a *A. fumigatus*, el cual puede crecer a más de 65 °C (86). A pesar de ello, las especies crípticas son especialmente relevantes en clínica porque son frecuentemente resistentes a azoles, alcanzando hasta el 30 % de los casos. Por el contrario, es relativamente poco frecuente que *A. fumigatus* sea resistente a azoles. Concretamente, en torno al 5 % de los aislados clínicos de *A. fumigatus* son resistentes (86, 88).

La sección ***Flavi*** abarca otra serie de especies relevantes para la biotecnología, la alimentación y la salud, siendo las especies más destacadas *A. flavus* y *A. parasiticus* (89, 90). La mayoría de las especies de la sección *Flavi* son patógenos de plantas y son capaces de producir micotoxinas de gran importancia para la seguridad alimentaria, como las aflatoxinas y las ocratoxinas. Por otro lado, existen especies incapaces de producir micotoxinas, como *A. oryzae* y *A. sojae*, siendo adecuadas para su uso alimentario (89, 90).

La sección ***Terrei*** se caracteriza sobre todo por la especie *A. terreus*, un hongo ubicuo que habita en el suelo y se encuentra en todo tipo de ecosistemas (91). Esta especie

Introducción

es capaz de infectar plantas y produce una amplia gama de micotoxinas perjudiciales para ser humano. Además, las especies de esta sección, especialmente *A. terreus*, son capaces de infectar a pacientes inmunocomprometidos (91).

En cuanto a la etiología de la AI, es causada mayoritariamente por la especie ***A. fumigatus***, que representa hasta el 92% de los aislados (92), mientras que otras especies como *A. flavus* (más predominante en algunas regiones de África o Asia), *A. terreus* o *A. niger* (sección *Nigri*) son aisladas con menor frecuencia en el resto de los casos (93-96). *A. terreus* solo representa el 4 % de los casos de AI. Sin embargo, antes de la aprobación de los azoles, fue un hongo de especial relevancia clínica porque posee una intrínseca sensibilidad disminuida a la anfotericina B, que fue el primer antifúngico aprobado (97-99) y *gold standard* en el tratamiento de las enfermedades fúngicas invasivas (100). Se necesitan concentraciones muy elevadas de este fármaco para inducir su muerte, lo que conlleva el riesgo de alcanzar la dosis tóxica para el hospedador (91, 101).

Otras especies patógenas representan menos del 3% de los casos de AI, e incluyen *A. nidulans* (sección *Nidulantes*), *A. calidoustus*, *A. glaucus*, *A. versicolor* y algunas especies crípticas (por ejemplo, *A. lentulus* y *A. udagawae*) (66).

1.2.3.2. Metabolismo secundario

Los hongos, al igual que muchos otros organismos, tienen la capacidad de sintetizar una gran variedad de metabolitos secundarios, que son moléculas pequeñas de bajo peso molecular. La síntesis de estos compuestos no se requiere para el crecimiento normal del microorganismo, pero pueden proporcionarle diferentes beneficios en cuanto a su adaptación e interacción con el entorno, confiriéndole capacidades como defensa o señalización (102). Pueden ser de naturaleza benigna, como los pigmentos o las moléculas utilizadas en la comunicación entre especies, pero también pueden ser malignos y presentar actividad antimicrobiana o tóxica con el objetivo de eliminar otros microorganismos competidores (103). Algunos de estos metabolitos secundarios microbianos han sido estudiados por científicos debido a su potencial beneficio para los humanos, y han proporcionado numerosos productos farmacéuticos, desde antibióticos hasta compuestos inmunosupresores (104).

Los metabolitos secundarios producidos por hongos son generalmente dependientes de la etapa de desarrollo del hongo y de factores ambientales como la concentración de nutrientes, la luz o la temperatura (103). Los hongos pertenecientes al género *Aspergillus* son especialmente capaces de producir una gran diversidad de

compuestos, que les permiten adaptarse a múltiples entornos. Los metabolitos secundarios producidos por *Aspergillus spp.* presentan actividad inmunomoduladora, oncogénica, antibacteriana, antiviral, y/o antifúngica. Estas propiedades les confieren a dichos hongos la capacidad de competir con el resto de los microorganismos que tratan de coexistir en el mismo ambiente.

Cada especie del género *Aspergillus* genera un perfil de metabolitos distinto, probablemente debido a que no todas las especies contienen los mismos clústeres de síntesis de metabolitos secundarios (105). En las regiones próximas a los telómeros es donde se encuentran los genes que intervienen en las rutas biosintéticas de los diferentes metabolitos secundarios, contenidos en sus pertinentes clústeres genéticos (106, 107).

La variabilidad genética en estas zonas es extremadamente alta, hasta el punto de que el grado de homología puede descender al 30% incluso entre hongos de la misma sección (106-108). Por el contrario, la parte del genoma que no se ve involucrada en el metabolismo secundario está altamente conservado (más de un 80%) en especies de *Aspergillus* (106). A pesar de que algunos metabolitos secundarios son producidos por diferentes secciones, lo más común es que los clústeres homólogos se hayan diferenciado durante los diferentes eventos de especiación, dando lugar a un metabolito secundario diferente o incluso a su pérdida (106-108).

En general, la estructura de los clústeres de síntesis de metabolitos secundarios es similar. Existe una enzima principal, normalmente una péptido sintasa no ribosomal (NRPS), una policétido sintasa (PKS) o una enzima mixta NRPS/PKS que sintetiza la estructura del metabolito secundario. Por otro lado, una serie de enzimas accesorias modifican la estructura inicial hasta generar el metabolito final. Y finalmente, uno o varios factores de transcripción controlan la expresión de los genes del clúster o el transporte del metabolito hacia el exterior (106). Por norma general, cada genoma de *Aspergillus* contiene varios clústeres que codifican NRPSs y PKSs (109, 110).

En concreto, *A. fumigatus* secreta más de 226 metabolitos secundarios que pueden clasificarse en 24 familias biosintéticas. Estos incluyen policétidos, péptidos cíclicos, alcaloides y sesquiterpenoides, que son frecuentemente objeto de estudio (111). Entre los metabolitos secundarios producidos por *A. fumigatus*, se encuentran las epipolítiodioxopiperazinas (ETPs), siendo la gliotoxina (GT) y su derivado la bismetilgliotoxina (bmGT) los miembros mejor caracterizados (112).

1.2.3.2.1. Gliotoxina

La GT es uno de los compuestos mejor conocidos de la clase de metabolitos fúngicos llamados ETPs. La GT fue descubierta por Weindling y Emerson (113), cuando en 1936 estudiaron su producción por *Trichoderma lignorum* y su eficacia antibiótica *in vitro* frente a bacterias grampositivas y gramnegativas. White publicó un trabajo en *Science* en 1940 (114) en el cual se describían las propiedades bactericidas de los filtrados de cultivos de *A. flavus*, lo cual dio origen a una verdadera fiebre de investigación en torno a esta especie (115). La GT fue descrita en 1943 (116) y su estructura elucidada en 1966 (117). La estructura de las ETPs, y concretamente de la GT (Figura 7), se caracteriza por la presencia de un enlace disulfuro interno en un anillo de piperazina, formado por la adición de dos átomos de azufre al enlace cíclico de los aminoácidos L-Serina (L-Ser) y L-Fenilalanina (L-Phe).

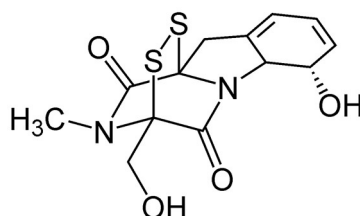


Figura 7. Estructura química de la GT. Imagen obtenida de Wikimedia Commons.

1.2.3.2.1.1. Biosíntesis de la gliotoxina

Al igual que muchos otros metabolitos secundarios producidos por hongos, la GT es sintetizada por diversas enzimas que están codificadas en un clúster de genes. El clúster de genes de la GT de *A. fumigatus* (clúster *gli*), que es uno de los más estudiados, contiene 13 genes necesarios para la biosíntesis de GT (118) (Figura 8). El estudio de estos genes ha dado a conocer la relevancia de la GT como compuesto tóxico y factor de virulencia durante la AI.

En uno de los extremos del grupo de genes, se localiza *gliZ*, que codifica un factor de transcripción con dedos de zinc responsable de la expresión de varios genes del clúster (119). Uno de esos genes, *gliP*, codifica una NRPS responsable del primer paso en la biosíntesis de GT. GliP es un NRPS de tres módulos (A1-T1-C1-A2-T2-C2-T3) capaz de sintetizar un péptido cíclico uniendo L-Phe y L-Ser (120). Aunque carece de un dominio tioesterasa que permita la liberación del ciclopéptido, los estudios *in vitro* han demostrado que esta liberación sucede lentamente en una cepa *A. fumigatus* Af293 GliP *wild type* (120). Las cepas deficientes en GliP no son capaces de producir GT al no poder catalizar la unión de los dos aminoácidos. Además, se correlaciona con una menor virulencia del hongo

debido a la deficiencia en la producción de GT, por lo que la GT ha sido identificada como factor de virulencia (121).

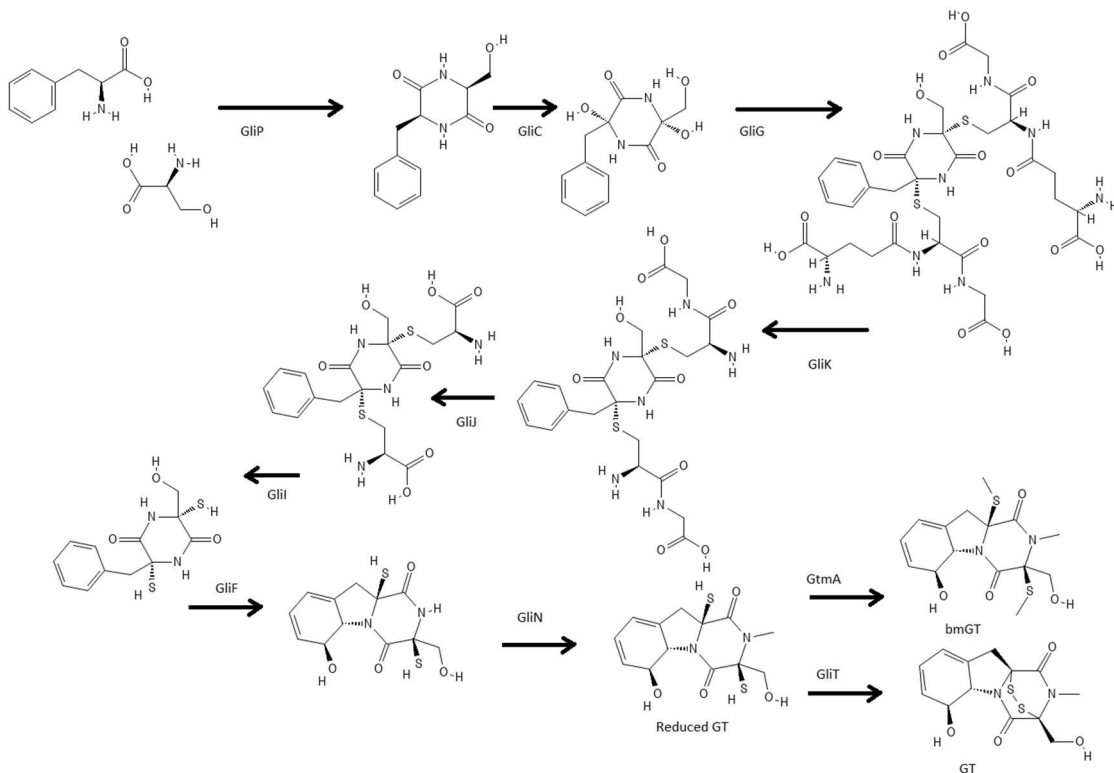


Figura 8. Biosíntesis de GT y genes implicados en el proceso (122).

Una vez que se sintetiza la estructura base de la dicetopiperazina, el citocromo P450 GliC cataliza el siguiente paso mediante la hidroxilación de los carbonos alfa de L-Phe y L-Ser (123). Posteriormente, se añaden dos átomos de azufre mediante la sustitución de los grupos hidroxilo unidos por dos moléculas de glutatión. Esta reacción la lleva a cabo la enzima gliG (124). La eliminación completa de las moléculas de glutatión se realiza mediante tres proteínas llamadas GliK, GliJ y GliI (125). GliK, una gamma glutamil ciclotransferasa, libera residuos de gamma-glutamilo del glutatión unidos a la estructura principal; GliJ, una dipeptidasa, escinde las glicinas formando el glutatión restante; y GliI, una liasa, rompe el enlace C-S eliminando completamente el resto de la estructura inicial (125). De esta manera, el producto liberado (3-bencil-3,6-ditio-6-(hidroximetil)-dicetopiperazina o ditiolgliotoxina) contiene los dos grupos tiol característicos de la GT.

El siguiente paso en la biosíntesis de GT es considerablemente importante. Como los grupos tiol pueden reaccionar con muchas otras moléculas, la GliT (tioreductasa) oxida ambos grupos formando un enlace disulfuro, transforma la GT de su forma reducida (ditiol) a la forma oxidada (disulfuro). Esta transformación, en microorganismos como *Saccharomyces cerevisiae* y *Aspergillus nidulans*, configura a GliT como una enzima clave

Introducción

en la autoprotección frente a la GT, minimizando la autotoxicidad inducida por este metabolito intermediario de la biosíntesis de GT (126-128).

Especies como *A. fumigatus* y *A. flavus* poseen un gen (*gtmA*) que codifica una enzima críptica conocida como GtmA con actividad S-adenosilmetionina bismetiltransferasa dependiente de GT. Este gen no se encuentra dentro del clúster de genes *gli*, pero codifica una enzima encargada de la metilación de los dos grupos sulfhidrilo de la GT reducida, generándose un derivado conocido como bisdetiobis(metiltio)gliotoxina, bismetiltiogliotoxina o bismetilgliotoxina (bmGT), carente de actividad. La función de la bmGT es atenuar la síntesis de GT mediante el desvío de moléculas de GT hacia la formación de este derivado metilado (129, 130). En concreto se ha descrito la enzima GtmA/TmtA como la principal responsable de la metilación e inactivación de GT (129, 131), lo que supone un mecanismo de protección del hongo (132).

Una vez que se produce GT, ésta inducirá la expresión del grupo *gli* (133) o será secretada por *A. fumigatus* a través de GliA. El gen GliA codifica un transportador encargado de transportar la GT al exterior del hongo (evitando así la generación de ROS en su interior). Un mutante deficiente de GliA es incapaz de transportar eficientemente la GT al exterior, acumulándose en el interior del hongo (134).

Debido a la alta reactividad de GT, *A. fumigatus* ha desarrollado estos mecanismos de autoprotección, siendo los dos más importantes la formación de bmGT que dará lugar a una toxina inactiva, y la secreción que reducirá la expresión de *gli* y la producción de GT (129).

Se ha sugerido que varias especies de hongos sintetizan GT, como *Trichoderma lignorum* (135), *Trichoderma virens* (136), *Penicillium obscuro* (136), *Gliocladium fimbriatum* (116), *Candida albicans* (137) y diferentes *Aspergillus spp.*, como *A. fumigatus* (138), *A. terreus*, *A. niger* y *A. flavus* (139), entre otros. Sin embargo, se han publicado hallazgos contradictorios con respecto a la síntesis de GT en algunas de estas familias de hongos (136, 139-142).

Dado que la producción de GT por diferentes especies de hongos podría ser relevante para diversos campos, incluidos la salud, la alimentación y la agricultura, resulta de suma importancia investigar qué hongos ambientales podrían ser productores significativos de GT.

1.2.3.2.1.2. Relevancia y aplicaciones de la gliotoxina

La alta reactividad del enlace disulfuro contenido en la estructura de GT, le confiere la capacidad de interactuar con otras moléculas biológicas mediante la formación de un puente disulfuro y de generar especies reactivas de oxígeno (ROS) mediante ciclos redox (112, 143). Gracias a los azufres que contiene la molécula, puede alternar entre una forma ditiol completamente reducida y una forma disulfuro (Figura 9) (124). Además, GT en su forma reducida puede reaccionar con otros grupos tiol accesibles en proteínas, inactivándolas, inhibiéndolas o desnaturalizándolas. En efecto, la mayoría de sus propiedades biológicas se deben a su estructura y reactividad (144).

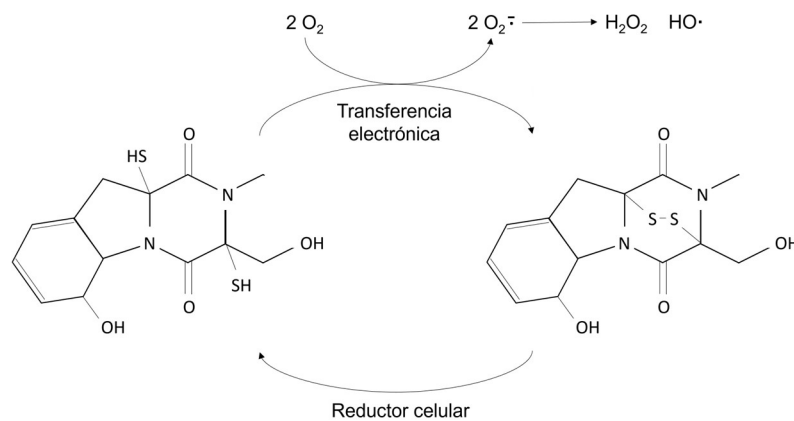


Figura 9. Ciclo redox entre la forma reducida (ditiol, izquierda) y oxidada (disulfuro, derecha) de GT. La oxidación de la GT y presumiblemente de otras ETPs genera especies reactivas de oxígeno, ya que es capaz de reducir el oxígeno molecular. Imagen adaptada de (112).

Cabe destacar el efecto de la GT en animales, especialmente mamíferos. La GT se secreta durante la infección por *A. fumigatus* para evitar la inmunidad del hospedador y ayudar a la colonización. La GT ha demostrado ser un factor de virulencia de este hongo, que mejora la invasión de los pulmones de ratón, lo que conduce a una mayor mortalidad (119, 121, 143, 145, 146). La GT se secreta durante los primeros pasos de la germinación de los conidios y la invasión de las hifas del hongo, ya que puede detectarse en los pulmones de ratones infectados desde el primer día después de la infección (147).

Esta evidencia sugiere que la GT se produce durante las primeras fases de la AI y que su actividad inmunosupresora contribuye y exagera los efectos patológicos observados en el hospedador infectado (148). La GT secretada durante la infección podría desempeñar un papel fundamental en la supervivencia de los hongos porque inhibe la fagocitosis de los macrófagos, provoca la muerte celular apoptótica en macrófagos y monocitos, inhibe el ensamblado de la NADPH oxidasa en fagocitos (149), bloquea la producción de citoquinas mediante la inhibición del factor de transcripción NF- κ B y

Introducción

previene la desgranulación de mastocitos (35, 143, 150-155). Además de su papel como factor de virulencia y su posible función durante la invasión del hospedador, se ha demostrado que la GT induce apoptosis en una amplia gama de células tumorales, como las células de fibroblastos L929 de ratón, células estrelladas hepáticas humanas y de rata, células de leucemia linfocítica crónica (156), células de cáncer de cuello uterino humano o de cáncer colorrectal (157-160). En cuanto al mecanismo por el cual la GT es capaz de inducir la apoptosis, se ha demostrado que, en células de ratón, activa la vía mitocondrial a través de Bak (proteína pro-apoptótica) (161). La activación de Bak podría deberse a la formación de enlaces transitorios entre el puente disulfuro de la GT y los residuos individuales de cisteína presentes en Bak. Una vez activado, Bak desencadena la generación de ROS, que es crucial para la formación eficaz de poros en la membrana mitocondrial, favoreciendo la liberación de factores apoptogénicos (como citocromo c y factor inductor de apoptosis) y la activación de caspasa-3, conduciendo finalmente a la muerte celular (Figura 10).

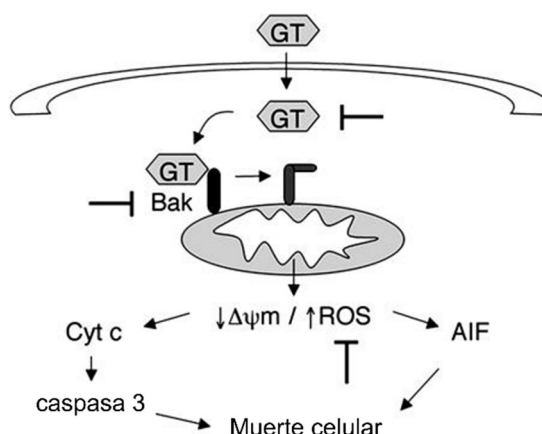


Figura 10. Mecanismo propuesto para la muerte celular inducida por GT. Después de la producción por *A. fumigatus*, la GT entraría en las células por un mecanismo redox-dependiente e induciría directamente un cambio conformacional en Bak, lo que llevaría a la despolarización mitocondrial y la producción de ROS. La producción de ROS desencadena la liberación mitocondrial de proteínas apoptogénicas como citocromo c (cyt c) y factor inductor de apoptosis (AIF). De esta forma, tanto los procesos dependientes de las caspasas como los independientes se activarían para inducir la muerte celular. Figura adaptada de (161).

Por lo tanto, la GT se ha propuesto como un potencial fármaco anticancerígeno, aunque debido a su toxicidad contra células y tejidos sanos, se requerirían formas selectivas de administración a las células cancerosas antes de que pueda probarse *in vivo*, como se propuso recientemente (162). Por otro lado, debido a su alta actividad biológica y sus posibles efectos nocivos contra los mamíferos y otras especies, la contaminación de alimentos destinados al consumo humano y animal por GT ha suscitado una creciente preocupación en los últimos años (163-165).

También se ha descubierto que la GT ejerce actividad biocida contra un gran número de especies pertenecientes a casi todos los reinos biológicos, y también contra algunos virus. En el caso de las bacterias, GT inhibe el crecimiento *in vitro* de *Mycobacterium tuberculosis*, *Pseudomonas aeruginosa*, *Enterobacter aerogenes*, *Escherichia coli*, *Microsporium gypseum*, *Staphylococcus aureus* y *Bacillus subtilis* (166-169). En lo que respecta al reino de los hongos, se ha estudiado que la GT previene el crecimiento *in vitro* del patógeno humano *Candida albicans* (104). Protozoos como *Plasmodium falciparum*, responsable de la enfermedad de malaria, también son susceptibles a la presencia de GT (170). Además, se ha descrito la actividad antiviral de GT frente a una amplia gama de virus como el virus Nipah, el virus Hendra, el virus de la polio, el virus del herpes simple, el virus coxsackie o el virus de la influenza A (171-174) entre otros.

Ante la urgente necesidad de desarrollar nuevas terapias contra bacterias resistentes, el estudio de nuevos antibióticos cobra una gran relevancia, siendo particularmente prometedora la investigación del potencial de la gliotoxina en este campo.

1.2.3.2.2. Bismetilgliotoxina

Como se ha mencionado anteriormente, la bmGT es el derivado inactivo de la GT. La estructura química de la bmGT es el resultado de la metilación de los dos grupos sulfhidrilo de la GT reducida (Figura 11).

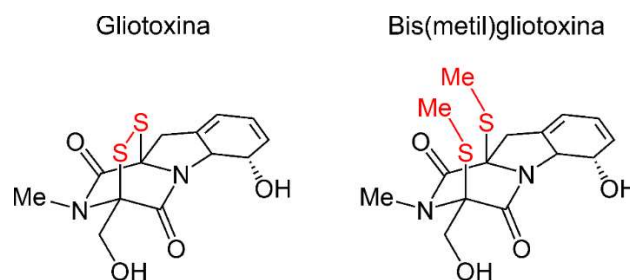


Figura 11. Estructura química de la GT (izquierda) y bmGT (derecha). Las diferencias estructurales están señaladas en color rojo (175).

Se ha sugerido que estos metabolitos secundarios se liberan en el inicio de la AI. También es importante destacar que ambos compuestos se producen después de que los conidios germinen y formen hifas. Consecuentemente, no se liberan en la población sana cuando se exponen a conidios en el aire. Por lo tanto, a priori, la GT podría ser considerada

Introducción

un marcador potencial para el diagnóstico de AI. Sin embargo, la GT se asocia rápidamente a las células durante la incubación *in vitro* de sobrenadantes de hongos con una variedad de tipos de células (176), lo que sugiere que la GT biológicamente activa también puede asociarse rápidamente con las células *in vivo* y eliminarse de los fluidos corporales por los tejidos adyacentes y las células circulantes y, por lo tanto, puede no ser detectable.

Por el contrario, su derivado inactivo bmGT, que también es producida por *A. fumigatus in vitro* (177), no es tóxica y no es captada por las células, por lo que a priori es un mejor marcador de la presencia de hongos. Varios grupos han aportado evidencia experimental del uso de micotoxinas producidas por *Aspergillus spp.*, incluida la GT, como marcadores para detectar la presencia de hongos. Esto se ha aplicado para estudiar el deterioro de los alimentos y la infección por hongos en tejidos animales y humanos (178-181). Sin embargo, debido a las diferencias en la reactividad de GT y bmGT (112, 144, 182), cabe esperar una mayor estabilidad de la bmGT en muestras biológicas comparada con la GT. Domingo *et al.* (148) confirmaron la elevada recuperación de GT y bmGT adicionados en suero y plasma. Por otro lado, en sangre, solo se recuperó con éxito la bmGT. Estos resultados corroboraron que, *in vitro*, las células sanguíneas eliminan rápidamente la GT, pero no la bmGT.

Por lo tanto, la alta reactividad y la rápida asociación de GT a las células, conlleva la disminución de la GT libre y detectable en los fluidos biológicos. De hecho, la presencia y las concentraciones de bmGT son mayores que las de GT en muestras de humanos con riesgo de AI. Por consiguiente, se considera que la bmGT es un mejor marcador que GT para monitorizar la presencia de hongos en pacientes inmunocomprometidos (148). Sin embargo, se necesita más trabajo, incluidos estudios prospectivos, para evaluar si la detección de bmGT es un indicador confiable de AI.

La incorporación de nuevos biomarcadores en el diagnóstico de AI podría complementar las técnicas actuales, brindando protección adicional contra falsos positivos y negativos en casos de aspergilosis potencialmente mortales.

1.3. COVID-19

1.3.1. Descripción y epidemiología del COVID-19

La enfermedad por coronavirus, COVID-19 es una enfermedad infecciosa causada por el virus SARS-CoV-2. La mayoría de los pacientes que la padecen experimentan una enfermedad respiratoria leve o moderada y se recuperan sin requerir un tratamiento especial. Sin embargo, algunos individuos enferman gravemente y requieren atención médica. Las personas de avanzada edad y las que padecen otras enfermedades subyacentes (como enfermedades cardiovasculares, diabetes, enfermedades respiratorias o cáncer) pueden desarrollar una enfermedad grave potencialmente mortal (183).

El origen del SARS-CoV-2 se sitúa en diciembre de 2019 en Wuhan, China, cuando se informó de un grupo de casos de neumonía debida a un patógeno desconocido. El Centro Chino para el Control y la Prevención de Enfermedades (CDC China), junto con científicos del Instituto Nacional para el Control y la Prevención de Enfermedades Virales (IVDC), llevaron a cabo una investigación epidemiológica y etiológica que demostró que el agente causante del brote era un coronavirus del linaje Betacoronavirus (184, 185).

El Grupo de Investigación de Coronavirus (CSG) del Comité Internacional para la clasificación de virus descubrió que el nuevo coronavirus mostraba una estrecha relación con el anterior virus SARS, por lo que lo renombró como SARS-CoV-2 (186). Al comienzo del brote de SARS-CoV-2, la mayoría de los casos estaban relacionados con la exposición al mercado mayorista de mariscos de Huanan, y se sugirió que el pangolín era el reservorio del virus desde el cual el SARS-CoV-2 se transmitió por primera vez a los humanos (187). Aun así, en poco tiempo se informó la transmisión de este nuevo coronavirus de persona a persona, y solo unos pocos pacientes tuvieron contacto directo con el comercio de animales vivos (188).

Inicialmente, se subestimó la rápida transmisión del virus, lo que permitió que el SARS-CoV-2 se propagara por todo el mundo. Aunque rápidamente se desarrollaron pruebas diagnósticas específicas, como PCR con transcriptasa inversa (RT-PCR), la transmisión asintomática del SARS-CoV-2 fue determinante en la difusión global, ya que dificultaba el rastreo de la transmisión (189). El 31 de enero de 2020 el SARS-CoV-2 fue declarado emergencia de salud pública de importancia internacional por la OMS, pasando a categoría de alto riesgo el 29 de febrero. El 11 de marzo, el SARS-CoV-2 y la enfermedad que induce, denominada enfermedad por coronavirus 2019 o COVID-19, adquirieron el estatus de pandemia (190).

Introducción

Los comités internacionales contra las enfermedades emergentes establecieron diferentes pautas a seguir a escala global. Estas medidas de control incluían estrictas normas de viaje, cuarentenas obligatorias y confinamientos. Sin embargo, debido a la falta de concienciación, muchos países no cumplieron las regulaciones de manera adecuada ni a tiempo, lo que contribuyó a la crisis sanitaria mundial. La mayoría de los hospitales enfrentaban escasez de equipos de protección individual (EPI) y falta de camas en las UCI. La sociedad dependía principalmente de los profesionales de la salud (191). Para evitar el colapso de los sistemas sanitarios, se establecieron medidas de confinamiento global, permitiendo únicamente trabajar al personal esencial. El distanciamiento social, el aislamiento y el saneamiento fueron claves para frenar la propagación del virus. Los procesos globales de confinamiento y cuarentena fueron medidas excepcionales para controlar la rápida transmisión del virus, pero no pudieron implementarse indefinidamente. A principios del verano de 2020 la mayoría de los países comenzaron a aliviar gradualmente las restricciones, en parte para disminuir mayores consecuencias económicas. Durante los meses siguientes, los diferentes países y regiones establecieron de manera independiente las medidas de desescalada dependiendo de las tendencias de las tasas de infección, tratando de alcanzar la “nueva normalidad” (192).

La aparición de nuevas variantes del SARS-CoV-2 con propiedades de transmisión más eficientes aumentó el número de casos. Al mismo tiempo, se estaba produciendo una carrera sin precedentes por el desarrollo y validación de vacunas, cuya administración se autorizó en profesionales sanitarios y poblaciones de riesgo a principios de 2021 (193). El papel de las vacunas ha sido fundamental para reducir la propagación, la gravedad y la muerte por COVID-19 (194). A pesar de ello, la desigualdad en la distribución de vacunas (sobre todo en países de bajos ingresos) ha producido una disminución progresiva de la inmunidad, lo que permitió el surgimiento de variantes con otros mecanismos de escape inmunológico que siguen circulando entre la población mundial actual (195).

Hasta la actualidad (noviembre de 2024), se han informado más de 776,9 millones de casos confirmados de infección por SARS-CoV-2, incluyendo 7,1 millones de muertes en todo el mundo. En la figura 12 se muestran los casos totales de COVID-19 (A), así como las muertes producidas por esta enfermedad (B) en todo el mundo (196). Se han administrado un total de 5.47 mil millones de dosis de vacunas (196).

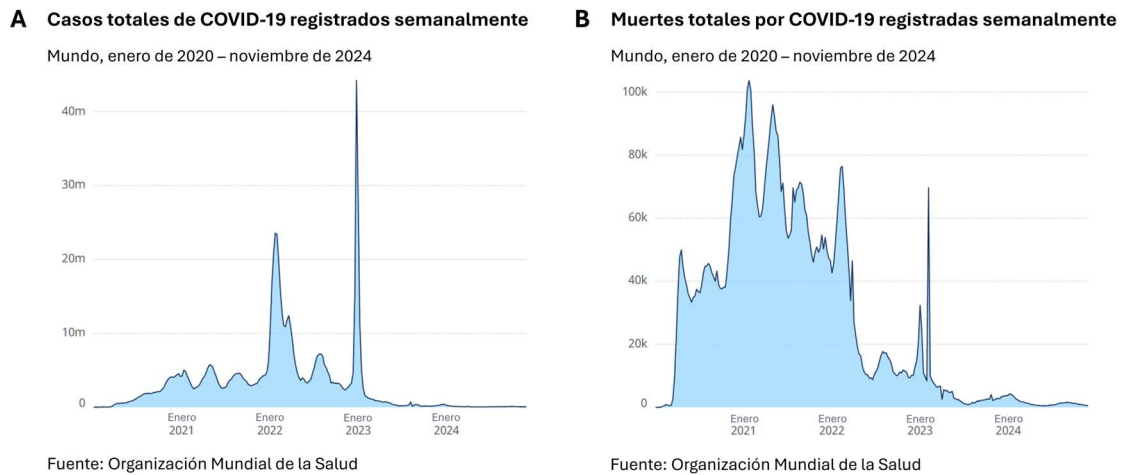


Figura 12. (A) Casos totales de COVID-19 en todo el mundo, registradas semanalmente desde enero de 2020 hasta noviembre de 2024. **(B)** Muertes totales causadas por COVID-19 en todo el mundo, registradas semanalmente desde enero de 2020 hasta noviembre de 2024. Fuente: <https://data.who.int/dashboards/covid19/deaths?n=0>

A pesar de ello, dado que se ha reducido el número de casos graves de COVID-19 y ya no es obligatorio informar un resultado positivo de la prueba, es probable que el número de casos confirmados esté subestimado. Además, en los países con menor acceso a las pruebas de antígenos también se subestima el número de casos registrados. Pese a estos hechos, el 4 de mayo de 2023 la OMS determinó que el COVID-19 ya no constituye una emergencia de salud pública de importancia internacional, pasando a una gestión a largo plazo de la pandemia.

La pandemia de SARS-CoV-2 ha afectado a todos los ámbitos de la vida en todo el mundo, incluida la economía, la cultura y el transporte, entre otros, pero también ha afectado en gran medida a las relaciones interpersonales y a nuestro estilo de vida.

1.3.2. Respuesta fisiológica a SARS-CoV-2

La patogénesis de la infección por SARS-CoV-2 no sólo depende de la replicación viral, ya que la respuesta inmune del hospedador está fuertemente implicada en el deterioro y la gravedad del COVID-19. Una activación disfuncional de la respuesta inflamatoria es responsable de la mayor parte del daño en los casos moderados a graves, mientras que una respuesta inmune regulada es capaz de controlar el patógeno y limitar el daño tisular en los casos leves de COVID-19 (197).

Comprender la respuesta fisiológica a SARS-CoV-2 es esencial para controlar la enfermedad, así como para desarrollar nuevas técnicas de diagnóstico y pronóstico.

- La inflamación y los cambios metabólicos producidos en pacientes con COVID-19 proporcionan prometedores biomarcadores con los que estudiar tanto el diagnóstico como el pronóstico de la enfermedad.

1.3.2.1. Inflamación

Una parte esencial de la respuesta inmune innata es la inducción de inflamación, que engloba una amplia gama de vías cuyo propósito fisiológico es defenderse de la infección (198). La primera fase de la inflamación incluye mecanismos controlados para prevenir el establecimiento inicial de la infección y su posterior propagación, con el reclutamiento de células efectoras de la respuesta inmune tanto innata como adaptativa contra el SARS-CoV-2 (199). La segunda fase incluye la eliminación y reparación del tejido dañado mediante un proceso cuidadosamente controlado, que concluye con la restauración de la homeostasis (200). Durante la infección por SARS-CoV-2, las partículas virales de las células dañadas y muertas actúan como inductores de inflamación (201). Éstos desencadenan cambios en el epitelio vascular de las vías respiratorias, con un aumento de la permeabilidad de los vasos sanguíneos y, por lo tanto, del flujo sanguíneo. En consecuencia, hay fuga vascular de proteínas plasmáticas y reclutamiento de células inmunes, previamente restringidas a la circulación, que acceden a los tejidos en el sitio de la infección (202). Los componentes celulares inducen aún más la respuesta inflamatoria con la liberación de mediadores de la inflamación, como enzimas proteolíticas, mediadores lipídicos, el sistema del complemento, moléculas vasoactivas, citoquinas y quimioquinas (203). Estos mediadores desencadenan el reclutamiento de otras células y terminan de construir una respuesta inmune antiviral proinflamatoria adecuada que implica desencadenar la muerte de las células infectadas por el virus. Además, los mediadores de la inflamación también actúan sobre células no inmunes, incluidas las células endoteliales y epiteliales de las vías respiratorias (202). En este sentido, dado que la respuesta inflamatoria en el sitio de infección no se limita a partículas virales y células infectadas, el daño colateral y la muerte del tejido sano del hospedador son inevitables (204). En los casos asintomáticos y leves de COVID-19, el daño colateral inducido por la respuesta inmune se repara durante la fase de resolución de la inflamación (205). Por lo tanto, las células inmunitarias antiinflamatorias favorecen la muerte celular de las células proinflamatorias y la eliminación de restos celulares y mediadores solubles. También se produce la reparación de tejidos y activación de la fibrosis (206). Sin embargo, la infección

por SARS-CoV-2 puede progresar hacia una hiperinflamación desregulada en el tracto respiratorio inferior y evolucionar hacia un síndrome de dificultad respiratoria aguda (SDRA) potencialmente mortal (207, 208).

1.3.2.1.1. Inmunomoduladores solubles

Tanto en pacientes con COVID-19 como con otras enfermedades críticas, la fisiopatología inflamatoria es uno de los principales factores que favorecen la muerte (209). La activación inflamatoria masiva en respuesta a infecciones implica la denominada tormenta de citoquinas (210-212), que se describe como una cascada de activación excesiva de las células inmunes y la regulación anormal, sobreexpresión y liberación incontrolada de varias citoquinas proinflamatorias que rompen el equilibrio del sistema inmunológico (213-215).

Las **citoquinas** y las **quimioquinas** son proteínas de señalización encargadas de la comunicación entre células inmunitarias. Entre sus funciones, inducen el reclutamiento y la activación celular, pero también modulan la inhibición celular una vez que se ha resuelto la infección. Por lo tanto, muestran un rango dinámico de funciones, lo que agrega complejidad a las vías con las que interactúan (216). Las citoquinas incluyen, entre otras, la familia de las interleuquinas (IL), cuya función principal es la inmunomodulación mediante la activación de la inflamación, la diferenciación y la proliferación celular (217). Las quimioquinas son un tipo de citoquinas que regulan principalmente la migración celular y el reclutamiento al sitio de infección debido a gradientes en su concentración (218).

Durante la primera etapa de la infección por SARS-CoV-2, se liberan citoquinas y quimioquinas para reclutar células inmunes de la circulación al tracto respiratorio (219). Es esperable que se produzca un aumento controlado de mediadores proinflamatorios durante una infección, ya que es necesaria una respuesta fisiológica adecuada. Sin embargo, en los casos graves de COVID-19, este equilibrio se altera, resultando en una liberación desregulada de citoquinas y daño tisular (220).

En comparación con los controles sanos, los pacientes con COVID-19 muestran una regulación positiva de una amplia gama de citoquinas y quimioquinas en plasma y en el sitio principal de infección (Tabla 3). El líquido de LBA proporciona una muestra más representativa de los mediadores solubles y las subpoblaciones celulares en el tracto respiratorio inferior durante la infección por SARS-CoV-2. No obstante, la extracción de

Introducción

LBA se limita principalmente a los casos graves, mientras que las muestras de plasma brindan información a nivel sistémico.

Tabla 3. Principales citoquinas y quimioquinas reguladas positivamente en pacientes con COVID-19. Se detallan sus fuentes celulares y sus respectivas dianas, así como sus funciones más relevantes (217, 218, 221, 222).

	ORIGEN	DIANA	FUNCIONES
CITOQUINAS			
IL-2	Linfocitos T y B, NK, monocitos, macrófagos	Linfocitos T y B, NK, macrófagos	Proliferación, aumento de la toxicidad, secreción de IFN γ , producción de anticuerpos, activación de macrófagos, inmunidad adaptativa
IL-4	Th, mastocitos	Linfocitos T y B, monocitos, macrófagos	Proliferación, diferenciación de Th2, inhibición de la inmunidad mediada por células, inmunidad adaptativa
IL-6	Macrófagos, linfocitos T, fibroblastos, otros	Múltiple	Proinflamatoria y coestimulante, induce proliferación y diferenciación, regula la respuesta de fase aguda
IL-7	Médula ósea	Linfocitos T y B	Homeostasis, diferenciación, supervivencia
IL-10	Th, células T reguladoras, linfocitos B, células dendríticas	Macrófagos, linfocitos T y B, células dendríticas	Supresión inmunitaria, disminuye la presentación de antígeno y la expresión de moléculas de clase II del complejo mayor de histocompatibilidad de las células dendríticas, regula negativamente la respuesta de las células T auxiliares, promueve la inmunidad humoral, antiinflamatoria
IL-12	Macrófagos, células dendríticas, linfocitos B, neutrófilos	Linfocitos T, NK	Diferenciación y proliferación, promoción de Th1 y citotoxicidad
IL-17	Th17, otros	Células epiteliales y endoteliales	Inmunidad proinflamatoria y protectora en el pulmón, promueve la movilización de neutrófilos y la producción de citoquinas
IL-18	Macrófagos, múltiple	Th1, NK, linfocitos B	Proinflamatoria, inducción de IFN γ
TFN	Múltiple	Neutrófilos, macrófagos, monocitos, endotelio, NK, linfocitos T	Proinflamatoria, promueve la activación y producción de proteínas de fase aguda, mejora la citotoxicidad, daña el endotelio, aumenta la fagocitosis de neutrófilos, ligando de muerte
QUIMIOQUINAS			
CCL5 (RANTES)	Linfocitos T y B, monocitos, epitelio, fibroblastos	Múltiple	Migración de NK y macrófagos, interacciones entre linfocitos T y células dendríticas
CXCL1	Macrófagos, neutrófilos, endotelio	Neutrófilos	Reclutamiento de neutrófilos
IL8 (CXCL8)	Macrófagos, monocitos, fibroblastos, células epiteliales y endoteliales	Neutrófilos, otros	Quimioatrayente de neutrófilos, basófilos y linfocitos T, activación de neutrófilos, amplificación de la respuesta inflamatoria
CXCL9	Macrófagos, células dendríticas, fibroblastos, células endoteliales	Linfocitos T, NK, algunos tipos de células dendríticas	Migración de linfocitos T activados y NK al sitio de infección/inflamación
CXCL10	Monocitos, macrófagos, linfocitos T	Monocitos, NK, linfocitos T	Respuesta Th1, CD8, tráfico de NK

IL: Interleuquina; NK: células asesinas naturales; Th: linfocitos t efectores; IFN γ : interferón gamma; CCL5: ligando 5 de quimioquina con motivo C-C; RANTES: quimioquina de regulación por activación expresada y secretada por los linfocitos T (del inglés, *regulated on activation, normal T-cell expressed and secreted*); CXCL: ligando 10 de quimioquina con motivo C-X-C.

Asimismo, numerosos compuestos relacionados con la adhesión celular y la coagulación juegan roles críticos en la respuesta inflamatoria, la disfunción endotelial y la

trombosis, todas ellas características clave de la patogénesis del COVID-19. La activación y la disfunción del endotelio vascular son eventos importantes que contribuyen a la gravedad de la enfermedad, incluyendo el SDRA y las complicaciones tromboembólicas (208). La hipercoagulabilidad y la disfunción endotelial desempeñan un papel central en las formas graves de infecciones por COVID-19 (223). A pesar de que los mecanismos moleculares implicados no están del todo claros, se ha demostrado que las siguientes glucoproteínas están involucradas en las formas graves de COVID-19:

- **Factor de von Willebrand (vWF-A2):** Es una glucoproteína fundamental en la hemostasia, que facilita la adhesión de plaquetas al sitio de lesión endotelial. En COVID-19, se ha observado que los pacientes pueden desarrollar trastornos de la coagulación, como trombosis, y los niveles elevados de vWF-A2 pueden estar asociados con un mayor riesgo de eventos trombóticos (224, 225).
- **E-selectina (e-sel):** Es una molécula de adhesión que se expresa en las células endoteliales en respuesta a citoquinas proinflamatorias como IL-1 y el factor de necrosis tumoral alfa (TNF- α , por sus siglas en inglés, *tumor necrosis factor*). En COVID-19, la infección puede provocar una fuerte respuesta inflamatoria (tormenta de citoquinas), lo que lleva a una mayor expresión de e-sel y, por lo tanto, a la adhesión y migración de leucocitos hacia el tejido pulmonar afectado (226).
- **Molécula de adhesión intercelular 1 (ICAM-1, del inglés, *Intercellular Adhesion Molecule 1*):** Es una glucoproteína expresada en células endoteliales y del sistema inmune. Facilita la adhesión de leucocitos al endotelio y su migración hacia los sitios de inflamación. En COVID-19, la inflamación sistémica y la lesión endotelial inducen la expresión de ICAM-1, contribuyendo a la infiltración de células inmunes en los pulmones y otros órganos afectados (226, 227).
- **Molécula de adhesión vascular 1 (VCAM-1, del inglés, *vascular cell adhesion molecule-1*):** Similar a ICAM-1, VCAM-1 facilita la adhesión y migración de leucocitos. La elevada expresión de VCAM-1 en respuesta a la inflamación en COVID-19 contribuye al reclutamiento de células inflamatorias, exacerbando la respuesta inflamatoria y la lesión tisular (226, 227).
- **Trombospondina (TSP):** Es una proteína multifuncional que regula la angiogénesis y la inflamación. En COVID-19, puede estar involucrada en la remodelación vascular y en la respuesta inflamatoria. Se ha demostrado que un desequilibrio en la expresión de TSP se relaciona con la gravedad de los pacientes, aunque su papel específico en la enfermedad aún está en investigación (228).

- **Factor tisular (TF, del inglés, *tissue factor*):** Es una glucoproteína de membrana crucial en la iniciación de la coagulación. En COVID-19, se ha observado un estado procoagulante, y la expresión aumentada del factor tisular puede contribuir a la trombosis intravascular diseminada (TID) y a otros eventos trombóticos observados en pacientes graves. En conjunto, diversos estudios sugieren que el TF desempeña un papel central en la trombosis en COVID-19. El TF puede ser un marcador pronóstico útil y un objetivo terapéutico para reducir la trombosis y la inflamación (229, 230).
- **Trombomodulina (TM):** Es una glucoproteína con potente acción anticoagulante, presente en la membrana de las células del endotelio vascular. TM ejerce su acción al unirse a la trombina, formando el complejo trombina-trombomodulina, que tiene el efecto contrario al de la trombina. En COVID-19, la disfunción endotelial puede llevar a una disminución de TM, contribuyendo a un estado procoagulante y a la formación de trombos (223, 231).

Entre otros tipos de proteínas relacionadas con la respuesta inmunitaria y la citotoxicidad, encontramos las granzimas (Gzms). Se trata de una familia de proteasas de serina contenidas en gránulos citotóxicos de las células asesinas naturales (NK) y linfocitos T citotóxicos (CTL), capaces de inducir la muerte celular mediante la escisión de diferentes sustratos (232). De las diferentes Gzms descritas en humanos (A, B, H, K y M), **GzmA** y **GzmB** son las más expresadas y mejor caracterizadas (233). Aparte de su papel principal induciendo la muerte celular de las células diana (que en la actualidad se atribuye predominantemente a GzmB) las Gzms pueden ejercer funciones intracelulares independientemente de la muerte celular, como la inactivación de patógenos (234). Además, pueden liberarse al espacio extracelular ejerciendo actividades adicionales no citotóxicas, incluyendo la regulación de la inflamación y la remodelación de la matriz extracelular, alterando la permeabilidad vascular, entre otras (235).

La cascada de exocitosis de los gránulos implica la liberación de GzmA y GzmB, que ejercen principalmente funciones inflamatorias y citotóxicas, respectivamente (236). Se han observado niveles elevados de GzmA en múltiples condiciones de inflamación desproporcionada (237-239), así como en pacientes infectados con una amplia gama de virus, como el virus de Epstein-Barr, el citomegalovirus (CMV) o el Virus de la Inmunodeficiencia Humana (VIH) (240-242). Sin embargo, durante la infección por SARS-CoV-2, los niveles de expresión de GzmA varían entre los estudios, probablemente debido al momento del muestreo y la gravedad de la enfermedad. Algunos estudios señalan un aumento (243, 244), mientras que otros encuentran una reducción durante el COVID-19

(245, 246). Por otro lado, también se han encontrado niveles elevados de expresión de GzmB en las células NK de muestras de pacientes con COVID-19 (247, 248). Sorprendentemente, respecto a la expresión de GzmB en los CTLs durante la infección por SARS-CoV-2, algunos estudios han encontrado un aumento (244, 249), mientras que otros han mostrado una reducción (243, 250).

En cuanto a la gravedad del COVID-19, los pacientes graves o críticos muestran una disminución tanto de GzmA como de GzmB respecto a los pacientes leves (pero significativamente aumentada en comparación con los controles sanos) (243, 251, 252). Por el contrario, también se han demostrado expresiones significativamente superiores de GzmB en pacientes con COVID-19 grave en la tercera semana de enfermedad en comparación con los pacientes leves (253).

La combinación de estos factores solubles, junto con el momento (cinética) y su concentración (dinámica) guía las respuestas fisiológicas e interacciones celulares en COVID-19 (254).

1.3.2.2. Cambios en el metabolismo: Compuestos orgánicos volátiles expirados en aire

Los procesos biológicos complejos en el cuerpo humano, así como sus microbiomas, generan múltiples metabolitos y productos de degradación. La función respiratoria es una vía esencial para la excreción de gases y numerosos compuestos volátiles (255). Entre otros, los compuestos orgánicos volátiles (COVs) son moléculas gaseosas generadas continuamente por el cuerpo humano (256) y liberadas en el aire exhalado durante la respiración (257-259).

El aliento exhalado contiene una gran variedad de COVs cuyo origen proviene tanto de fuentes exógenas como endógenas. Tras su producción, ingesta o inhalación, los COVs se excretan al torrente sanguíneo, tras lo cual se difunden a los pulmones, donde se exhalan (260). La composición y concentración de los COVs es detectable y medible en muestras de aire expirado (261) mediante un procedimiento no invasivo. Los COVs detectados en aire exhalado incluyen alcoholes, ésteres de ácido carboxílico, cetonas, hidrocarburos alifáticos, hidrocarburos aromáticos, ácidos carboxílicos, terpenos y fenoles, entre otros (Tabla 4).

Introducción

Tabla 4. Descripción de la fuente, origen metabólico y metabolismo alterado de los COVs excretados en aire expirado. Los grupos de compuestos mostrados son los más estudiados en clínica. BHT: Butilhidroxitolueno.

Grupo	Fuente principal	Origen metabólico / Fuente de exposición	Metabolismo alterado
Alcoholes	Endógeno y exógeno (262)	<ul style="list-style-type: none"> ☞ Metabolismo microbiano de carbohidratos en el intestino delgado. El etanol puede ser producido especialmente por <i>S. aureus</i> (263) y <i>E. coli</i> (264). ☞ Pueden surgir en células de mamíferos a través de la reducción de aldehídos (265). ☞ Se generan endógenamente como subproductos del metabolismo de los hidrocarburos (266), que produce aldehídos o cetonas con subproductos alcohólicos (267). 	<ul style="list-style-type: none"> ☞ Exposición a factores ambientales estresantes, actividad humana, estados patológicos (268) y sus consecuencias, como alteraciones y disbiosis de la microbiota intestinal (269).
Ésteres de ácido carboxílico	Endógeno y exógeno (270, 271)	<ul style="list-style-type: none"> ☞ El acetato de etilo existe en todas las células eucariotas, desde las levaduras hasta los humanos (272). ☞ Fuentes externas: principalmente alimentos y plantas (272). 	<ul style="list-style-type: none"> ☞ Alteración de la microbiota gastrointestinal (273). ☞ Se han asociado a diversas enfermedades como hígado graso no alcohólico y a trastornos metabólicos congénitos, como la celiaquía (272).
Cetonas	Endógeno y exógeno (274, 275)	<ul style="list-style-type: none"> ☞ Se producen cuando la tasa de oxidación de los ácidos grasos se acelera debido a una modificación de las circunstancias metabólicas (276). ☞ La formación de acetona está relacionada con el metabolismo de la glucosa y el proceso de lipólisis (262) (se produce a partir de los depósitos de grasa cuando la glucosa no está disponible) (277). ☞ Varios compuestos presentes en el hígado pueden sufrir una descarboxilación no enzimática para producir acetona (278). ☞ Pueden producirse a través del metabolismo de las proteínas (278). ☞ Fuentes externas: alimentos o contaminantes ambientales (274, 275). 	<ul style="list-style-type: none"> ☞ Aumento de la β-oxidación de los ácidos grasos y del metabolismo proteico \rightarrow sobreproducción de cetonas (277). ☞ Disbiosis de la microbiota intestinal (279, 280) \rightarrow intestino permeable \rightarrow aumento de la captación de glucosa \rightarrow glucólisis hepática y celular \rightarrow producción de acetona (281)
Hidrocarburos alifáticos saturados (alcanos)	Endógeno y exógeno (282, 283)	<ul style="list-style-type: none"> ☞ Pueden producirse por oxidación de proteínas o proceder de la flora intestinal (262). ☞ Los alcanos de cadena larga se producen como consecuencia del estrés oxidativo (284). ☞ Inhalación de vapores de combustibles o de la industria química (282). 	<ul style="list-style-type: none"> ☞ Los compuestos que contienen decano tienen propiedades antivirales (285) \rightarrow pueden reflejar un componente de la respuesta inmune del hospedador a la infección viral (286). ☞ La infección viral regula positivamente el decano producido por las células epiteliales de las vías respiratorias (286). ☞ El decano y otros compuestos alcanos de cadena larga se correlacionan con la magnitud de la respuesta inmune antiviral (286).
Hidrocarburos aromáticos	Endógeno y exógeno (277, 287)	<ul style="list-style-type: none"> ☞ Algunos derivan de la peroxidación lipídica y los procesos inflamatorios (288). ☞ Se han descrito compuestos como el fenol y el tolueno en <i>E. coli</i> y <i>S. aureus</i> (262). ☞ El estireno es una toxina ambiental conocida y se ha relacionado con daño en el ADN (277). 	<ul style="list-style-type: none"> ☞ Se ha informado ampliamente de su presencia en una variedad de afecciones respiratorias (288).
Ácidos carboxílicos	Endógeno (262)	<ul style="list-style-type: none"> ☞ Son los principales subproductos del metabolismo anaeróbico (262). Se forman durante la fermentación bacteriana de carbohidratos y son especialmente abundantes en el intestino humano (289). 	<ul style="list-style-type: none"> ☞ Los patógenos que producen principalmente ácido acético son <i>S. aureus</i> (263). ☞ Disbiosis de la microbiota intestinal (279).
Terpenos	Exógeno (290)	<ul style="list-style-type: none"> ☞ Son producidos por varios tipos de plantas, en especial las coníferas. Componentes de aceites esenciales. ☞ La inhalación es la fuente exógena principal. ☞ Pueden ser absorbidos a través del consumo de alimentos y bebidas que los contienen (291). ☞ Tienen propiedades anticancerígenas, antimicrobianas, antiinflamatorias, antioxidantes y antiulcerígenas (292). 	<ul style="list-style-type: none"> ☞ Niveles influenciados por los hábitos nutricionales. ☞ El limoneno es un antioxidante \rightarrow bajos niveles encontrados en pacientes con enfermedades respiratorias por su consumo acelerado o catabolización en respuesta al estrés oxidativo (290).
Fenoles	Exógeno (290)	<ul style="list-style-type: none"> ☞ Aditivo en numerosos procesos industriales (producción de alimentos, cosméticos, fármacos, plásticos...). ☞ Se detecta en el aire de ambientes interiores. ☞ Tienen propiedades antioxidantes (290). 	<ul style="list-style-type: none"> ☞ Altos niveles de estrés oxidativo en pacientes con enfermedades respiratorias \rightarrow agotamiento de BHT o catabolización a otro compuesto (290).

Se ha demostrado que la composición del perfil de COVs cambia como consecuencia de perturbaciones metabólicas en todo el cuerpo o en tejidos específicos como resultado de varias condiciones y procesos patológicos (293). Los COVs endógenos se producen por procesos biológicos que incluyen el estrés oxidativo y la inflamación en el cuerpo humano (294, 295). Cuando estos mecanismos se ven alterados, inducen variaciones en la composición de los COVs excretados por el órgano afectado y, en consecuencia, en el perfil de COVs exhalado.

Algunos COVs también se originan como productos metabólicos de la microbiota normal del organismo, así como por el metabolismo de microorganismos invasores y patógenos (296). Durante el transcurso de la enfermedad, diferentes patógenos causantes e interacciones huésped-patógeno pueden cambiar la composición de los COVs que se encuentran en el aliento exhalado, lo que proporciona un objetivo potencial para el diagnóstico (260, 279).

Múltiples enfermedades, entre las que se incluyen las infecciones respiratorias y en especial el COVID-19, causan disbiosis de la flora intestinal y alteran la integridad de la barrera intestinal, lo que conduce a un intestino permeable (297, 298). Tanto los cambios en la microbiota como sus consecuencias pueden alterar numerosas rutas metabólicas, y por lo tanto la producción de ciertos metabolitos, incluidos los expirados en aire (269).

Todas estas alteraciones generan variaciones específicas, distintivas y medibles en el patrón de COVs procedente del aliento exhalado (299). Estudios previos han demostrado que existen perfiles o “huellas” de COVs característicos de diferentes enfermedades, desde infecciones pulmonares hasta cáncer (279, 300-302). En el caso de las infecciones por SARS-CoV-2, también se ha estudiado el perfil de diferentes COVs expirados en aire con fines diagnósticos. En la tabla 5 se muestra un resumen de los COVs más utilizados en la literatura, sobre todo para el diagnóstico de enfermedades respiratorias.

Tabla 5. Principales COVs estudiados en el diagnóstico de diversas enfermedades, mayoritariamente respiratorias, entre las que se destaca el COVID-19. Las flechas indica que, en la enfermedad señalada, los pacientes tienen aumentada (↑) o disminuida (↓) la concentración de ese compuesto en aire exhalado, en comparación con individuos sanos. El símbolo ≠ indica que la concentración del compuesto ha mostrado diferencias significativas entre pacientes enfermos y sanos, o se ha utilizado con éxito para distinguir pacientes, pero no se indica en la literatura si aumenta o disminuye. *: Estudios en los que los controles fueron pacientes con otras enfermedades respiratorias no-COVID-19.

Grupo	Compuesto	↑	↓	≠
Alcoholes	Etanol	<ul style="list-style-type: none"> COVID-19 (303-305) Cáncer de pulmón (277, 306) 		

Introducción

Grupo	Compuesto	↑	↓	≠
		<ul style="list-style-type: none"> • NAVM (260) • Esteatosis hepática severa (307) 		
	Isopropanol	<ul style="list-style-type: none"> • COVID-19 (269, 308, 309) 	<ul style="list-style-type: none"> • COVID-19* (310) 	
	Butanol	<ul style="list-style-type: none"> • COVID-19 (303) 		
Ésteres de ácido carboxílico	Acetato de etilo	<ul style="list-style-type: none"> • COVID-19 (311) 		
	Acetato de propilo	<ul style="list-style-type: none"> • Infecciones virales (influenza, rinovirus) (299, 312) 		
Cetonas	Acetona	<ul style="list-style-type: none"> • COVID-19 (269, 313) • EPOC (314) 	<ul style="list-style-type: none"> • COVID-19 (309, 315) • Cáncer de pulmón (316) 	<ul style="list-style-type: none"> • Cáncer de pulmón (277)
	2-butanona	<ul style="list-style-type: none"> • COVID-19 (303, 313, 317) 		
Hidrocarburos alifáticos saturados (alcanos)	Hexano	<ul style="list-style-type: none"> • COVID-19 (311) • Asma (318) 		
	<i>i</i> -octano	<ul style="list-style-type: none"> • Colitis ulcerosa (319) 	<ul style="list-style-type: none"> • COVID-19* (320) 	<ul style="list-style-type: none"> • Cáncer de pulmón (277) • SDRA (321)
	Decano	<ul style="list-style-type: none"> • COVID-19 (322, 323) • EPOC (290, 324) • Cáncer de pulmón (325) 		<ul style="list-style-type: none"> • COVID-19* (320) • EPOC + Rinovirus (286) • Tuberculosis (299)
Hidrocarburos aromáticos	Tolueno	<ul style="list-style-type: none"> • Cáncer de pulmón (326) 	<ul style="list-style-type: none"> • COVID-19 (327) 	<ul style="list-style-type: none"> • AI crónica (328)
	Etilbenceno	<ul style="list-style-type: none"> • COVID-19 (323) • NAVM (260) 		<ul style="list-style-type: none"> • Cáncer de pulmón (277) • AI crónica (328)
	<i>o</i> -xileno	<ul style="list-style-type: none"> • Tuberculosis (en orina) (329) 		
	Estireno			<ul style="list-style-type: none"> • Influenza (300) • Cáncer de pulmón (277)
Ácidos carboxílicos	Ácido acético	<ul style="list-style-type: none"> • Colitis ulcerosa (319) • Fibrosis quística (330) • EPOC (331) 		<ul style="list-style-type: none"> • COVID-19 (315) • Enfermedad de Crohn (280)
	Ácido propanoico	<ul style="list-style-type: none"> • Síndrome de intestino irritable (280) 		
Terpenos	Alfa-pineno	<ul style="list-style-type: none"> • Colitis ulcerosa (319) • Malaria (332) 		<ul style="list-style-type: none"> • Función pulmonar deteriorada (333)
	Limoneno	<ul style="list-style-type: none"> • Cirrosis y encefalopatía hepáticas (334) 	<ul style="list-style-type: none"> • EPOC (290) 	<ul style="list-style-type: none"> • AI crónica (328)
	3-careno	<ul style="list-style-type: none"> • Malaria (332) 		
Fenoles	Butilhidroxitolueno		<ul style="list-style-type: none"> • EPOC (290) 	

NAVM: neumonía asociada a ventilación mecánica; EPOC: Enfermedad pulmonar obstructiva crónica; SDRA: Síndrome de dificultad respiratoria aguda; AI: Aspergilosis invasiva.

El análisis metabolómico de los COVs en el aliento exhalado es una alternativa no invasiva prometedora a los métodos actuales para diagnosticar las LRI (335). En general, las combinaciones de varios biomarcadores en un perfil de COVs o en huellas respiratorias obtenidas con dispositivos electrónicos demuestran una elevada sensibilidad y especificidad (336).

1.3.3. Diagnóstico de COVID-19

La pandemia de COVID-19 ha presentado un desafío sin precedentes para los sistemas de salud a nivel mundial. Hoy en día, el diagnóstico de COVID-19 puede implicar criterios clínicos, pruebas de laboratorio o radiológicas (337) (Figura 13).

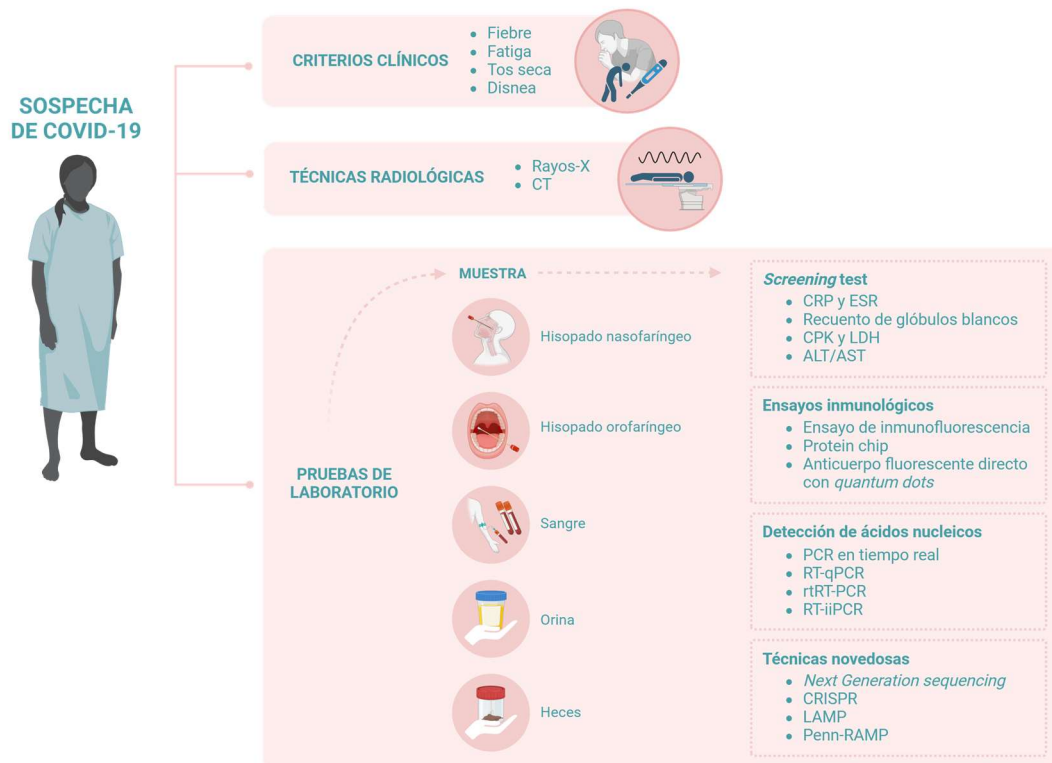


Figura 13. Diferentes técnicas actualmente disponibles para el diagnóstico de COVID-19 (338, 339).

Las pruebas de laboratorio incluyen estudios de diagnóstico, como la detección de ARN, antígenos, o anticuerpos del SARS-CoV-2. Otras pruebas de laboratorio estudian cambios en el recuento de glóbulos blancos, niveles elevados de lactato deshidrogenasa y niveles elevados de ferritina (340). El examen radiológico incluye principalmente la TC y examen de rayos X (337). A continuación, se detallan las técnicas más utilizadas.

☼ **Métodos de diagnóstico basados en la detección de ácidos nucleicos:** En la actualidad, el diagnóstico clínico de COVID-19 implica la detección de ácidos nucleicos en muestras nasofaríngeas y orofaríngeas mediante la reacción en cadena de la polimerasa cuantitativa con transcripción inversa (RT-qPCR). Debido a su alta especificidad, se ha utilizado como prueba universal para el diagnóstico de COVID-19. Sin embargo, a veces, debido a la contaminación de la muestra o a fallos técnicos, se pueden obtener resultados FNs. También cabe destacar que el procedimiento a seguir para realizar la prueba puede aumentar el riesgo de exposición a gotitas virales procedentes de las muestras (341-343).

☼ **Pruebas serológicas e inmunológicas:** Se han desarrollado varias pruebas de diagnóstico para detectar COVID-19 en pacientes que están o estuvieron infectados. Estas pruebas se basan en el análisis de muestras de sangre para detectar la presencia de anticuerpos IgM e IgG. Los anticuerpos IgM proporcionan una indicación

de infección temprana, mientras que los anticuerpos IgG son indicativos de infección actual o previa (344, 345). Esta prueba, que proporciona una evaluación de las respuestas de anticuerpos a corto y largo plazo, así como de la abundancia y diversidad de anticuerpos, desempeña un papel importante en la comprensión de la epidemiología del COVID-19 y en el desarrollo de vacunas contra la enfermedad. Además, se caracteriza por ser un método sencillo y rápido, por lo que es un procedimiento de alto rendimiento para diagnosticar infecciones virales (346, 347).

☼ **Métodos de detección de anticuerpos de unión:** La proteína de la espícula purificada del SARS-CoV-2 se puede utilizar para técnicas de detección de unión a anticuerpos. Estas pruebas se pueden realizar rápidamente utilizando reactivos específicos. En comparación con los métodos de detección molecular, los métodos de detección de anticuerpos tienen las ventajas de una respuesta rápida, bajo coste y amplia disponibilidad (348). La detección de anticuerpos de unión se divide en dos categorías principales. Una de ellas son las pruebas en el lugar de atención, en las que las muestras de sangre obtenidas de hisopos dactilares o hisopos nasofaríngeos y orofaríngeos se procesan rápidamente para detectar anticuerpos totales o anticuerpos IgG e IgM en muestras de saliva, plasma, suero o sangre total, y pueden ser utilizadas para pruebas de emergencia de SARS-CoV-2 en escuelas, hospitales y centros de pruebas (349). Sin embargo, la validez y eficiencia de estas pruebas son bajas y, a menudo, se necesita una validación secundaria mediante métodos basados en RT-qPCR. La segunda categoría de detección son las pruebas de laboratorio, que incluyen ensayos inmunoabsorbentes ligados a enzimas (ELISA), inmunoensayos de quimioluminiscencia e inmunoensayos de flujo lateral. Estas pruebas son muy sensibles y rápidas, pero requieren técnicos altamente capacitados; por lo tanto, a menudo también requieren una validación secundaria (350).

☼ **Exploración por TC de tórax:** La TC es un método de diagnóstico ampliamente utilizado para COVID-19, especialmente en la etapa inicial (351). Los pacientes con COVID-19 presentan síntomas típicos de neumonía en las primeras etapas de la infección. En algunos pacientes, los hisopos nasofaríngeos y orofaríngeos y las muestras de heces daban positivo para SARS-CoV-2, mientras que las muestras de sangre resultaban negativas (352). Sin embargo, en estos casos, los hallazgos resultantes de TCs de tórax proporcionaban un diagnóstico definitivo de COVID-19, incluyendo infiltrados bilaterales, opacidades en parches o en vidrio deslustrado (áreas de los pulmones que están parcialmente llenas de líquido o inflamación, pero no de forma tan densa como

los infiltrados) (80%) y lesiones subpleurales intrapulmonares unilaterales (20%), que pueden ayudar a confirmar el diagnóstico de COVID-19 (337).

Una vez que se ha diagnosticado COVID-19 positivo, se debe clasificar al paciente con una herramienta de triaje y evaluación estandarizada para valorar la **gravedad** de la enfermedad. Desde el inicio de la pandemia se han desarrollado e implementado diferentes métodos para estratificar la gravedad de los pacientes, creando diversas escalas y scores para clasificarlos en base a los resultados de las pruebas de laboratorio, la observación de los síntomas y signos vitales, y la consideración de los factores de riesgo (353). Esta evaluación permite a los facultativos tomar decisiones informadas sobre el tratamiento, como la administración de oxígeno, el uso de antivirales, corticosteroides, o la necesidad de ingreso en la UCI. Habitualmente se siguen guías de práctica clínica locales para la prevención y control de infecciones (354). A lo largo de la pandemia, importantes instituciones, tanto a nivel internacional como nacional, han establecido diferentes clasificaciones de enfermedad: COVID-19 leve, moderado, grave y crítico (Tabla 6). Cabe destacar que, en general, uno de los criterios más importantes para la estratificación de la gravedad de los pacientes es la necesidad de aporte de oxígeno.

Tabla 6. Criterios para el diagnóstico de la gravedad de la enfermedad en pacientes con COVID-19 según las pautas de la Organización Mundial de la Salud (OMS) (354), los Institutos Nacionales de Salud (NIH, por sus siglas en inglés, *National Institutes of Health*) (355), la Sociedad de Enfermedades Infecciosas de América (IDSA, del inglés, *Infectious Diseases Society of America*) (356) y el Sistema Nacional de Salud (SNS) siguiendo las pautas establecidas por el Ministerio de Sanidad, Gobierno de España (357, 358).

Gravedad de la enfermedad	OMS	NIH	IDSA	SNS España
Leve	Sintomáticos sin evidencia de hipoxia o neumonía	Sintomáticos sin dificultad para respirar, disnea o imágenes anormales de tórax	SpO ₂ >94%, no requieren oxigenoterapia	Sin signos de gravedad. SpO ₂ >90%. CURB65 ≤1
Moderada	Signos clínicos de neumonía (no grave), SpO ₂ >90%	Evidencias de enfermedad respiratoria inferior y SpO ₂ >94%		
Grave	Signos clínicos de neumonía (fiebre, tos, disnea, respiración rápida) + uno de los siguientes: - Frecuencia respiratoria > 30 respiraciones/minuto - Dificultad respiratoria grave - SpO ₂ <90%	Frecuencia respiratoria >30 respiraciones por minuto, SpO ₂ ≤94%, PaO ₂ /FiO ₂ <300, o infiltraciones pulmonares >50%	SpO ₂ ≤94%, incluidos los pacientes con oxigenoterapia	Fallo de ≥1 órgano o SpO ₂ <90% o frecuencia respiratoria de ≥30
Crítica	SDRA, sepsis, shock séptico, trombosis aguda o síndrome inflamatorio multisistémico en niños	Insuficiencia respiratoria, shock séptico y/o disfunción de múltiples órganos	Ventilación mecánica y ECMO. Disfunción orgánica, sepsis/shock séptico, SDRA	Insuficiencia respiratoria, SDRA, sepsis, shock séptico. Respuesta inflamatoria excesiva. Necesidad de ventilación mecánica u otro criterio de necesidad de UCI

SpO₂: porcentaje de saturación de oxígeno de la hemoglobina en sangre arterial; ECMO: Oxigenación por membrana extracorpórea (del inglés, *Extracorporeal Membrane Oxygenation*); SDRA: síndrome de dificultad respiratoria aguda; UCI: Unidad de Cuidados Intensivos; PaO₂/FiO₂: relación de la presión parcial arterial de oxígeno a la fracción de oxígeno inspirado.

Tras establecer el diagnóstico y la gravedad inicial de la enfermedad, es fundamental considerar el pronóstico, que permite anticipar la evolución clínica del paciente y evaluar la probabilidad de que desarrolle formas más severas de COVID-19.

1.3.4. Pronóstico de COVID-19

El pronóstico del COVID-19, es decir, la estimación de la gravedad y el desenlace clínico de la enfermedad ha sido objeto de intensa investigación desde el inicio de la pandemia y sigue siendo un área de principal preocupación para la comunidad científica contemporánea. Uno de los aspectos más críticos en la gestión de las infecciones respiratorias, y especialmente del COVID-19, ha sido la capacidad de predecir su evolución en pacientes individuales. A pesar de que el COVID-19 ha sido estudiado exhaustivamente, hoy en día todavía es impredecible cómo se manifestará la enfermedad en un paciente en particular (359). Cada vez hay más evidencias que demuestran que las manifestaciones del COVID-19 sintomático pueden variar ampliamente. La mayoría de los pacientes manifiestan síntomas clínicos leves o moderados. Sin embargo, la enfermedad sigue un curso incierto y puede progresar a una enfermedad grave, como insuficiencia respiratoria o enfermedades sistémicas potencialmente mortales (360).

Los métodos utilizados para pronosticar la evolución del COVID-19 son variados y han evolucionado con el tiempo a medida que se ha profundizado en la comprensión de la enfermedad. Además, se tienen en cuenta factores de riesgo que pueden predisponer a una evolución más grave (361). Estos factores pronósticos clave (Figura 14) se utilizan para estratificar a los pacientes y determinar el nivel de vigilancia y las intervenciones necesarias.

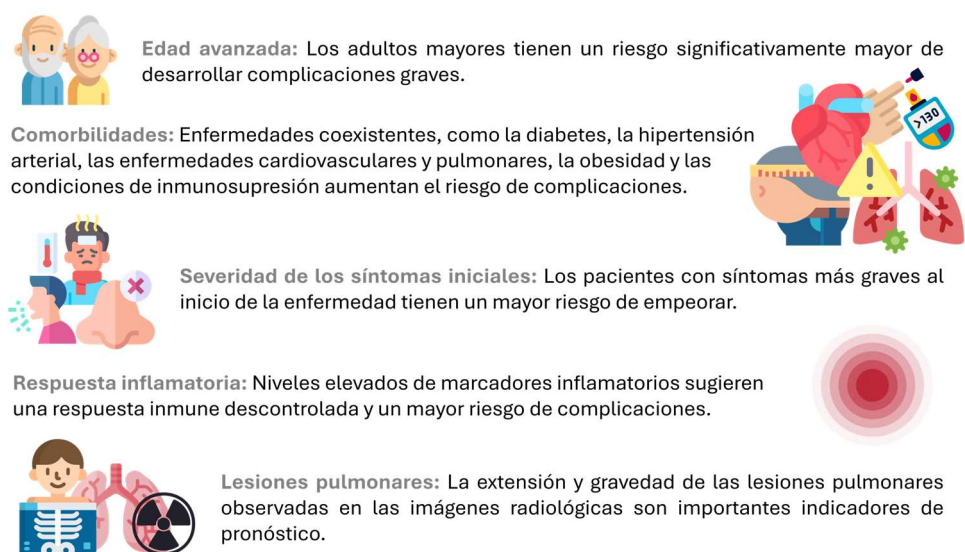


Figura 14. Factores pronósticos clave que se han asociado con un peor pronóstico en pacientes con COVID-19 (353, 358, 362-364). Iconos de Flaticon.

En la actualidad, el pronóstico del COVID-19 se realiza mediante una combinación de pruebas diagnósticas, evaluaciones clínicas y análisis de datos epidemiológicos. Los médicos y especialistas utilizan diversas herramientas para evaluar la probabilidad de que un paciente evolucione hacia una forma grave de la enfermedad o, en cambio, se recupere con síntomas leves. Entre las herramientas más comunes se encuentran:

- ☼ **Evaluación clínica:** Se basa en la evaluación de los síntomas del paciente. Los síntomas más comunes del COVID-19 son fiebre, tos, dificultad para respirar, fatiga y pérdida del olfato y/o del gusto (365). También se consideran factores como la edad, el sexo, el índice de masa corporal (IMC) y la presencia de comorbilidades (otras enfermedades coexistentes). El enfoque para evaluar el pronóstico de la enfermedad se centra en la identificación de signos de complicaciones, como la dificultad respiratoria, saturación de oxígeno baja y alteraciones en los niveles de consciencia. La principal causa de muerte es la insuficiencia respiratoria por el SDRA (366).
- ☼ **Pruebas de laboratorio:** El diagnóstico inicial del COVID-19 se basa en pruebas de laboratorio explicadas en el punto 1.2.2. Estas pruebas son fundamentales para confirmar la infección y determinar el manejo clínico del paciente. En cuanto al estudio del pronóstico relacionado con la función pulmonar y renal del paciente, se miden los niveles de marcadores inflamatorios, como la proteína C reactiva (CRP, por sus siglas en inglés, *C-reactive protein*) y la IL-6. Tanto en el COVID-19 grave como en otras enfermedades críticas, la fisiopatología inflamatoria es uno de los principales contribuyentes a la muerte (209, 367).
- ☼ **Criterios radiológicos:** Las radiografías y las TCs de tórax permiten visualizar las lesiones pulmonares causadas por el virus, revelando así la extensión del daño pulmonar, lo que ayuda a prever complicaciones graves (368, 369).
- ☼ **Marcadores de gravedad:** Existen diferentes marcadores que han demostrado ser indicadores de un pronóstico desfavorable (358). La carga viral de SARS-CoV-2 se ha identificado como un marcador pronóstico independiente de mortalidad (370). Por otro lado, se ha demostrado que el síndrome hiperinflamatorio, característico tanto de COVID-19 como de otras infecciones, conduce al deterioro de los pacientes (215, 371, 372). Se ha observado una elevación progresiva de los niveles de neutrófilos, indicadores de inflamación y de daño miocárdico con la progresión de la enfermedad (373). Los perfiles elevados de determinados biomarcadores característicos de la tormenta de citoquinas se correlacionan directamente con la inflamación pulmonar,

la lesión tisular y el mal pronóstico de los casos graves de COVID-19 (374, 375). En consecuencia, la mayoría de los marcadores investigados en el pronóstico de COVID-19 pertenecen a las vías inmunoinflamatorias y de coagulación (359). Diferentes estudios demuestran elevaciones de indicadores inflamatorios en sangre como la IL-6 (374, 376-379), IL-2 (380-382), IL-4 (382), IL-7 (382), IL-8 (376, 377, 383), IL-10 (374, 384), IL-12 (382), IL-17 (382), IL-18 (197), TNF- α (376, 380, 384), ligando 10 de quimioquina con motivo C-X-C (CXCL10) (377, 382), CXCL9 (385), interferón gamma (IFN- γ) (382), CRP (386) y dímero-D (387, 388), entre otros.

☀ **Modelos predictivos:** Para anticipar la evolución de la enfermedad, se han desarrollado numerosos modelos estadísticos y de aprendizaje automático para predecir el riesgo de desarrollar una enfermedad grave o de fallecer a partir de datos clínicos, de laboratorio y radiológicos. Estos modelos utilizan algoritmos complejos para identificar patrones en los datos y hacer predicciones (252, 323, 389).

La predicción del pronóstico es muy relevante en la práctica clínica, ya que permite identificar de manera precoz aquellos pacientes con peor pronóstico y actuar en consecuencia. Diferentes escalas o *scores* como COVID-GRAM (390), NEWS2 (*National Early Warning Score 2*) (391, 392), MuLBSTA (393), CURB65 (394), PSI (395), SOFA (*Sequential Organ Failure Assessment*) (396, 397), qSOFA (*Quick SOFA*) (398), SIRS (399) o SCAP (400) han sido validados en pacientes con COVID-19 y han demostrado ser competentes predictores de la gravedad de los pacientes. La capacidad de detectar signos tempranos de deterioro clínico influye en las decisiones sobre el tratamiento terapéutico adecuado, el ingreso en UCI, el uso de medidas avanzadas de soporte vital, o, incluso para los casos más graves, en la planificación de cuidados paliativos. Este enfoque permite a los profesionales de la salud manejar de manera más efectiva los casos de COVID-19, reduciendo la mortalidad y mejorando los resultados clínicos en los pacientes afectados.

A pesar de los avances en el manejo del COVID-19, el pronóstico sigue presentando desafíos significativos. La variabilidad en la respuesta de los pacientes, la aparición de nuevas variantes del virus y la presencia de comorbilidades dificultan la predicción precisa de los resultados clínicos (401). Actualmente, los modelos pronósticos pueden verse limitados por la falta de datos personalizados y precisos en tiempo real, lo que puede afectar la capacidad de los profesionales de la salud para prever complicaciones graves y adaptar los tratamientos de manera óptima (402).

Existe una necesidad urgente de mejorar las herramientas de diagnóstico y pronóstico de las infecciones respiratorias mediante el uso de tecnologías avanzadas como la inteligencia artificial (IA) y el análisis de grandes volúmenes de datos, para proporcionar predicciones más exactas y permitir una intervención temprana y personalizada.

1.4 *Machine learning* en diagnóstico y pronóstico clínico

Impulsados por un aumento en la potencia computacional, almacenamiento, memoria y la generación de abundantes volúmenes de datos, los ordenadores se están utilizando para realizar una amplia gama de tareas complejas con una precisión impresionante. El *Machine Learning* (ML) o aprendizaje automático es una disciplina del campo de la IA que, a través de algoritmos, dota a los ordenadores de la capacidad de identificar patrones en datos masivos y elaborar predicciones (análisis predictivo). Se emplea tanto en el ámbito académico como en la industria para impulsar el desarrollo de "productos inteligentes" con la capacidad de realizar predicciones precisas utilizando diversas fuentes de datos (403). En el campo médico, el diagnóstico y la predicción de resultados pueden beneficiarse de la aplicación de técnicas de ML (404), ya que permite el análisis de conjuntos de datos grandes y complejos y, por lo tanto, tiene el potencial de mejorar la atención médica (405).

1.4.1 Preparación de los datos

En aplicaciones de ML, es fundamental la integración, la anotación, la preparación y el procesamiento de los datos recopilados. El preprocesamiento y la limpieza de datos son tareas importantes que se deben llevar a cabo para que un conjunto de datos se pueda utilizar para el entrenamiento de modelos. Los datos sin procesar son a menudo ruidosos, no confiables e incompletos. El uso de estos datos para el modelado puede producir resultados engañosos (406).

Se recopilan datos del mundo real de varios orígenes y procesos y pueden contener irregularidades que comprometen la calidad del conjunto de datos. Los problemas de calidad de datos más habituales que surgen son:

- ❁ **Incompletos:** en los datos no hay atributos o contienen valores que faltan.
- ❁ **Ruidosos:** los datos contienen registros erróneos o valores atípicos.
- ❁ **Incoherentes:** los datos contienen discrepancias o registros en conflicto.

Los datos de calidad son un requisito previo para desarrollar modelos predictivos de calidad. Para evitar la entrada y salida de elementos no utilizados y mejorar la calidad de los datos y, por tanto, el rendimiento del modelo, es fundamental llevar a cabo un mantenimiento de datos para detectar problemas de datos al principio del estudio y decidir acerca de los pasos de limpieza y preprocesamiento de datos correspondientes (407). Algunas de las tareas principales de preprocesamiento de datos (408) son:

- ❁ **Limpieza de datos:** rellenar los valores incompletos, detectar y eliminar los valores atípicos y los datos con ruido.
- ❁ **Transformación de datos:** normalizar datos para reducir el ruido y las dimensiones. Esta transformación de datos solo es necesaria para aplicar algunos tipos de algoritmos, por ejemplo, para *K-Nearest Neighbors* (KNN). La normalización de datos escala los valores numéricos a un intervalo especificado. Entre los métodos de normalización de datos más conocidos se incluyen:
 - ❁ **Normalización mínima-máxima:** transformar linealmente los datos a un intervalo, por ejemplo, entre 0 y 1, donde el valor mínimo se escala a 0 y el máximo a 1.
 - ❁ **Normalización de puntuación z o estandarización z-score:** escalar los datos en función de la desviación estándar (SD) y la media, dividiendo la diferencia entre los datos y la media por la SD:

$$z = \frac{(x - \mu)}{\sigma}, \text{ siendo } x \text{ el dato original, } \mu \text{ la media de los datos y } \sigma \text{ la desviación estándar.}$$

Aunque la normalización se usa tradicionalmente para la clasificación de algoritmos como KNN, puede que no siempre sea la forma más adecuada de cambiar la escala de las entidades. Debido a que los valores estandarizados z-score no tienen un mínimo y un máximo predefinidos, los valores extremos no se comprimen hacia el centro. Uno podría sospechar que, con una enfermedad infecciosa, podríamos ver algunos valores atípicos muy extremos. Por lo tanto, podría ser razonable permitir que los valores atípicos tengan una mayor ponderación en el cálculo de la distancia. La media de una variable estandarizada siempre debe ser cero y el rango debe ser bastante compacto. Una puntuación z superior a 3 o inferior a -3 indica un valor extremadamente raro.

- ❁ **Escalado decimal:** escala los datos moviendo la coma decimal del valor del atributo.
- ❁ **Reducción de datos:** atributos o registros de datos de ejemplo para un control de datos más sencillo.
- ❁ **Discretización de datos:** convertir atributos continuos en atributos de categorías para facilitar su uso con determinados métodos de aprendizaje automático.

- ❁ **Limpieza de texto:** eliminar caracteres incrustados que puedan ocasionar errores en la alineación de los datos, por ejemplo, pestañas incrustadas en un archivo de datos separado por tabulaciones o nuevas líneas incrustadas que pueden dividirse en registros.

Para tratar los valores incompletos, es recomendable identificar el motivo por el que no existen y así poder abordar mejor el problema. Los métodos típicos de control de datos incompletos son:

- ❁ **Eliminación:** descartar los registros cuyos datos estén incompletos.
- ❁ **Sustitución ficticia:** reemplazar los valores no registrados por un valor ficticio; por ejemplo, “desconocido” para categorías o 0 para valores numéricos.
- ❁ **Sustitución media:** si los datos que faltan son numéricos, reemplazar los valores que faltan por la media.
- ❁ **Sustitución frecuente:** si los datos que faltan son de categoría, cambiar los valores que faltan por el elemento más frecuente.
- ❁ **Sustitución de regresión:** reemplazar los valores ausentes por valores obtenidos con métodos de regresión (409).

1.4.2 Algoritmos de *machine learning* para crear modelos de clasificación

Los modelos de clasificación son algoritmos que se utilizan en informática para analizar conjuntos complejos de puntos de datos y categorizarlos adecuadamente en diferentes clases (410). Los tipos de implementación de ML pueden englobarse en tres categorías diferentes dependiendo de la naturaleza de los datos que recibe:

- ❁ **Aprendizaje supervisado:** necesita conjuntos de datos etiquetados, es decir, le decimos al modelo qué es lo que queremos que aprenda.
- ❁ **Aprendizaje no supervisado:** trabaja con datos que no han sido etiquetados. No existe una etiqueta que predecir. Estos algoritmos se usan principalmente en tareas donde es necesario analizar los datos para extraer un nuevo conocimiento o agrupar entidades por afinidad.
- ❁ **Aprendizaje por refuerzo:** es un método de aprendizaje automático que se basa en recompensar los comportamientos deseados y penalizar los no deseados. Aplicando este método, un agente es capaz de percibir e interpretar el entorno, ejecutar acciones y aprender a través de prueba y error. Es un aprendizaje que fija objetivos a largo plazo para obtener una recompensa general máxima y lograr una solución óptima (411).

Introducción

Los modelos desarrollados con técnicas de aprendizaje supervisado “aprenden” las propiedades de las clases y son capaces de clasificar automáticamente un nuevo ejemplo. Cada técnica de ML emplea diferentes algoritmos de aprendizaje para construir modelos con buena capacidad de generalización, es decir, modelos que predican con precisión las etiquetas de clase de registros previamente desconocidos. Generalmente, la construcción del modelo de clasificación se realiza dividiendo el conjunto de datos disponible en un conjunto de entrenamiento, que se utilizará en la fase de construcción del clasificador, y un conjunto de prueba para la validación del modelo (412).

En el ámbito del diagnóstico clínico, es interesante hablar del aprendizaje supervisado, ya que los algoritmos trabajan con datos etiquetados, intentando encontrar una función que, dadas las variables de entrada, les asigne la etiqueta de salida adecuada. Esta etiqueta es el parámetro que nos interesa predecir; por ejemplo, el diagnóstico o el pronóstico real del paciente. El algoritmo se entrena con un set de datos y “aprende” a asignar la etiqueta de salida adecuada a un nuevo ejemplo, es decir, predice el valor de salida (por ejemplo, predice el diagnóstico de un nuevo paciente) (413).

Los modelos creados mediante aprendizaje supervisado pueden clasificarse en **paramétricos** y **no paramétricos**, dependiendo de cómo esos modelos representan la relación entre las variables de entrada y salida, y cómo manejan los datos en el proceso de aprendizaje. Los algoritmos paramétricos (como por ejemplo las redes neuronales artificiales (ANN, por sus siglas en inglés, *Artificial Neural Networks*) y la regresión logística (RL)) dependen de un conjunto fijo de parámetros que se ajustan durante el proceso de entrenamiento. Es decir, el modelo tiene una estructura o forma funcional predefinida (por ejemplo, una línea recta en la regresión lineal). Estos algoritmos paramétricos suelen hacer ciertas suposiciones sobre la distribución de los datos o la relación entre las variables. Por el contrario, los algoritmos no paramétricos (como los árboles de decisión, Support Vector Machines (SVM) y KNN) no hacen suposiciones sobre la distribución de probabilidad de los datos de la muestra. Es una forma de análisis en el que el predictor no tiene una forma predeterminada, sino que se construye de acuerdo con la información derivada de los datos. Los algoritmos no paramétricos resultan más flexibles y dan generalmente mejor resultado, a costa de requerir más datos para su entrenamiento y resultando más lentos que sus equivalentes paramétricos. Sin embargo, son más proclives al sobreentrenamiento y más difíciles de interpretar (414).

A continuación, se detallan las características de los modelos de ML utilizados en este trabajo.

1.4.2.1 Regresión logística

La RL es un algoritmo de clasificación que se utiliza principalmente para describir datos y explicar la relación entre una variable dependiente categórica (binaria, multinomial u ordinal) y una o más variables independientes que pueden ser nominales, ordinales, intervalos o escalas. En base a los datos de entrada, el algoritmo puede predecir si una nueva instancia pertenece a una de varias clases (415).

El modelo de RL “aprende” la relación entre las características de entrada y sus etiquetas de clase correspondientes. Se puede utilizar este tipo de modelos para determinar la probabilidad de que una muestra dada pertenezca a una de dos o más clases posibles.

El resultado de la ejecución de este algoritmo es un resultado de probabilidad: obtenemos un porcentaje de probabilidad de que esa determinada instancia pertenezca a cada clase. En la clasificación binaria, la regresión logística asigna los datos de entrada a una probabilidad entre 0 y 1, lo que indica la probabilidad de uno de dos resultados posibles. Por ejemplo: si existe un 10 % de probabilidad de que los valores de una muestra correspondan a la clase 0 y un 90 % de que correspondan a la clase 1, el modelo predecirá que la muestra corresponde a la clase 1. En el contexto de la clasificación multinomial, el algoritmo de RL extiende este enfoque para dar cabida a múltiples clases, asignando probabilidades a cada clase para una entrada determinada (416).

En la RL, la operación principal es transformar los datos de entrada en un valor de probabilidad, lo que se logra utilizando la función sigmoidea. Esta función, también conocida como función logística, es una ecuación matemática que transforma cualquier número de valor real en un valor entre 0 y 1. Se expresa mediante la ecuación mostrada en la figura 15. Este proceso posibilita el objetivo de la RL: predecir la probabilidad de un resultado binario (sí/no, verdadero/falso, 1/0), es decir, representa la probabilidad de que una muestra pertenezca a cada una de las clases (417). A diferencia de la regresión lineal, que genera valores continuos, la RL se utiliza cuando el resultado es categórico.

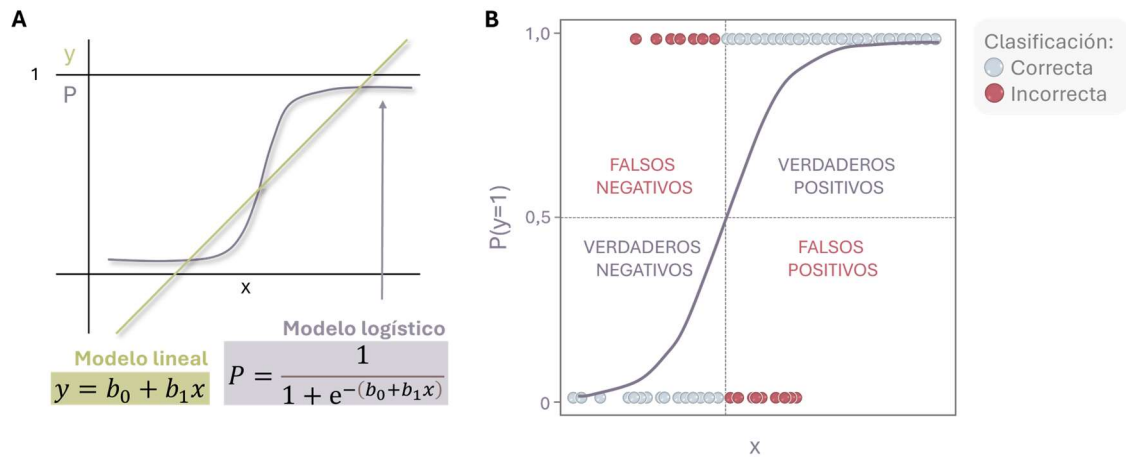


Figura 15. (A) Representación gráfica de los modelos y ecuaciones de regresión lineal y logística. (B) Representación de un caso simple de regresión logística con un solo predictor (x) y probabilidad P de clasificar y como y=1. Cualquier punto de datos con un valor x a la derecha de la línea discontinua tendrá una probabilidad predicha mayor o igual a 0,5 y se clasificará como y=1. Por el contrario, los puntos de datos con valores de x a la izquierda de la línea discontinua se clasificarán como 0 (418).

Al realizar una RL múltiple (es decir, con más de una variable independiente), existen varios enfoques para introducir las variables independientes o predictoras en el modelo. El enfoque más simple introduce todos los predictores en el modelo en un solo paso, se conoce como el método de regresión “estándar”. Otro enfoque consiste en introducir los predictores en pasos determinados, realizando una “regresión jerárquica”, que es adecuada cuando se tiene un número elevado de predictores (419). La regresión *stepwise* o por pasos, es un tipo especial de regresión jerárquica en el que los algoritmos estadísticos determinan qué predictores forman parte del modelo final. Esta técnica tiene tres variantes básicas: selección hacia adelante, eliminación hacia atrás y por pasos (*stepwise*) (420). En la primera, el modelo comienza sin predictores y se introducen predictores significativos sucesivamente hasta alcanzar un criterio estadístico de detención. En la eliminación hacia atrás, el modelo comienza con todos los predictores posibles y elimina los predictores no significativos sucesivamente hasta alcanzar un criterio de detención. El método de regresión *stepwise* combina los dos enfoques anteriores, añadiendo y eliminando predictores a medida que construye el modelo (421).

1.4.2.2 Árboles de decisión

Los árboles de decisión son algoritmos estadísticos que nos permiten construir modelos predictivos de analítica de datos basados en su clasificación según ciertas características o propiedades, o en la regresión mediante la relación entre distintas variables para predecir el valor de otra. La finalidad es crear un modelo que prediga el valor

de una variable objetivo mediante el aprendizaje de reglas de decisión simples deducidas de las características del resto de variables de entrada (422).

Los árboles de decisión clasifican los datos utilizando particionamiento recursivo, que es un método estadístico para el análisis multivariable (423). La partición recursiva crea un árbol de decisiones que se esfuerza por clasificar correctamente a los miembros de la población al dividirlos en subpoblaciones basadas en diversas variables independientes dicotómicas. El proceso se denomina recursivo porque cada subpoblación puede, a su vez, dividirse un número indefinido de veces hasta que el proceso de división termina después de alcanzar un criterio de detención particular (Figura 16).

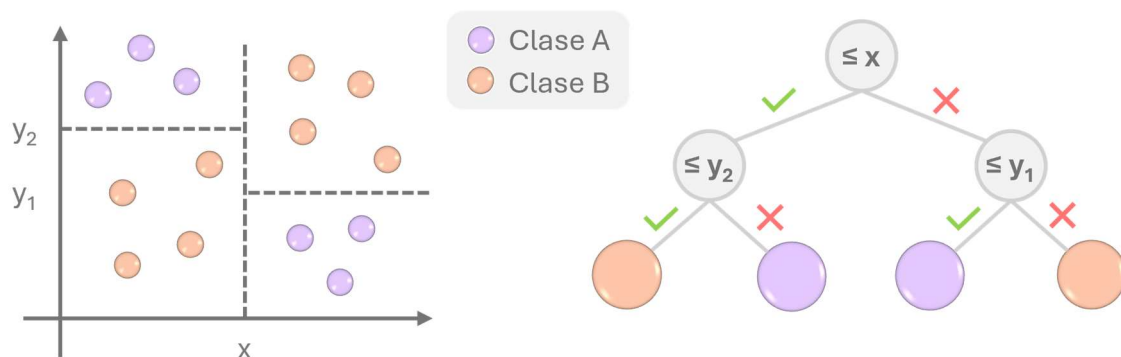


Figura 16. Generalización del funcionamiento de un árbol de decisión para dos clases (morado A y naranja B) con dos variables predictoras (x e y).

En la elaboración de un árbol de decisión es importante determinar qué atributo es el mejor, o más predictivo, para dividir los datos en función de la característica objetivo, por lo que el algoritmo elige la característica más predictiva en función de la cual se deben dividir los datos. La predictibilidad se basa en la disminución de la «impureza» de los nodos. Un nodo en el árbol se considera «puro» si, en el 100% de los casos, los nodos son clasificados en una categoría específica del campo objetivo. De hecho, el método utiliza el particionamiento recursivo para dividir los registros en segmentos minimizando la «impureza» en cada paso. La «impureza» de los nodos se calcula mediante la entropía de los datos en el nodo, que es la cantidad de desorden de la información, o la cantidad de aleatoriedad en los datos. La entropía en un nodo depende de la cantidad de datos aleatorios que se encuentran en ese nodo y se debe calcular para cada uno de ellos. En los árboles de decisión, se buscan los árboles que tengan la entropía más pequeña en sus nodos (424).

Los árboles resultantes de estos algoritmos se pueden visualizar, por lo que son fáciles de entender e interpretar. Además, la creación de estos modelos requiere poca preparación de datos en comparación con otras técnicas que a menudo requieren aplicar normalización a los datos (KNN), eliminar valores en blanco (SVM), etc.

Los métodos bien conocidos de partición recursiva en árboles de decisión incluyen el algoritmo ID3 de Ross Quinlan (424) y sus sucesores, C4.5 y C5.0 y los árboles de clasificación y regresión (CART, por sus siglas en Inglés: *Classification and Regression Trees*) (425). Una crítica común a estos métodos es su vulnerabilidad al sobreajuste de los datos (426). Para solucionar este problema se pueden utilizar diferentes algoritmos y combinar sus resultados, utilizando métodos de aprendizaje por conjuntos como *Random Forest* (RF) o bosques aleatorios (sección 1.4.4.2.1.1 – *Random Forest*).

1.4.2.3 Support Vector Machines

SVM son un conjunto de algoritmos de aprendizaje supervisado propiamente relacionados con problemas de clasificación y regresión. Dado un conjunto de muestras de entrenamiento con clases etiquetadas, se construye un modelo que predice la clase de una nueva muestra. SVM es un modelo que representa los puntos de las muestras en el espacio, separando las clases a 2 espacios lo más amplios posibles mediante un hiperplano de separación definido como el vector entre los 2 puntos (vectores de soporte) más cercanos de cada clase (Figura 17).

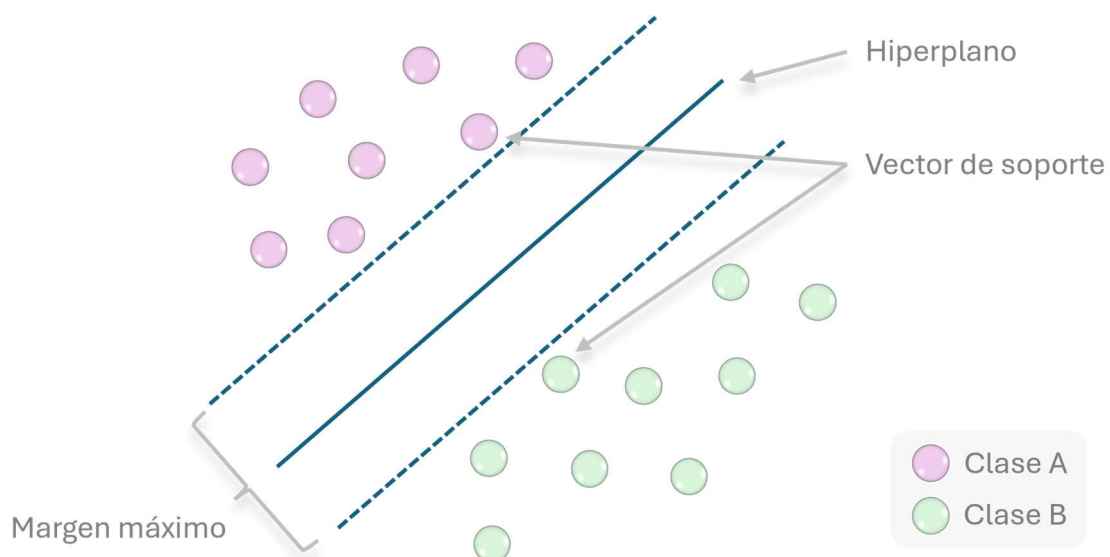


Figura 17. Ejemplo simple de clasificación de dos clases (rosa A y verde B) basado en *support vector machines*.

Cuando las nuevas muestras se ponen en correspondencia con dicho modelo, en función de los espacios a los que pertenezcan, pueden ser clasificadas a una u otra clase.

SVM construye un hiperplano o conjunto de hiperplanos en un espacio de dimensionalidad muy alta (o incluso infinita) que puede ser utilizado en problemas de clasificación o regresión. Una buena separación entre las clases permitirá una clasificación correcta (427).

Los algoritmos basados en SVM utilizan un conjunto de funciones matemáticas que se definen como *kernel*. La función del *kernel* es tomar datos como entrada y transformarlos en la forma requerida. Diferentes algoritmos SVM usan diferentes tipos de funciones del *kernel*, que pueden ser de varios tipos, por ejemplo: lineal, no lineal, polinomial, función de base radial (RBF, del inglés, *radial basis function*), tangente hiperbólica o sigmoide.

1.4.2.4 K-Nearest Neighbors

KNN es un algoritmo de tipo supervisado de ML que puede utilizarse para clasificar nuevas muestras (valores discretos) o para predecir valores continuos (regresión) (428). Sirve esencialmente para clasificar valores buscando los puntos de datos “más similares” (por cercanía) aprendidos en la etapa de entrenamiento y haciendo conjeturas de nuevos puntos basado en esa clasificación (429). Es un método que simplemente busca en las observaciones más cercanas a la que se está tratando de predecir y clasifica el punto de interés basado en la mayoría de los datos que le rodean (Figura 18).

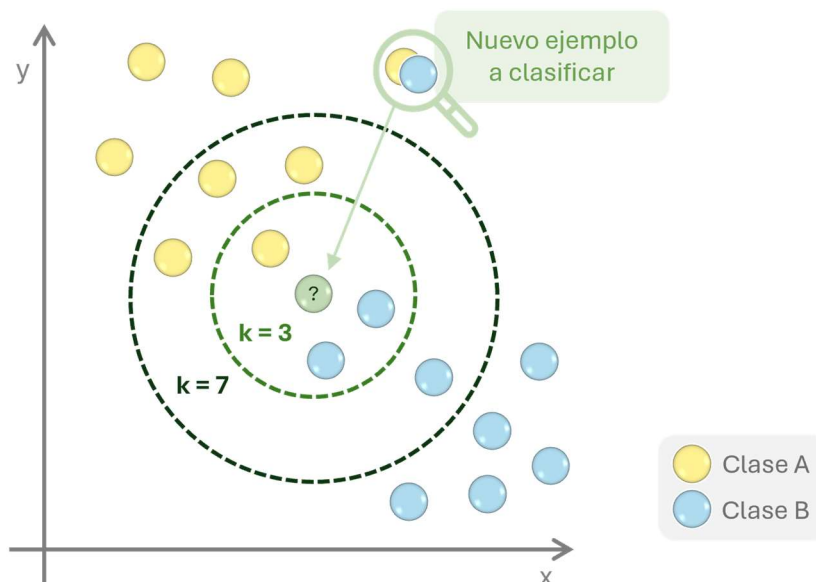


Figura 18. Representación gráfica del algoritmo de clasificación KNN. Simplificación con dos clases (amarillo A y azul B), dos variables predictoras (x e y) y dos valores de k (3 y 7). En este ejemplo, si desarrollamos un modelo con k=3 vecinos, el nuevo ejemplo a clasificar será adjudicado a la clase B, pero si desarrollamos un modelo con k=7 vecinos, el nuevo ejemplo a clasificar será adjudicado a la clase A.

KNN está basado en instancia, es decir, no aprende explícitamente un modelo (como por ejemplo en árboles de decisión). En cambio, memoriza las instancias de entrenamiento que son usadas como “base de conocimiento” para la fase de predicción.

El algoritmo KNN toma su nombre del hecho de que usa información sobre los k vecinos más cercanos de un ejemplo para clasificar ejemplos no etiquetados. La letra k es un término variable que indica el número de vecinos más cercanos que se utilizan para el algoritmo de clasificación.

El algoritmo requiere un conjunto de datos de entrenamiento compuesto por ejemplos que ya están clasificados en varias categorías, según la etiqueta de una variable categórica. Para cada instancia nueva no etiquetada, KNN identifica los registros k en los datos de entrenamiento que son los “más cercanos” en similitud. A esa instancia no etiquetada se le asigna la clase de la mayoría de los k vecinos más cercanos (429).

1.4.2.5 Redes neuronales artificiales

Una ANN es un tipo de técnica de IA que se inspiró en la estructura y la forma en que las neuronas biológicas actúan unidas para identificar fenómenos, sopesar opciones y llegar a conclusiones. Los algoritmos basados en ANN permiten tratar problemas complejos orientados a patrones, tanto de categorización como de series temporales (análisis de tendencias) (430).

Las ANN están compuestas por capas de nodos o neuronas. Poseen una capa de entrada, una o varias capas ocultas y una capa de salida. La cantidad de nodos en cada capa se determina según la estructura del problema. Estos nodos están conectados por flechas que simulan las sinapsis cerebrales. Los datos de entrada se facilitan en la primera capa, y las “señales” se van propagando a otros nodos de la capa siguiente, que a su vez las procesa y las envía a otras neuronas conectadas. Al final, se envía un resultado desde la capa de salida. La “señal” es un número real, y la salida de cada nodo se calcula a través de una función no lineal de la suma de sus entradas. Esta función de propagación establece relaciones matemáticas entre los valores de entrada, su peso y un sesgo para el cálculo del valor de salida del nodo. También existe una función de transferencia, que adapta o modula el valor calculado de salida del nodo (431). La fuerza de cada señal está determinada por un peso, que se ajusta durante el proceso de aprendizaje (432). En el entrenamiento de una ANN, las señales pasan a través de los nodos y simultáneamente modifican sus pesos y las funciones de transferencia para minimizar el error de la red y que

los resultados sean más fiables. Este proceso se repite hasta que se alcanza una salida. (430). La figura 19 representa la estructura tradicional de una ANN.

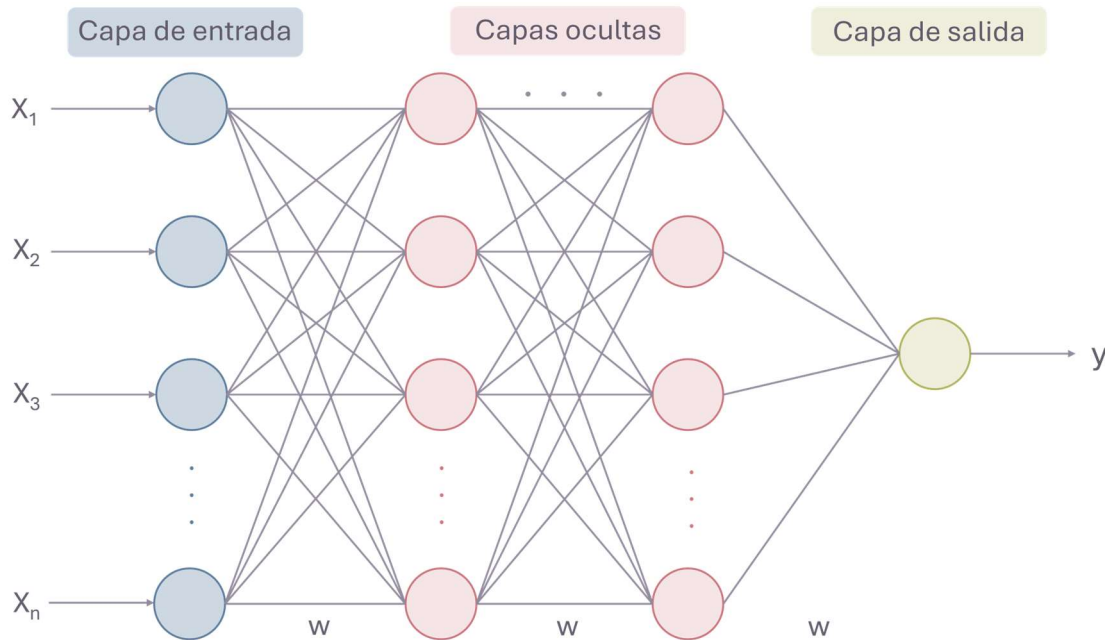


Figura 19. Representación simple de la estructura de una red neuronal artificial, que incluye nodos (representados con círculos) en las capas de entrada, ocultas y de salida. “X” son los datos de entrada que se introducen en la primera capa, “y” es el valor de predicción final y “w” son los pesos de los valores transferidos entre nodos (433).

Cada nodo que se conecta a los demás tiene su propia ponderación y umbral asociados. Si la salida de cualquier nodo individual está por encima del valor umbral especificado, ese nodo se activa y envía datos a la siguiente capa de la red. De lo contrario, no se pasa ningún dato a la siguiente capa de la red. La función de activación es la que determina si la neurona está activa, es decir, si produce un valor de salida o no (431). Las redes neuronales se basan en datos de entrenamiento para aprender y mejorar su precisión con el tiempo. Al inicio del entrenamiento todas las ponderaciones son aleatorias, pero la red aprende examinando las instancias individuales, generando una predicción para cada muestra y realizando ajustes a las ponderaciones cuando realiza una predicción incorrecta. Este proceso se repite múltiples veces y la red continúa mejorando sus predicciones hasta alcanzar el criterio estadístico de detención establecido. Una vez perfeccionadas, las ANN se convierten en potentes herramientas en informática e IA, que nos permiten clasificar y agrupar datos a gran velocidad (434).

1.4.3 Evaluación de los modelos

Se debe evaluar cada modelo desarrollado para determinar si realizará un buen trabajo de predicción para futuros nuevos datos. Para poder evaluar un modelo

Introducción

correctamente, se utiliza una muestra del set de datos completo que no ha sido utilizada para el entrenamiento (set de datos de prueba o testeo).

Se debe comprobar si las predicciones del modelo son correctas, es decir, si para nuevas instancias (que no han sido utilizadas en el entrenamiento), el algoritmo es capaz de asignar correctamente la etiqueta que corresponde basándose en los valores de las variables de entrada y el algoritmo de clasificación. Por lo tanto, para testar la predicción del modelo, se comparan las predicciones generadas por el modelo frente a las etiquetas originales de esas mismas instancias del set de datos de prueba.

Para llevar a cabo la evaluación de los modelos de manera individual, se crean matrices de confusión similares a la de la figura 20, en la que se enfrentan los resultados de la predicción frente a los datos reales conocidos. Los números contenidos en la tabla indican el número de pacientes que han sido clasificados en cada una de las cuatro categorías. En la celda superior izquierda, están los resultados verdaderos positivos (VPs). Estos valores indican casos que son positivos, y el algoritmo de clasificación lo identifica correctamente como tal. La celda inferior derecha indica los resultados verdaderos negativos (VNs), donde el clasificador y la etiqueta de los datos originales concuerdan.


		Valores reales		
		Negativo	Positivo	
Valores predicción	Negativo	VN Verdadero negativo	FN Falso negativo	Valor predictivo negativo $\frac{VN}{VN + FN}$
	Positivo	FP Falso positivo	VP Verdadero positivo	Valor predictivo positivo $\frac{VP}{VP + FP}$
		Especificidad $\frac{VN}{VN + FP}$	Sensibilidad $\frac{VP}{VP + FN}$	Exactitud $\frac{VP + VN}{N}$

Figura 20. Matriz de confusión y parámetros derivados de ella. VN: Verdadero negativo; VP: Verdadero positivo; FP: Falso positivo; FN: Falso negativo; N: número total de predicciones.

Las celdas de la diagonal opuesta contienen recuentos de ejemplos en los que la predicción no concuerda con el valor verdadero. La celda en la parte superior derecha contiene los resultados FPs. Estos valores ocurren cuando el modelo clasifica como

positivo, y el diagnóstico real es negativo. Los ejemplos en la celda inferior izquierda son resultados FNs; en este caso, el valor predicho es negativo, pero la etiqueta real es positiva. Los últimos errores pueden ser extremadamente costosos, ya que, en aplicaciones en diagnóstico clínico, pueden llevar a creer que un paciente está libre de la enfermedad, cuando en realidad la enfermedad está presente. En consecuencia, este paciente no será tratado farmacológicamente o no recibirá las recomendaciones clínicas para su enfermedad, pudiendo acarrear riesgos muy peligrosos e incluso la muerte.


Para comparar los modelos se utilizan diferentes parámetros, entre los que destacan:

-  **Exactitud (*accuracy*):** Porcentaje total de elementos clasificados correctamente. Es la medida más directa de la calidad de los clasificadores. Es un valor entre 0 y 1, siendo 1 la mayor precisión.


$$Exactitud = \frac{VP + VN}{N}$$

Siendo N el número total de predicciones.


*Cuando las clases a categorizar están muy desequilibradas este método resulta poco útil.

-  **Precisión balanceada:** Precisión que tiene en cuenta en el cálculo si las clases no están balanceadas.

$$Precisión\ balanceada = \frac{Sensibilidad + Especificidad}{2}$$

-  **Sensibilidad:** Número de elementos identificados correctamente como positivos del total de positivos verdaderos. Es decir, la proporción de pacientes que se identificaron correctamente por tener una condición, VP, sobre el número total de pacientes que realmente presentan esa condición.

$$Sensibilidad = \frac{VP}{VP + FN}$$

-  **Especificidad:** Proporción de pacientes que se identificaron correctamente por no tener una condición, VN, sobre el número total de pacientes que realmente no presentan esa condición.

$$Especificidad = \frac{VN}{VN + FP}$$

- ❁ **Valor predictivo positivo (VPP):** Proporción de VPs respecto a todas las predicciones positivas.

$$VPP = \frac{VP}{VP + FP}$$

- ❁ **Valor predictivo negativo (VPN):** Proporción de VNs respecto a todas las predicciones negativas.

$$VPN = \frac{VN}{VN + FN}$$

1.4.4 Procedimientos para la mejora de los modelos de clasificación

1.4.4.1 Validación cruzada con ajuste de parámetros automatizado

La validación cruzada o *cross-validation* es una técnica utilizada para evaluar los resultados de un modelo y garantizar que son independientes de la partición entre datos de entrenamiento y prueba. Consiste en repetir y calcular la media aritmética obtenida de las medidas de evaluación sobre diferentes particiones. Se utiliza en entornos donde el objetivo principal es la predicción y se quiere estimar la precisión de un modelo que se llevará a cabo a la práctica (435). Mediante validación cruzada con ajuste de parámetros (o *autotuning*) podemos entrenar un modelo utilizando un conjunto de datos y probándolo con un conjunto diferente. El ajuste de parámetros es el proceso de seleccionar los valores de los parámetros de un modelo que maximizan la precisión del modelo. Además de mejorar el rendimiento en observaciones no vistas y en entornos con limitaciones de datos, puede ser una herramienta eficaz para el entrenamiento de modelos con un conjunto de datos más pequeño.

Se establece un valor predeterminado de validación cruzada de n veces para evaluar los resultados. Todo el set de datos de prueba se divide en n subconjuntos de datos. La validación cruzada itera a través de los subconjuntos y en cada iteración usa uno de los subconjuntos n como conjunto de validación mientras usa todos los demás subconjuntos como set de entrenamiento. Este proceso se repite hasta que cada subconjunto de datos se haya utilizado como conjunto de validación (ejemplo con $n = 5$ en la figura 21). Al entrenar y testar el modelo n veces en diferentes subconjuntos de los mismos datos de entrenamiento, se obtiene una representación más precisa del funcionamiento del modelo con datos que no ha visto antes. Se califica el modelo después

de cada iteración y se calcula el promedio de todos los puntajes para obtener una mejor representación de cómo funciona el modelo.

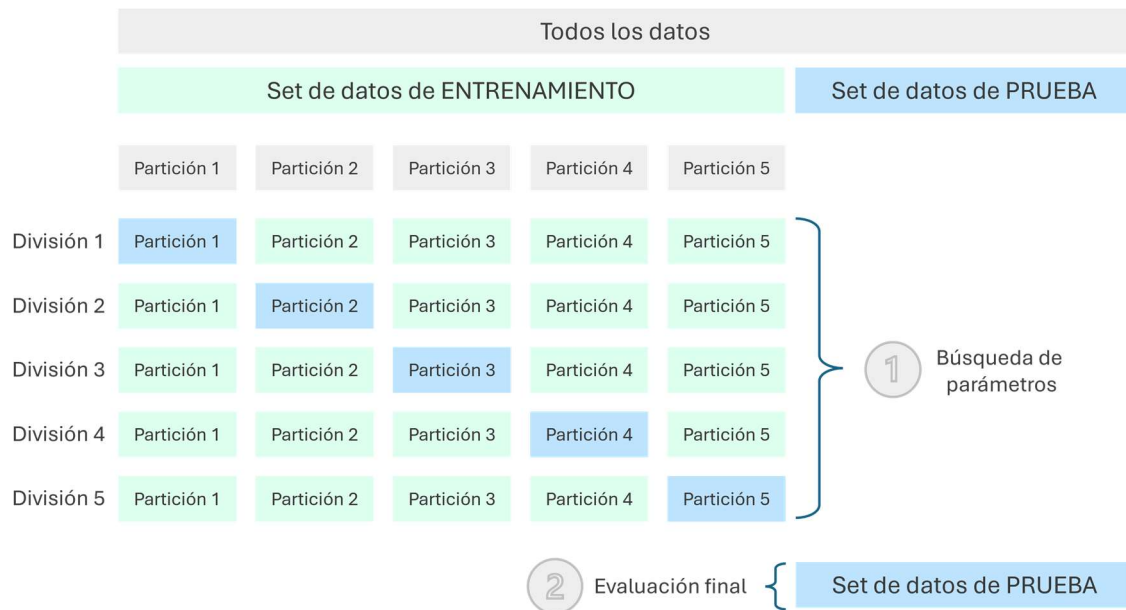


Figura 21. Representación de un proceso de validación cruzada de 5 iteraciones con ajuste de parámetros automatizado (436).

En cuanto a la selección de parámetros del modelo, no existe una solución única para encontrar los valores óptimos. Un conjunto de parámetros que funciona bien para un problema de ML puede funcionar mal en otro. Gracias al proceso de validación cruzada con ajuste de parámetros automatizado somos capaces de iterar este proceso cambiando los parámetros que seleccionemos de cada modelo en concreto. Finalmente se obtiene el modelo con los parámetros que muestran mejor funcionamiento (427).

1.4.4.2 Ensemble learning

En estadística y ML, los métodos de aprendizaje por conjuntos o *ensemble learning* utilizan múltiples algoritmos de ML diferentes para generar predicciones individuales débiles para después fusionar los resultados de predicción mediante mecanismos de votación (437). De esta manera, se mejora la precisión y se obtiene un mejor rendimiento predictivo que el que se podría obtener de cualquiera de los algoritmos de aprendizaje constituyentes por sí solos (438-440). Un conjunto de ML consiste en una agrupación finita concreta de modelos alternativos, pero generalmente permite que exista una estructura mucho más flexible entre esas alternativas.

Existen diferentes tipos de *ensemble learning*, pero en esta tesis nos centramos en *bagging* y *boosting*. Ambos son similares en el sentido de que son técnicas de conjunto,

donde un conjunto de modelos débiles se combina para crear un modelo robusto que obtiene un mejor rendimiento que uno solo de los modelos individuales del conjunto (437).

1.4.4.2.1 Bagging

Bagging (también llamado *Bootstrap aggregating*), es un meta-algoritmo de conjunto de ML diseñado para mejorar la estabilidad y precisión de los algoritmos utilizados para clasificación y regresión estadística. También reduce la varianza y ayuda a evitar el sobreajuste. Aunque normalmente se aplica a métodos de árboles de decisión, se puede utilizar con cualquier tipo de método. Es un caso especial del promediado de modelos. *Bagging* utiliza subconjuntos de variables combinando las filas o instancias de los datos originales. Realiza el entrenamiento para cada subconjunto de datos, construyendo una larga colección de modelos no correlacionados y posteriormente los promedia para generar un modelo final, lo que da como resultado mejores propiedades de generalización (441, 442).

El primer paso en el proceso es la generación de lo que se denominan conjuntos de datos de *bootstrap* o remuestreo (Figura 22).

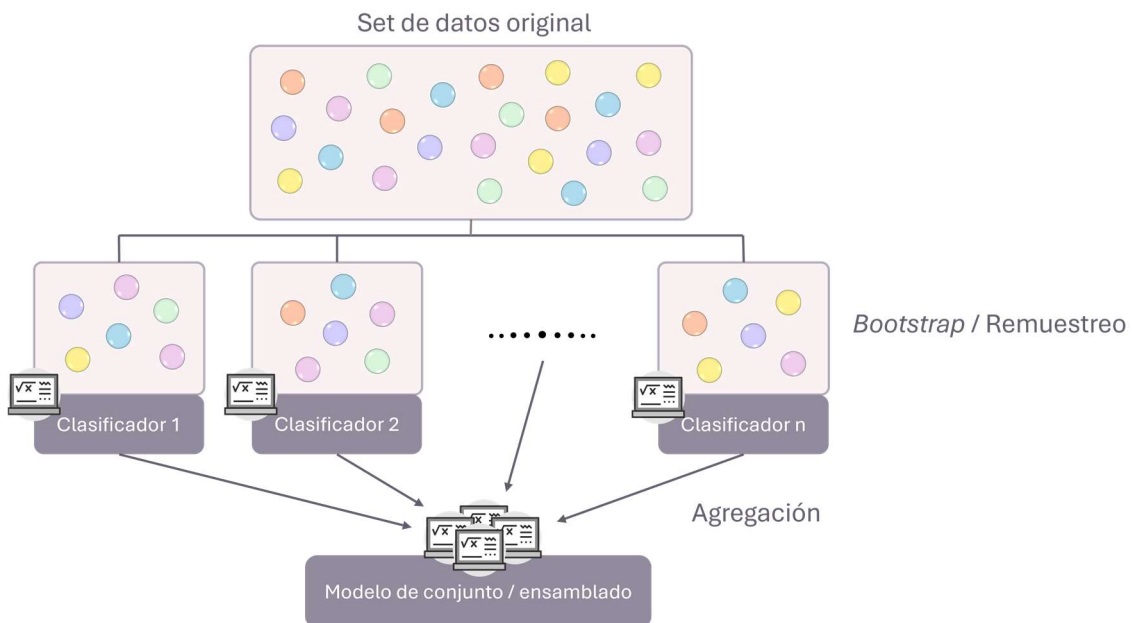


Figura 22. Ilustración del concepto *Bagging*. Imagen adaptada de Medium (443).

Para cada conjunto de *bootstrap*, la cantidad de elementos seleccionados es la misma que la del conjunto de datos de entrenamiento original, pero los elementos se eligen aleatoriamente con reemplazo. Por lo tanto, una muestra determinada del conjunto de entrenamiento original puede aparecer cero, una o varias veces en un conjunto de *bootstrap* determinado. El proceso de *bagging* también genera un conjunto de datos *out-of-*

bag, que se compone de todos los elementos del conjunto de entrenamiento original que no están en un conjunto de *bootstrap* específico. Como tal, cada conjunto de datos tendrá un conjunto de *out-of-bag*, incluso si el conjunto de *out-of-bag* es el conjunto vacío. Después de crear los subconjuntos de datos, cada uno se asigna a uno de los modelos del conjunto (por ejemplo, un árbol de decisión). Dicho modelo utiliza el conjunto de *bootstrap* para el entrenamiento, y una vez entrenado se convertirá en un modelo clasificador del conjunto de modelos. Dado que cada conjunto de *bootstrap* se selecciona al azar, los conjuntos tendrán variedad y, como resultado, los modelos entrenados tendrán cada uno una perspectiva diferente del conjunto de entrenamiento original (444, 445).

La variación de la información local en los conjuntos de *bootstrap* y la selección de características utilizadas (por ejemplo, nodos de un árbol de decisión) promueven la diversidad entre los modelos del conjunto, y, de acuerdo con la teoría del conjunto (446), pueden fortalecer el conjunto. Para utilizar la fuerza de esta diversidad, se emplea la agregación, que es la forma en que un conjunto se traduce de una serie de evaluaciones individuales a una sola evaluación colectiva de una muestra. El proceso de agregación de un conjunto implica recopilar las valoraciones individuales de cada uno de los modelos del conjunto. Finalmente, se cuentan las evaluaciones individuales y la conclusión alcanzada por la mayoría de los modelos del conjunto se considera la evaluación general del conjunto.

Una aplicación común del proceso de agregación de *bootstrap* son RF, que son colecciones de árboles de decisión en un conjunto para lograr una mayor precisión de clasificación que un árbol de decisión individual (447).

1.4.4.2.1 Random Forest

Los bosques aleatorios o RF son una variante de *bagging* específicamente diseñados para trabajar con árboles de decisión. También se introduce aleatoriedad en las variables, no sólo en las observaciones. Para evitar dependencias, los posibles predictores se seleccionan al azar en cada nodo.

El algoritmo de RF (Figura 23) surge como la agrupación de varios árboles de clasificación; selecciona de manera aleatoria una cantidad de variables con las cuales se construye cada uno de los árboles individuales, y se realizan predicciones con estas variables, que posteriormente serán ponderadas a través del cálculo de la clase más votada de los árboles que se generaron, para finalmente hacer la predicción por RF (448).

Introducción

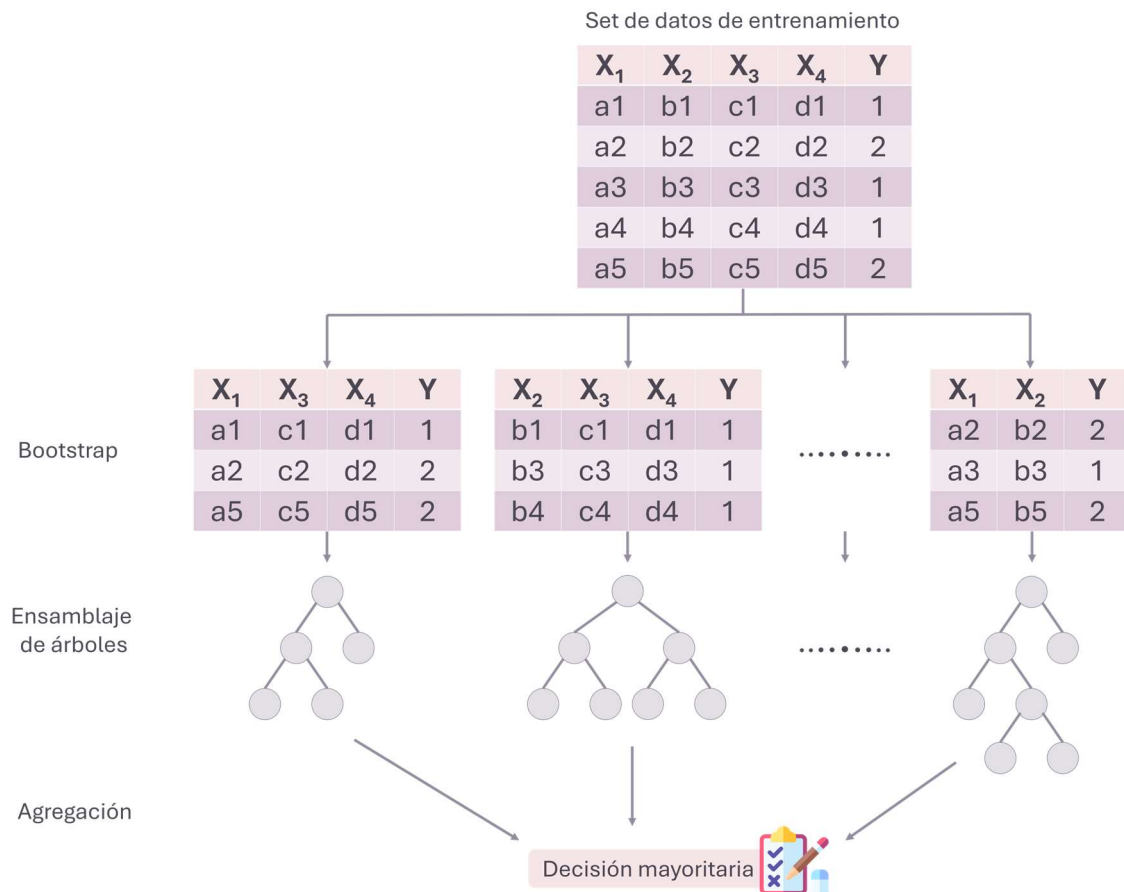


Figura 23. Representación del algoritmo RF. Las columnas X_1 - X_4 representan las variables, la columna Y es la variable a predecir, y las filas son las observaciones o instancias del set de datos.

Las muestras *bootstrap* (remuestreo con reemplazo) que se generan al hacer *bagging* introducen un elemento de aleatoriedad que en la práctica provoca que todos los árboles sean distintos, pero en ocasiones no son lo suficientemente distintos. Es decir, es frecuente que los árboles tengan estructuras muy similares, especialmente en la parte alta, aunque después se vayan diferenciando según se desciende por ellos. Esta característica se conoce como correlación entre árboles y se da cuando el árbol es un modelo adecuado para describir la relación entre los predictores y la respuesta, y también cuando uno de los predictores es muy fuerte, es decir, es especialmente relevante, con lo cual casi siempre va a estar en el primer nodo del árbol. Esta correlación entre árboles se va a traducir en una correlación entre sus predicciones.

Promediar variables altamente correlacionadas produce una reducción de la varianza mucho menor que si promediamos variables no correlacionadas. La solución pasa por añadir aleatoriedad al proceso de construcción de los árboles, para que estos dejen de estar correlacionados. En la construcción de cada uno de los árboles que finalmente constituirán el bosque, se van haciendo cortes binarios, y para cada corte hay

que seleccionar una variable predictora. La modificación introducida suele ser que, antes de hacer cada uno de los cortes, se seleccionan al azar predictores de entre todas las variables predictoras que van a ser los candidatos para el corte.

Los bosques aleatorios son computacionalmente más eficientes que *bagging* porque, aunque requieren más árboles, la construcción de cada árbol es mucho más rápida al evaluarse sólo unos pocos predictores en cada corte.

Estos métodos dificultan la interpretación, pero se puede medir la importancia de las variables (índices de importancia). Por ejemplo, para cada árbol se suman las reducciones en el índice de Gini correspondientes a las divisiones de un predictor y posteriormente se promedian los valores de todos los árboles (449, 450). La disminución media del coeficiente de Gini (MDG, del inglés, *Mean Decrease Gini*) es una medida de cómo cada variable contribuye a la homogeneidad de los nodos y hojas en el algoritmo resultante. Cuanto mayor sea el valor de la puntuación de Gini de disminución media, mayor será la importancia de la variable en el modelo.

En la figura 24 se puede observar la diferencia entre el modelo aprendido por un algoritmo CART (izquierda) y RF (derecha) cuando resuelven el mismo problema de regresión, siendo el error de generalización del RF mucho menor.

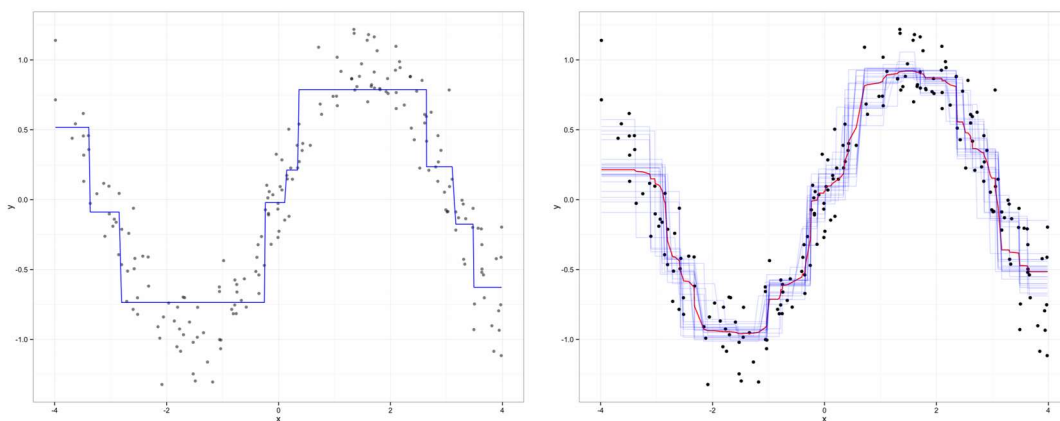


Figura 24. Comparación entre un modelo entrenado mediante un árbol de decisión simple, CART (izquierda) y RF (derecha). Los puntos representan cada muestra y las líneas azules representan la predicción de cada árbol. En la gráfica derecha se muestran 25 árboles seleccionados al azar (azul) que dan como resultado el algoritmo de RF (la predicción se muestra en rojo) (451).

Este hecho se debe a la combinación de los múltiples modelos (representados con líneas azules), que da como resultado un modelo final (línea roja) que disminuye la varianza de las predicciones, ajustando varios modelos y promediando sus predicciones para obtener una predicción regularizada, minimizando así el riesgo de sobreajuste. Se puede

Introducción

observar que la aproximación es mucho más “suave” en comparación con la aproximación de cualquier árbol individual (451).

Cada árbol se entrena con distintas muestras de datos para un mismo problema. De esta forma, al combinar sus resultados, unos errores se compensan con otros y se obtiene una predicción que generaliza mejor.

1.4.4.2.2 Boosting

Al igual que *bagging*, *boosting* también es una técnica para mejorar el rendimiento de los modelos de ML mediante el uso de múltiples modelos individuales. Aunque ambos son métodos de *ensemble learning*, difieren en cómo combinan los modelos para mejorar la precisión y minimizar el error de predicción (Figura 25).

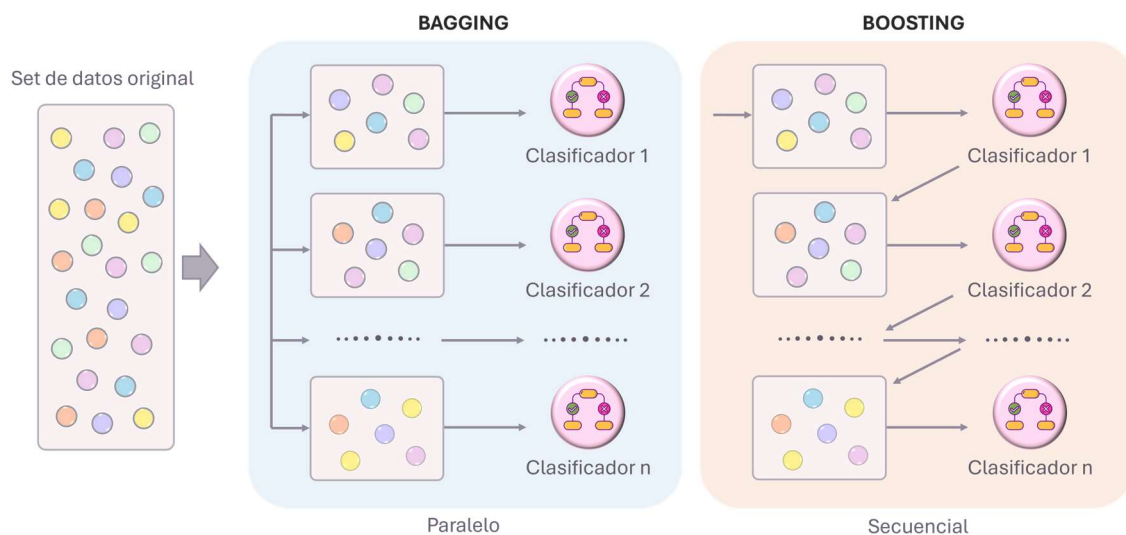


Figura 25. Representación de la diferencia entre los algoritmos *bagging* (izquierda) y *boosting* (derecha) (452).

En los algoritmos de *boosting*, varios modelos simples son utilizados secuencialmente, es decir, cada modelo simple se ejecuta delante o detrás de otro modelo simple. Cada modelo trata de corregir los errores cometidos por los modelos anteriores. El rendimiento general puede ser mejorado haciendo que un modelo simple posterior les dé más importancia a los errores cometidos por un modelo simple previo. Este método itera el mismo clasificador una y otra vez, enfocándose en los errores del anterior, forzando al algoritmo a clasificarlos correctamente y evitando que clasifique incorrectamente las instancias que ya clasificaba de manera correcta (437). Las predicciones de cada modelo simple se combinan por medio de una votación (para problemas de clasificación) o por medio de una suma ponderada (para problemas de regresión) para producir la predicción final. En resumen, la diferencia entre *bagging* y *boosting* es que en *boosting* los algoritmos

no se entrenan independientemente, sino que se ponderan según los errores de los anteriores.

1.4.5 Explicabilidad de los modelos de caja negra

Los modelos de caja negra son aquellos algoritmos de ML que dan un resultado o toman una decisión sin explicar o mostrar cómo llegaron dicho resultado o predicción. Los procesos internos llevados a cabo y los múltiples factores tenidos en cuenta permanecen desconocidos o difícilmente comprensibles para quienes los ejecutan.

Un claro ejemplo de este tipo de modelos es RF, donde el usuario no ve el funcionamiento interno del algoritmo. El desarrollador solamente es capaz de ver las características de entrada y de salida del modelo.

Mejorar la interpretabilidad de estos modelos es particularmente importante en entornos clínicos, donde es fundamental justificar las decisiones de los modelos para garantizar la seguridad y confiabilidad de los mismos. Por este motivo, además de evaluar el rendimiento y el ajuste a los datos del estudio, es recomendable utilizar técnicas avanzadas para explicar las predicciones de los modelos de caja negra (Figura 26).

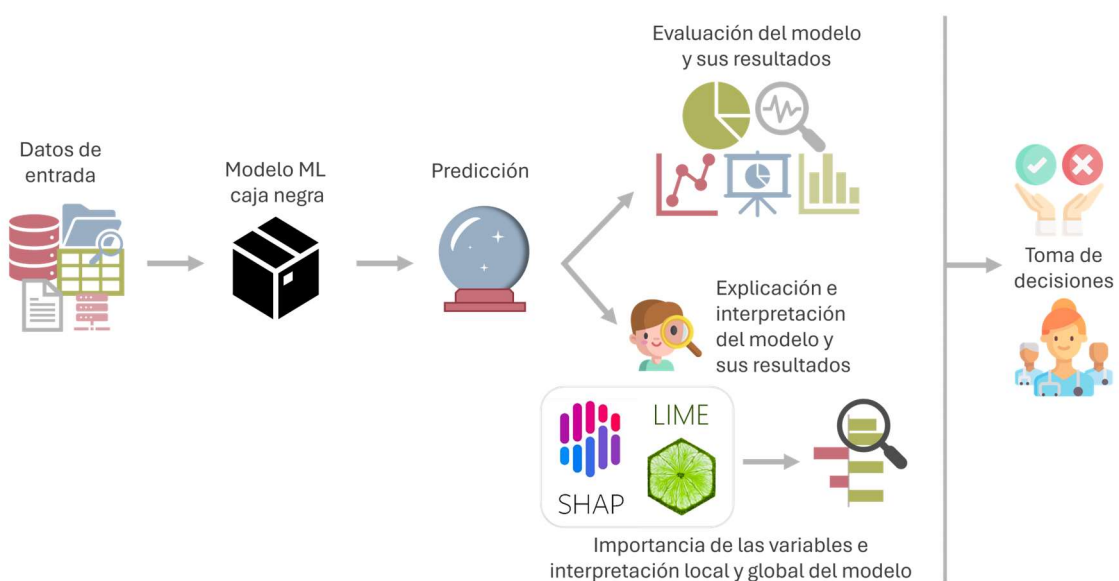


Figura 26. Ilustración que representa la limitación de los modelos de caja negra, los cuales deben ser evaluados, explicados e interpretados con el objetivo de alcanzar una comprensión adecuada de los mismos que permita tomar decisiones, sobre todo en entornos clínicos. ML: machine learning; SHAP: Shapley Additive exPlanations; LIME: Local Interpretable Model-agnostic Explanations.

Explicar una predicción se refiere a la presentación de resultados escritos o visuales que permiten un conocimiento cualitativo de la relación entre los componentes de la instancia y la predicción del modelo (453). En la literatura se han utilizado diferentes

técnicas para interpretar los modelos, entre las que destacan *Shapley Additive exPlanations* (SHAP) (454, 455) y *Local Interpretable Model-agnostic Explanations* (LIME) (453).

1.4.5.1 Shapley Additive exPlanations

El método SHAP o explicaciones aditivas Shapley, desarrollado por Lundberg y Lee en 2017 (455), tiene como objetivo interpretar las predicciones de modelos de ML a través de valores Shapley. Estos valores Shapley son un concepto de la Teoría de Juegos, que sirve para explicar cómo cada uno de los “jugadores” que intervienen en un “juego colaborativo” contribuyen en el éxito de la partida (456). Una predicción puede explicarse suponiendo que cada valor de las características de la instancia es un “jugador” en un juego donde la predicción es la “ganancia”. Utilizando esta técnica, los valores Shapley establecen cómo distribuir equitativamente la “ganancia” entre las características (457). Cada valor Shapley representa el impacto que genera en la predicción la variable a la que está asociado. De manera simplificada, un valor Shapley es la contribución marginal promedio de un valor de una característica entre todas las coaliciones posibles (Figura 27) (458).

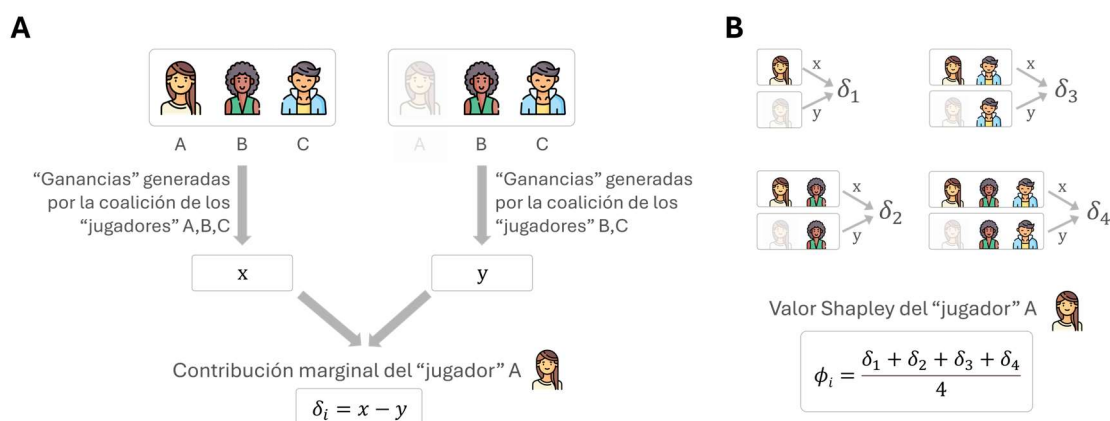


Figura 27. Representación simple del cálculo de los valores de Shapley según la Teoría de Juegos, donde los “jugadores” son las variables que constituyen un modelo de ML y las “ganancias” son las predicciones que realiza dicho modelo. Para calcular el valor de Shapley de una variable concreta, **(A)** se calcula la diferencia entre la utilidad que se genera cuando la variable está presente con respecto a la utilidad que se genera cuando está ausente (dicha diferencia es la contribución marginal de la variable a la coalición actual). **(B)** El proceso A se realiza para cada uno de los subgrupos (coaliciones) que se pueden generar donde esté la variable de interés. La media de las diferencias obtenidas (la contribución marginal media) es el valor de Shapley. Iconos obtenidos de Freepik, figura basada en (458).

El algoritmo utiliza una combinación de líneas de base, funciones de importancia local y el Teorema de Valor de Shapley para calcular la importancia de cada variable (medida como el cambio promedio en la predicción del modelo cuando varía el valor de la variable) en una predicción individual. SHAP también permite estudiar interpretaciones globales obteniendo el promedio de las contribuciones de cada variable para cada

predicción del modelo (459). De esta manera, SHAP facilita la explicación local y global de los resultados de un modelo, es decir, explica la influencia de cada variable en las observaciones del modelo, y la importancia de cada variable en los resultados globales del modelo (460).

Los valores Shapley asignan una puntuación a cada variable, mostrando la magnitud (que representa la estimación de la importancia de esa característica) y dirección (signo) de su impacto en la predicción del modelo. Las variables con valores Shapley positivos contribuyen a que el modelo elija una determinada clase, mientras que las variables con signo negativo contribuyen de manera negativa a la predicción de esa misma clase (461).

Cuando se aplica SHAP para interpretar un modelo de clasificación, el resultado que se obtiene es una descomposición detallada de la predicción del modelo en contribuciones aditivas de cada característica. Este resultado permite identificar cómo influye cada variable en las predicciones.

1.4.5.2 Local Interpretable Model-agnostic Explanations

La técnica LIME fue propuesta en 2016 por Marco Tulio Ribeiro, Sameer Singh y Carlos Guestrin (453), con el objetivo de responder al “por qué” de cada predicción y del modelo completo.

El método LIME interpreta las predicciones de modelos individuales basándose en la aproximación local del modelo en torno a una predicción dada. Es decir, LIME se fundamenta en generar explicaciones locales para las predicciones individuales de un modelo.

Este proceso se lleva a cabo aproximando el comportamiento del modelo complejo original mediante la construcción de un modelo local simplificado y fácilmente interpretable (mediante, por ejemplo, una regresión lineal) en una pequeña región alrededor del ejemplo que se quiere explicar (Figura 28). Este modelo aproximado permite entender qué características influyen más en la predicción específica de esa instancia (453).

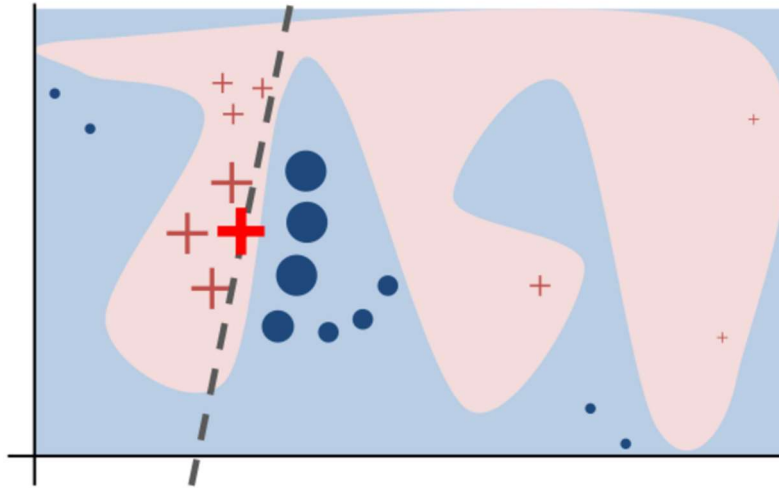


Figura 28. Ilustración que representa LIME de manera intuitiva (453). La función de decisión compleja f del modelo de caja negra (desconocida para LIME) está representada por el fondo azul/rosa, que no se puede aproximar con un modelo lineal. Las áreas coloreadas representan las regiones de decisión del clasificador binario, mostrando cómo el modelo asigna una clase a la variable dependiente según los valores de dos variables explicativas continuas, representadas en los ejes. Cada zona coloreada indica las combinaciones de estos valores para las que el modelo clasifica la observación en una de las dos clases posibles. La cruz roja en negrita es la instancia de interés (ejemplo a predecir). LIME utiliza un conjunto de datos ligeramente modificados alrededor del punto de interés, obtiene predicciones usando f y las pondera por la proximidad a la instancia que se explica (representada en función del tamaño). La línea discontinua representa un modelo lineal simple que se ajusta a los datos localmente pero no globalmente. El modelo simple explica el comportamiento local del modelo de caja negra alrededor de la instancia de interés (453, 462).

Esta técnica permite la explicación de resultados a nivel local, es decir, la explicación de los resultados de una instancia concreta en particular, a partir de la información de otros casos similares (459).

Cuando se aplica LIME a una predicción individual de un modelo de ML, se obtiene un conjunto de explicaciones locales que permiten entender qué variables del conjunto de datos influyen más en esa predicción concreta. A modo de resultado, se obtiene una lista de características con su correspondiente peso o contribución relativa a la predicción, visualizadas generalmente en gráficos (463). Estas contribuciones permiten interpretar el impacto de variables específicas en la predicción, lo que es especialmente útil en contextos clínicos para justificar decisiones automatizadas basadas en datos del paciente.

Contribuciones experimentales

Capítulo

I

Estudio de la actividad antibacteriana de la gliotoxina frente a *Staphylococcus aureus*

Índice

Antecedentes.....	71
Material y métodos.....	74
Resultados.....	82
Discusión.....	90

2 CONTRIBUCIONES EXPERIMENTALES

2.1 CAPÍTULO 1 – Estudio de la actividad antibacteriana de la gliotoxina frente a *Staphylococcus aureus*

2.1.1 ANTECEDENTES

Actualmente apenas existen evidencias científicas sobre el empleo de la GT como antifúngico o antibiótico. Como se ha descrito en la introducción, la GT posee un amplio espectro de actividades biológicas que incluyen algunos efectos antibacterianos y antivirales. También es un potente agente inmunomodulador (144) y un inductor de muerte celular apoptótica en varios tipos celulares (464). Algunos estudios *in vitro* demuestran que las ROS producidas por la GT son capaces de provocar daños en el ADN (465-467). Un estudio realizado por Nieminen *et al.* (468) evaluó la genotoxicidad a nivel de ADN de la GT frente a *E. coli*. En el ensayo de reparación, la GT produjo zonas de inhibición de crecimiento bacteriano más amplias en la cepa deficiente en reparación de ADN que en la competente, lo que sugiere que la GT podría provocar de alguna forma daño en el ADN, tal y como se ha descrito en células eucariotas.

Aunque se conoce que es muy tóxica frente a células de mamífero, se conoce muy poco sobre la respuesta de otras bacterias y hongos frente a GT. Wan-Ling *et al.* (168) estudiaron la acción antibacteriana de GT frente a *E. coli* y *S. aureus*. Sus resultados muestran la eficacia antibacteriana del compuesto frente a ambos microorganismos, pero de manera más efectiva frente a *S. aureus*, ya que la GT muestra una potente inhibición frente a dos cepas de *S. aureus*, una de ellas resistente a meticilina (ATCC29213, concentración mínima inhibitoria (CMI) = 12,20 μM ; R3708, CMI = 1,53 μM). Sin embargo, se desconoce el efecto que la GT podría ejercer frente a *S. aureus* resistentes o con sensibilidad reducida a otros fármacos ampliamente utilizados en clínica, como la vancomicina.

S. aureus es una bacteria gram(+) que a menudo forma parte de la microbiota normal del cuerpo humano (469, 470). Se trata de un patógeno oportunista que, después de superar la barrera cutánea, puede causar una variedad de infecciones sistémicas y piógenas, infecciones agudas y crónicas y síndromes mediados por toxinas (471). Las infecciones por *S. aureus* varían desde infecciones de la piel y los tejidos blandos hasta neumonía necrotizante grave, endocarditis potencialmente mortal o bacteriemia en adultos y niños (472-475). Durante la era moderna de los antibióticos, *S. aureus* ha desarrollado la capacidad de adquirir resistencia a la mayoría de los antibióticos. Las cepas de *S. aureus* resistentes a meticilina (SARM) surgieron en la década de 1960 (476),

Antecedentes

siendo bacterias prevalentes y de gran relevancia, que actualmente se han extendido de manera global. Se ha convertido en una de las principales causas de infecciones bacterianas tanto nosocomiales como adquiridas en la comunidad (477). Las infecciones por SARM se asocian con tasas de mortalidad más altas que las infecciones causadas por cepas de *S. aureus* sensibles a meticilina (SAMS) (478). El tratamiento de elección actual para el tratamiento de infecciones graves por SARM es la vancomicina, un antibiótico glucopéptido que inhibe la biosíntesis de la pared celular (479, 480). Sin embargo, en los últimos 20 años han surgido aislados clínicos de *S. aureus* con resistencia intermedia a la vancomicina (SARIV, CMI = 4-8 µg/mL), cuya susceptibilidad reducida a la vancomicina se asocia con infecciones persistentes, fracaso del tratamiento y malos resultados clínicos. Además, aunque con menor frecuencia, se han descrito cepas de *S. aureus* con resistencia completa a la vancomicina (SARV, CMI ≥ 16 µg/mL) (481-483).

Estos organismos resistentes a múltiples fármacos plantean un desafío en el sistema sanitario actual. Por lo tanto, es necesario desarrollar nuevas estrategias de tratamiento para combatir las infecciones causadas por *S. aureus* resistente. Resulta indispensable centrar los esfuerzos en la investigación y el desarrollo de nuevos antibióticos, o la combinación de estos con tratamientos existentes para los que están surgiendo resistencias bacterianas.

Hipótesis: *Se plantea que la GT podría tener actividad antibacteriana significativa contra cepas de S. aureus de interés clínico por su resistencia a múltiples fármacos y que, en combinación con fármacos utilizados actualmente en clínica, potenciaría su efecto antibiótico, permitiendo una reducción en las dosis necesarias para el tratamiento.*

Por lo tanto, el **objetivo** principal de este capítulo es el estudio *in vivo* e *in vitro* de la actividad antibacteriana de GT en monoterapia y en combinación con otros antibióticos frente a *S. aureus*. Para llevarlo a cabo se han establecido una serie de **objetivos intermedios**:

- Estudiar la susceptibilidad antimicrobiana de cepas de *S. aureus* sensibles y resistentes a meticilina y con sensibilidad reducida a la vancomicina, frente a la GT.
- Estudiar la sinergia antimicrobiana de la GT en combinación con antimicrobianos utilizados actualmente en clínica.
- Determinar la concentración de prevención de mutaciones (CPM) de todos los antibióticos estudiados y la GT para *S. aureus*.

- Desarrollar un modelo de infección *in vivo* en *C. elegans*, optimizando su cultivo y sincronización, así como la posterior infección de larvas en estadio L4 con *S. aureus*.
- Estudiar el efecto *in vivo* de la GT en el modelo de infección desarrollado en *C. elegans*. Determinar la eficacia antibiótica y toxicidad de la GT en monoterapia y en combinación con otros antibióticos utilizados en clínica.

2.1.2 MATERIALES Y MÉTODOS

2.1.2.1 Cepas de *Staphylococcus aureus*

Las cepas SAMS ATCC 29213 y SARM/SARIV ATCC 700699 fueron proporcionadas por el Hospital Universitario Miguel Servet (Zaragoza). Se utilizaron para estudiar el efecto de la GT (Enzo Life Sciences, Inc., Farmingdale, NY, USA) a través de ensayos de susceptibilidad microbiana, sinergia y un modelo de infección *in vivo*.

El inóculo bacteriano se preparó independientemente para cada experimento. Los cultivos de *S. aureus* se prepararon en Luria-Bertani (LB, Sigma-Adrich, St. Louis, MO, USA), siendo incubados durante la noche a 37 ± 2 °C y agitación a 170 revoluciones por minuto (rpm) (Innova® 44 Incubator Shaker Series, Eppendorf SE, Germany). Se centrifugó a 1160 g durante 10 min (Beckman Coulter Allegra X-15) y se lavó con tampón fosfato salino (PBS), para posteriormente resuspenderlo en PBS.

La carga bacteriana requerida para cada experimento se ajustó determinando la densidad óptica a 600 nm (DO600) con un espectrofotómetro (DO600 DiluPhotometer, IMPLEN, Munich, Germany), en base a nuestra propia línea de unidades formadoras de colonias (UFC)/mL-DO600 calculada para estas cepas.

2.1.2.2 Ensayos de susceptibilidad antimicrobiana

Se siguieron los procedimientos clínicos microbiológicos recomendados por *The European Committee on Antimicrobial Susceptibility Testing* (EUCAST) (484) y por el *Clinical and Laboratory Standards Institute* (CLSI) (485) para realizar el método de difusión en agar. Se eligió un antibiótico con eficacia antimicrobiana conocida para cada cepa particular como control antimicrobiano. Se utilizaron estos antibióticos contenidos en discos comerciales (Thermo Scientific™ Oxoid™, Thermo Fisher Scientific, Waltham, MA, EE. UU.). Para la cepa sensible a meticilina se utilizó un control con cefoxitina (30 µg). Este control también se utilizó, además de ácido fusídico (10 µg), para la cepa resistente a meticilina. En el caso de la GT, se utilizaron discos no impregnados (Suministros Clínicos Lanau, Zaragoza, España). Éstos se depositaron en la placa y se les añadió 10 µg de GT (Enzo Life Sciences, Inc.; 10 µl de una solución a 1000 µg/ml). También se utilizaron discos no impregnados para añadirles 10 µl de la mezcla en la que se diluye la GT (agua:metanol 1,2:1) como control para descartar el efecto del metanol sobre los microorganismos. Los experimentos se llevaron a cabo por triplicado para cada cepa.

Se realizó un ensayo de microdilución en medio líquido para determinar la CMI de GT para *S. aureus* ATCC 29213 y ATCC 700699, por triplicado y de acuerdo con CLSI (486). Para identificar la concentración más baja requerida para que un antibiótico dado inhiba el crecimiento bacteriano, se introdujo una cantidad idéntica de bacterias en pocillos de medio líquido que contenían concentraciones progresivamente menores del fármaco. Se realizó utilizando diluciones seriadas de GT desde 125 hasta 0,49 µg/mL en LB o Mueller-Hinton (MH), agregando la suspensión bacteriana a una concentración final de 5×10^5 UFC/mL. Dado que se encontraron resultados similares usando LB y MH, se empleó LB para el resto de los experimentos. Las placas de microdilución se incubaron a 37 °C durante 24 horas. Se midió la DO630 de cada pocillo mediante un espectrofotómetro (BioTek Synergy HT) para determinar la concentración que exhibe inhibición del crecimiento de *S. aureus in vitro*. La CMI se determinó por triplicado utilizando cloxacilina (0,04 - 2 µg/ml), linezolid (0,13 - 8 µg/ml), ácido fusídico (0,01 - 0,5 µg/ml) y vancomicina (0,07 - 8 µg/ml). mL) como controles antibióticos (Sigma-Aldrich).

Los resultados se analizaron estadísticamente utilizando GraphPad Prism para Windows, (versión 5.0, GraphPad Software, San Diego, CA, EE. UU.). Los datos de absorbancia obtenidos en cada pocillo se compararon con los controles. A cada valor de absorbancia se restó el valor de DO630 obtenido en el control negativo, que solo contenía el medio, y se calculó el porcentaje de crecimiento o viabilidad con respecto al crecimiento en ausencia de antimicrobiano. También se realizó un análisis de varianza (ANOVA, por sus siglas en inglés, *ANalysis Of VAriance*) unidireccional de los datos correspondientes a GT respecto al control positivo.

2.1.2.3 Estudio de sinergia antimicrobiana

El efecto sinérgico de GT con fármacos antiestafilocócicos (cloxacilina, vancomicina, ácido fusídico y linezolid) contra *S. aureus* se probó mediante un ensayo de tablero de ajedrez (Figura 29). Esta prueba se utilizó para determinar el impacto sobre la potencia de la combinación de antibióticos en comparación con sus actividades individuales.

El ensayo de sinergia en tablero de ajedrez se realizó mediante el método de microdilución por triplicado, como se describió anteriormente (487). Para ello se utilizó una matriz bidimensional de concentraciones en serie de los compuestos de prueba.

Materiales y métodos

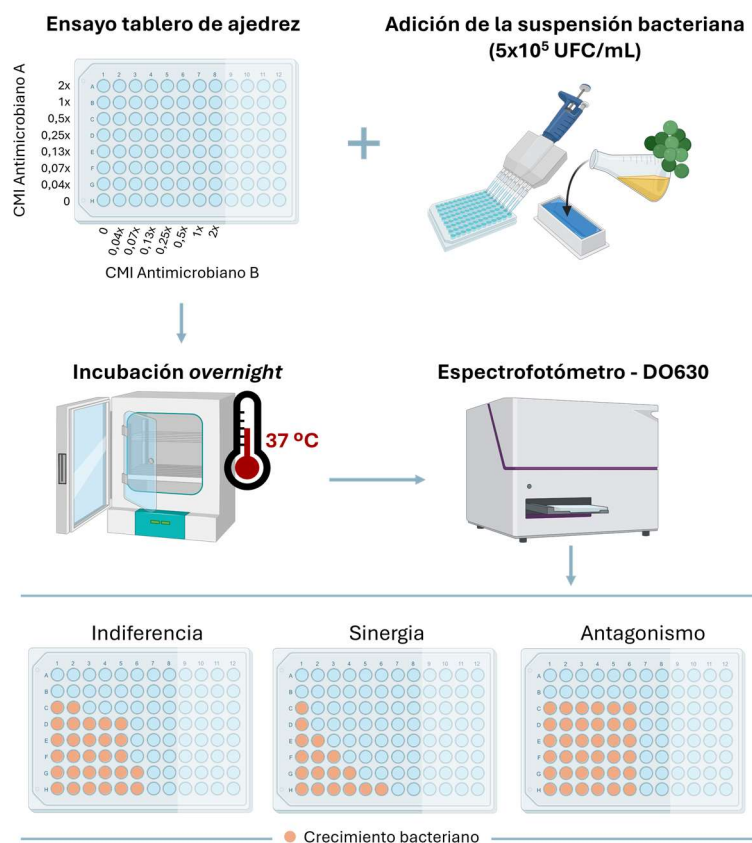


Figura 29. Representación gráfica de la técnica del tablero de ajedrez para el estudio de la sinergia entre dos antimicrobianos. En la parte inferior se muestra de manera visual y a grandes rasgos, los tres resultados posibles del estudio. CMI: Concentración Mínima Inhibitoria, DO630: Densidad óptica (DO) a 630 nm.

Se utilizaron tanto GT como antibióticos para probar la sinergia en diferentes combinaciones y concentraciones: se inició con dos o cuatro veces la CMI, y se prepararon diluciones seriadas en los pocillos correspondientes, creando un gradiente de concentraciones para cada compuesto:

- Para la cepa *S. aureus* ATCC 700699: GT (0,13 - 8 µg/ml) se combinó con vancomicina (0,13 - 8 µg/ml), linezolid (0,13 - 8 µg/ml) y ácido fusídico (0,01 - 0,5 µg/ml).
- Para la cepa *S. aureus* ATCC 29213: GT se combinó con vancomicina (0,07 - 4 µg/ml), linezolid (0,13 - 8 µg/ml), ácido fusídico (0,01 - 0,5 µg/ml) y cloxacilina (0,02 - 1 µg/ml).

La suspensión bacteriana se añadió a una concentración final de 5×10^5 UFC/mL. En cada placa se incluyeron controles de crecimiento y de esterilidad. Las placas de microtitulación se incubaron a 37 °C durante la noche antes de medir el valor de absorbancia mediante el espectrofotómetro a DO630. El porcentaje de crecimiento bacteriano medido en los experimentos se empleó como base para analizar el efecto de

los fármacos. Para ello se utilizó el software CalcuSyn (versión 2.1 para Windows, Biosoft: Chou, 1996–2007), que está diseñado para evaluar interacciones farmacológicas como sinergia o antagonismo. El programa CalcuSyn se basa en el método Chou-Talalay para combinación de fármacos (488), que calcula el índice de combinación (IC). El IC es una representación cuantitativa de la interacción farmacológica que tiene en cuenta tanto la potencia como la forma de la curva dosis-respuesta. El IC fue generado por el software CalcuSyn en un rango de niveles de fracción de células afectadas correspondientes a distintos grados de inhibición del crecimiento y se interpretó de acuerdo con la Tabla 7.

Tabla 7. Valores del índice combinatorio, símbolos recomendados y descripciones para clasificar el sinergismo o el antagonismo utilizando el método Chou-Talalay.

Rango del Índice de Combinación	Símbolo	Descripción
<0.1	+++++	Sinergia muy fuerte
0.1-0.3	++++	Sinergia fuerte
0.3-0.7	+++	Sinergia
0.7-0.85	++	Sinergia moderada
0.85-0.90	+	Sinergia leve
0.90-1.10	±	Casi aditivo
1.10-1.20	-	Antagonismo leve
1.20-1.45	--	Antagonismo moderado
1.45-3.3	---	Antagonismo
3.3-10	----	Antagonismo fuerte
>10	-----	Antagonismo muy fuerte

Fuente: CalcuSyn manual, Biosoft, 2006 (489)

Los criterios definidos más ampliamente sugieren que el índice de combinación, IC < 1, = 1 y > 1 indican sinergismo, efecto aditivo y antagonismo, respectivamente.

2.1.2.4 Determinación de la concentración de prevención de mutaciones

La CPM fue descrita por Dong *et al.* (490) como una medición *in vitro* de la susceptibilidad a los antimicrobianos y tiene en cuenta la probabilidad de que subpoblaciones mutantes estén presentes en poblaciones bacterianas de alta densidad. Para analizar dicha CPM se utilizó el método descrito anteriormente (491) ligeramente modificado.

Se realizó un método de microdilución en placas de 96 pocillos similar al ensayo de tablero de ajedrez para crear un gradiente de concentraciones en LB para GT y vancomicina solas o en combinación. Las concentraciones finales de los antibióticos fueron realizadas de manera seriada desde 256 veces la CMI de vancomicina y 128 veces la CMI de GT. Se

Materiales y métodos

añadieron a cada pocillo 50 µl de cultivo de *S. aureus* ATCC 700699 que contenía aproximadamente 2×10^9 UFC. Se utilizó un control de crecimiento positivo sin antimicrobiano y un control negativo solo con LB.

Las placas se incubaron a 37 °C durante 24 horas. A continuación, la CPM se registró como la concentración más baja para cada tratamiento individual y en combinación que impedía el crecimiento bacteriano. El crecimiento bacteriano se evaluó sembrando cada pocillo en placas de agar LB e incubándolo a 37 °C durante 24 horas.

La determinación del CPM para cada antimicrobiano y sus combinaciones fue la concentración más baja que no mostró crecimiento bacteriano en ninguno de los tres experimentos realizados.

2.1.2.5 Cultivo y sincronización de *Caenorhabditis elegans*

Se utilizaron gusanos de la cepa *C. elegans* SS104 para desarrollar un modelo de infección *in vivo*. La cepa SS104 es muy útil para producir grandes poblaciones de gusanos que carecen de una línea germinal debido a la mutación *glp-4*, que impide que los gusanos puedan generar descendencia cuando se cultivan a 25 °C (492). El mutante se obtuvo del *Caenorhabditis Genetics Centre* (<http://www.cbs.umn.edu/research/resources/cgc>).

Los nemátodos se propagaron en placas de agar con medio de crecimiento para nemátodos (NGM Lite, US Biological Life Sciences, Swampscott, MA, EE. UU.) suplementadas con estreptomicina 100 µg/ml (Sigma-Aldrich) y kanamicina 50 µg/ml (Sigma-Aldrich). Se cultivaron a 20 °C, utilizando *E. coli* OP50 (*Caenorhabditis Genetics Center*) como fuente de alimento.

Por lo general, los gusanos se encuentran en diferentes etapas de desarrollo en el cultivo. Por lo tanto, deben sincronizarse para poder utilizar una población de gusanos que estén en el mismo estadio de crecimiento y, por consiguiente, sea una cohorte homogénea y reproducible. El proceso de sincronización consiste en eliminar todos los nemátodos y obtener únicamente huevos viables. Para ello es necesario dejar crecer los gusanos durante el tiempo suficiente como para tener una población elevada de gusanos adultos que generen una gran cantidad de huevos, como se puede observar en la figura 30 (a). Los gusanos cultivados en placas de agar NGM se lavaron con suficiente tampón M9 (Na₂HPO₄, 6 g; NaCl, 5 g; KH₂PO₄, 3 g; H₂O destilada, 1 L; y 1 ml de MgSO₄ 1 M) para recoger todos los gusanos y despegar los huevos de las placas de agar. La suspensión de gusanos en tampón M9 se transfirió a tubos Falcon de 15 ml. La suspensión se centrifugó (Beckman Coulter

Allegra X-15) a 650 g durante 2 minutos. Se descartó el sobrenadante y se conservaron 2 ml de tampón M9 con una mezcla de gusanos en diferentes estadios de crecimiento, así como los huevos. Mediante el método de sincronización, la cutícula de las larvas y los gusanos adultos se debilita para liberar los huevos de los gusanos grávidos. Para ello, se preparó una solución que contenía 600 μl de NaOH 1 M y 600 μl de NaClO (lejía comercial) y se añadió a los gusanos. Esta mezcla se agitó durante diez segundos mediante un vortex y se tomó una alícuota para examinar los gusanos bajo un microscopio de disección (Leica DMi1 u Olympus IX81).

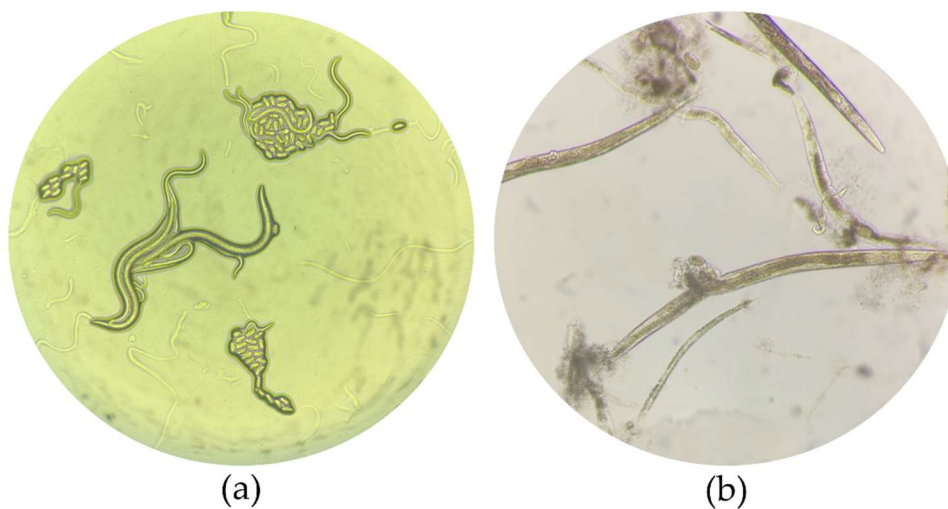


Figura 30. Fotografías reales de diferentes estados de *C. elegans* a través del microscopio. **(a)** Estado ideal de la placa de cultivo de gusanos antes de la sincronización. Nótese la gran cantidad de huevos. **(b)** Proceso de sincronización. Desintegración de la cutícula de lombrices grávidas y liberación de los huevos de su interior.

Si los gusanos permanecieran vivos o la cutícula de la mayoría de los gusanos no se hubiera roto, podrían ser necesarias más repeticiones de 10 segundos agitando en el vórtex hasta un tiempo máximo de contacto de 10 min. Es posible que se necesite una mayor concentración de NaOH para romper la cutícula si la suspensión contiene demasiados gusanos. Observar los gusanos regularmente en un microscopio proporcionó la mejor información para saber cuándo se debe detener el proceso. Se pudo observar cómo los gusanos más jóvenes se desintegraban y los huevos de las lombrices grávidas se liberaban cuando su cutícula se debilitaba y se rompía (Figura 30 (b)). En este punto, el tubo Falcon de 15 ml se relleno con tampón M9 y se centrifugó durante 2 minutos a 650 g. Este proceso se repitió dos veces más para reducir la cantidad de NaOH y NaClO restantes. El

sobrenadante del lavado M9 se descartó hasta que solo quedasen 0,5 ml con un pellet de huevos en el fondo del tubo.

El pellet de huevos se resuspendió y se sembró en una placa de agar NGM sin *E. coli* OP50 para permitir que los huevos eclosionen y reducir las diferencias de desarrollo en nuevas larvas debido a las diferencias en el tiempo de desarrollo de los huevos. A partir de ese momento, las placas se incubaron a 25 °C para que los gusanos no desarrollaran una línea germinal cuando crecieran y no fueran capaces de procrear. Se añadió *E. coli* OP50 a la placa de agar NGM 24 horas después. Como resultado de la sincronización, se obtuvo un conjunto de gusanos desarrollados y envejecidos de manera similar.

2.1.2.6 Cultivo e infección de larvas L4 de *Caenorhabditis elegans*

Las larvas L1 obtenidas mediante la sincronización se cultivaron a 25 °C hasta que los gusanos se desarrollaron hasta el estadio L4. Se optimizó el método para obtener larvas de *C. elegans* L4 infectadas con *S. aureus* mediante un ensayo en líquido, como se describió anteriormente (493).

Se llenó una placa de 96 pocillos de fondo plano de con *S. aureus* en medio líquido (80 % de tampón M9, 20 % de cultivo de *S. aureus* ATCC700699) y se transfirieron aproximadamente entre 15 y 20 nematodos adultos jóvenes sincronizados en cada pocillo.

El cultivo de *S. aureus* incubado durante la noche en LB contenía aproximadamente 1×10^9 UFC/ml. Se centrifugó a 1160 g durante 10 min para descartar el sobrenadante y se resuspendió en 4 mL de PBS. Este proceso fue repetido dos veces.

2.1.2.7 Ensayo de supervivencia de *Caenorhabditis elegans*

Se agregaron diferentes concentraciones de GT, vancomicina y combinaciones de ambos compuestos a las placas de 96 pocillos que contenían *C. elegans* infectado o no infectado para evaluar la efectividad y toxicidad de los tratamientos, respectivamente. Los pocillos que contenían únicamente *C. elegans*, *E. coli* OP50 y tampón M9 se utilizaron como controles. Las placas se incubaron a 25 °C y se controló la supervivencia de los gusanos cada 24 horas durante 7 días después de la exposición al patógeno y al tratamiento. Se realizaron cuatro experimentos independientes. En cada experimento, el número total de gusanos contados para cada grupo fue mayor o igual a 30.

La tasa de supervivencia se calculó contando el número de nemátodos vivos que quedaban en cada grupo en el momento correspondiente con un microscopio (Leica DMi1

u Olympus IX81). Distinguir los nematodos vivos de los muertos es, en esencia, un proceso simple. Sin embargo, su ejecución resultó considerablemente laboriosa y demandante. Los gusanos vivos se movían constantemente y adoptaban formas redondeadas (Figura 31 (a)). Cuando los gusanos comenzaron a alimentarse de *S. aureus* y este comenzó a colonizar su tracto intestinal, se pudo observar un engrosamiento y arrugamiento de la cutícula (Figura 31 (b)). Los nematodos muertos fueron fácilmente identificables, ya que son oscuros y rígidos, y a menudo adoptan forma de ángulo recto (Figura 31 (c)).

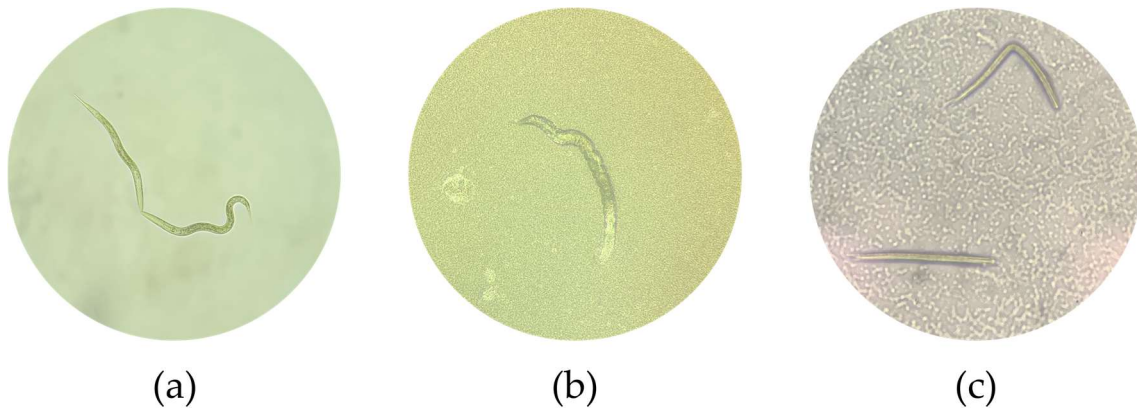


Figura 31. Fotografías reales de diferentes estados de *C. elegans* a través del microscopio. **(a)** Nematodos vivos, control alimentado con *E. coli* OP50. Forma fisiológica redondeada. **(b)** *C. elegans* comienza a ser infectado por *S. aureus*. Forma de cutícula rugosa y color más oscuro en el tracto intestinal. **(c)** Dos gusanos muertos, con las inconfundibles formas y ángulos rectos que caracterizan la muerte de estos nemátodos.

El análisis estadístico para el ensayo de supervivencia de *C. elegans* se realizó utilizando curvas de Kaplan-Meier y pruebas de rango logarítmico y de Gehan-Wilcoxon empleando el software GraphPad Prism. En este estudio fue necesario realizar comparaciones múltiples debido al hecho de que estábamos comparando varias curvas de supervivencia a la vez. Se utilizó ANOVA unidireccional seguido de pruebas de comparaciones *post-hoc* mediante la prueba de Bonferroni.

2.1.3 RESULTADOS

2.1.3.1 Susceptibilidad antimicrobiana de *Staphylococcus aureus* frente a gliotoxina

2.1.3.1.1 Efecto de la gliotoxina en el crecimiento de *Staphylococcus aureus*

A partir del resultado de los antibiogramas, se calculó la media de los halos de inhibición obtenida para cada disco impregnado (Figura 32). Todos los antibióticos utilizados como controles produjeron un halo de inhibición dentro del rango establecido por EUCAST 2020 específico para cada cepa. El halo de inhibición de los discos que contenían agua y metanol fue de 0 mm en todos los casos. Por lo tanto, estos resultados muestran que el vehículo para disolver GT no ejerce acción antimicrobiana en ninguno de los microorganismos bajo las condiciones testadas.

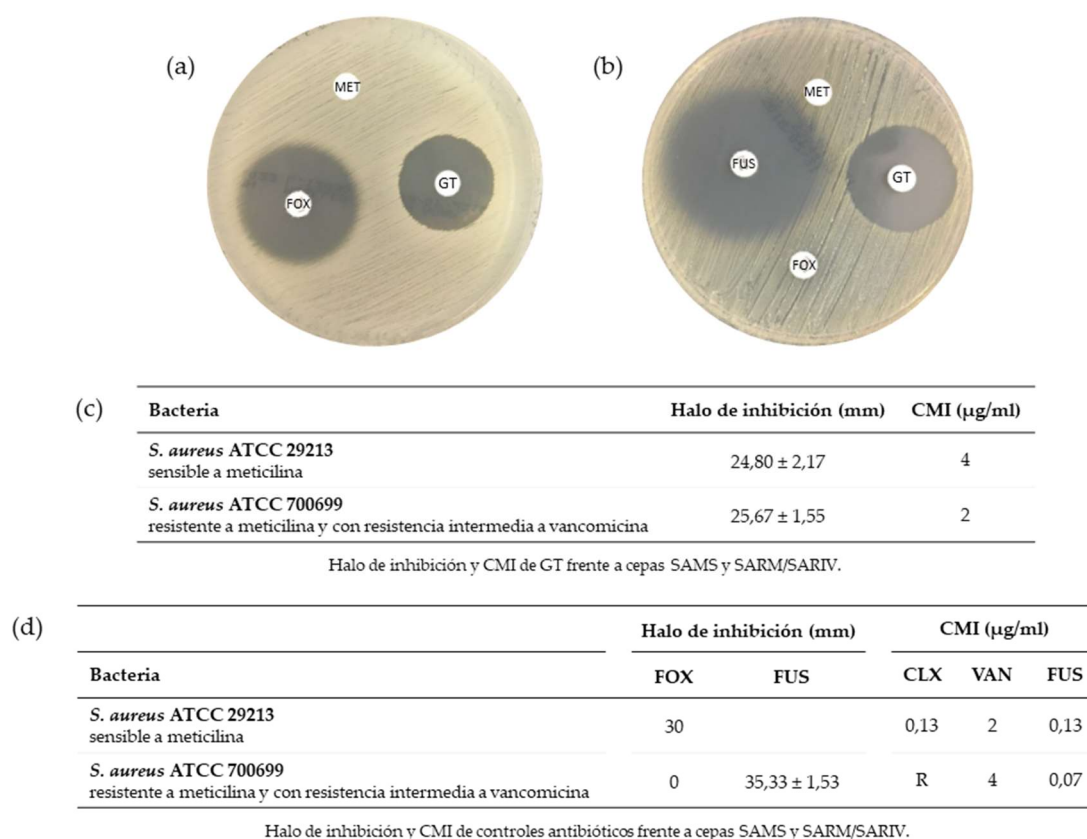


Figura 32. Actividad antimicrobiana de la GT frente a cepas de *S. aureus* sensibles a meticilina (SAMS) y resistentes a meticilina / con resistencia intermedia a vancomicina (SARM/SARIV). **(a)** Antibiograma de la cepa *S. aureus* ATCC 29213. **(b)** Antibiograma de la cepa *S. aureus* ATCC 700699. Un disco estéril que contiene la solución de metanol (MET) se colocó en la parte superior de ambas placas. El disco que contiene el control de antibiótico se colocó en la parte izquierda de la placa: (a) cefoxitina (FOX) 30 μg y (b) ácido fusídico (FUS) 10 μg . En el lado derecho, se colocó el disco impregnado con GT (10 μg). En el caso de la cepa resistente de *S. aureus* (b), se colocó un cuarto disco en la parte inferior de la placa, que contenía 30 μg de FOX. $n \geq 3$. **(c)** Tabla que contiene la media \pm la desviación estándar (SD) del halo de inhibición de crecimiento (diámetro en mm), así como la concentración mínima inhibitoria (CMI) ($\mu\text{g/ml}$) de GT. **(d)** Tabla que muestra la media \pm SD del halo de inhibición de crecimiento (diámetro en mm), así como la CMI ($\mu\text{g/ml}$) de los antibióticos control para cada cepa: cloxacilina (CLX), vancomicina (VAN) y FUS. $n = 3$.

En los antibiogramas se observaron claramente halos de inhibición alrededor de los discos que contenían GT, indicando que el compuesto es capaz de afectar negativamente al crecimiento de estas cepas. Es importante recalcar el efecto de inhibición del crecimiento de GT sobre la cepa resistente, donde el halo de inhibición ($25,67 \pm 1,55$) fue superior al halo obtenido con la cepa sensible ($24,80 \pm 2,17$).

2.1.3.1.2 Concentración mínima inhibitoria de gliotoxina y antibióticos frente a *Staphylococcus aureus*

En la figura 32 (c) se muestran los valores obtenidos para la CMI de GT en las dos cepas de *S. aureus*. El valor de la CMI de la cepa SARM/SARIV ($2 \mu\text{g/ml}$) fue menor que el valor de la CMI para la cepa SAMS ($4 \mu\text{g/ml}$), lo cual corrobora los resultados de los antibiogramas, donde el halo de inhibición de la cepa SARM fue mayor que el de la cepa SAMS para las mismas condiciones del experimento.

Como se muestra en la figura 32 - d, todos los antimicrobianos utilizados como control para la cepa *S. aureus* ATCC 29132 dieron como resultado una CMI que concuerda con el rango establecido por EUCAST (*Routine and extended internal quality control for MIC determination and disk diffusion as recommended by EUCAST, version 10.0, 2020*, <http://www.eucast.org>). En cuanto a la cepa *S. aureus* ATCC 700699, la susceptibilidad obtenida con los antibióticos utilizados como controles, corroboró que es una cepa SARM (mostró resistencia a la cloxacilina) y SARIV (CMI = $4 \mu\text{g/ml}$ frente a vancomicina).

2.1.3.2 Sinergia antimicrobiana de gliotoxina en combinación con fármacos antiestafilocócicos

Como se muestran en las figuras 33 y 34, varias de las combinaciones de GT y antibióticos probadas fueron sinérgicas, siendo superior a la que presentan los compuestos de manera individual.

Para la cepa SAMS, los resultados muestran actividad sinérgica significativa de GT con todos los antibióticos testados excepto con cloxacilina (Figura 33). En el caso de la combinación con vancomicina, se observó una disminución drástica en el porcentaje de crecimiento para todas las combinaciones, resultando en una sinergia fuerte o muy fuerte. En las combinaciones testadas de GT con linezolid, la reducción del porcentaje de crecimiento se observó en menor medida, pero nuevamente, todas las concentraciones probadas mostraron efectos sinérgicos, desde casi aditivos hasta sinergia muy fuerte, siendo la mejor combinación $\frac{1}{2}$ CMI de GT ($2 \mu\text{g/ml}$) en conjunto con linezolid en concentraciones de $0,13$ a $2 \mu\text{g/ml}$.

Resultados

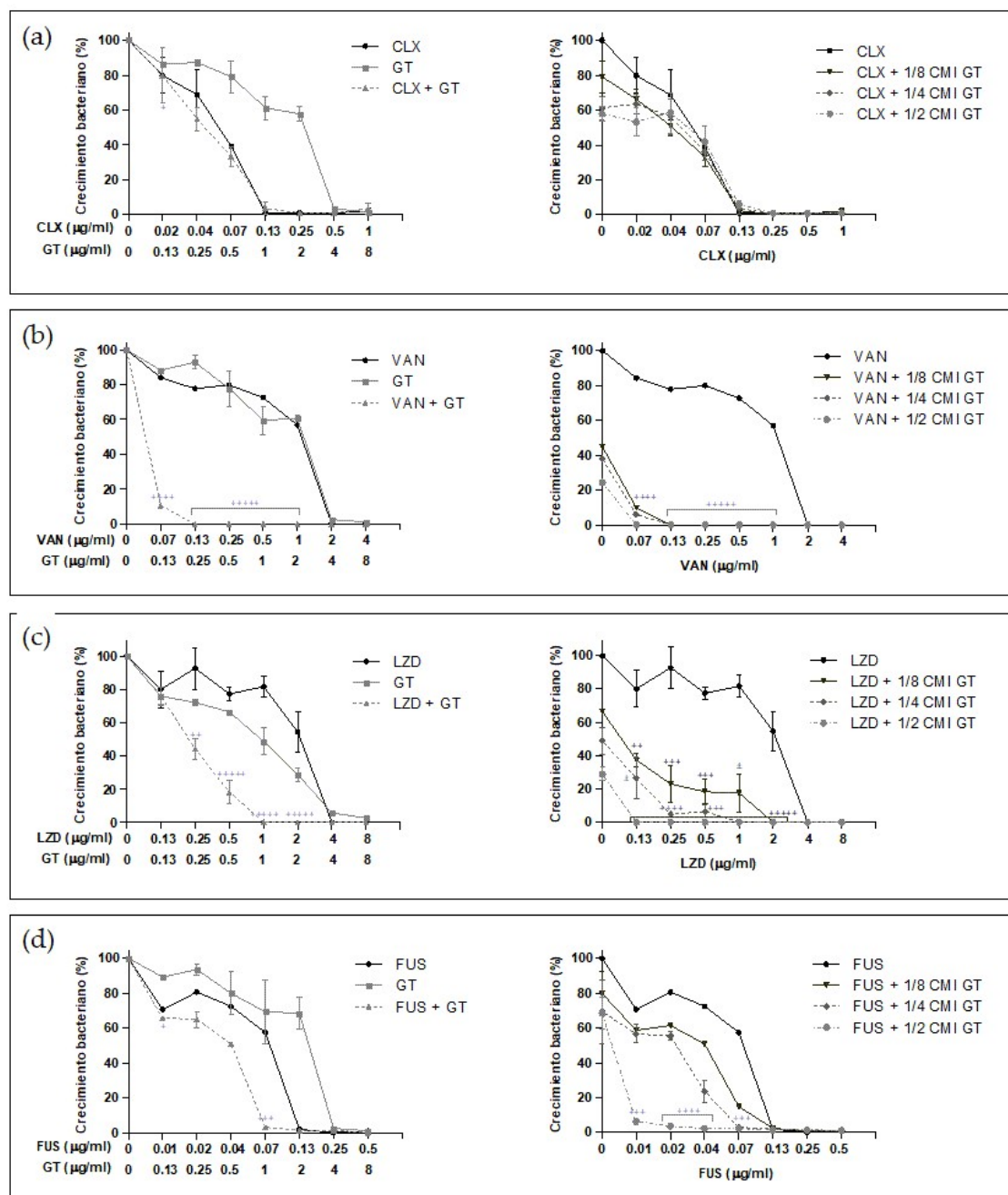


Figura 33. Resultados del ensayo de sinergia *in vitro* para la cepa SAMS ATCC 29213. El efecto de gliotoxina (GT) y los diferentes antibióticos en monoterapia o en combinación se testó contra *S. aureus*. **(a)** Combinación de GT y cloxacilina (CLX). **(b)** Combinación de GT y vancomicina (VAN). **(c)** Combinación de GT y linezolid (LZD). **(d)** Combinación de GT y ácido fusídico (FUS). En cada panel se muestra la concentración de cada compuesto y la media \pm desviación estándar (SD) del porcentaje del crecimiento bacteriano. Los símbolos corresponden a los diferentes niveles de sinergia calculados utilizando el software CalcuSyn: (\pm) Casi aditivo, (+) Sinergia leve, (++) Sinergia moderada, (+++) Sinergia, (++++), and (+++++) Sinergia fuerte y muy fuerte. CMI de GT = 4 $\mu\text{g/ml}$ para *S. aureus* ATCC 29213. $n = 3$.

La GT combinada con ácido fusídico mostró sinergia cuando se usaron 0,07 $\mu\text{g/ml}$ de ácido fusídico con 1 y 2 $\mu\text{g/ml}$ de GT. También mostró sinergia al combinar 2 $\mu\text{g/ml}$ de GT con 0,01 $\mu\text{g/ml}$ de ácido fusídico, donde se vio una gran reducción en el porcentaje de crecimiento. Cabe destacar que una concentración tan baja de ácido fusídico por sí solo

mostró alrededor del 30% de inhibición del crecimiento. Finalmente, 2 µg/ml de GT combinada con 0,02 y 0,04 µg/ml de ácido fusídico mostró una fuerte sinergia.

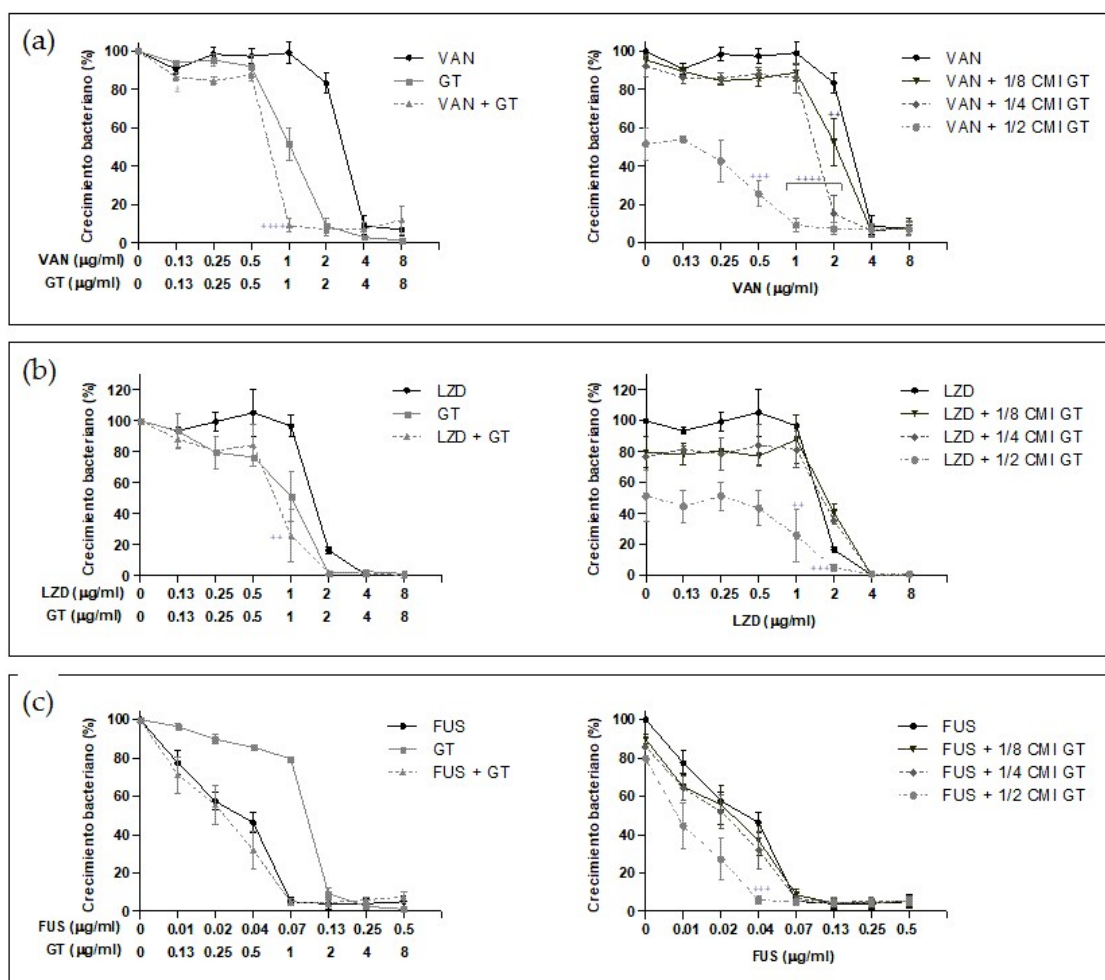


Figura 34. Resultados del ensayo de sinergia *in vitro* para la cepa SARM/SARIV ATCC 700699. El efecto de gliotoxina (GT) y los diferentes antibióticos en monoterapia o en combinación se testó contra *S. aureus*. (a) Combinación de GT y vancomicina (VAN). (b) Combinación de GT y linezolid (LZD). (c) Combinación de GT y ácido fusídico (FUS). En cada panel se muestra la concentración de cada compuesto y la media ± desviación estándar (SD) del porcentaje del crecimiento bacteriano. Los símbolos corresponden a los diferentes niveles de sinergia calculados utilizando el software CalcuSyn: (++) Sinergia moderada, (+++) Sinergia, (++++) Sinergia fuerte, and (+++++) Sinergia muy fuerte. CMI de GT = 2 µg/ml para *S. aureus* ATCC 700699. n = 3.

Respecto a la cepa SARM/SARIV, se demostraron los efectos sinérgicos de GT en combinación con vancomicina, linezolid y ácido fusídico (Figura 34). En este caso, el ácido fusídico a 0,04 µg/ml mostró sinergia al combinarlo con GT a una concentración ½ de la CMI para esta cepa (1 µg/ml). La GT a 1 µg/ml presentó sinergia moderada al combinarla con 1 µg/ml de linezolid y sinergia cuando se combinó con 2 µg/ml (que es la CMI de linezolid para esta cepa). Para la cepa SARM/SARIV, la combinación de compuestos con el mayor efecto sinérgico fue GT + vancomicina. Se observó sinergia al combinar 0,5 µg/ml de vancomicina con 1 µg/ml de GT, aunque el porcentaje de crecimiento excedió el 20%. A

Resultados

mayores concentraciones de vancomicina, se observó una sinergia muy fuerte y una gran reducción en el porcentaje de crecimiento, combinando 1 µg/ml de GT con 1 µg/ml de vancomicina y también combinando 0,5-1 µg/ml de GT con 2 µg/ml de vancomicina.

2.1.3.3 Concentración de prevención de mutaciones de gliotoxina y vancomicina solas o en combinación frente a *Staphylococcus aureus*

A continuación, se determinó la CPM de GT y vancomicina tanto individualmente como en combinación. Estos experimentos se llevaron a cabo utilizando *S. aureus* ATCC 700699, GT y vancomicina, debido a que GT ha demostrado ser un compuesto antiestafilocócico frente a estas cepas resistentes a meticilina y con resistencia intermedia a vancomicina y muestra sinergia en combinación con vancomicina. Cabe destacar que el estudio de cepas SARM y SARIV es muy relevante para el ámbito clínico y farmacéutico.

Todas las determinaciones de la CPM se realizaron por triplicado y los resultados obtenidos fueron idénticos. La Tabla 8 muestra las CPMs de los antimicrobianos individualmente y en combinación con vancomicina frente a la cepa SARM/SARIV.

Tabla 8. Concentración de prevención de mutaciones obtenida para GT, vancomicina y la combinación de ambas frente a *S. aureus* ATCC 700699. n = 3.

	Monoterapia		Combinación
Vancomicina	64	0	32
Gliotoxina	0	256	2

Concentración de Prevención de Mutaciones (µg/ml)

La CPM de vancomicina se redujo a la mitad cuando se combinó con una concentración muy baja de GT, específicamente la CMI de GT para esta cepa (2 µg/ml).

2.1.3.4 Eficacia antibiótica *in vivo* de gliotoxina como monoterapia y en combinación con vancomicina frente a *Staphylococcus aureus* en un modelo de *Caenorhabditis elegans*

La eficacia antibiótica de GT como monoterapia y en combinación con vancomicina frente a SARM se evaluó utilizando un modelo de infección en *C. elegans*. Los tratamientos seleccionados para este estudio de supervivencia fueron aquellos que obtuvieron resultados de actividad sinérgica en los experimentos *in vitro* con SARM.

El método de sincronización se optimizó para obtener un elevado número de nemátodos en el mismo estadio de desarrollo para realizar el ensayo de supervivencia. Para este propósito, se utilizaron placas de cultivo que contenían una gran concentración

de gusanos grávidos y huevos. El mejor resultado se obtuvo cuando se sincronizaron 2 placas, distribuyendo el buffer M9 con los gusanos en varios tubos Falcon de 15 ml, de manera que, una vez centrifugados, éstos contenían un pellet de 3 mm aproximadamente. La adición de la solución de lejía y NaOH a este pellet es suficiente si el tiempo de exposición es controlado.

Se observó una reducción significativa en el ratio de supervivencia de *C. elegans* infectados con SARM, utilizando ensayos en líquido (Figura 35) y obteniendo aproximadamente un 40% de supervivencia. En cambio, se observó mayor supervivencia cuando se utilizaron los tratamientos de GT y vancomicina, tanto en monoterapia como en combinación.

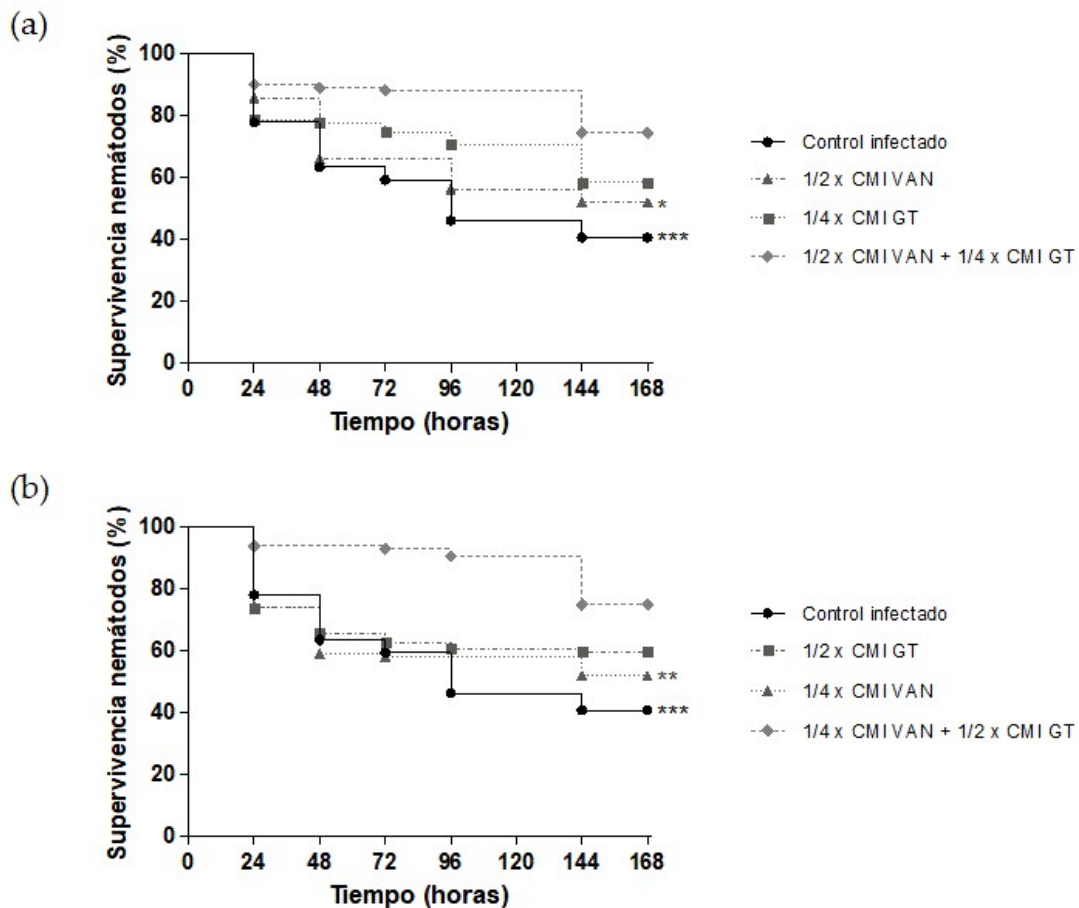


Figura 35. Curvas de supervivencia el ensayo in vivo del modelo de infección con SARM en *Caenorhabditis elegans*, que fue tratado con GT, vancomicina y combinaciones de ambas. El porcentaje de supervivencia de los nemátodos se representa por cada tratamiento y para el control infectado cada 24h durante 7 días. **(a)** Resultados del tratamiento en monoterapia de 1/2 CMI de vancomicina, 1/4 CMI de GT y la combinación de ambas concentraciones. **(b)** Resultados del tratamiento en monoterapia con 1/4 CMI de vancomicina, 1/2 CMI de GT y la combinación de estas dos concentraciones. *, **, *** Diferencias significativas ($p < 0,05$; $0,01$; $0,001$, respectivamente). $n = 4$.

Resultados

Las curvas de supervivencia se obtuvieron utilizando el programa GraphPad Prism (Figura 35). Se observó que el efecto sinérgico visto *in vivo* para dos de las combinaciones produjo un mayor porcentaje de supervivencia de los nemátodos comparado con el efecto de los antimicrobianos en monoterapia. La combinación de ½ CMI de vancomicina (2 µg/ml) con ¼ CMI de GT (0,5 µg/ml) resultó en un 74,55% de supervivencia de los gusanos, siendo 58,58% para GT y 52,11% para vancomicina a las mismas concentraciones en monoterapia. Por otro lado, la combinación de ¼ CMI de vancomicina (1 µg/ml) con ½ CMI de GT (1 µg/ml) resultó en 74,80% de supervivencia de los nemátodos. Las mismas concentraciones en monoterapia dieron como resultado un 59,60% de supervivencia para GT y un 51,79% para vancomicina. El análisis estadístico del estudio de supervivencia reveló que los resultados fueron estadísticamente significativos cuando comparamos la combinación de GT + vancomicina con el control infectado y con el efecto observado a la misma concentración de vancomicina en monoterapia.

El efecto tóxico de la GT y la vancomicina individualmente o en combinación se estudió también en *C. elegans* con las mismas concentraciones que han sido testadas en el modelo de infección (Figura 36).

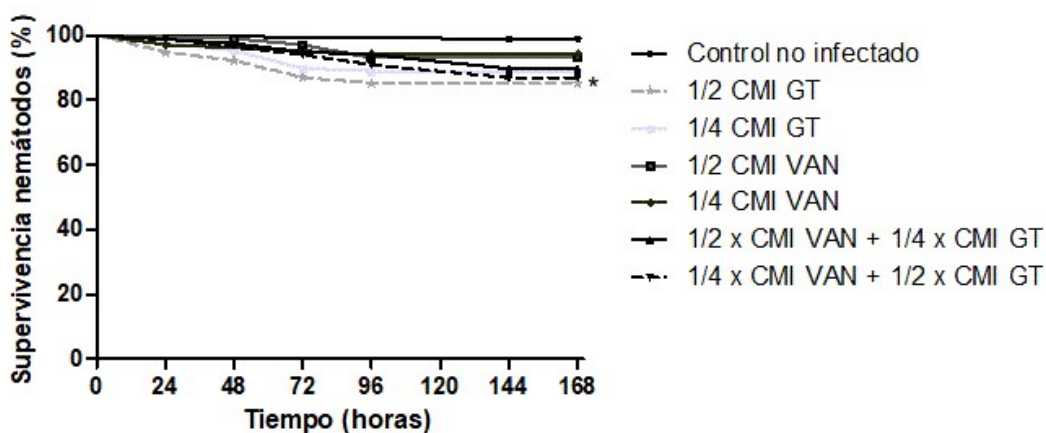


Figura 36. Test de toxicidad en *C. elegans* no infectados y tratados con GT, vancomicina y combinaciones de ambas. El porcentaje de supervivencia de los nemátodos se representa para cada tratamiento y para el control no infectado cada 24h durante 7 días. Resultados de los tratamientos en monoterapia de ½ CMI vancomicina, ½ CMI de GT, ¼ CMI vancomicina, ¼ CMI de GT y sus combinaciones. *, Diferencias significativas ($p < 0,05$) respecto al control no infectado. $n = 4$.

El porcentaje de supervivencia para la vancomicina fue del 93,00 % y 94,74 % para 2 y 1 µg/ml, respectivamente. En cuanto a la GT en monoterapia, la supervivencia se vio ligeramente disminuida, 85,22 % y 88,98 % para 1 y 0,5 µg/ml, respectivamente. aunque no fue significativamente diferente de los resultados de vancomicina. Además, respecto a las combinaciones, se mostró un 89,74 % de supervivencia para 2 µg/ml de vancomicina + 0,5

$\mu\text{g/ml}$ de GT. Adicionalmente, el 86,87 % de los nemátodos sobrevivieron en el caso de la combinación de 1 $\mu\text{g/ml}$ de vancomicina con 1 $\mu\text{g/ml}$ de GT. Por consiguiente, estos resultados indican que la GT podría mostrar un leve efecto tóxico en las concentraciones testadas, aunque este efecto no es significativamente diferente del observado para la vancomicina.

2.1.4 DISCUSIÓN

S. aureus es uno de los principales causantes de infecciones de la piel y tejidos blandos en todos los grupos de edad. También existe evidencia de un aumento en las infecciones por cepas SARM adquiridas en la comunidad (SARM-AC) en todo el mundo (494). El manejo de las infecciones causadas por SARM-AC pueden ser un reto debido a la resistencia a los antibióticos. Por lo tanto, es necesario estudiar nuevos antibióticos, tanto en monoterapia como en combinación con tratamientos existentes, con el fin de poder abordar eficazmente estas infecciones.

Desde su descubrimiento, la GT ha despertado interés en la comunidad científica con el objetivo de explotar sus actividades antimicrobianas. Ha sido testada frente a diferentes patógenos, como virus, hongos y bacterias (104, 468, 495, 496). Nuestros hallazgos muestran la actividad antimicrobiana de GT contra cepas de *S. aureus*, tanto sensibles a meticilina como resistentes a meticilina con susceptibilidad reducida a vancomicina (es decir, cepas SARM/SARIV). En cuanto a la cepa SAMS, el único estudio realizado con la misma cepa encontrado en la literatura (168) resultó en una CMI = 3,98 µg/ml, un valor prácticamente equivalente al nuestro (CMI = 4). Algunos investigadores han estudiado el efecto de la GT contra diferentes cepas de SARM, obteniendo un rango de CMI entre 0,5-4 µg/ml (168, 497, 498), confirmando la validez de los resultados que obtuvimos en este estudio para la cepa SARM ATCC 700699 (CMI = 2 µg/ml). Es importante destacar que nuestros resultados proporcionan observaciones novedosas y amplían los análisis anteriores al mostrar que la GT presenta actividad no solo contra SARM, sino también contra una cepa SARM/SARIV.

La frecuencia del uso de vancomicina como el principal fármaco para tratar las infecciones por SARM es una consecuencia directa de la propagación mundial de clones de SARM resistentes a múltiples fármacos durante las últimas décadas. Este aumento de la presión selectiva ha dado lugar a la aparición de aislados de SARM con susceptibilidad reducida a la vancomicina y, más recientemente, a la aparición de cepas SARIV con un alto nivel de resistencia a antibióticos (499). En la mayoría de los casos, las cepas SARIV han surgido en pacientes con infecciones por SARM sometidos a tratamientos prolongados con vancomicina, lo cual a menudo ha resultado en un fracaso del tratamiento (500-503). Actualmente, se ha informado de un gran número de cepas clínicas de *S. aureus* con susceptibilidad reducida a la vancomicina, incluidas las cepas SARIV heterorresistentes

(subpoblaciones resistentes entre la población bacteriana total de la cepa, que pueden seleccionarse mediante tratamiento) (504).

Por todas estas razones, la cepa que seleccionamos para los ensayos de este estudio fue una cepa SARM/SARIV, ATCC 700699. Los resultados obtenidos son novedosos, ya que no existe evidencia científica que determine el efecto de la GT para las cepas SARIV. Aunque los determinantes moleculares de la resistencia no están completamente determinados, está claro que la adquisición secuencial de mutaciones puntuales puede conducir a la resistencia (502). En este estudio demostramos que la combinación de GT con vancomicina disminuye su CPM (concentración mínima que inhibe el crecimiento del mutante bacteriano más resistente generado en un solo paso mutacional), reduciendo así la concentración más baja del fármaco requerida para bloquear el crecimiento de la célula menos susceptible presente en poblaciones bacterianas de alta densidad.

Para pacientes con infecciones recurrentes por SARM, se debe considerar la minimización de la exposición a vancomicina para prevenir la aparición de una infección por SARIV. Las alternativas podrían ser el desbridamiento agresivo (eliminación exhaustiva de tejido necrótico, infectado o desvitalizado de la herida) o la administración de antimicrobianos no glicopeptídicos, cuando sea apropiado (501). Por lo tanto, se podría argumentar que una terapia combinada que utilice una concentración más baja de vancomicina puede ser apropiada en un intento de evitar la selección de mutantes resistentes en infecciones graves por *S. aureus*. La vancomicina administrada en dosis adecuadas junto con otros agentes activos aún puede presentar una opción relativamente segura para el tratamiento de pacientes con infecciones por SARM una vez que el aislado clínico se haya probado adecuadamente (502). Además de disminuir el riesgo de la aparición de resistencias asociadas al uso prolongado o a altas dosis de antibióticos, también se podrían disminuir los efectos secundarios en los pacientes al administrar una dosis reducida de dichos fármacos.

Se estudió el efecto sinérgico de la combinación de GT con diferentes antibióticos empleados en clínica para el tratamiento de infecciones por *S. aureus*. Los resultados *in vitro* muestran que la GT es capaz de potenciar el efecto de estos fármacos, permitiendo reducir la concentración a la que son eficaces para las cepas SAMS y SARM/SARIV. Los resultados son especialmente alentadores cuando se combina GT con vancomicina, ya que se obtiene un potente efecto sinérgico al reducir la concentración de vancomicina a la

Discusión

mitad o incluso a un cuarto de la CMI, obteniendo un potente efecto antibiótico. Los resultados *in vivo* del modelo de infección en *C. elegans* validan los hallazgos *in vitro* y demuestran el efecto potenciador de la GT cuando se combina con vancomicina, lo que da como resultado un efecto antibiótico significativamente superior cuando se combinan ciertas concentraciones. El mecanismo de este efecto sinérgico es actualmente desconocido y requerirá pruebas experimentales adicionales. Se ha descubierto que varias dianas de *S. aureus* son afectadas por la vancomicina, alterando la síntesis de la pared celular, la permeabilidad y la síntesis de ARN, y se han propuesto diferentes mutaciones en genes involucrados en estos procesos como mediadores de la resistencia (499). Se ha descubierto que la GT se une directamente a los residuos Cys de las proteínas, modificando su función tanto en células eucariotas como procariotas. Además, las ROS inducidas por la GT pueden afectar a las proteínas, los lípidos y el ADN. Por lo tanto, es posible que el efecto de la GT sobre estas moléculas module la resistencia a la vancomicina en *S. aureus* y, por tanto, promueva un efecto sinérgico entre ambos antimicrobianos. Sin embargo, esta es una hipótesis que requerirá validación experimental.

La GT, en las concentraciones estudiadas, apenas es tóxica para *C. elegans*. Sin embargo, dado que se ha demostrado que la GT es tóxica para diferentes tipos de células de mamíferos (161), se requiere un estudio toxicológico detallado que involucre células de la piel (el tejido principal donde ocurren infecciones graves por *S. aureus* resistente (494)), así como modelos *in vivo* de mamíferos. Se necesitarán más estudios para respaldar la viabilidad de la aplicación clínica de las dosis efectivas más bajas de GT combinadas con vancomicina. Además, podría ser importante considerar el desarrollo de formas farmacéuticas que no atraviesen la capa córnea de la epidermis o que tengan poca capacidad de penetración y, por otro lado, el uso de vehículos para la administración de fármacos dirigidos al sitio de acción específico, y así reducir la entrada de GT en la circulación sistémica.

A la espera de resolver estas posibles limitaciones para el futuro desarrollo clínico, nuestros resultados muestran que la GT presenta actividad antimicrobiana *in vitro* e *in vivo* contra una cepa SARM/SARIV clínicamente relevante. Cabe destacar que su capacidad para mejorar sinérgicamente el efecto de la vancomicina presenta perspectivas prometedoras para el desarrollo de nuevos fármacos que ayuden en el tratamiento de cepas resistentes a los medicamentos y reduzcan la aparición de cepas resistentes a la vancomicina.

Capítulo

II

Estudio de la producción de gliotoxina y bismetilgliotoxina por diferentes especies fúngicas y su utilidad como biomarcador en aspergilosis invasiva

Índice

Antecedentes.....	95
Material y métodos.....	98
Resultados.....	107
Discusión.....	126

2.2 CAPÍTULO 2 – Estudio de la producción de gliotoxina y bismetilgliotoxina por diferentes especies fúngicas y su utilidad como biomarcador en aspergilosis invasiva

2.2.1 ANTECEDENTES

Como se ha descrito anteriormente, la GT es un metabolito secundario fúngico capaz de generar un impacto en la salud y la agricultura, por lo que resulta relevante estudiar qué especies de hongos son capaces de producirla. Actualmente existe una gran controversia sobre la producción de GT en diferentes especies fúngicas (136, 139-142). De las diversas secciones de *Aspergillus*, se ha demostrado que solo las asociadas con aspergilosis como *A. fumigati*, *A. terreus*, *A. flavus* y *A. niger*, son productoras de GT (139, 179). Más del 90% de los aislados en un estudio de casos de AI en centros de cáncer de atención terciaria fueron productores de GT (179). Además, la cantidad de GT producida por *A. fumigatus* es sustancialmente mayor que la cantidad producida por otras especies menos frecuentes de *Aspergillus* patógenas, incluidos *A. terreus* y *A. flavus* (139, 179). Además de *Aspergillus spp.*, también se ha sugerido que varias especies de hongos sintetizan GT, como *T. lignorum* (135), *T. vires* (136), *P. obscurum* (136), *G. fimbriatum* (116) y *C. albicans* (137), aunque cuál de ellos es capaz de producirlo en diferentes condiciones es un tema controvertido.

Como muchos otros metabolitos secundarios producidos por hongos, las enzimas que sintetizan la GT están codificadas en un grupo de genes. Como se ha expuesto en la introducción, *gliP* es uno de los genes más relevantes en la producción de GT, ya que codifica una NRPS responsable del primer paso en la biosíntesis de GT (120) y las cepas deficientes en *gliP* no son capaces de producirla (121).

Hipótesis: Aunque los clústeres de genes implicados en la síntesis y regulación de GT varían entre especies de hongos en términos de número de genes, sintenia y disposición, existen similitudes conservadas que sugieren una homología funcional entre ellos. Se puede esperar que las NRPS con alta homología de secuencia sean ortólogas a *GliP* y realicen funciones biosintéticas similares o idénticas. Asumiendo esta homología, es probable que las enzimas auxiliares codificadas en los genes del clúster conserven tanto su secuencia como su función en cuanto a la producción de micotoxinas.

Algunos de estos hongos son los causantes de graves infecciones que representan una complicación preocupante en pacientes inmunocomprometidos. En ausencia de síntomas específicos de la enfermedad, es importante establecer nuevos indicadores biológicos que se puedan utilizar durante la profilaxis activa contra el hongo (505). En los

Antecedentes

últimos años se han producido avances significativos en el diagnóstico de AI en pacientes hematológicos. La AI se clasifica de acuerdo con los criterios EORTC/MSK en las categorías de enfermedad probada, probable o posible, en base a factores del hospedador, criterios clínicos y criterios micológicos (67). En las últimas décadas, se han desarrollado varias técnicas de diagnóstico serológico y molecular para detectar la presencia de infección por *Aspergillus spp.* (506). Algunos de los biomarcadores utilizados son GM (componente de la pared celular de hongos del género *Aspergillus* liberado durante su crecimiento), BDG (componente de la pared de los hongos), además de la PCR para detectar la presencia de *Aspergillus* (71) o algunos marcadores en fase de estudio como CMV (la infección causada por este virus supone un factor de riesgo para padecer AI (507, 508)). Sin embargo, su precisión es limitada y por el momento no existe un método diagnóstico temprano y preciso. Debido a la elevada morbimortalidad de la AI, es fundamental concentrar los esfuerzos en su diagnóstico precoz, especialmente en pacientes inmunocomprometidos.

La bmGT (metabolito inactivo de la GT) ha demostrado ser un biomarcador estable para la AI, siendo posible detectarla y cuantificarla mediante cromatografía en capa fina de alto rendimiento (HPTLC, del inglés, *high-performance thin layer chromatography*) con una alta sensibilidad y especificidad (148). Aunque ambos compuestos se pueden detectar en cultivos del hongo, solo la bmGT es estable en sangre completa, lo que indica que puede ser un mejor biomarcador para la detección *in vivo*. De hecho, en algunos casos, la bmGT se ha detectado antes de que se obtuviera evidencia micológica de infección (148). Sin embargo, el uso de bmGT como biomarcador podría presentar algunas limitaciones, principalmente debido a la falta de un protocolo universalmente aceptado para su cuantificación, los costes y disponibilidad de la tecnología para ello, la sensibilidad del método de detección y el impacto que el tratamiento profiláctico podría ejercer en la liberación del compuesto.

De la misma manera, los biomarcadores utilizados actualmente en clínica como GM o BDG también presentan limitaciones, puesto que exhiben FPs y FNs con asiduidad debido a diversas situaciones, entre ellas, los tratamientos administrados, infecciones concomitantes o factores clínicos de los pacientes (69, 76-79, 509).

Por otro lado, se ha demostrado que los niveles de IL-8 están significativamente aumentados en sueros de pacientes que padecen aspergilosis pulmonar invasiva (510-

512), lo que resalta su potencial como biomarcador para el diagnóstico y monitorización de esta enfermedad.

Habitualmente, se recopilan grandes cantidades de datos durante el tratamiento de pacientes hematológicos de alto riesgo y estos datos pueden ser utilizados de manera combinada para realizar pronósticos más precisos del diagnóstico de AI. Algunos investigadores han utilizado técnicas de ML como árboles de decisión para crear modelos predictivos del riesgo de padecer AI (513, 514).

Hipótesis: *Debido a su simplicidad y sensibilidad, una tecnología de cuantificación de bmGT basada en HPTLC podría implementarse en los laboratorios clínicos para ayudar en el diagnóstico de AI. Tanto bmGT como IL-8 podrían ser unos biomarcadores útiles en combinación con otros marcadores utilizados actualmente en clínica para el diagnóstico temprano de esta infección fúngica potencialmente mortal.*

Por consiguiente, el **objetivo** principal de este capítulo es el análisis de la producción de GT y bmGT en diversas especies fúngicas, así como la aplicación de bmGT junto con otros biomarcadores utilizados actualmente en clínica (GM, BDG) o en fase de estudio (CMV) para el diagnóstico de AI en pacientes hematológicos. Para desarrollar este objetivo, se han establecido una serie de **objetivos intermedios**:

- 🌱 Aplicar estrategias bioinformáticas para la búsqueda y estudio de las estructuras proteicas más similares al clúster *gli*, en concreto, al *gliP* de *A. fumigatus*, en varias especies de hongos de interés clínico.
- 🌱 Llevar a cabo un estudio filogenético con las secuencias de las proteínas homólogas.
- 🌱 Analizar y cuantificar la producción de GT y bmGT en varias cepas fúngicas en cultivos *in vitro*.
- 🌱 Detectar y cuantificar bmGT en suero de pacientes mediante su extracción y posterior análisis en HPTLC. Además, detectar y cuantificar IL-8 en esos mismos sueros.
- 🌱 Procesar y unificar los datos clínicos y analíticos disponibles para obtener un conjunto de datos adecuado y consistente que permita un análisis riguroso.
- 🌱 Realizar un análisis exploratorio y estadístico de los datos: En base a ello, analizar la capacidad de los biomarcadores estudiados para predecir el diagnóstico de AI.

2.2.2 MATERIALES Y MÉTODOS

2.2.2.1 Análisis de la producción de gliotoxina y bismetilgliotoxina en diferentes especies fúngicas

2.2.2.1.1 Estudio bioinformático

2.2.2.1.1.1 Búsqueda de proteínas homólogas de Af293 GliP de *Aspergillus fumigatus* en otras especies

La búsqueda de genes y proteínas homólogas de GliP, una NRPS de *A. fumigatus* Af293, en otras especies relacionadas se realizó utilizando las herramientas de alineamiento de secuencias en línea BLASTn y BLASTp (515) (*Basic Local Alignment Search Tool*, <https://blast.ncbi.nlm.nih.gov/Blast.cgi>, consultado el: 26 de marzo de 2020).

La selección de las posibles proteínas homólogas de GliP en otras especies se definió por el *query coverage* > 90 % (porcentaje de la longitud total de la secuencia de consulta que está alineada con una secuencia en la base de datos) y la identidad de la secuencia de aminoácidos > 35 % (porcentaje de residuos que son exactamente iguales entre las secuencias dentro de la región alineada). Después de la búsqueda, se eligieron 22 secuencias homólogas de GliP pertenecientes a diferentes especies de hongos. Se realizó un alineamiento de secuencias múltiples con las secuencias seleccionadas utilizando Clustal Omega (<http://www.clustal.org>, consultado el: 26 de marzo de 2020). Esta herramienta en línea utiliza árboles guía y técnicas de perfil-perfil basadas en modelos probabilísticos *Hidden Markov Models* (HMM) para comparar perfiles de secuencias, generando así alineamientos entre tres o más secuencias (516).

Para comprender fácilmente los resultados, los alineamientos de secuencias se visualizaron con el software Jalview (517). Este software pudo mostrar gráficamente los residuos y regiones conservadas entre los diferentes NRPS seleccionados.

2.2.2.1.1.2 Análisis filogenético

Los análisis filogenéticos facilitan la comprensión de los roles funcionales de los dominios conservados en las secuencias. En este estudio se analizaron las proteínas previamente seleccionadas por ser similares al NRPS de *A. fumigatus* Af293.

Se desarrolló un árbol filogenético utilizando el software *Molecular Evolutionary Genetics Analysis across computing platforms* (MEGA X) (518). Este análisis involucró 22 secuencias de aminoácidos. Hubo un total de 2379 posiciones (longitud total del alineamiento incluyendo espacios insertados) en el conjunto de datos final. Se seleccionó el método de máxima verosimilitud y el modelo Le_Gascuel_2008 (519) para inferir la

historia evolutiva. Se midió la bondad de ajuste de varios modelos a nuestros datos y finalmente se eligió el modelo Le_Gascuel_2008 en función de los valores del criterio de información bayesiano (BIC, por sus siglas en inglés, *Bayesian information criterion*) (520) y el criterio de información de Akaike (AIC, por sus siglas en inglés, *Akaike information criterion*) (521). Se utilizó una distribución Gamma discreta para modelizar las diferencias en las tasas evolutivas entre sitios (+G, parámetro = 1,6080). El modelo de variación de la tasa permitió que algunos sitios fueran evolutivamente invariables ([+I], 5,76 %).

El software MEGA X elaboró un árbol inicial obtenido automáticamente al aplicar los algoritmos Neighbor-Join y BioNJ a una matriz de distancias por pares calculada utilizando el modelo Jones-Taylor-Thornton (JTT). Después, se realizó una búsqueda heurística y se seleccionó la topología que presentaba la mayor verosimilitud (*log-likelihood* más alto), que con mayor probabilidad representaba correctamente las relaciones evolutivas.

Una de las pruebas más utilizadas para medir la fiabilidad de un árbol inferido es la prueba *bootstrap* de Felsenstein (522), que se evalúa utilizando la técnica de remuestreo de Efron (523). Este método *bootstrap* se realizó con 100 réplicas. Los valores *bootstrap* en cada clado indican cuántas veces de cada 100 (en nuestro caso) se encuentran los mismos taxones agrupados al repetir la reconstrucción filogenética en un conjunto de datos remuestreados. En general, si el valor *bootstrap* para una rama interior dada fue del 95 % o superior, entonces la topología en esa rama se consideró correcta (524).

2.2.2.1.2 Análisis y cuantificación de la producción de gliotoxina y bismetilgliotoxina por diferentes cepas fúngicas *in vitro*

2.2.2.1.2.1 Cepas fúngicas y condiciones de cultivo

A. fumigatus 1631562 y *A. lentulus* 353 fueron cedidos por el servicio de microbiología del hospital Miguel Servet. *A. fumigatus* B5223 y *P. expansum* MD-8 fueron proporcionados por J. A. Sugui y A. R Ballester, respectivamente. *A. pseudofischeri* CBS 404.67, *A. fischeri* CBS 420.96, *A. turcosus* CBS 140371, *Rhizodiscina lignyota* CBS 133067, *Trichoderma parareesei* CBS 125925, *Colletotrichum fructicola* CBS 120005, *Penicillium flavigenum* CBS 110407, *Trichoderma harzianum* CBS 226.95, *Trichoderma reesei* CBS 383.78, *Penicillium zonata* CBS 506.65, *Trichoderma virens* CBS 249.59 y *Elsinoe ampelina* CBS 208.25 se adquirieron en el Centro de Biodiversidad Fúngica (CBS) de los Países Bajos.

Materiales y métodos

Las cepas de hongos se cultivaron en agar de glucosa Sabouraud con 50 mg/L de cloranfenicol (Merk, Darmstadt, Alemania) durante 7 días en diferentes condiciones para cada especie con el fin de producir una cantidad significativa de conidios como se describe en la tabla 9. Las esporas se recolectaron con un hisopo impregnado con Tween-20 (Thermo-Fischer, Waltham, MA, EE. UU.) y se resuspendieron en agua esterilizada. Las suspensiones de conidios se ajustaron a 2 McFarland y se vertió 1 mL de esta suspensión en matraces de cultivo celular que contenían 9 mL de medio Czapek-Dox (Thermo-Fischer, Waltham, MA, EE. UU.) o 9 mL de medio *Roswell Park Memorial Institute* 1640 (RPMI 1640) con 20 mM de ácido 4-(2-hidroxietil)-1-piperazinaetanosulfónico (HEPES) y sin NaHCO₃ (Merk, Darmstadt, Alemania). Se cultivaron durante 4 días a 37 °C. Después de 4 días, los sobrenadantes se filtraron con filtros de 0,22 µm y se almacenaron a -20 °C hasta su uso posterior. Si una especie, como *A. lentulus* o *E. ampelina*, no podía esporular en el tiempo o medio indicados, se cortaba un pequeño trozo de micelio de 1 × 1 cm de la placa de agar y se insertaba en matraces de cultivo celular.

Tabla 9. Condiciones de cultivo utilizadas para las cepas estudiadas. Si no se especifica, las cepas se cultivaron en la oscuridad. Las condiciones del cultivo en líquido fueron las mismas para el medio RPMI 1640 y el medio Czapek-Dox. TA: temperatura ambiente.

Cepa fúngica	Condiciones de cultivo sólido	Condiciones de cultivo líquido
<i>A. fumigatus</i> B5233	37 °C	37 °C
<i>A. fumigatus</i> 1631562	37 °C	37 °C
<i>A. fischeri</i> CBS 420.96	37 °C	37 °C
<i>A. lentulus</i> 353	37 °C	37 °C
<i>A. turcosus</i> CBS 140371	37 °C	37 °C
<i>A. pseudofischeri</i> CBS 404.67	37 °C	37 °C
<i>P. flavigenum</i> CBS 110407	30 °C	30 °C
<i>P. expansum</i> MD-8	30 °C	30 °C
<i>B. victoriae</i> CBS 174.57	30 °C	30 °C
<i>E. ampelina</i> CBS 208.25	30 °C	30 °C
<i>C. cylindrospora</i> CBS 449.70	30 °C	30 °C
<i>C. fructicola</i> CBS 120005	30 °C	30 °C
<i>T. virens</i> CBS 249.59	TA, luz	TA, luz
<i>P. zonata</i> CBS 506.65	TA, luz	TA, luz
<i>T. reesei</i> CBS 383.78	TA, luz	TA, luz
<i>R. ligniota</i> CBS 133067	30 °C	30 °C

Cepa fúngica	Condiciones de cultivo sólido	Condiciones de cultivo líquido
<i>T. harzianum</i> CBS 226.95	TA, luz	TA, luz
<i>T. parareesei</i> CBS 125925	TA, luz	TA, luz

2.2.2.1.2.2 Extracción química de gliotoxina y bismetilgliotoxina

Se añadieron 3 ml del sobrenadante del cultivo fúngico a 10 ml de diclorometano (DCM) en un frasco de vidrio de 20 ml y se agitó en un vórtex durante 30 segundos. La fase orgánica que contenía las toxinas se transfirió a otro frasco y se evaporó bajo flujo de nitrógeno. El residuo sólido se disolvió en 1 ml de DCM y se evaporó en un vial de vidrio de 2 ml bajo flujo de nitrógeno nuevamente. Finalmente, el residuo sólido se disolvió en 55 % de agua MiliQ y 45 % de metanol para su posterior análisis y cuantificación.

2.2.2.1.2.3 Análisis y cuantificación mediante HPLC

Se utilizó cromatografía líquida de alta eficacia (HPLC, del inglés, *High Performance Liquid Chromatography*, Alliance e2695, Waters, Milford, MA, EE. UU.) para la cuantificación de GT y bmGT en extractos de cultivo como se describió anteriormente (139, 178, 179, 525). Se utilizó una columna C-18 (XBridge® C18 3.5 μ m 4.6 \times 100 mm Column, Waters). Se ajustó la temperatura a 30 °C. El análisis HPLC de las muestras se realizó utilizando una elución de gradiente con agua y metanol: 0-10 min: 55% de agua; 10-11 min: 40% de agua; 11-20 min: 40% de agua; 20-22 min: 55% de agua. La velocidad de flujo se estableció en 0,8 mL/min y el volumen de inyección fue de 10 μ L. Se utilizó un detector de señal ultravioleta/visible (2489 UV/Vis Detector, Waters), monitorizando la absorbancia a 273 nm. El tiempo de retención fue de 8,1 minutos para la GT y de 10,2 minutos para la bmGT (Anexo 1 – Figura S1 - A). Se obtuvieron curvas estándar para los patrones de toxinas que oscilaban entre 0,1 y 35 μ g/mL (Anexo 1 - Figura S1 – B, C). Como control positivo, se utilizó un estándar de GT y bmGT extraído de medio RPMI o Czapek-Dox. La recuperación total dependió de la concentración inicial, alcanzando hasta un 72% y un 91% de recuperación de GT y bmGT, respectivamente. El límite de detección de la técnica fue de 16,7 ng/mL tanto para GT como para bmGT.

2.2.2.2 Estudio de la combinación de bismetilgliotoxina con otros biomarcadores para el diagnóstico de aspergilosis invasiva en pacientes hematológicos

2.2.2.2.1 Pacientes y diseño del estudio

Se incluyeron en el estudio pacientes hematológicos ingresados o atendidos en el Hospital Universitario Miguel Servet (Zaragoza). Se seleccionaron aquellos a quienes, por

su riesgo de desarrollar AI, se les solicitó detección de GM entre enero de 2019 y diciembre de 2022. Se reclutaron si cumplían los criterios de inclusión: enfermedad subyacente (neoplasia hematológica u oncológica y receptores de trasplante de órgano sólido) o factores del hospedador según los criterios EORTC/MSG (neutropenia durante más de 10 días, trasplante alogénico de células madre, uso prolongado de corticosteroides, tratamiento con inmunosupresores de células T) (67).

Los médicos desconocían los resultados de bmGT e IL-8. El estudio siguió un protocolo doble ciego donde los investigadores involucrados en el análisis de bmGT e IL-8 fueron diferentes de los involucrados en la revisión de la historia clínica. La recolección de datos fue retrospectiva; si la historia clínica no estaba disponible, los pacientes fueron excluidos de la investigación. El estudio fue aprobado por el Comité de Ética de la Investigación de la Comunidad Autónoma de Aragón (CEICA) con la referencia PI12/0041. Se siguieron los estándares para el informe de precisión diagnóstica (STARD, del inglés, *STAndards for the Reporting of Diagnostic accuracy*) (526).

2.2.2.2 Detección y cuantificación de biomarcadores en suero de pacientes

La detección y cuantificación de biomarcadores utilizados actualmente en clínica para el diagnóstico de AI se llevó a cabo en el hospital de manera rutinaria. Se realizaron multitud de análisis para cada uno de los pacientes incluidos en el estudio. La detección de GM se realizó mediante el kit Platelia®Aspergillus (Bio-Rad) y las muestras se consideraron positivas cuando el índice de lectura fue igual o superior a 0,5. El análisis de BDG se realizó con el kit β -Glucan Test (Fujifilm, Wako), considerando un resultado negativo si el resultado fue menor o igual que 2,672 pg/ml, positivo si fue mayor que 11 pg/ml e indeterminado entre ambos valores. También se llevaron a cabo inmunoensayos para la determinación cuantitativa de anticuerpos IgG contra CMV (utilizando el kit Elecsys® CMV IgG, Roche), considerando negativos los resultados por debajo de 0,9 UI/ml, positivos los resultados por encima de 1,1 UI/ml e indeterminado entre ambos valores. Además, se realizaron PCR en tiempo real para la detección de especies de *Aspergillus* (con el kit AspID, IMMY MDx), nombrado como PCRASP en este estudio. El personal del hospital conservó congeladas muestras de esos mismos sueros para proporcionárnoslas con el objetivo de cuantificar la bmGT e IL-8 en nuestros laboratorios.

De la misma manera, se cuantificó bmGT e IL-8 en diversos sueros de cada uno de los pacientes. El procedimiento para la detección y cuantificación de bmGT (esquemático en la figura 37) se realizó mediante HPTLC como describieron Domingo et

al. (148). Se siguieron los protocolos previamente establecidos en nuestro laboratorio, optimizados para maximizar la estabilidad de la toxina y su porcentaje de recuperación en sueros congelados.



Figura 37. Protocolo utilizado para la detección y cuantificación de bmGT en muestras de sueros de pacientes susceptibles de padecer AI.

La bmGT se extrajo de las muestras añadiendo 1 ml de DCM a 280 μ l de suero. Se agitó cada muestra para posteriormente dejarla reposar hasta que se apreciaran dos fases claramente diferenciadas. La fase orgánica (donde se encuentra el analito diluido) se evaporó en corriente de N_2 , y posteriormente se adicionaron 100 μ l de DCM. Las muestras preparadas se inyectaron sobre placas de gel de sílice de HPTLC (Merck KGaA, Darmstadt, Alemania) utilizando un aplicador de muestras Linomat IV con pulverizador de banda (Camag, Muttenz, Suiza), realizando bandas de 4 mm. La separación se realizó con la mezcla tetrahidrofurano/7-heptano/acetonitrilo (29:20:1 (v/v/v)), eluyendo durante 25 min en una cámara de desarrollo horizontal (Camag). Se utilizó densitometría de barrido UV para la determinación de bmGT (TLC Scanner 3, Camag; $\lambda = 280$ y 367 nm; barrido lineal). Los datos del área de pico se recopilaron, mostraron y almacenaron utilizando el software winCATS (Planar Chromatography Manager, versión 1.4.3.6336). Cualquier resultado de bmGT por encima del límite de detección establecido en nuestro laboratorio para la técnica (0,0357 μ g/ml) se consideró positivo (148, 527).

Se utilizó espectrometría de masas por electrospray de infusión directa con un sistema MDS SCIEX 3200 Q TRAP LC/MS/MS (Applied Biosystems, Foster City, CA, EE. UU.) para confirmar la identidad de los picos observados de bmGT. Para ello, previamente se empleó una interfaz TLC-MS (CAMAG® TLC-MS Interface 2, Suiza) y metanol como disolvente para extraer cada pico detectado del gel de sílice de la placa de HPTLC.

Por otro lado, se utilizó el kit Human IL-8 Uncoated ELISA (Invitrogen) para cuantificar IL-8 en las mismas muestras de suero.

2.2.2.2.3 Análisis de datos mediante técnicas estadísticas y *machine learning*

Este estudio se desarrolló utilizando principalmente el lenguaje R con la versión “4.3.0 (2023-04-21 ucrt) -- Already Tomorrow”, plataforma “x86_64-w64-mingw32/x64 (64-bit)” (528) en Windows 10 x64 (build 19043). Se utilizó el entorno de desarrollo RStudio, versión 2022.12.0+353. Se eligió el lenguaje de programación R, de código abierto, porque es uno de los que más se aplican en las investigaciones científicas y ciencia de datos, ya que es efectivo para el análisis de datos estadísticos y gráficos, y es potente para realizar modelización. En cuanto al entorno de desarrollo RStudio, es un programa de interfaz amigable que permite a los usuarios mayor facilidad en cada una de las tareas de programación, así como el análisis de los datos. Los principales paquetes utilizados, así como las funciones más empleadas se detallan en el anexo 2 – Tabla S1.

Retrospectivamente, se obtuvieron los resultados de todos los biomarcadores analizados entre enero de 2019 y diciembre de 2022 en el hospital. Se desarrolló un algoritmo para seleccionar aquellos análisis comprendidos entre la fecha de ingreso y de alta para cada uno de los pacientes incluidos en el estudio.

Además de los resultados de los biomarcadores, se incluyeron en el estudio datos demográficos y clínicos de los pacientes, así como información del ingreso, profilaxis, tratamientos y diagnóstico clínico.

Uno de los pasos más críticos de este estudio fue conseguir conjuntos de datos útiles en un formato apropiado para su análisis. Se trata de la parte más costosa y ardua del tratamiento de datos, que ocupó una gran parte del tiempo de trabajo, al tratarse de datos reales generados por distintas fuentes. Esta fase fue un punto clave para poder trabajar en los pasos posteriores. Fue necesario realizar una limpieza exhaustiva, ya que los datos proporcionados no presentaban un formato uniforme, y muchos de ellos estaban registrados en formato “comentario” sin ningún tipo de criterio que los unificase para poder automatizar los procesos. Se homogeneizaron los datos en cuanto a formato (para obtener columnas con variables numéricas, booleanas o categóricas con todos o casi todos los datos completos), eliminando observaciones que no interesaban (como los comentarios de los clínicos o variables que estaban incompletas o vacías para la mayoría de los pacientes) y almacenando las más útiles para el estudio (por ser información relevante y estar completa para todos o la mayoría de los casos).

Se utilizaron diferentes metodologías de visualización de datos, así como pruebas estadísticas para evaluar las diferencias de las variables demográficas y clínicas entre los

grupos de diagnóstico. Se utilizó la prueba de Kruskal-Wallis en el caso de las variables continuas y Chi-cuadrado o Fisher para las variables categóricas (dependiendo de si las frecuencias esperadas eran mayores o menores de 5, respectivamente). En cuanto a los resultados de los biomarcadores, se realizaron pruebas de proporciones mediante Chi-cuadrado en el caso de las variables cualitativas. También se realizaron análisis de correlación entre las variables cuantitativas, utilizando Spearman por la naturaleza no paramétrica de los datos.

Posteriormente, se procedió a evaluar la capacidad diagnóstica de los diferentes biomarcadores (GM, BDG, CMV, bmGT e IL-8) de manera independiente mediante curvas Característica Operativa del Receptor (ROC, del inglés, *Receiver Operating Characteristic*). Se calculó el área bajo la curva (AUC, del inglés, *Area Under the Curve*), su error estándar y su intervalo de confianza del 95 % (IC 95 %), el valor p , la sensibilidad, la especificidad, los valores predictivos positivo y negativo (VPP, VPN). El valor de corte (*cut off* en inglés) para cada biomarcador se calculó mediante el índice de Youden (529, 530), maximizando la sensibilidad y especificidad para distinguir entre los grupos.

Como los criterios EORTC/MSG incluyen GM dentro de la definición de IA probable, esto introduce un sesgo para la comparación directa con bmGT. Del mismo modo que describieron previamente Barnes *et al.* (531) y Vidal-García *et al.* (527), resolvimos este problema mediante dos estrategias diferentes: Primero, modificando los criterios EORTC/MSG para excluir el GM de la definición y tomando como casos probables aquellos con PCR positiva para *Aspergillus* spp. (nombrado como EORTC-GM) y segundo, añadiendo bmGT a los criterios EORTC/MSG, dando el mismo valor a bmGT que a GM (nombrado como EORTC+BMGT).

Una vez estudiada la capacidad diagnóstica de los biomarcadores por separado, se estudió su capacidad en combinación bajo los diferentes umbrales de diagnóstico. Se utilizó RL para combinar todas las variables y calcular los resultados de AUC, su error estándar e IC 95%, así como el valor p , la sensibilidad, especificidad, VPP y VPN.

Por último, se utilizaron técnicas de ML para desarrollar árboles de decisión simples mediante la combinación de las variables que mejor resultado ofrecieron en la RL, añadiendo o no los resultados de PCRASP. Se utilizó el algoritmo C5.0 (532), que elige las variables para cada nodo del árbol en función de su capacidad de diferenciación entre las clases objetivo. Este algoritmo utiliza un criterio de ganancia de información de forma recursiva, midiendo la reducción de incertidumbre en la clasificación de los datos al

Materiales y métodos

dividirlos según cada variable. También incluye un proceso de poda o *pruning* para simplificar el árbol, eliminando los nodos que no contribuyen significativamente a la precisión general, reduciendo así el riesgo de sobreajuste (*overfitting*) y mejorando su capacidad de generalización (425, 533). Los árboles se entrenaron con el 80% de los datos y se testaron con el 20% restante, obteniendo las matrices de confusión y diferentes métricas que evalúan el comportamiento de los algoritmos en los datos de prueba: exactitud (*accuracy* en inglés), exactitud balanceada, valor p (que evalúa la diferencia entre la *accuracy* y el porcentaje de datos que corresponden a la clase más numerosa), AUROC, Kappa, índice F1, sensibilidad, especificidad, VPP y VPN.

En todo el estudio se utilizó un nivel de significancia $\alpha = 0,05$, por lo que se consideraron estadísticamente significativos valores $p < 0,05$.

2.2.3 RESULTADOS

2.2.3.1 Producción de gliotoxina y bismetilgliotoxina en diferentes cepas fúngicas

2.2.3.1.1 Estudio bioinformático

2.2.3.1.1.1 Búsqueda de proteínas homólogas de GliP de *Aspergillus fumigatus* Af293 en otras especies

Alrededor de un tercio de las especies obtenidas mediante BLAST pertenecieron a la sección *Fumigati*. También se encontraron dos cepas de *Penicillium*, cuatro cepas del género *Trichoderma*, así como otras especies menos comunes como *Coleophoma* or *Colletotrichum*.

Se puede esperar que los NRPS con alta homología de secuencia sean ortólogos a GliP y realicen funciones biosintéticas similares, o incluso las mismas que GliP. Asumiendo homología, también es probable que las secuencias y funciones del resto de las enzimas auxiliares codificadas en los genes del cluster estén conservadas.

En este estudio, hemos seleccionado manualmente 20 NRPS de diferentes especies con un grado relativamente alto de similitud (query coverage > 90 %, identidad de secuencia de aminoácidos > 35 % y valor e = 0).

Como se muestra en la tabla 10, los mejores resultados de BLAST para la búsqueda de la proteína gliP de *A. fumigatus* Af293 resultaron ser gliPs de otras cepas de *A. fumigatus*, lo que valida nuestro enfoque.

Los siguientes resultados con mayor puntuación fueron obtenidos por otras especies de la sección *Fumigati*, como *A. fischerii*, *A. turcosus*, *A. novofumigatus* o *A. lentulus*, entre otras. Se espera que estos NRPS pertenecientes a especies de la sección *Fumigati* sean ortólogos de GliP de *A. fumigatus* debido a la alta proximidad genética entre ellos. Cabe señalar que no se obtuvieron especies de *Aspergillus* que no fueran de la sección *Fumigati* al utilizar BLAST, lo que sugiere que solo las especies de la sección *Fumigati* podrían ser capaces de producir GT. De hecho, los mejores resultados BLAST para *A. terreus* (sección *Terrei*) y *A. flavus* (sección *Flavi*) son respectivamente AtaP y AclP, unas proteínas NRPS cuya identidad es inferior al 35% (datos no mostrados).

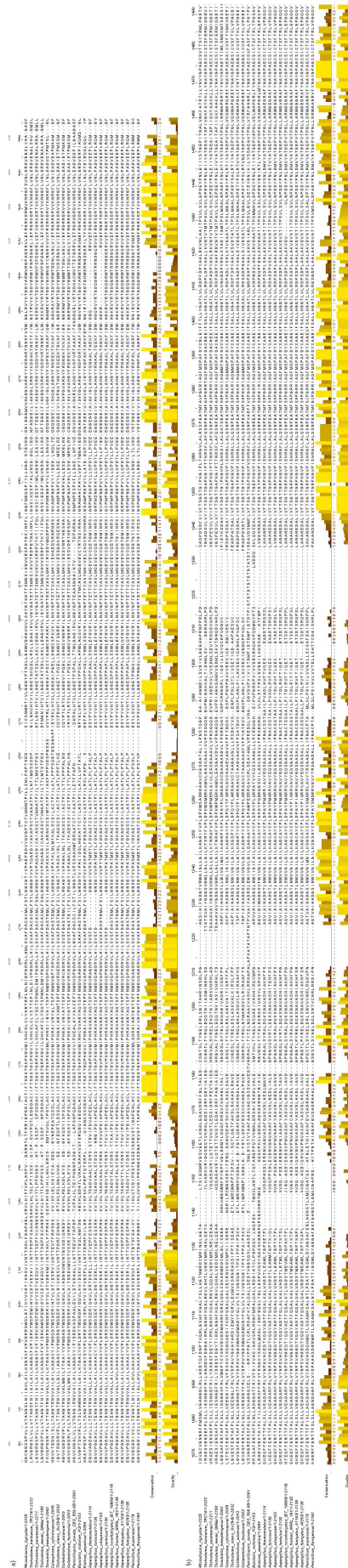
Tabla 10. 21 mejores resultados de BLAST para la proteína GliP de *A. fumigatus* Af293. Las entradas se ordenan por puntuación, aunque también se muestran el *query coverage*, la identidad y el valor esperado E (parámetro que describe la cantidad de resultados que se pueden “esperar” ver por casualidad al buscar en una base de datos de un tamaño determinado). Estos resultados son el propio *A. fumigatus* Af293 y otros *A. fumigatus*, varias secciones de *Fumigati* como *A. fischerii*, *A. novofumigatus* o *A. turcosus*, 4 especies de *Trichoderma*, 2 especies de *Penicillium* y otras especies menos relacionadas de géneros como *Coleophoma*, *Bipolaris* o *Rhizodiscinia*, entre otros.

Resultados

Cepa fúngica	Locus Tag	Acceso	Puntuación	Query Coverage (%)	Valor E	Identidad (%)
<i>Aspergillus fumigatus</i> Af293	AFUA_6G09660	EAL88817	4396	100	0.0	100.00
<i>Aspergillus fumigatus</i> A1163	AFUB_075710	EDP49541	4386	100	0.0	99.77
<i>Aspergillus fischerii</i> NRRL 181	NFIA_055350	XP_001258083	4152	100	0.0	94.75
<i>Aspergillus lentulus</i> IFM 54703	ALT_0509	GAQ03188	4101	100	0.0	93.13
<i>Aspergillus novofumigatus</i> IBT 16806	P174DRAFT_511160	XP_024683983	4012	100	0.0	92.06
<i>Aspergillus udagawae</i> IFM 46973	AUD_6972	GAO88012	3920	100	0.0	89.30
<i>Aspergillus turcosus</i> HMR AF 1038	CFD26_102651	RLL93873	3896	100	0.0	88.42
<i>Aspergillus thermomutatus</i> HMR Af 39	CDV56_101444	XP_026610442	3779	100	0.0	87.07
<i>Penicillium flavigenum</i> IBT 14082	PENFLA_c013G03821	OQE22222	2652	99	0.0	60.95
<i>Penicillium expansum</i> MD-8	PEX2_011780	XP_016603809	2644	99	0.0	61.24
<i>Bipolaris victoriae</i> F13	COCVIDRAFT_32734	EUN32662	1969	95	0.0	47.29
<i>Elsinoe ampelina</i> CECT 20119	BDZ85DRAFT_305103	KAF2219165	1712	99	0.0	44.93
<i>Coleophoma cylindrospora</i> BP6252	BP6252_07505	RDW70942	1642	95	0.0	44.21
<i>Colletotrichum fructicola</i> CGMCC3.17371	CGMCC3_g11493	KAE9572546.1	1640	94	0.0	44.00
<i>Colletotrichum asianum</i> ICMP 18580	GQ607_003908	KAF0328883	1627	94	0.0	44.09
<i>Trichoderma virens</i> Gv29-8	TRIVIDRAFT_78708	EHK22005	1621	94	0.0	43.99
<i>Torrubiella hemipterigena</i> BCC 1449	VHEMI02393	CEJ82322	1718	94	0.0	41.43
<i>Penicillium zonata</i> CBS 506.65	ASPZODRAFT_160119	OJJ45943.1	1481	96	0.0	42.54
<i>Trichoderma reesei</i> QM6a	TRIREDRAFT_24586	EGR52474.1	1461	98	0.0	39.94
<i>Rhizodiscina lignyota</i> CBS 133067	NA57DRAFT_81988	KAF2092841	1461	98	0.0	39.34
<i>Trichoderma harzianum</i> TR274	CI102_1861	PKK53446	1453	99	0.0	38.55
<i>Trichoderma parareesei</i> CBS 125925	A9Z42_0047170	OTA04143	1433	99	0.0	39.13

Sin embargo, en las cepas que no pertenecen a la sección *Fumigati* de *Aspergillus*, la identidad de secuencia disminuyó al 60-35 %, aunque aún se conservaron altos porcentajes de *query coverage*. Por otro lado, las tres especies de *Penicillium* lograron sus mejores resultados de identidad de alrededor del 60%. Entre los resultados de BLAST con peor puntuación, cabe mencionar el GliP Gv29-8 de *T. virens*, que posee un 44% de identidad.

El alineamiento múltiple de secuencias de las 22 NRPS realizado mediante Clustal Omega y visualizado por Jalview permitió determinar los fragmentos de secuencia o aminoácidos conservados entre todas las secuencias. El análisis visual de los alineamientos mostró que la mayoría de las regiones conservadas se encontraron en las secuencias de los dominios de adenilación (Figura 38).



Resultados

Figura 38. Alineamiento de las secuencias proteicas de (a) los primeros dominios de adenilación pertenecientes a los NRPS fúngicos estudiados y (b) los segundos dominios de adenilación de los mismos NRPS. Las barras de color marrón a amarillo debajo de cada residuo representan el grado de conservación y la calidad de cada posición. Las barras amarillas expresan un alto grado de conservación, mientras que las barras marrones indican una baja conservación.

2.2.3.1.1.2 Árbol filogenético

El árbol filogenético se construyó a partir de las secuencias similares a GliP de *A. fumigatus* Af293. Se seleccionó el árbol con mayor verosimilitud ($\log\text{-likelihood} = 52146,32$) (Figura 39). En la mayoría de los casos, cepas pertenecientes a la misma especie fúngica se agruparon en el mismo clado. El ejemplo más claro fue el del género *Aspergillus*, localizado en la parte superior del árbol filogenético.

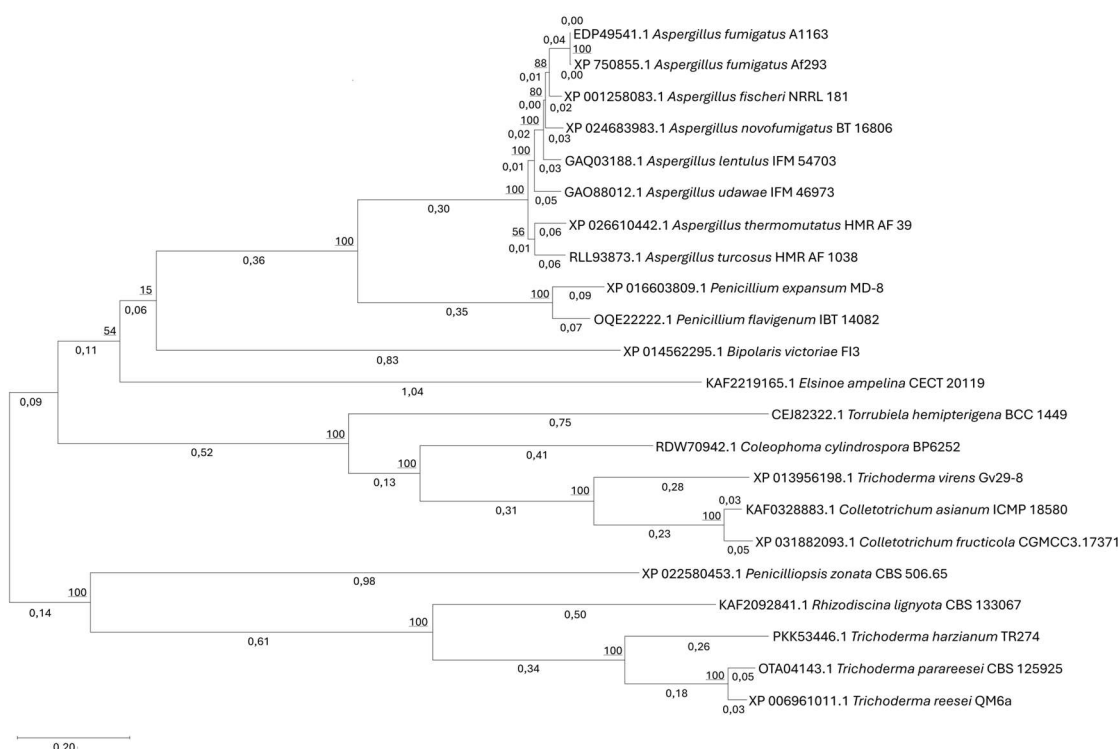


Figura 39. Árbol filogenético inferido mediante el software MEGA X. La historia evolutiva está compuesta por 22 secuencias de aminoácidos similares a GliP de *A. fumigatus* Af293. El árbol está dibujado a escala, con longitudes de ramas medidas en número de sustituciones por sitio (junto a las ramas). Los valores *bootstrap* se indican mediante el número subrayado junto a cada clado. Este número indica el porcentaje de árboles que realizaron la misma agrupación en los mismos clados del total de árboles creados mediante el método *bootstrap* (utilizando 100 remuestreos).

El valor *bootstrap* del clado que engloba a todos los taxones de *Aspergillus* tuvo un valor de 100. Este hecho indica que las 100 veces que se realizó el algoritmo de inferencia del árbol, estas secuencias fueron clasificadas en el mismo clado. Por lo tanto, la topología de esta rama fue altamente reproducible dando un alto valor de confianza. De igual manera, las proteínas que fueron seleccionadas para estudiar las cepas de *Penicillium*,

han sido clasificadas en el mismo clado con un valor *bootstrap* de 100. A su vez, estas proteínas provienen del mismo ancestro común que las del género *Aspergillus*, con un valor *bootstrap* de 100. Este resultado sugiere alta similitud entre las secuencias de estas proteínas.

La longitud de la rama entre las cepas de *A. fumigatus* fue de 0, lo que significa que no hay sustitución por sitio de una proteína a otra. Entre las cepas de *A. fumigatus* y el resto de las especies del género *Aspergillus* encontramos distancias pequeñas (mínima 0,00, máxima 0,15). Por otro lado, entre las cepas de *Penicillium* la distancia es ligeramente mayor (0,16), mostrando más diferencias evolutivas entre ellas. Además, la distancia genética entre el clado de *Aspergillus* y el de *Penicillium* es de 0,65, lo que demuestra diferencias entre estas proteínas.

En su mayoría, las cepas del género *Trichoderma* también se agruparon con un mismo ancestro común, pero en este caso las distancias evolutivas fueron mayores, por lo que se pudo deducir que estas proteínas presentan más diferencias entre ellas y respecto al género *Aspergillus*. Generalmente, el resto de las cepas pertenecientes a la misma especie también han sido agrupadas por el algoritmo, como *Colletotrichum*.

Los valores *bootstrap* para la mayoría de los clados generados fueron 100 o próximos a 100, demostrando una alta fiabilidad del árbol creado.

2.2.3.1.2 Análisis y cuantificación de la producción de gliotoxina y bismetilgliotoxina por varias cepas fúngicas

Para comprobar la validez del enfoque *in silico*, se analizó mediante HPLC la presencia de GT y bmGT (que deriva de la dimetilación de GT) en sobrenadantes de cultivos fúngicos de especies de hongos seleccionadas utilizando el previo estudio bioinformático. Como se muestra en la tabla 11, la producción de GT y bmGT se encontró principalmente en la sección *Fumigati*, excepto para *A. lentulus*. Aparte de la sección *Fumigati*, solo la cepa *T. virens* fue capaz de sintetizar GT y bmGT. Para todas estas cepas, se detectaron las toxinas en las especies cultivadas tanto en medio Czapek-Dox como en RPMI. Bajo las condiciones utilizadas en este estudio, el resto de las especies analizadas no fueron capaces de producir GT o bmGT. Las concentraciones de GT y bmGT obtenidas de los sobrenadantes de cultivo variaron de 0,085 a 11,067 µg/mL y de 0,222 a 11,131 µg/mL respectivamente.

Resultados

Tabla 11. Diferentes concentraciones de GT y bmGT obtenidas en sobrenadantes de cultivo de los hongos estudiados. Las cepas fúngicas se cultivaron tanto en medio Czapek-Dox como en medio RPMI 1640. Se cuantificaron las concentraciones de las toxinas mediante HPLC. Las concentraciones se expresan en $\mu\text{g/mL}$ \pm SD o ND si no se detectan. n = 3.

Cepa fúngica	Czapek-Dox		RPMI 1640	
	GT ($\mu\text{g/mL}$)	bmGT ($\mu\text{g/mL}$)	GT ($\mu\text{g/mL}$)	bmGT ($\mu\text{g/mL}$)
<i>A. fumigatus</i> B5233	11.067 \pm 14.286	2.595 \pm 3.225	8.373 \pm 2.663	11.131 \pm 15.538
<i>A. fumigatus</i> 1631562	3.913 \pm 4.278	0.845 \pm 0.805	4.154 \pm 7.442	0.519 \pm 0.052
<i>A. fischeri</i> CBS 420.96	0.135 \pm 0.077	0.222 \pm 0.206	0.062	0.311 \pm 0.108
<i>A. lentulus</i> 353	ND	ND	ND	ND
<i>A. turcosus</i> CBS 140371	0.085 \pm 0.0539	1.215 \pm 0.884	0.491 \pm 0.585	1.795 \pm 1.732
<i>A. pseudofischeri</i> CBS 404.67	0.840 \pm 0.028	0.518 \pm 0.297	0.295 \pm 0.092	0.253 \pm 0.226
<i>P. flavigenum</i> CBS 110407	ND	ND	ND	ND
<i>P. expansum</i> MD-8	ND	ND	ND	ND
<i>E. ampelina</i> CBS 208.25	ND	ND	ND	ND
<i>C. cylindrospora</i> CBS 449.70	ND	ND	ND	ND
<i>C. fructicola</i> CBS 120005	ND	ND	ND	ND
<i>T. virens</i> CBS 249.59	0.840 \pm 0.824	0.518 \pm 1.762	0.330 \pm 0.178	2.417 \pm 1.697
<i>P. zonata</i> CBS 506.65	ND	ND	ND	ND
<i>T. reesei</i> CBS 383.78	ND	ND	ND	ND
<i>R. ligniota</i> CBS 133067	ND	ND	ND	ND
<i>T. harzianum</i> CBS 226.95	ND	ND	ND	ND
<i>T. parareesei</i> CBS 125925	ND	ND	ND	ND

2.2.3.2 Utilidad de la combinación de bismetilgliotoxina con otros biomarcadores para el diagnóstico de aspergilosis invasiva en pacientes hematológicos

En total, se analizaron datos de 140 pacientes, de los cuales, 6 fueron pacientes hospitalizados en 2 episodios diferentes, por lo que se decidió incluirlos como independientes, obteniendo un total de 146 registros. De estos 146 episodios de pacientes, un 69,9% (n = 102) obtuvieron un diagnóstico negativo para AI, un 17,8% (n = 26) AI posible y un 12,3% (n = 18) AI probable (Figura 40-A). En la figura 40-B y 40-C también se muestra la distribución de factores clínicos y enfermedades subyacentes, respectivamente, en la cohorte de pacientes del estudio. Los factores clínicos más frecuentes fueron la neutropenia y el tratamiento con inmunosupresores. Por otro lado, las enfermedades subyacentes más prevalentes fueron la leucemia mieloide aguda y el linfoma para los tres grupos de estudio. En el caso de los pacientes con AI negativa o posible, el mieloma se sitúa como la tercera enfermedad subyacente más predominante.

Sin embargo, para los pacientes con AI probable, la tercera enfermedad subyacente más recurrente es la leucemia linfoblástica aguda.

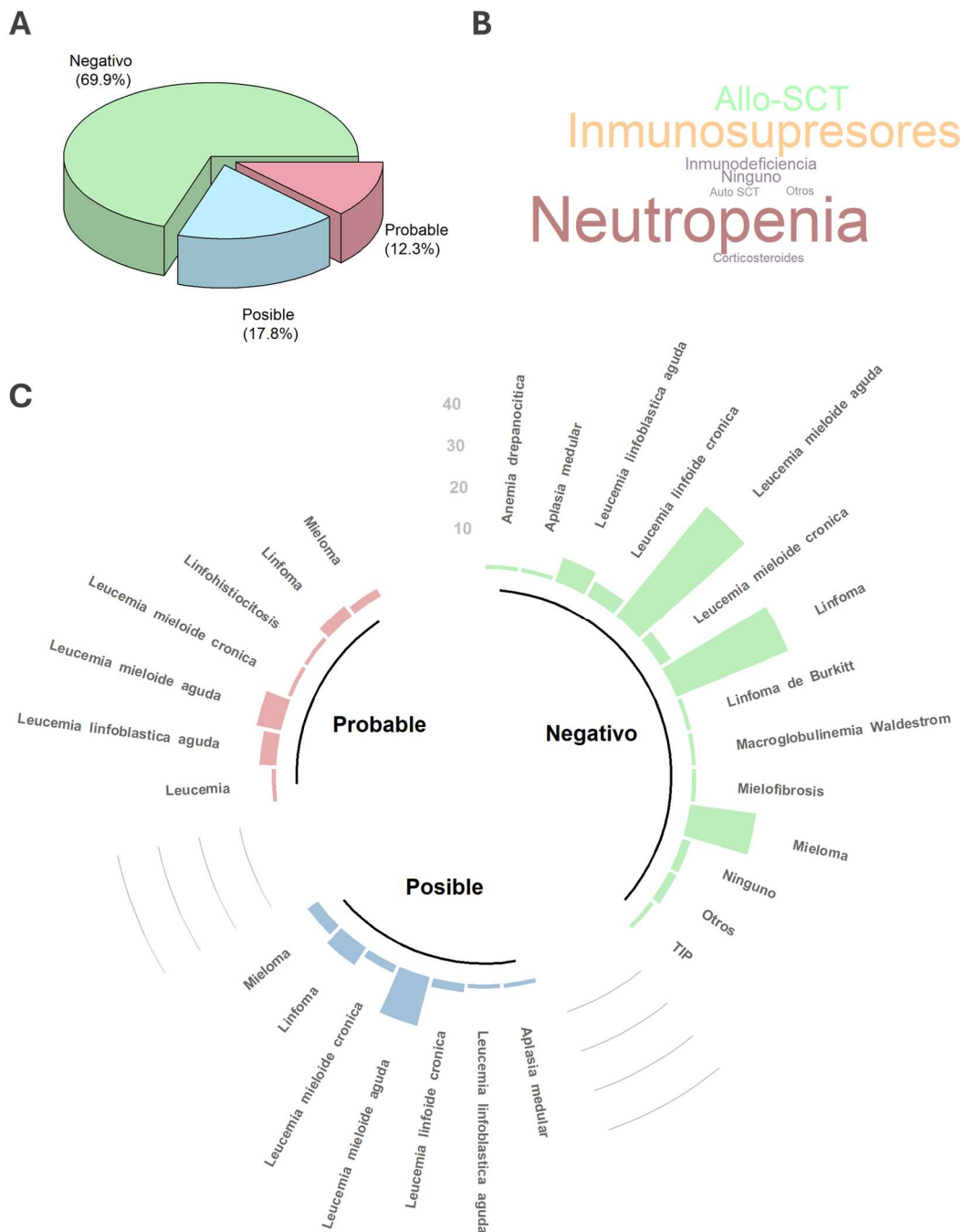


Figura 40. Visualización del conjunto de datos. **(A)** Diagrama sectorial que representa el porcentaje de pacientes según su diagnóstico de AI en nuestro conjunto de datos. **(B)** Nube de palabras que representa la frecuencia de los factores clínicos de los pacientes. El tamaño de cada palabra es proporcional al número de pacientes que presentan dicho factor clínico. **(C)** Gráfico de barras circular que muestra la distribución de enfermedades subyacentes entre los pacientes del estudio. Cada barra representa un tipo de enfermedad subyacente, y su longitud refleja la cantidad de pacientes que padecen esa enfermedad. El círculo está dividido en tres secciones, correspondientes a las tres clases de diagnóstico de AI presentes en nuestros datos: negativo, posible y probable.

Resultados

Las comparaciones entre los tres grupos de diagnóstico en base a sus características demográficas y clínicas se muestran en la tabla 12.

Tabla 12. Características demográficas y clínicas de los pacientes involucrados en el estudio según su diagnóstico. Se muestra la mediana y rango intercuartílico (IQR, del inglés, *Interquartile Range*) para la edad, y el número de pacientes (n) y porcentaje respecto al total de pacientes para el resto de las variables. También se muestra el valor *p*, resultante de la comparación entre grupos mediante Kruskal-Wallis para la variable continua y, en el caso de las variables categóricas, resultante del análisis mediante Chi-cuadrado o Fisher si las frecuencias son mayores o menores que 5, respectivamente. Se consideraron estadísticamente significativos valores $p < 0,05$.

	Negativo (n = 102)	Posible (n = 26)	Probable (n = 18)	Valor <i>p</i>
Edad , mediana años (IQR)	61,5 ± 16,5	62,5 ± 23,25	65 ± 15,25	0,286
Sexo , n (%)				0,452
Femenino	43 (29,5%)	9 (6,2%)	5 (3,4%)	
Masculino	59 (40,4%)	17 (11,6%)	13 (8,9%)	
Factores clínicos , n (%)				0,384
Allo-SCT	23 (15,8%)	2 (1,4%)	1 (0,7%)	
Auto SCT	1 (0,7%)	0 (0%)	0 (0%)	
Corticosteroides	3 (2,1%)	0 (0%)	0 (0%)	
Inmunodeficiencia	4 (2,7%)	3 (2,1%)	0 (0%)	
Inmunosupresores	28 (19,2%)	6 (4,1%)	8 (5,5%)	
Neutropenia	36 (24,7%)	14 (9,6%)	9 (6,2%)	
Ninguno	6 (4,1%)	1 (0,7%)	0 (0%)	
Otros	1 (0,7%)	0 (0%)	0 (0%)	
Enfermedad subyacente , n (%)				0,534
Anemia drepanocítica	1 (0,7%)	0 (0%)	0 (0%)	
Aplasia medular	1 (0,7%)	1 (0,7%)	0 (0%)	
Leucemia	0 (0%)	0 (0%)	1 (0,7%)	
Leucemia linfoblástica aguda	6 (4,1%)	1 (0,7%)	4 (2,7%)	
Leucemia linfocítica crónica	4 (2,7%)	2 (1,4%)	0 (0%)	
Leucemia mieloide aguda	34 (23,3%)	12 (8,2%)	6 (4,1%)	
Leucemia mieloide crónica	3 (2,1%)	2 (1,4%)	1 (0,7%)	
Linfomatosidad	0 (0%)	0 (0%)	1 (0,7%)	
Linfoma	29 (19,9%)	5 (3,4%)	3 (2,1%)	
Linfoma de Burkitt	1 (0,7%)	0 (0%)	0 (0%)	
Macroglobulinemia Waldenström	1 (0,7%)	0 (0%)	0 (0%)	
Mielofibrosis	1 (0,7%)	0 (0%)	0 (0%)	
Mieloma	16 (11%)	3 (2,1%)	2 (1,4%)	
Ninguno	2 (1,4%)	0 (0%)	0 (0%)	
Otros	2 (1,4%)	0 (0%)	0 (0%)	
TIP	1 (0,7%)	0 (0%)	0 (0%)	
Fiebre , n (%)				0,035
Fiebre	71 (48,6%)	24 (16,4%)	15 (10,3%)	
No Fiebre	31 (21,2%)	2 (1,4%)	3 (2,1%)	
Candidiasis posible , n (%)				1
No	97 (66,4%)	25 (17,1%)	17 (11,6%)	
Si	5 (3,4%)	1 (0,7%)	1 (0,7%)	
COVID-19 , n (%)				0,705
Negativo	94 (64,4%)	25 (17,1%)	16 (11%)	
Positivo	8 (5,5%)	1 (0,7%)	2 (1,4%)	

No se encontraron diferencias significativas entre los grupos de diagnóstico, excepto en la presencia de fiebre ($p < 0.05$). Se realizó una comparación *post-hoc* mediante

Fisher, donde los valores p ajustados no resultaron significativos para ninguna de las combinaciones (Negativo Vs Posible, valor $p = 0,067$; Negativo Vs Probable, valor $p = 0,819$; Posible Vs Probable, valor $p = 1$).

Las terapias farmacológicas administradas a los pacientes, tanto en profilaxis como en tratamiento, fueron diferentes y personalizadas en base a las condiciones de cada paciente. Los antifúngicos más utilizados, tanto individualmente como en combinación, fueron anfotericina (12 pacientes en profilaxis y 29 en tratamiento), fluconazol (7 pacientes en profilaxis y 10 en tratamiento), voriconazol (7 pacientes en profilaxis y 10 en tratamiento) y posaconazol (6 pacientes en profilaxis y 9 en tratamiento). Otros antimicóticos utilizados fueron isavuconazol, caspofungina y micafungina. Cabe destacar la gran cantidad de pacientes con diagnóstico negativo para AI que fueron tratados farmacológicamente con antifúngicos (Figura 41) tanto en profilaxis ($n = 7$) como en tratamiento ($n = 16$) o incluso con ambos ($n = 7$). En total, a un 29,4% de los pacientes que fueron diagnosticados como AI negativa se les administraron fármacos antifúngicos.

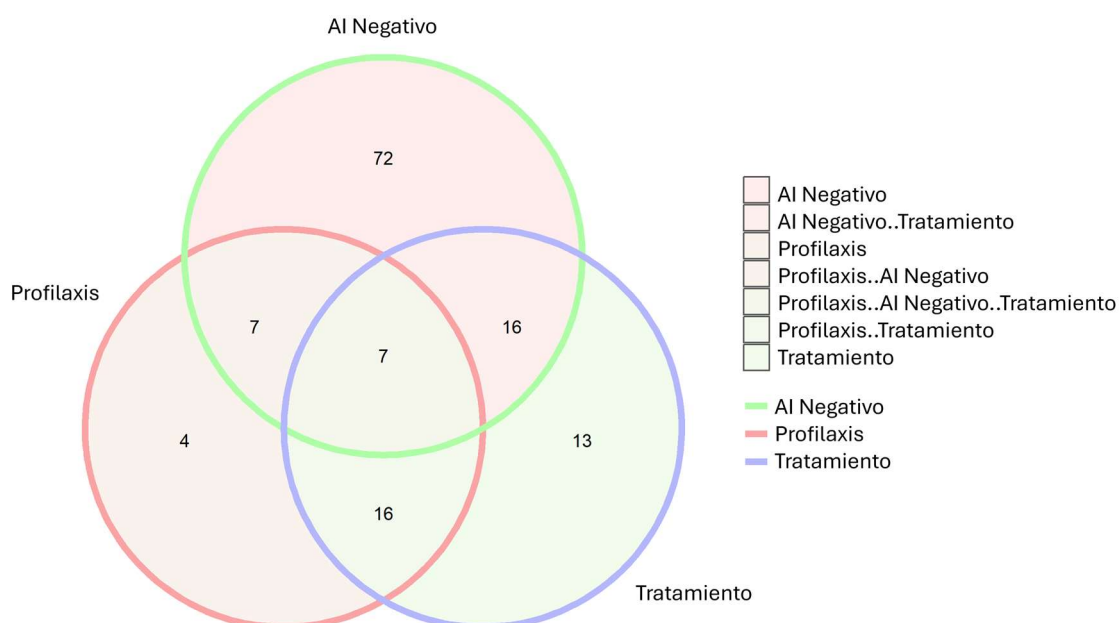


Figura 41. Diagrama de Venn. Representa el número de pacientes diagnosticados como AI negativa que recibieron una intervención farmacológica mediante profilaxis y/o tratamiento con antifúngicos.

Respecto a los biomarcadores utilizados actualmente en clínica para evaluar el diagnóstico y progreso de la AI (PCRASP, GM, BDG y CMV), se obtuvieron 4045 instancias en total. Es decir, hubo 4045 mediciones individuales de estos biomarcadores entre todos

Resultados

los pacientes durante sus respectivos episodios. Con esta información, se crearon variables de interés para cada biomarcador y cada paciente, como el valor del primer resultado, el valor del primer resultado positivo, el valor máximo y la media de todos los resultados durante su episodio. Se siguió el mismo procedimiento para los resultados de la cuantificación de bmGT e IL-8.

Se estudió el porcentaje de pacientes que dieron al menos un resultado positivo de PCRASP, GM, BDG, CMV y bmGT durante su episodio (Figura 42). Cabe destacar que en los casos de GM y bmGT todos los pacientes tuvieron al menos un suero analizado. Sin embargo, existen numerosos pacientes diagnosticados como AI negativa a los que no se les realizaron análisis PCRASP (29%) ni BDG (16%). En el caso de CMV, hubo un gran porcentaje de pacientes no analizados.

El 56% de los pacientes diagnosticados como probables, el 19% de los clasificados como posibles, y solo el 1% de los negativos obtuvieron al menos un resultado positivo en PCRASP. Este patrón creciente según el grado de diagnóstico se observa también en los resultados de BDG, donde el 11% de los pacientes negativos, el 15% de los posibles y el 44% de los probables para AI dieron al menos un resultado positivo. Del mismo modo, en los resultados de CMV, los porcentajes de pacientes que obtuvieron al menos un positivo fueron 21%, 27% y 33%, respectivamente. En cuanto al GM, un 6% de pacientes negativos dieron como resultado al menos un positivo, todos los pacientes posibles obtuvieron negativo en sus pruebas de GM, y un 44% de pacientes probables obtuvieron al menos un resultado positivo. Este hecho se debe a que, según los criterios EORTC que actualmente se siguen en clínica, la obtención de un resultado positivo en GM en pacientes clasificados como posibles cambia su diagnóstico a probable. Por último, respecto a la bmGT, se obtuvieron un 11% de pacientes negativos, 46% posibles y 39% probables con al menos un resultado positivo de esta toxina en suero. La representación utilizando un diagrama de Venn (Figura 42-F) de los pacientes que habían dado al menos un positivo en PCRASP, BDG, GM y bmGT nos permitió percibir la poca concordancia entre los resultados de los diferentes biomarcadores. Solamente 2 pacientes obtuvieron resultados positivos en todos los biomarcadores, siendo relevante mencionar que ambos eran pacientes diagnosticados como probables. No obstante, 5 pacientes dieron positivo en PCRASP, 11 en BDG, 8 en GM y 23 en bmGT, siendo negativo en todos los demás análisis del resto de marcadores para esos mismos pacientes.

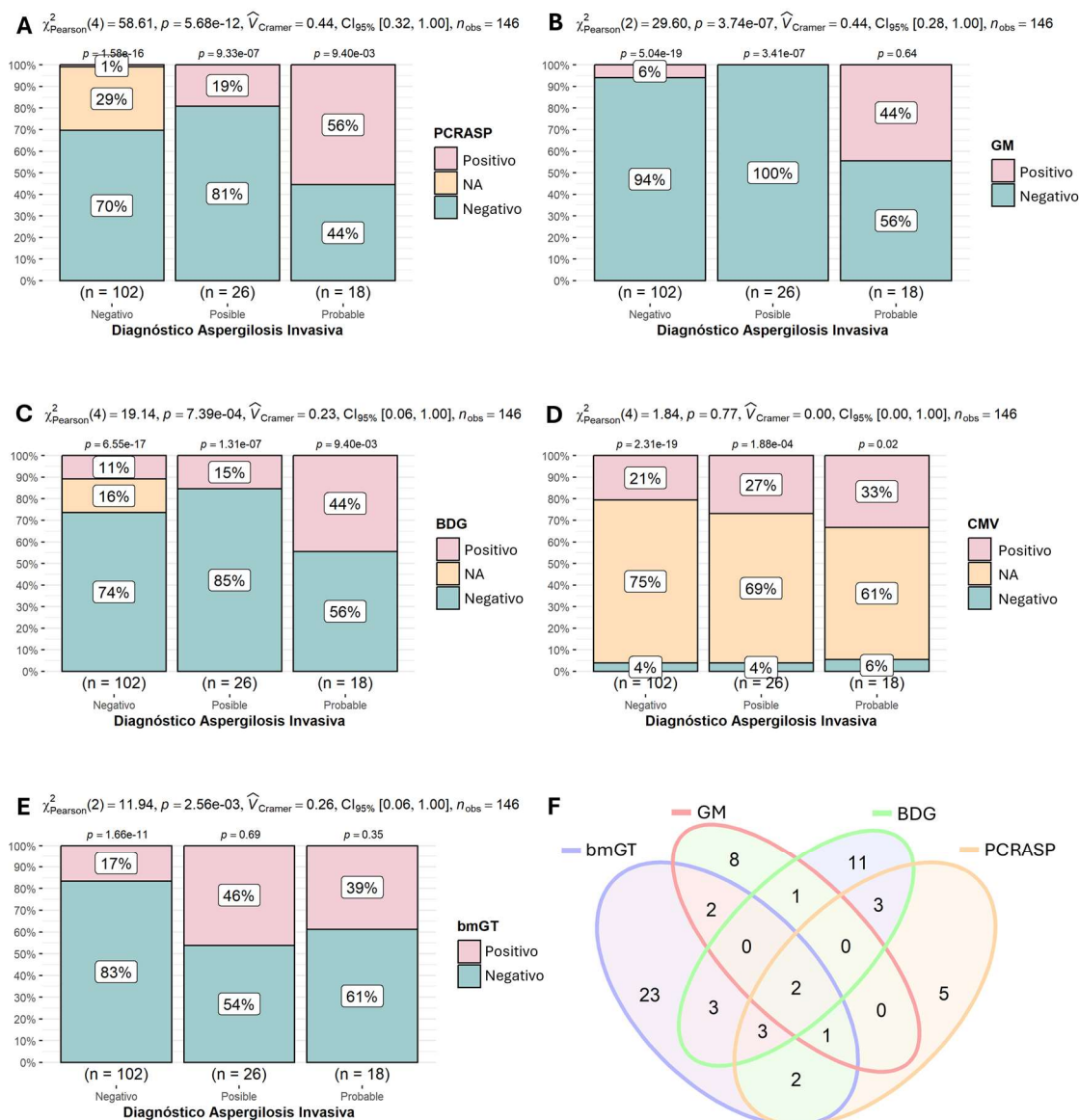


Figura 42. Prueba de proporciones mediante Chi-cuadrado para los biomarcadores (A) PCRASP, (B) GM, (C) BDG, (D) CMV y (E) bmGT, donde se comparan los porcentajes de pacientes que han dado al menos un resultado positivo de cada uno de los análisis en los tres grupos de diagnóstico: Al negativo, posible o probable. Se consideraron estadísticamente significativos valores $p < 0,05$. (F) Diagrama de Venn que solapa los pacientes que obtuvieron al menos un resultado positivo en los biomarcadores bmGT, GM, BDG y PCRASP. Se representan $n = 64$ pacientes porque los 82 restantes obtuvieron resultados negativos en todos los biomarcadores comparados. NA: No analizado.

También se estudió la correlación entre la cuantificación de los diferentes biomarcadores según su resultado en el primer análisis realizado a cada paciente, de manera global y dividiendo los grupos según su diagnóstico (Figura 43). El GM se excluyó del análisis de correlación porque la mayoría de sus resultados para el primer análisis eran negativos. La bmGT mostró correlaciones de magnitud leve con el resto de los biomarcadores. La IL-8 mostró una correlación negativa con el BDG, siendo de $-0,404$ (valor $p < 0,10$) en el caso de los pacientes diagnosticados como negativos. IL-8 también

Resultados

mostró correlación negativa con CMV en los pacientes posibles (-0,659, valor $p < 0,10$) y positiva en los probables (0,429). No obstante, de manera general, ninguna correlación estudiada resultó estadísticamente significativa.

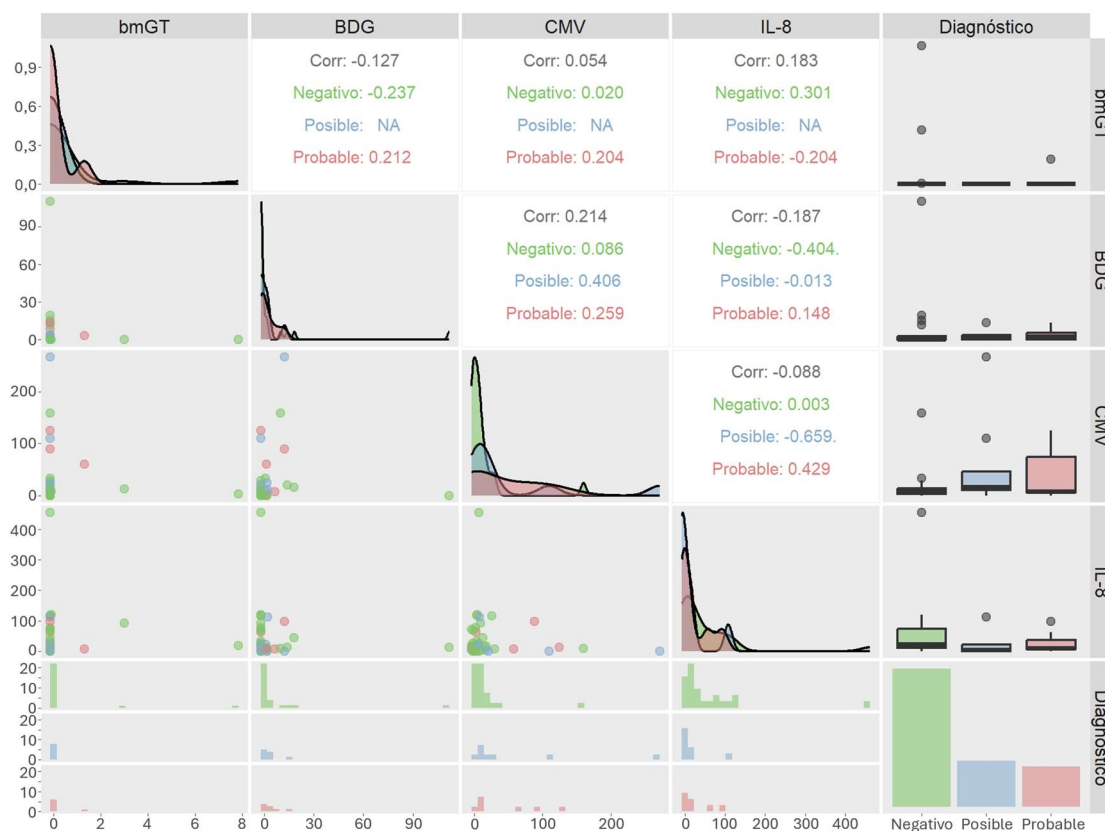


Figura 43. Análisis de correlación de Spearman (por la naturaleza no paramétrica de los datos) entre los resultados de cuantificación de los primeros análisis de bmGT, BDG, CMV e IL-8 realizados a los pacientes. En la diagonal se muestra la distribución de los resultados de cada biomarcador. Por encima de la diagonal se muestran los valores de la prueba de correlación para todos los pacientes (Corr) y de manera individual para los pacientes diagnosticados como AI negativo, posible y probable. Símbolos tras el resultado de correlación: “*” valor $p < 0,05$; “.” valor $p < 0,10$; “” valor $p \geq 0,05$. Por debajo de la diagonal se muestran los gráficos de correlación para todas las combinaciones de biomarcadores. En la columna de la derecha se muestran los gráficos de cajas de la cuantificación de cada biomarcador según el diagnóstico. En la fila inferior se muestra la distribución de los resultados de cada biomarcador, en forma de barras y dividido en filas según el diagnóstico.

En cuanto a la evaluación de la capacidad diagnóstica de los biomarcadores, se realizaron diversos estudios, comenzando por la elaboración de curvas ROC y el cálculo de puntos de corte o *cut off* para evaluar si nuestra cohorte de pacientes cumplía con los criterios establecidos actualmente.

Considerando el limitado set de datos que obtuvimos, así como la necesidad de eliminar de él los pacientes con datos faltantes, se decidió unificar grupos, basándonos en los procedimientos aplicados por otros investigadores en cohortes similares para la evaluación de nuevos biomarcadores en el diagnóstico de AI (527, 531). Agrupamos los

pacientes negativos y posibles para utilizar un solo grupo “No AI” frente a los pacientes con AI probable.

Se realizaron curvas ROC para cada biomarcador (GM, BDG, CMV, IL-8, bmGT) utilizando el valor máximo de los resultados de cada uno de ellos durante el episodio de cada paciente (Tabla 13). Este valor máximo suele presentar el momento más crítico de todo el ingreso del paciente y puede facilitar la diferenciación entre casos “No AI” y probables. Al utilizar para todos los pacientes su valor máximo, se asegura que el biomarcador esté midiendo su "pico" potencial, que puede ser relevante en infecciones o condiciones dinámicas, como es la AI.

Tabla 13. Resultados de las curvas ROC para cada biomarcador utilizando el valor máximo del episodio de cada paciente. El punto de corte o *cut off* se calculó mediante el índice de Youden. El error estándar y el intervalo de confianza del 95% (IC 95%) pertenecen al área bajo la curva (AUC del inglés, *Area Under the Curve*). Sens: sensibilidad, Espec: especificidad; VPP: Valor Predictivo Positivo; VPN: Valor Predictivo Negativo. Se consideraron estadísticamente significativos valores $p < 0,05$. *Las unidades del *cut off* para cada biomarcador fueron: GM (índice de lectura), bmGT ($\mu\text{g/ml}$), BDG (pg/ml), CMV (U/ml) e IL-8 (pg/ml).

Biomarcador	Cut off*	Índice Youden	AUC	Error estándar	IC 95%	valor p	Sens (%)	Espec (%)	VPP (%)	VPN (%)
GM	0,513	0,688	0,875	0,074	0,729 - 1,000	0,002	100	68,8	61,5	100
bmGT	2,663	0,33	0,635	0,133	0,375 - 0,896	0,865	57,1	75,9	36,4	88
BDG	21,694	0,445	0,717	0,083	0,555 - 0,879	0,009	53,8	90,7	63,6	86,7
CMV	46,65	0,379	0,596	0,142	0,318 - 0,874	0,236	50	87,9	42,9	90,6
IL-8	30,784	0,262	0,595	0,08	0,439 - 0,751	0,904	50	76,2	23,7	91,2

El *cut off* que se obtuvo del GM fue un índice de lectura de 0,513, calculado con un Índice de Youden por encima de 0,5 (0,688). Esta prueba logró una combinación relativamente alta de sensibilidad (100%) y especificidad (68,8%). El resultado AUC = 0,875 también muestra una buena capacidad discriminativa para diferenciar entre los dos grupos de diagnóstico, aunque no es perfecta. Su valor p resultó significativo (0,002). El otro resultado significativo corresponde al BDG, que obtuvo un AUC = 0.717 (valor p = 0,009) y un *cut off* de 21,694 pg/ml . En el caso de CMV e IL-8, se obtuvieron AUC relativamente inferiores y, pese a que los VPN eran elevados (90,6 y 91,2%, respectivamente), los VPPs fueron especialmente bajos. Respecto a la bmGT, se obtuvo un AUC = 0,635 no significativo con un 57,1 de sensibilidad y 75,9 de especificidad.

Aunque utilizamos el valor máximo para evaluar la validez de nuestra cohorte y poder establecer comparaciones según los criterios de diagnóstico actuales, a partir de

Resultados

este punto se utilizó como variable de estudio el valor del primer análisis para cada biomarcador. Este interés en el primer análisis radica en su aplicabilidad en clínica, dada la importancia de detectar con la máxima prontitud la mínima sospecha de una posible IFI en pacientes susceptibles de padecerla. Evaluar qué biomarcador o biomarcadores pueden ofrecernos resultados de manera rápida y eficaz puede ser un punto clave en la mejora de la toma de decisiones relativas al manejo de los pacientes.

Con el objetivo de comparar los nuevos biomarcadores con los ya utilizados en clínica, y como se comentó en la sección de materiales y métodos, se modificaron los criterios EORTC/MSK para eliminar el sesgo de GM de la definición de diagnóstico. Se compararon los resultados del valor del primer análisis de cada biomarcador para las 3 variantes del criterio de diagnóstico (Tabla 14).

Tabla 14. Resultados de las curvas ROC para cada biomarcador utilizando el primer resultado del episodio de cada paciente. El error estándar y el intervalo de confianza del 95% (IC 95%) pertenecen al área bajo la curva (AUC, del inglés, *Area Under the Curve*). Sens: sensibilidad, Espec: especificidad; VPP: Valor Predictivo Positivo; VPN: Valor Predictivo Negativo. Se consideraron estadísticamente significativos valores $p < 0,05$.

Biomarcador	AUC	Error estándar	IC 95%	Valor p	Sensibilidad (%)	Especificidad (%)	VPP (%)	VPN (%)
EORTC								
GM	0,544	0,039	0,468 - 0,621	0,028	11,1	99,2	66,7	88,8
bmGT	0,57	0,054	0,463 - 0,677	0,059	27,8	89,1	26,3	89,8
BDG	0,622	0,065	0,495 - 0,749	0,019	50	75	24,3	90,3
CMV	0,468	0,15	0,174 - 0,761	0,402	57,1	66,7	26,7	88
IL-8	0,572	0,064	0,448 - 0,697	0,842	72,2	55,5	18,6	93,4
EORTC-GM								
GM	0,481	0,008	0,464 - 0,497	0,79	100	0	11	0
bmGT	0,631	0,065	0,504 - 0,758	0,003	37,5	90,8	33,3	92,2
BDG	0,586	0,068	0,452 - 0,719	0,083	31,2	88,6	27,8	90,2
CMV	0,649	0,16	0,337 - 0,962	0,113	57,1	93,9	66,7	91,2
IL-8	0,694	0,052	0,591 - 0,796	0,995	87,5	56,2	19,7	97,3
EORTC+BMGT								
GM	0,508	0,018	0,473 - 0,543	0,315	4,5	98	50	70,4
bmGT	0,739	0,038	0,664 - 0,813	<0,001	47,7	100	100	81,6
BDG	0,558	0,044	0,472 - 0,645	0,087	36,4	75,6	43,2	69,9
CMV	0,49	0,101	0,291 - 0,688	0,462	18,8	100	100	64,9
IL-8	0,493	0,052	0,39 - 0,595	0,444	54,5	53,9	33,8	73,3

Siguiendo los criterios EORTC, se obtuvo un AUC de 0,544 ($p = 0,028$) para GM y un AUC de 0,622 ($p = 0,019$) para BDG. A pesar de ser resultados estadísticamente significativos, se obtuvieron unas AUC muy inferiores a las obtenidas con el valor máximo. En el caso de la bmGT se obtuvo un AUC = 0,57 ($p = 0,059$) con una elevada especificidad (89,1%) pero baja sensibilidad (27,8%). El *cut off* del primer análisis de la bmGT en pacientes clasificados según los criterios EORTC fue 0,601 $\mu\text{g/ml}$. La IL-8 mostró un AUC similar (0,572), pero en este caso la sensibilidad (72,2%) fue superior a la especificidad (55,5%).

Cuando se alteró el criterio EORTC para eliminar el sesgo del GM (EORTC-GM) y se clasificaron los pacientes en los grupos “No AI” y “Probable” según sus resultados de PCRASP, se obtuvo únicamente una AUROC significativo, el correspondiente a la bmGT, siendo AUC = 0,631 (valor $p = 0,003$) con una especificidad del 90,8%. Los valores AUC también mejoraron para CMV e IL-8, como se puede visualizar en la figura 44. Sin embargo, los resultados de GM y BDG fueron considerablemente inferiores. El VPP del GM resultó en 0 porque en base a ese biomarcador, y eliminando su sesgo de la clasificación, no se pudieron identificar correctamente los casos negativos.

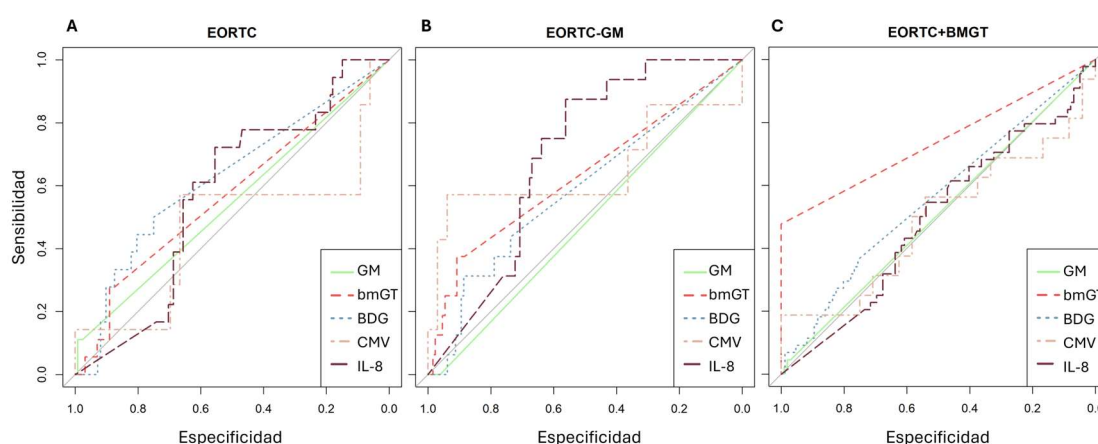


Figura 44. Análisis ROC (*Receiver Operating Characteristic*) de las primeras muestras analizadas de cada biomarcador para cada paciente, que discrimina a los individuos “No AI” de los “probables” según el criterio de diagnóstico (A) EORTC (B) EORTC-GM y (C) EORTC+BMGT.

Finalmente, alterando los criterios para añadir bmGT dándole el mismo valor que a GM (EORTC+BMGT), se obtuvo un AUC = 0,739 ($p < 0,001$) para bmGT, con una especificidad del 100% y una sensibilidad del 47,7%, alcanzando su mejor capacidad predictiva (Figura 45). En el caso del GM, también se obtuvo un AUC significativo, pero con un valor considerablemente menor al de la bmGT (0,508) y una sensibilidad deficiente (4,5%). El resto de los biomarcadores no mostraron resultados significativos.

Resultados

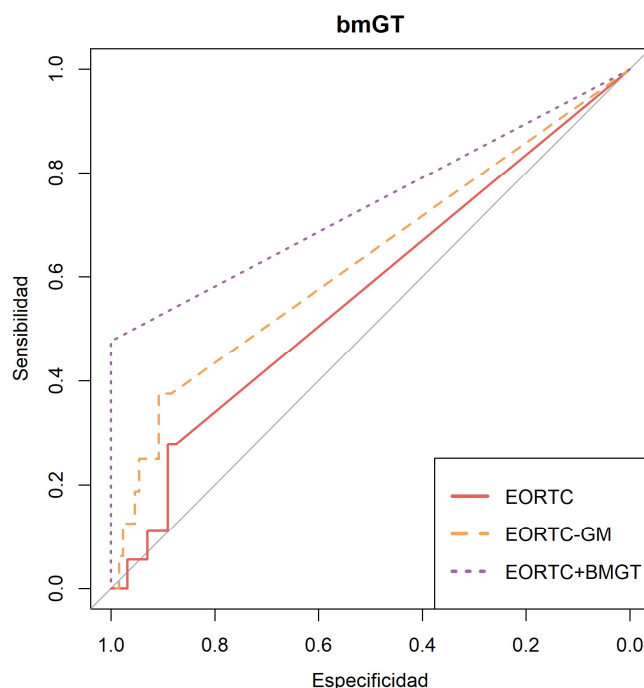


Figura 45. Comparativa entre las curvas ROC (*Receiver Operating Characteristic*) calculadas con los resultados de las primeras muestras de bmGT para cada paciente, que discrimina a los individuos “No AI” de los “probables” según el criterio de diagnóstico EORTC, EORTC-GM y EORTC+BMGT.

Una vez estudiada la capacidad diagnóstica de las variables de manera individual, también estudiamos su capacidad en combinación, mediante el uso de RL.

Como análisis preliminar y a modo comparativo, en la tabla 15 se muestran los resultados de una RL utilizando únicamente el GM, que es el utilizado actualmente en clínica para clasificar a los pacientes según los criterios EORTC. Se obtuvo un AUC relativamente bajo, aunque significativo, con una elevada especificidad (99,2 %).

Tabla 15. Resultados del análisis de regresión logística para la variable GM utilizando los valores del primer análisis de los pacientes. Se muestran el área bajo la curva (AUC, del inglés, *Area Under the Curve*), su error estándar e intervalo de confianza del 95% (IC 95%), valor p , sensibilidad, especificidad, valor predictivo positivo (VPP) y valor predictivo negativo (VPN). Se consideraron estadísticamente significativos valores $p < 0,05$.

AUC	Error estándar	IC 95%	Valor p	Sensibilidad (%)	Especificidad (%)	VPP (%)	VPN (%)
0,544	0,039	0,468 - 0,621	7,78E-15	11,1	99,2	66,7	88,8

Se pudo observar una mejora sustancial en los AUC generados cuando se combinaron los biomarcadores (Tabla 16) en todas las variantes de criterios EORTC estudiadas. Cabe destacar que se evaluaron todas las combinaciones posibles de biomarcadores, siempre utilizando los valores del primer análisis de cada uno. No obstante, en la tabla 16 solamente se muestran aquellas combinaciones que

proporcionaron los mejores resultados. Todas las combinaciones mostradas superaron al GM en solitario, que es el biomarcador más utilizado en el diagnóstico clínico actual.

Tabla 16. Resultados del análisis de regresión logística para las combinaciones de variables predictoras (biomarcadores) que mostraron mejores resultados, utilizando los valores del primer análisis de los pacientes. Se muestran el área bajo la curva (AUC, del inglés, *Area Under the Curve*), su error estándar e intervalo de confianza del 95 % (IC 95 %), valor *p*, sensibilidad, especificidad, valor predictivo positivo (VPP) y valor predictivo negativo (VPN). Se consideraron estadísticamente significativos valores $p < 0,05$.

Variable Dependiente	Variables Predictoras	AUC	Error estándar	IC 95 %	Valor <i>p</i>	Sensibilidad (%)	Especificidad (%)	VPP (%)	VPN (%)
EORTC	bmGT GM IL-8 CMV	0,649	0,107	0,439 – 0,860	0,011	71,4	60,6	27,8	90,9
EORTC-GM	bmGT GM IL-8 CMV	0,900	0,053	0,791 – 1,000	0,238	100	72,7	43,8	100
	bmGT GM IL-8	0,741	0,051	0,642 – 0,840	<0,001	87,5	64,6	23,3	97,7
EORTC+BMGT	bmGT GM BDG	0,781	0,043	0,696 – 0,866	<0,001	50,0	98,8	95,7	79,4

El resultado de la combinación de bmGT, IL-8 y CMV con GM supuso una mejora en la capacidad de predicción del diagnóstico respecto a la capacidad individual de GM, con un AUC = 0,649 (valor $p = 0,011$). En el caso de las variantes de los criterios EORTC, también se produjo una mejora en la discriminación de los pacientes. Respecto a los criterios sin el sesgo de GM, la combinación de bmGT, GM e IL-8 produjo un AUC = 0,741 con un valor p significativo ($< 0,001$), lo que sugiere una relación estadísticamente relevante entre esta combinación de biomarcadores y el diagnóstico clínico. Incluyendo CMV a este conjunto se obtuvo un AUC de 0,900, un valor notablemente alto, aunque no estadísticamente significativo. Finalmente, en lo que respecta a los criterios EORTC+BMGT, la combinación óptima fue bmGT, GM y BDG con un AUC = 0,781 (valor $p < 0,001$). Es relevante mencionar que todas las combinaciones que obtuvieron los mejores resultados incluyen tanto la bmGT como el GM.

Una vez estudiada la capacidad diagnóstica de la combinación de variables, se decidió investigar su capacidad predictiva mediante el uso de ML. En este caso se decidió utilizar un clasificador simple, que crease un algoritmo fácilmente interpretable, como los árboles de decisión. Se introdujeron en el modelo de predicción las variables que, en los análisis previos con RL, dieron el mejor resultado para la clasificación de pacientes según los criterios EORTC, es decir, los biomarcadores bmGT, GM, CMV e IL-8. Se utilizó la

Resultados

información del primer análisis y la información de si los pacientes habían obtenido algún resultado positivo o no durante su episodio. Se desarrollaron dos árboles de decisión que se diferenciaron en la inclusión o exclusión de los resultados de PCRASP.

Los algoritmos resultantes de entrenar los modelos con el 80% de los datos se muestran en la figura 46-A y 46-B (excluyendo e incluyendo los datos de PCRASP, respectivamente).

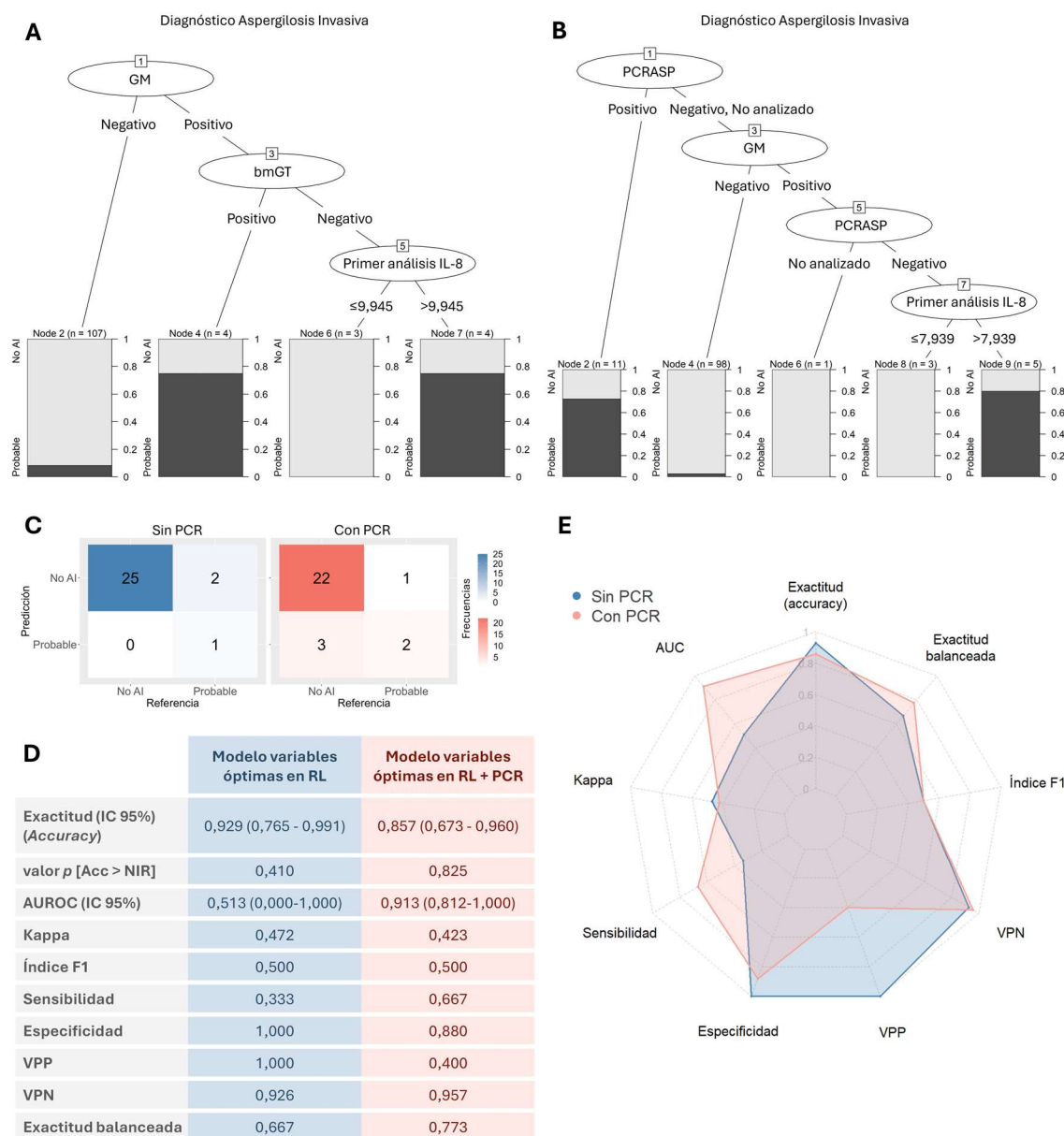


Figura 46. Resultados del desarrollo de modelos de clasificación de diagnóstico de AI basados en árboles de decisión. Esquema de los árboles de decisión resultantes de entrenar los modelos con el 80% de los datos (**A**) excluyendo o (**B**) incluyendo PCRASP. (**C**) Matrices de confusión que muestran los resultados de la predicción de los modelos utilizando el 20% restante de los datos. (**D**) Métricas de evaluación del comportamiento de los modelos en el set de datos de validación y (**E**) gráfico de radar que compara visualmente el desempeño de ambos modelos. Se muestran la exactitud (*accuracy*) y su intervalo de confianza del 95% (IC 95%), el valor *p* (relativo a si la exactitud es superior al *Non Information Rate* (NIR), que es el porcentaje de individuos que

pertenecen a la clase más numerosa de las clases estudiadas en el set de datos de prueba), área bajo la curva ROC (AUROC, del inglés, *area under the receiver operating characteristic*), kappa, índice F1, sensibilidad, especificidad, valor predictivo positivo (VPP) y valor predictivo negativo (VPN) y la exactitud balanceada (que toma en consideración si las clases presentan una distribución asimétrica). Se consideraron estadísticamente significativos valores $p < 0,05$.

El primero de ellos, utiliza las variables más importantes en la clasificación de pacientes con ese set de datos concreto, que son si los pacientes han dado al menos un resultado positivo en GM, bmGT y el valor del primer análisis de IL-8 (estableciendo un punto de corte en 9,945 pg/ml). Los porcentajes del uso de atributos fueron 100% para GM, 9,32% para bmGT y 5,93% para IL-8. La evaluación de la predicción en los datos de entrenamiento (118 casos) dio un error del 9,3%. A continuación, se realizó la evaluación del modelo utilizando datos que no fueron usados en el entrenamiento, es decir, con el 20% restante. La predicción de diagnóstico de estos pacientes se muestra en la matriz de confusión azul de la figura 46-C, donde se pudo observar la presencia de 2 FNs. Las métricas que evaluaron el comportamiento de este algoritmo en los datos de prueba se muestran en la figura 46-D, y de manera visual en la figura 46-E, donde se muestra una exactitud muy elevada (0,929), con un VPP del 100% y un VPN del 92,6%. La sensibilidad fue del 33% debido a la presencia de los dos FNs. La exactitud balanceada fue del 0,667.

Por otro lado, el árbol de decisión entrenado con los datos que incluían los resultados de PCRASP (Figura 46-B), utilizó como variables más importantes la existencia de al menos un resultado positivo en PCRASP, GM y el valor del primer análisis de IL-8 (estableciendo un punto de corte en 7,939 pg/ml). Los porcentajes del uso de atributos fueron 100%, 90,68% y 16,10%, respectivamente, con un error del 4,2%. En cuanto a la evaluación del modelo utilizando los datos de validación, se obtuvo una exactitud del 0,857, con la presencia de un FN y tres FPs. En este caso, la sensibilidad y especificidad están más balanceadas, siendo 66,7 y 88% respectivamente. Sin embargo, el VPP es solamente del 40% y la exactitud balanceada fue del 0,773.

En ninguno de los dos árboles de decisión desarrollados se obtuvo un valor p significativo relativo a la exactitud.

2.2.4 DISCUSIÓN

La GT es una micotoxina y factor de virulencia de *A. fumigatus*. Se ha demostrado que es producida por otras especies de *Aspergillus* y otros géneros de hongos con gran relevancia para la salud humana y animal debido a su capacidad para producir infecciones o contaminar alimentos humanos y animales. En este estudio, hemos desarrollado un análisis *in silico* de géneros y especies de hongos que potencialmente pueden producir GT, que ha sido validado en el laboratorio mediante el análisis de la producción de GT y su metabolito inactivo, bmGT, por especies de hongos seleccionadas. Además, se ha valorado la capacidad de bmGT como biomarcador de AI en pacientes inmunocomprometidos.

A pesar de que otros investigadores han demostrado la producción de GT por *A. niger*, *A. terreus* y *A. flavus* (aunque con frecuencias y concentraciones menores que *A. fumigatus*) (139, 179), nuestra búsqueda de secuencias homólogas a GliP no ofreció resultados significativos para estas especies. Los mejores resultados BLAST para las especies *A. terreus* y *A. flavus* no superaron las puntuaciones ni la identidad de los resultados mostrados. Las secuencias que presentaron esta coincidencia fueron proteínas NRPS involucradas en la formación de metabolitos secundarios. En concreto, ataP de *A. terreus* es necesario para la biosíntesis de la acetilaranotina (534) y AclP presente en *A. flavus* para la producción de aspiroclorina (535).

En cuanto al análisis filogenético, éste resulta esencial para el estudio comparativo de secuencias de proteínas. Sus resultados tienen muchas aplicaciones en el estudio de la evolución y funciones de las proteínas, así como en la predicción de la función de los genes que las codifican, la identificación, construcción y descubrimiento de familias de genes y la anotación del genoma (536). Por estas razones, es más apropiado utilizar secuencias de proteínas para analizar la filogenia de especies que utilizar secuencias de ADN (537, 538). Las proteínas con alta identidad de secuencia tienden a mostrar relaciones evolutivas y similitudes en función, lo que indica una función biológica conservada (539). El análisis filogenético proporcionó resultados representados por un árbol filogenético, en el que las secuencias se agruparon en función de las similitudes de secuencia.

El hecho de que las proteínas de las cepas de los géneros *Aspergillus* y *Penicillium* se ubicasen evolutivamente cerca en el árbol filogenético corroboró la similitud de secuencias que ya obtuvimos con el análisis BLAST. El resto de las proteínas también

mostraron coherencia entre el porcentaje de *query coverage* e identidad obtenida a partir de BLAST y la taxonomía del árbol filogenético inferido.

En cuanto al análisis realizado *in vitro*, la mayoría de las especies de la sección *Fumigati* fueron capaces de producir GT y bmGT. La cepa *A. fumigatus* B5233 produjo mayores concentraciones tanto de GT como de bmGT en ambos tipos de cultivo, comparada con la producción de *A. fumigatus* 1631562. Nuestros resultados fueron consistentes con los descritos por C. Kupfahl *et al.*, que describieron las concentraciones de GT detectadas en cepas de *A. fumigatus* de origen clínico (máximo 21,35 μ /ml, media 5,75 μ /ml) y ambiental (máximo 26,25 μ /ml, media 5,27 μ /ml) (139).

Sin embargo, no pudimos detectar GT o bmGT en los sobrenadantes de *A. lentulus* utilizando nuestras condiciones experimentales. Estudios previos han demostrado que una frecuencia muy baja de cepas y aislados de *A. lentulus* son capaces de generar estos metabolitos secundarios (142). Cabe mencionar que otros estudios no pudieron detectar estas toxinas en otras cepas de *A. lentulus* (141, 540). Se requieren estudios adicionales para dilucidar por qué los perfiles de metabolitos secundarios de *A. lentulus* son muy diferentes de los de *A. fumigatus* como se mostró previamente (141, 540) a pesar de ser especies estrechamente relacionadas.

Aunque el hongo no patógeno *A. fischeri* había sido descrito como incapaz de producir GT (140, 541), trabajos recientes muestran que efectivamente es capaz de producirla (542). Knowles *et al.* descubrieron que las condiciones de cultivo que inducen la síntesis de GT en *A. fumigatus* también inducen cantidades similares de GT en cultivos de *A. fischeri*. Sin embargo, nuestros resultados mostraron que se obtienen cantidades muy bajas de GT en cultivos de *A. fischeri* CBS 420.96 en comparación con *A. fumigatus*.

Respecto a las especies de otras secciones distintas de *Fumigati*, solo detectamos GT y bmGT en filtrados de cultivo de *T. virens* CBS 249.59, acorde con lo descrito anteriormente (543). Por otro lado, a pesar de que las cepas de *Penicillium* estudiadas hayan presentado alta homología de secuencia de GliP, no se ha detectado la producción de GT ni bmGT en ninguna de ellas. Sin embargo, otros investigadores han informado que algunas especies de *Penicillium* pueden producir GT (544-548). Esta discrepancia podría atribuirse a las condiciones de cultivo utilizadas. Estos resultados subrayan que la producción de metabolitos secundarios fúngicos puede verse afectada por las variables ambientales y destacan la importancia de optimizar las condiciones experimentales para potenciar el crecimiento de cepas concretas.

Discusión

La identificación de cepas productoras de estos metabolitos secundarios adquiere importancia en el contexto del manejo de las enfermedades producidas por dichos microorganismos, como la AI.

Pese a los numerosos avances en medicina clínica, el diagnóstico de infecciones complejas como la AI continúa siendo un desafío. El problema radica en que las manifestaciones clínicas de la AI son, en su mayoría, inespecíficas de esta enfermedad y en la limitada precisión de las herramientas de diagnóstico disponibles (506, 549). A modo de ejemplo, los cultivos de LBA tienen una sensibilidad aproximada del 50% (506). Además, el diagnóstico definitivo de la AI consiste en el uso de procedimientos invasivos que no suelen ser recomendables para el tipo de pacientes susceptibles de padecer la infección, dado que una gran parte de ellos son pacientes inmunocomprometidos. Estos métodos invasivos consisten en la identificación del hongo en muestras del tejido infectado mediante un examen histopatológico. La captura de tejido para el diagnóstico conlleva riesgos significativos, y debido a esto, muchos pacientes reciben un tratamiento antimicótico empírico en lugar de someterse a una biopsia.

Como se evidencia en nuestros resultados, un alto porcentaje de los pacientes que fueron diagnosticados como AI negativa habían recibido terapias farmacológicas mediante profilaxis y/o tratamiento con antifúngicos durante su episodio (29,4%). Se trata de un hecho altamente preocupante, ya que el tratamiento antifúngico conlleva efectos secundarios y supone un elevado gasto económico (549). En el caso de pacientes inmunocomprometidos, la capacidad para tolerar efectos secundarios graves se ve reducida por su estado funcional basal y sus comorbilidades. La toxicidad hepática y renal causada por los antifúngicos, así como las interacciones que presentan con otros fármacos, son particularmente relevantes debido a la administración de tratamientos concomitantes en estos pacientes, como antibióticos, quimioterapia y otros inmunosupresores. Sin embargo, la incertidumbre en la definición de la enfermedad, sobre todo para pacientes inmunodeprimidos en estado crítico, es un factor clave que contribuye a la controversia sobre el inicio de la terapia antimicótica (506). Cabe destacar que el inicio rápido del tratamiento antimicótico empírico ha demostrado beneficios en términos de mortalidad en pacientes con AI. Además, el retraso en el inicio de la terapia antimicótica se asocia a un mal pronóstico de estas infecciones (549), lo que conlleva una mayor duración del ingreso y, en consecuencia, mayores costes hospitalarios (43).

Es por ello que se pone de manifiesto la necesidad del estudio de una técnica sensible y específica para diagnosticar de manera temprana esta infección potencialmente mortal, y por lo tanto, plantear una intervención farmacológica solo si es necesaria, que aumente la eficacia y reduzca los efectos secundarios, así como sus costes asociados (550). En las últimas décadas, se han desarrollado varias pruebas de diagnóstico serológico y molecular para detectar biomarcadores de *Aspergillus spp.*, que pueden facilitar el diagnóstico temprano (506). En este estudio, nos centramos en la detección temprana comparando los biomarcadores utilizados actualmente en clínica con bmGT e IL-8, así como con combinaciones de ellos.

Previo a estas comparaciones, nos cercioramos de disponer de una cohorte de pacientes adecuada. Pudimos demostrar que la población estudiada es representativa de los estándares establecidos por los criterios EORTC, donde el diagnóstico se guía, entre otros parámetros, por la detección de GM en suero o LBA (551). Nuestros resultados ofrecen, mediante el estudio de los máximos de cada biomarcador durante los episodios de los pacientes, una capacidad diagnóstica elevada y significativa del GM para discernir entre pacientes sin AI y pacientes con AI probable. El AUC resultante (0,875) fue similar al encontrado en estudios anteriores (0,900) (552). Por otro lado, nuestros resultados muestran una sensibilidad mayor y una especificidad menor a las indicadas en estudios similares (552, 553). Es importante señalar que se obtuvo un *cut off* de 0,513 (índice de lectura), que concuerda con los criterios EORTC para considerar la prueba GM positiva en suero ($\geq 0,5$) (553), y asignar el diagnóstico probable en base a ello. De la misma manera, la detección de BDG también demostró una capacidad diagnóstica significativa, con un elevado VPN y similar a la mostrada en estudios previos (506, 554).

Dada la importancia y aplicabilidad clínica de un diagnóstico temprano, en este estudio se dirigieron todos los esfuerzos a evaluar la capacidad diagnóstica de los biomarcadores en los primeros análisis realizados a cada paciente durante su ingreso. La primera muestra se toma en el contexto de una sospecha clínica clara, generalmente cuando los signos y síntomas de la infección comienzan a manifestarse. Analizar el valor diagnóstico de esta primera muestra permite una evaluación clínica realista, que responde a las necesidades urgentes de los pacientes y contribuye a mejorar los resultados terapéuticos al facilitar una intervención más temprana. Detectar la existencia de infección en esta etapa inicial puede ser crucial sobre todo en pacientes inmunocomprometidos, en los que la AI progresa rápidamente y puede ser letal si no se diagnostica y trata a tiempo. Los resultados obtenidos en cuanto a la capacidad

Discusión

diagnóstica de los biomarcadores en el primer análisis realizado a los pacientes ponen en valor la utilidad de la bmGT, siendo equiparable en términos de especificidad y VPN al GM y BDG (aunque que el AUC de bmGT no resultó significativo) y superior a ambos si se elimina el sesgo del GM. De hecho, la bmGT fue el único biomarcador que ofreció una capacidad diagnóstica significativa utilizando los criterios EORTC-GM. Estos resultados indican una vinculación significativa entre los resultados de la primera muestra de bmGT analizada en los sueros de pacientes y el diagnóstico de AI en la población estudiada.

Centrarse en los niveles individuales de biomarcadores puede ofrecer solo una visión parcial de una patología compleja como la AI. Esta aproximación podría no considerar la variabilidad de los perfiles clínicos asociados con las infecciones respiratorias. Tanto las pruebas de biomarcadores como las moleculares están lejos de ser perfectas para el diagnóstico de AI cuando se utilizan individualmente, y se requieren pruebas combinadas para superar las limitaciones de rendimiento (555). Este hecho destaca la importancia de utilizar una combinación de pruebas para mejorar el abordaje de enfermedades complejas. Nuestros resultados demostraron que la combinación de bmGT, GM, IL-8 y CMV como biomarcadores de AI supone una mejora sustancial en la discriminación de pacientes con AI probable respecto al GM en solitario. Estudios clínicos recientes han demostrado que el índice de GM es un punto final sustituto útil de AI. Sin embargo, es importante reconocer que el GM puede ser indetectable en muestras tempranas debido a la presencia de agentes antimicóticos, como se vio en varios estudios con modelos animales (556). Cabe destacar que todas las combinaciones de biomarcadores que mostraron los mejores resultados incluyeron tanto GM como bmGT.

A partir de esta información, se desarrollaron algoritmos basados en árboles de decisión mediante técnicas de ML, con el objetivo de evaluar la capacidad predictiva de la combinación de dichos biomarcadores. Se utilizó el algoritmo C5.0, que elige las variables que finalmente conformarán el modelo basándose en su capacidad para mejorar la separación de las clases. Además, realiza *pruning* para evitar sobreajuste y obtener un árbol más simple y manejable que mantenga un buen desempeño predictivo. Las variables seleccionadas tienden a ser relevantes para la diferenciación de grupos, aunque no siempre son las “más importantes” en términos absolutos, ya que el árbol representa una combinación óptima de decisiones. Se decidió utilizar este tipo de árboles de decisión simples como análisis preliminar y con el objetivo de crear un resultado visual similar al de un protocolo de actuación de cualquier guía de práctica clínica, que fuese claro y de fácil comprensión para los facultativos. Nuestros resultados mostraron que, para nuestro

conjunto de datos, la predicción del diagnóstico de AI se optimizó con el uso de los biomarcadores GM, bmGT e IL-8 (en ausencia de los resultados de PCRASP). Este algoritmo de decisión alcanzó una elevada exactitud, especificidad, VPN y VPP. Por otro lado, el algoritmo que combina los resultados de PCRASP con los biomarcadores, utiliza también GM e IL-8 para la toma de decisiones, obteniendo unos resultados similares en términos de exactitud y VPN, con una mayor sensibilidad, pero una menor especificidad y VPP que en el árbol sin PCRASP. Es relevante destacar que ambos árboles de decisión incluyeron en su algoritmo la variable del primer análisis de IL-8, estableciendo un *cut off* en cada caso para clasificar a los individuos. Estos resultados sugieren la gran utilidad de la cuantificación temprana de los valores de IL-8 en los pacientes de nuestra cohorte, ya que resulta fundamental en la predicción de su diagnóstico.

Sin embargo, la exactitud de los árboles no fue estadísticamente significativa. Este hecho se debe a que el valor *p* analizado estudia si la exactitud es estadísticamente superior al *Non Information Rate* (NIR), que es el porcentaje de individuos que pertenecen a la clase más numerosa de las estudiadas en el set de datos de prueba. En nuestro caso, la clase más numerosa es “No AI” ($n = 25$), que supone un 89,29% de los datos de validación ($n = 28$). Por lo tanto, es de suma dificultad obtener una exactitud que sea estadísticamente superior a este porcentaje. Sin embargo, cabe destacar que, aunque el valor *p* no sea significativo, la exactitud del árbol de decisión que excluye la PCRASP, pero incluye bmGT, está por encima del NIR. En estos casos también es adecuado valorar el resultado de exactitud balanceada, que toma en consideración si existe un desequilibrio marcado entre las clases. Dicho resultado ofreció valores relativamente más bajos que la exactitud no balanceada para ambos árboles. La sensibilidad del árbol de decisión sin PCRASP y el VPP del árbol de decisión con PCRASP también son considerablemente bajas, debido al bajo número de sujetos en el grupo de diagnóstico AI probable. Aunque se den resultados erróneos en un número pequeño de pacientes, ello supone un alto porcentaje de fallo porque la cantidad de pacientes en ese grupo es muy limitada. De nuevo, se trata de un problema de desequilibrio en el número de individuos en cada clase, lo cual es una de las limitaciones de este estudio.

También es relevante señalar otras limitaciones de esta investigación. En primer lugar, la muestra analizada es limitada y específica, ya que proviene de un único hospital y está compuesta por pacientes con características clínicas y demográficas muy definidas. Este hecho podría dificultar la generalización de los hallazgos a poblaciones más amplias o a otros contextos clínicos, donde las condiciones pueden ser significativamente

Discusión

diferentes. También se debe considerar que la utilidad clínica de cualquier prueba diagnóstica y, en general, el manejo de pacientes con AI, depende de las particularidades de cada institución, incluidas la prevalencia de AI, la población en riesgo, las estrategias para el uso de antimicóticos (profilaxis, terapia preventiva, terapia empírica), los estudios costo-económicos y la disponibilidad de dichas herramientas diagnósticas (527, 531). Además, cabe destacar que nuestro estudio tiene un carácter retrospectivo, lo que puede introducir sesgos relacionados con la calidad y disponibilidad de los datos registrados. Este inconveniente se agrava por la presencia de datos incompletos en algunas variables clave, lo que podría afectar la robustez de los análisis realizados. Por otra parte, la decisión de eliminar el criterio de GM en el diagnóstico de AI para evitar sesgos puede haber introducido nuevas fuentes de error, especialmente al reetiquetar a los pacientes. Esto podría comprometer la independencia y objetividad de los resultados. Todas estas limitaciones subrayan la necesidad de validar los hallazgos en estudios prospectivos, multicéntricos y con muestras más amplias y heterogéneas.

En este estudio se demostró la producción de GT y bmGT en sobrenadantes de cultivos de especies de la sección *Fumigati* y *T. virens*. Además, se estudió la utilidad de la bmGT como biomarcador de AI, mediante la comparación de la precisión diagnóstica de los biomarcadores utilizados actualmente en clínica con la de bmGT e IL-8 tanto individualmente como en combinación. Los resultados son prometedores y sugieren que la bmGT e IL-8 muestran una utilidad potencial como herramientas de diagnóstico de AI. Por otro lado, los hallazgos indican que la combinación de biomarcadores resulta de mayor utilidad para detectar específicamente la AI de manera temprana. Sin embargo, se requerirán estudios prospectivos adicionales que utilicen una cohorte de pacientes más numerosa para confirmar estas conclusiones, con el objetivo de demostrar su aplicabilidad en clínica y las posibles mejoras que ello supone en el diagnóstico y tratamiento de la AI.

Capítulo



Estudio de biomarcadores para el
pronóstico y diagnóstico de COVID-19 y
otras infecciones respiratorias

Índice

Antecedentes.....	135
Material y métodos.....	138
Resultados.....	142
Discusión.....	168

2.3 CAPÍTULO 3 – Estudio de biomarcadores para el pronóstico y diagnóstico de COVID-19 y otras infecciones respiratorias

2.3.1 ANTECEDENTES

Las infecciones respiratorias agudas suponen una de las principales causas de morbimortalidad a nivel mundial. Además, representan un gran desafío para los sistemas sanitarios debido a su elevada incidencia, su heterogénea presentación clínica y su riesgo de evolucionar hacia cuadros graves. En este contexto, la pandemia de COVID-19 ha acentuado la importancia de las infecciones respiratorias como una amenaza global. A pesar de los avances en el estudio de su patogénesis, el diagnóstico y pronóstico de COVID-19 aún representan un reto considerable. El diagnóstico temprano y preciso es una prioridad clave para un manejo adecuado de las infecciones respiratorias. No obstante, uno de los aspectos más críticos es la capacidad de predecir la evolución de la enfermedad en pacientes individuales. Por lo tanto, no solo el diagnóstico, sino también la mejora en la predicción del pronóstico es esencial para optimizar el uso de recursos y guiar las intervenciones terapéuticas.

Numerosos estudios han señalado factores asociados con un mal pronóstico en infecciones respiratorias, especialmente en COVID-19. Entre ellos, marcadores inflamatorios elevados, comorbilidades y características demográficas (183). Durante las infecciones respiratorias, los patógenos interactúan con el hospedador produciendo una serie de alteraciones en el metabolismo celular que resultan en cambios en la composición de los compuestos involucrados en las rutas afectadas. Dos claros ejemplos de ello son las alteraciones en la composición de los COVs expirados y los inmunomoduladores solubles en suero. Aunque muchos de estos biomarcadores se han testado previamente en pacientes con COVID-19, su utilidad exacta en el diagnóstico y el pronóstico sigue siendo un desafío (384).

Aunque existen evidencias experimentales que respaldan la aplicación del análisis del aire exhalado para diagnosticar diferentes infecciones respiratorias (299, 328, 557-559), el estudio de los perfiles de COVs exhalados (“huellas respiratorias”) es un enfoque novedoso y no invasivo para el diagnóstico y seguimiento de enfermedades (560-564). Algunos resultados prometedores han sugerido el potencial del análisis de COVs en el diagnóstico de infecciones virales como las producidas por influenza (300, 565, 566) y rinovirus (567, 568). Recientemente, algunos investigadores han utilizado el análisis de COVs para la detección de COVID-19 en entornos públicos (569, 570). Estos estudios involucran diferentes compuestos (por ejemplo, aldehídos, cetonas, alcoholes,

Antecedentes

hidrocarburos, ácidos carboxílicos, entre otros) (309, 313, 315, 327, 571-573). Además, se han presentado resultados exitosos en cuanto a la detección de COVID-19 mediante dispositivos nasales electrónicos (303, 569, 574-577). Sin embargo, en la mayoría de los casos se realizaron análisis comparando pacientes con COVID-19 y controles sanos, sin estudiar la relación de los COVs expirados en aire con la gravedad de la enfermedad ni con otras enfermedades respiratorias.

Por otro lado, estudios previos demostraron que los perfiles elevados de citoquinas séricas se correlacionan directamente con la inflamación pulmonar, la lesión tisular y el mal pronóstico de COVID-19, evolucionando a estados graves (374, 375, 578). En consecuencia, la mayoría de los biomarcadores investigados en el pronóstico de pacientes con COVID-19 pertenecen a las vías inmunoinflamatorias y de coagulación (359). Cabe destacar que muchas citoquinas reguladas positivamente en COVID-19 también aumentan en otras enfermedades respiratorias, como la infección grave por influenza (579, 580).

Centrarse en los niveles de biomarcadores de manera individual puede dar solo una visión parcial de una enfermedad compleja que podría no explicar los diversos perfiles clínicos asociados al COVID-19 (581). Cada vez hay más evidencia que sugiere que los enfoques de inteligencia artificial basados en parámetros bioquímicos pueden mejorar la calidad de la toma de decisiones clínicas (582, 583). En particular, los métodos de ML han sido reconocidos como una herramienta útil para el diagnóstico de COVID-19 mediante el estudio del perfil de COVs (309). Por otro lado, el pronóstico de COVID-19 se ha enfocado principalmente en el estudio de las características inmunológicas (584).

No obstante, el desarrollo de estrategias diagnósticas y pronósticas, así como herramientas robustas para la estratificación del riesgo en pacientes con infecciones respiratorias sigue siendo un área de investigación prioritaria. La capacidad de predecir la gravedad de la enfermedad podría optimizar el manejo clínico, reduciendo así la mortalidad asociada y el impacto en los sistemas sanitarios. La identificación de un perfil de biomarcadores preciso y rigurosamente validado, sumado al desarrollo de modelos predictivos avanzados es esencial no solo para la gestión del COVID-19, sino también para mejorar el manejo clínico de otras infecciones respiratorias en las que también sea aplicable.

Hipótesis: *Considerando los procesos y mecanismos de defensa activados en los hospedadores frente a una infección respiratoria, es razonable esperar que tanto los COVs*

expirados en aire como los inmunomoduladores solubles estén presentes en concentraciones y composiciones diferentes en pacientes dependiendo de la gravedad de su enfermedad. La combinación de estos perfiles de parámetros diferenciadores podría utilizarse para facilitar el pronóstico y diagnóstico de las enfermedades infecciosas respiratorias, especialmente COVID-19.

En consecuencia, el **objetivo** principal de este capítulo es combinar la información de la composición de COVs expirados en aire y biomarcadores séricos relacionados con el sistema inmunológico, para proponer un enfoque más amplio tanto en el diagnóstico como en la diferenciación y clasificación de la gravedad de pacientes con infecciones respiratorias, especialmente COVID-19. Para llevarlo a cabo, se han establecido una serie de **objetivos intermedios**:

- ☼ Determinar las concentraciones de COVs expirados en aire y biomarcadores relacionados con el sistema inmunológico en suero de pacientes.
- ☼ Procesar y unificar los datos clínicos y analíticos obtenidos de pacientes con COVID-19, con otras infecciones respiratorias no-COVID-19, y sanos, con el objetivo de obtener un conjunto de datos adecuado y consistente que permita un análisis riguroso.
- ☼ Realizar un análisis exploratorio y estadístico de los datos.
- ☼ Desarrollar modelos de decisión basados en inteligencia artificial, utilizando diferentes técnicas de ML para predecir la gravedad de los pacientes en base a múltiples biomarcadores (COVs y biomarcadores en suero).
- ☼ Evaluar y explicar los modelos de clasificación desarrollados. En función de ello, estudiar el potencial de los biomarcadores estudiados para pronosticar la gravedad de los pacientes que padecen enfermedades infecciosas respiratorias.

2.3.2 MATERIALES Y MÉTODOS

2.3.2.1 Participantes y diseño del estudio

Se realizó un estudio prospectivo desde noviembre de 2021 hasta julio de 2023 con pacientes del Hospital Clínico Universitario Lozano Blesa (Zaragoza, España), que fue aprobado por el CEICA con la referencia PI21/292. Todos los pacientes firmaron un consentimiento informado antes de ser implicados en la investigación. El estudio incluyó 72 donantes sanos, 86 pacientes con COVID-19 y 75 pacientes con otras infecciones respiratorias (no-COVID-19). Los diagnósticos de los pacientes fueron confirmados mediante RT-qPCR (VIASURE) en su admisión al hospital. En el grupo no-COVID-19 se incluyeron pacientes que ofrecieron resultados negativos en las pruebas de RT-qPCR específicas de SARS-CoV-2 y que tuvieron un diagnóstico clínico, radiológico o microbiológico de evidencia de otra infección del tracto respiratorio. Los microorganismos causantes de estas infecciones fueron principalmente influenza, *Haemophilus influenzae* y *Streptococcus pneumoniae*.

Se recogieron datos demográficos y clínicos de los participantes, incluyendo el sexo, la edad, los hábitos tabáquicos, el IMC, si fueron ingresados en la UCI o no, el nivel de linfocitos y la necesidad de soporte de oxígeno. Los pacientes se agruparon en función de la gravedad de su enfermedad, de acuerdo con sus requerimientos de soporte de oxígeno. Concretamente, 50 de los 75 pacientes no-COVID-19 (66,7 %) y 35 de los 86 pacientes con COVID-19 (40,7 %) fueron asignados al grupo grave, debido a que precisaron soporte de oxígeno. El resto de los pacientes fueron asignados al grupo leve (Tabla 17).

Tabla 17. Clasificación de los pacientes en graves o leves según sus requerimientos de soporte de oxígeno. Se muestra el número de pacientes en cada grupo y su porcentaje respecto al grupo no-COVID-19 o COVID-19.

Gravedad de la enfermedad	Pacientes no-COVID-19 (n = 75)	Pacientes COVID-19 (n = 86)
Leve	25 (33,3 %)	51 (59,3 %)
Grave	50 (66,7 %)	35 (40,7 %)

2.3.2.2 Cuantificación de biomarcadores

Las muestras se recolectaron durante la admisión de los pacientes en el hospital, y en cualquier momento en el caso de los donantes sanos (después de haber certificado que éstos no sufrían ninguna enfermedad en ese momento ni en los 7 días anteriores).

Se utilizaron bolsas Tedlar 0.5L con válvula y accesorio de soplado para recoger muestras de aire expirado de todos los individuos incluidos en el estudio (Figura 47). La

preparación de la muestra se realizó mediante microextracción en fase sólida y los COVs contenidos en ella se analizaron mediante cromatografía de gases acoplada a espectrometría de masas (SPME/GC-MS, por sus siglas en inglés, *solid phase microextraction associated with gas chromatography-mass spectrometry*). Se analizaron las concentraciones de 27 COVs: etanol, isopropanol, acetona, 2-metil propanal, 2-butanona, hexano, acetato de etilo, isobutanol, ácido acético, 1-butanol, *i*-octano, metil isobutil cetona (MIBK, del inglés, *methyl isobutyl ketone*), tolueno, acetato de propilo, hexanal, ácido propanoico, etilbenceno, *m,p*-xileno, *o*-xileno, estireno, alfa-pineno, beta-pineno, decano, 3-careno, limoneno, mentol y butilhidroxitolueno (BHT).

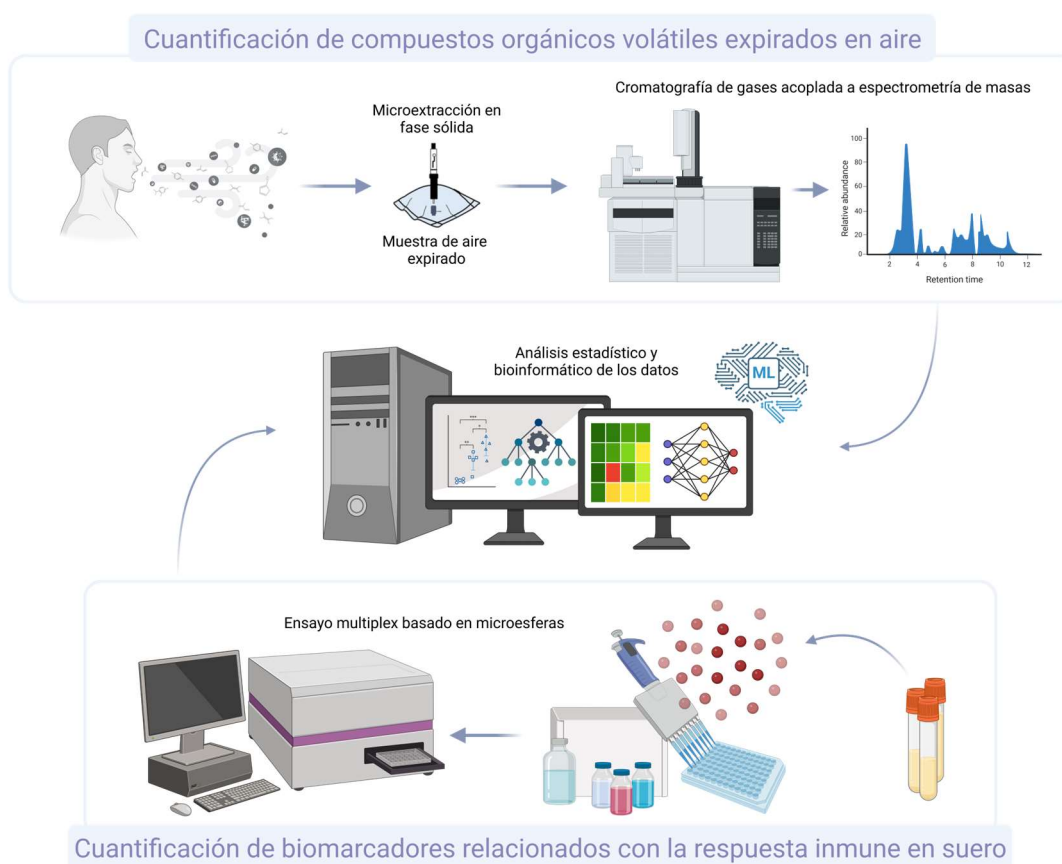


Figura 47. Protocolo utilizado para la detección y cuantificación de COVs en aire expirado y biomarcadores en muestras de suero.

También se cuantificaron las concentraciones de varios biomarcadores relacionados con la respuesta inmune en sueros de pacientes utilizando ensayos multiplex basados en microesferas (Luminex[®] Discovery Assay Human Premixed Multi-Analyte Kit, R&D Systems, Bio-Techne, Minneapolis, EE.UU.). Se analizaron citoquinas proinflamatorias como TNF- α , IL-2, IL-6, IL-8 e IL-17. También se cuantificó CXCL10. Por otro lado, se analizaron biomarcadores relacionados con la adhesión y coagulación como

vWF-A2, e-sel, ICAM-1, VCAM-1 y TSP. Asimismo, se cuantificaron proteínas relacionadas con el sistema inmune como GzmA, GzmB, TF y TM.

2.3.2.3 Análisis estadístico y bioinformático

Las características de los pacientes se mostraron como frecuencias (y sus porcentajes) para las variables categóricas y como medias \pm su SD o medianas (y su rango intercuartílico (IQR, del inglés, *interquartile range*)) para los datos continuos. Las diferencias estadísticas entre los grupos se analizaron mediante Chi-cuadrado o la prueba exacta de Fisher, y Mann-Whitney-Wilcoxon para los datos categóricos y continuos, respectivamente. Para estudiar la variabilidad de los COVs en los grupos estudiados se utilizó un análisis de componentes principales (PCA, del inglés, *Principal Component Analysis*) (585). Las diferencias en las concentraciones de los COVs entre poblaciones se analizaron mediante la prueba de Kruskal-Wallis, seguida del test de Dunn con corrección de Benjamini-Hochberg (586). Las diferencias en los parámetros de suero entre pacientes se calcularon utilizando la prueba de Mann-Whitney-Wilcoxon. También se realizaron mapas de calor o *heatmaps* para visualizar el agrupamiento y la correlación de las variables entre los grupos estudiados, utilizando el método de Ward para la generación de dendrogramas. Por otro lado, las relaciones entre los biomarcadores se estudiaron utilizando el coeficiente de correlación de Spearman, tanto para los grupos de pacientes COVID-19 como no-COVID-19 y sus respectivos subgrupos leves y graves. Adicionalmente, se estudió el potencial de todos los compuestos como biomarcadores de gravedad de la enfermedad mediante curvas ROC. Se calcularon el AUC, su error estándar y su IC 95%, el valor p , la sensibilidad, la especificidad, VPP y VPN. Los valores de corte o *cut off* se calcularon mediante el método del índice de Youden (529, 530), maximizando la sensibilidad y especificidad para distinguir entre los grupos.

Una vez estudiada la capacidad pronóstica de los biomarcadores por separado, se estudió su capacidad conjunta mediante la combinación de todas las variables (COVs y biomarcadores en suero) en el desarrollo de algoritmos de ML. Se construyeron modelos de predicción de gravedad para COVID-19 y no-COVID-19 individualmente. Se utilizaron diferentes algoritmos, incluyendo SVM, KNN, ANN, RL y RF. Además, se utilizaron técnicas de aprendizaje por conjuntos o *ensemble learning* como *bagging*, *boosting* y validación cruzada con ajuste de hiperparámetros para mejorar la generalización y robustez de los modelos. Los modelos se diseñaron utilizando el set de datos de entrenamiento (que consta de un 80 % de los datos) y el comportamiento del modelo fue validado utilizando el set de datos de prueba (con el 20 % restante que no ha sido utilizado en el desarrollo del

modelo). Solamente para el modelo que mejor se ajustó a nuestros datos (de acuerdo con varias métricas como exactitud y sensibilidad, entre otras), se realizó un análisis profundo y explicativo de su comportamiento tanto con los datos de entrenamiento como con los de prueba. Para abordar el hecho de que los modelos creados con RF ofrecen un algoritmo de caja negra, y mejorar su interpretabilidad, se realizaron análisis de importancia de las variables en la predicción de la gravedad de la enfermedad. Se utilizó el gráfico MDG para evaluar la importancia de las variables en los modelos desarrollados con los datos de entrenamiento. MDG mide cómo contribuye cada variable a la homogeneidad de los nodos y hojas del RF resultante. Muestra la importancia de cada compuesto en la clasificación de los datos de entrenamiento, expresando cuánta precisión pierde el modelo al excluir cada variable. Cuanto mayor sea el valor MDG, mayor será la importancia de la variable en el modelo (587, 588). También se utilizaron técnicas avanzadas para evaluar el comportamiento de los modelos y su ajuste a los datos de prueba. Se utilizó SHAP (454, 455) para explicar las predicciones globales y cuantificar la importancia de cada parámetro. En concreto, se empleó *Kernel Explainer*, un enfoque agnóstico, independiente del modelo que produce una regresión lineal ponderada en función de los datos, las predicciones y el modelo. Y por otro lado, se utilizó LIME (453) para explicar las predicciones individuales en dos ejemplos: un paciente leve y un paciente grave. Explicar una predicción se refiere a la presentación de herramientas escritas o visuales que permiten conocer cualitativamente la relación entre los componentes de la instancia y la predicción del modelo (389). Los modelos RF creados en este trabajo se utilizaron para predecir si un paciente tiene un caso grave de enfermedad, y los métodos SHAP y LIME resaltaron las variables que contribuyeron a esa predicción.

Se utilizaron *scripts* (fragmentos de código) desarrollados en lenguaje R con la versión “4.3.0 (2023-04-21 ucrt) -- Already Tomorrow”, plataforma “x86_64-w64-mingw32/x64 (64-bit)” (528) en Windows 10 x64 (build 19043). Se utilizó el entorno de desarrollo RStudio, versión 2022.12.0+353. Los paquetes y funciones utilizados se encuentran en el anexo 2 – Tabla S1. En este estudio se utilizó un nivel de significancia $\alpha = 0,05$, por lo que se consideraron estadísticamente significativos valores $p < 0,05$.

2.3.3 RESULTADOS

2.3.3.1 Diferencias entre sujetos con COVID-19, otras infecciones respiratorias no COVID-19 y sanos

Se incluyeron un total de 233 participantes en este estudio. Sus características clínicas y demográficas, así como las comparaciones entre grupos en base a ellas, se muestran en la tabla 18.

Tabla 18. Comparación de las características clínicas y demográficas entre los grupos COVID-19, no-COVID-19 y donantes sanos. Se realizó el análisis estadístico mediante Chi-cuadrado o la prueba exacta de Fisher para los datos continuos. La prueba no paramétrica Mann-Whitney-Wilcoxon se utilizó para las variables continuas. Se consideraron estadísticamente significativos los valores $p < 0,05$.

	Media \pm SD, mediana (IQR) o N (%)			Valor p		
	Controles sanos (n = 72)	Pacientes No-COVID-19 (n = 75)	Pacientes COVID-19 (n = 86)	Controles sanos vs No-COVID-19	Controles sanos vs COVID-19	No-COVID-19 vs COVID-19
Sexo, n (%)						
Femenino	55 (76,39%)	32 (42,67%)	38 (44,19%)	<0,001	<0,001	0,972
Edad (años)	39,72 \pm 11,23	70,83 \pm 19,83	66,67 \pm 17	<0,001	<0,001	0,069
Hábitos tabáquicos, n (%)						
Fumador	3 (4,17%)	13 (17,33%)	11 (12,79%)	0,014	0,032	0,862
Exfumador	19 (26,39%)	26 (34,67%)	41 (47,67%)	0,131	0,004	0,303
Desconocido		4 (5,33%)	1 (1,16%)			
IMC (kg/m ²)		29,14 \pm 5,81	28,83 \pm 6,66			0,530
UCI, n (%)						
Ingresado		1 (1,33%)	1 (1,16%)			1,000
Desconocido		2 (2,67%)				
Linfocitos (mil/mm ³)		1,28 \pm 0,66	1,1 (0,7)			0,600
Soporte de oxígeno, n (%)		50 (66,67%)	35 (40,7%)			0,001

SD: Desviación estándar; IQR: Rango intercuartílico; IMC: Índice de Masa Corporal; UCI: Unidad de Cuidados Intensivos

La media de edad del grupo de personas sanas fue más baja debido al perfil de los pacientes ingresados en el hospital. Los hábitos tabáquicos muestran diferencias estadísticamente significativas en sus comparaciones con los controles sanos. El número de admitidos en la UCI fue igual en ambos grupos de pacientes. En base a la necesidad de soporte de oxígeno, el 66,7 % de los pacientes no-COVID-19 y el 40,7 % de los pacientes COVID-19 fueron asignados al grupo de enfermedad grave.

2.3.3.1.1 Comparación entre los niveles de compuestos orgánicos volátiles en aire expirado

De los 27 COVs analizados, 7 no se detectaron en las muestras (2-metil propanal, 2-butanona, acetato de etilo, isobutanol, acetato de propilo, hexanal, y mentol), por lo que fueron eliminados del estudio.

Se realizó un PCA para identificar los perfiles de COVs asociados con el diagnóstico de enfermedad. La representación bidimensional de las observaciones (Figura 48-A), donde las dos primeras componentes principales (Dim1 y Dim2) explican el 30,3 % de la varianza total de los datos, mostró diferencias entre los grupos estudiados. Las elipses mostraron las áreas donde se concentraron los datos de cada grupo específico, basándose en la media y la dispersión. En general, los pacientes con COVID-19 mostraron mayor heterogeneidad comparados con los pacientes no-COVID-19 y controles sanos, ya que las elipses más amplias indican mayor variabilidad en la dispersión de los datos.

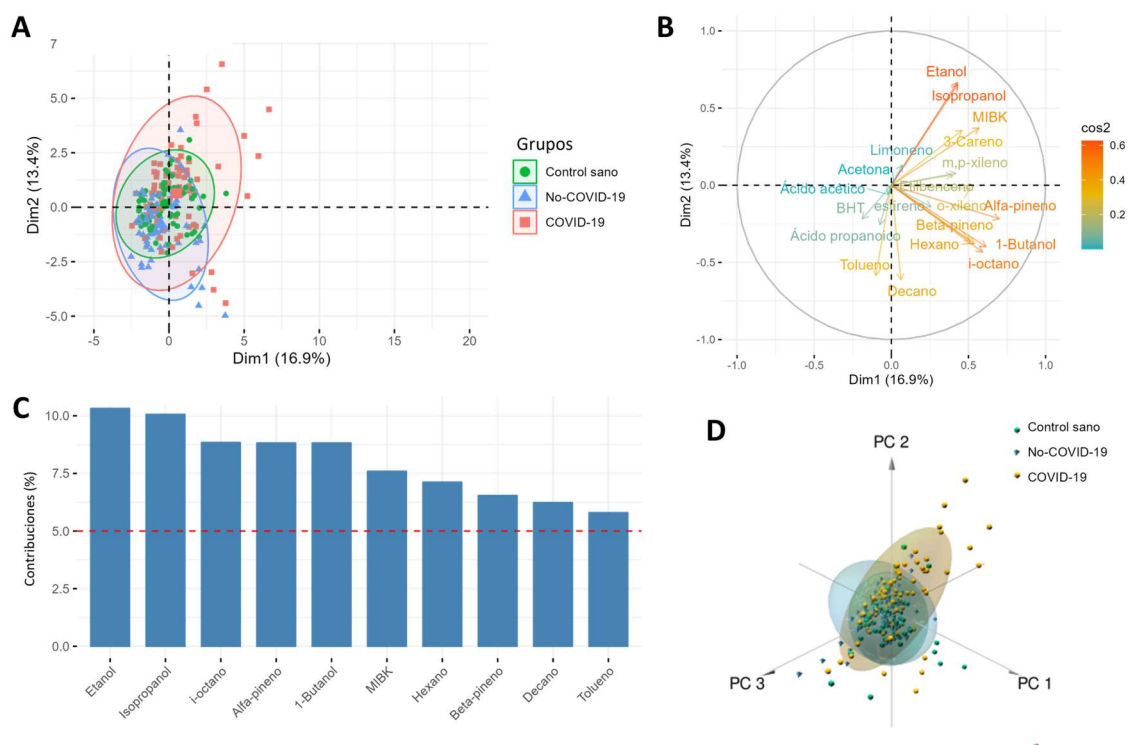


Figura 48. PCA del perfil de COVs expirados en aire de controles sanos y pacientes COVID-19 y no-COVID-19. **(A)** Representación de los individuos en las dimensiones (Dim) 1 y 2, agrupando las clases mediante elipses. **(B)** Gráfico del cuadrado del coseno (\cos^2) que muestra la calidad de representación de las variables en el espacio de las dimensiones 1 y 2. Un valor alto sugiere que la característica contribuye significativamente a esa dimensión. Las variables localizadas cerca del origen de coordenadas es posible que requieran más de dos componentes principales para su representación. **(C)** Variables que más contribuyen a las dimensiones 1 y 2. **(D)** Representación de las dimensiones 1-3 del PCA en un gráfico 3D. MIBK: Metil isobutil cetona; BHT: butilhidroxitolueno.

Alfa-pineno, 1-butanol e *i*-octano fueron los COVs que más contribuyeron al primer componente principal. Por otro lado, etanol, isopropanol y MIBK tuvieron una mayor contribución en el segundo componente principal (Figura 48-B, 48-C). Se identificaron tres perfiles de COVs distintos, que explican el 41,61 % de la varianza acumulada en un gráfico tridimensional (Figura 48-D).

Resultados

En cuanto a la comparación de las concentraciones de los COVs entre los grupos de participantes (Figura 49), todos presentaron diferencias significativas al menos entre dos grupos, excepto la acetona, el ácido acético, el 1-butanol y el BHT. MIBK y *m,p*-xileno fueron las variables más informativas, ya que mostraron diferencias significativas entre los tres grupos de diagnóstico.

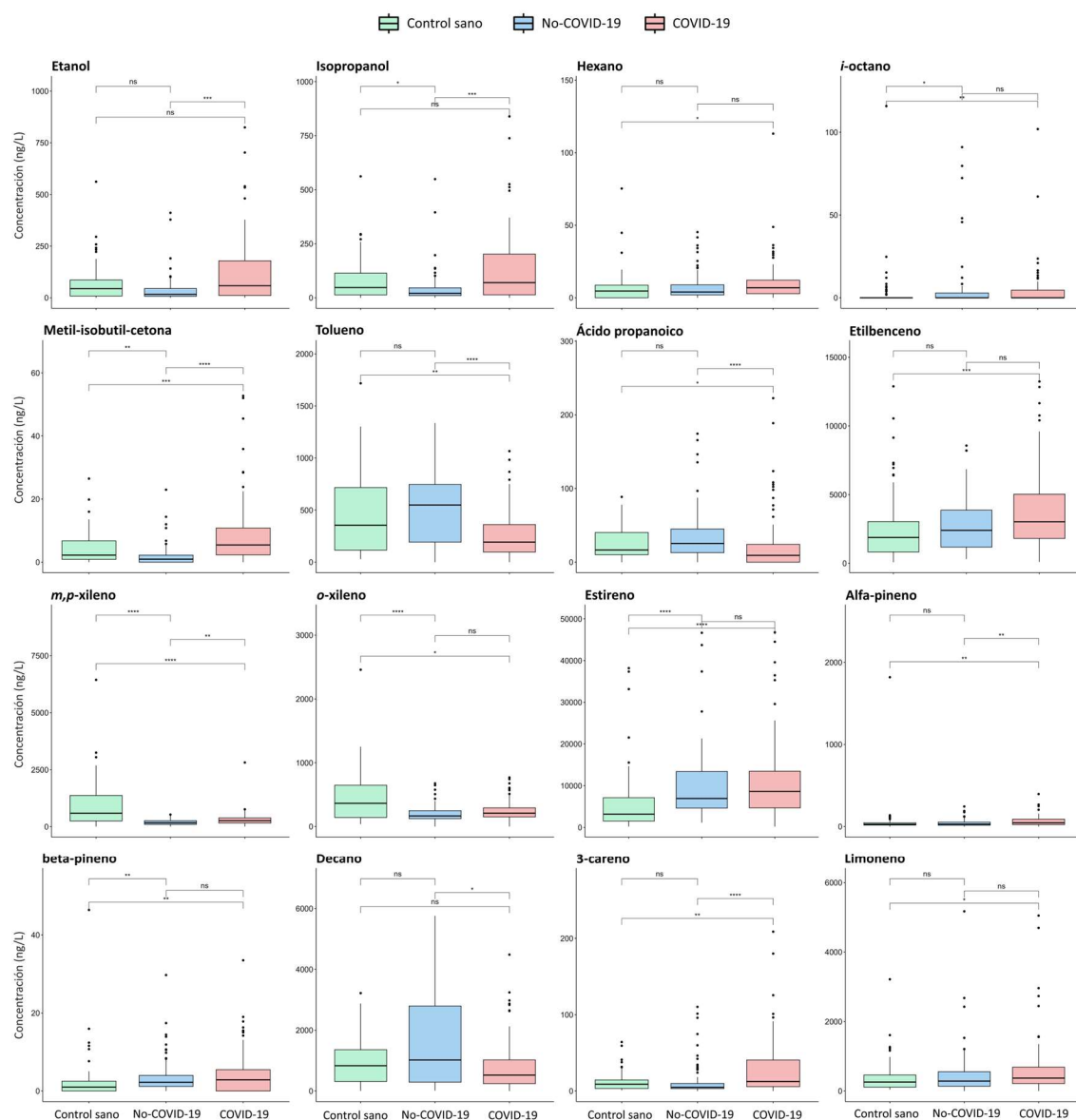


Figura 49. Comparación de las concentraciones de los COVs expirados en aire entre los donantes sanos, pacientes con COVID-19 y pacientes no-COVID-19. Se realizaron las pruebas Kruskal-Wallis y el post-test de Dunn para las comparaciones dos a dos. *, **, ***, ****, ns: Diferencias significativas ($p < 0,05$; $0,01$; $0,001$; $0,0001$; no significativo, respectivamente). Se consideraron estadísticamente significativos los valores $p < 0,05$.

Se encontraron niveles elevados de *i*-octano, estireno y beta-pineno en pacientes comparados con los sanos. Los niveles de *m,p*-xileno y *o*-xileno se vieron disminuidos, siendo *m,p*-xileno significativamente más alto en pacientes COVID-19 que en no-COVID-19. Los niveles de 3-careno, alfa-pineno y MIBK fueron superiores en los pacientes COVID-

19 respecto a los no-COVID-19. A su vez, las concentraciones de MIBK e isopropanol fueron significativamente menores en pacientes no-COVID que en sanos. Los pacientes con COVID-19 mostraron niveles más altos de hexano, etilbenceno y limoneno comparados con los controles sanos.

Los niveles de tolueno y ácido propanoico fueron significativamente menores en pacientes con COVID-19 comparado con el resto de los grupos. En cuanto a las concentraciones de decano, se vieron disminuidas en pacientes COVID-19 respecto a los no-COVID-19. El *heatmap* revela un patrón similar del decano y tolueno en todos los grupos (Figura 50). Los niveles de etanol e isopropanol fueron significativamente más altos en pacientes COVID-19, mostrando un perfil similar en el *heatmap*, que indicó mayores niveles de estos compuestos en uno de los clados formado principalmente por pacientes COVID-19.

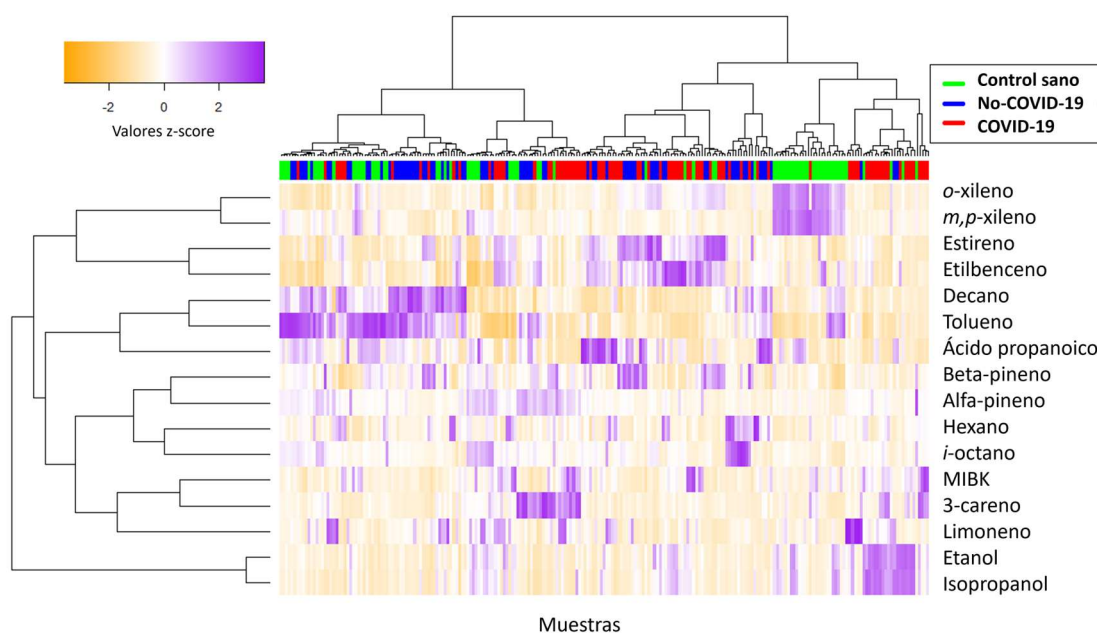


Figura 50. Comparación de los niveles de COVs expirados en aire en sujetos sanos, pacientes con COVID-19 y no-COVID-19. Mapa de calor o *heatmap* de las concentraciones de los COVs (estandarizadas mediante puntuación z o *z-score*) en aire expirado de los grupos de estudio. El dendrograma superior representa la distribución de los participantes y el dendrograma de la izquierda representa la clusterización de los COVs. MIBK: metil isobutil cetona.

2.3.3.2 Diferencias entre pacientes COVID-19 y no-COVID-19

2.3.3.2.1 Comparación entre los niveles de biomarcadores relacionados con la respuesta inmune

Se compararon las concentraciones de los biomarcadores en suero en los dos grupos de pacientes, COVID-19 y no-COVID-19. Los niveles de IL-2, IL-8, IL-17, TNF- α , e-sel y GzmA fueron significativamente más altos en pacientes no-COVID-19 que en

Resultados

pacientes COVID-19 (Figura 51). Sin embargo, las concentraciones de VCAM-1 fueron significativamente más elevadas en pacientes con COVID-19.

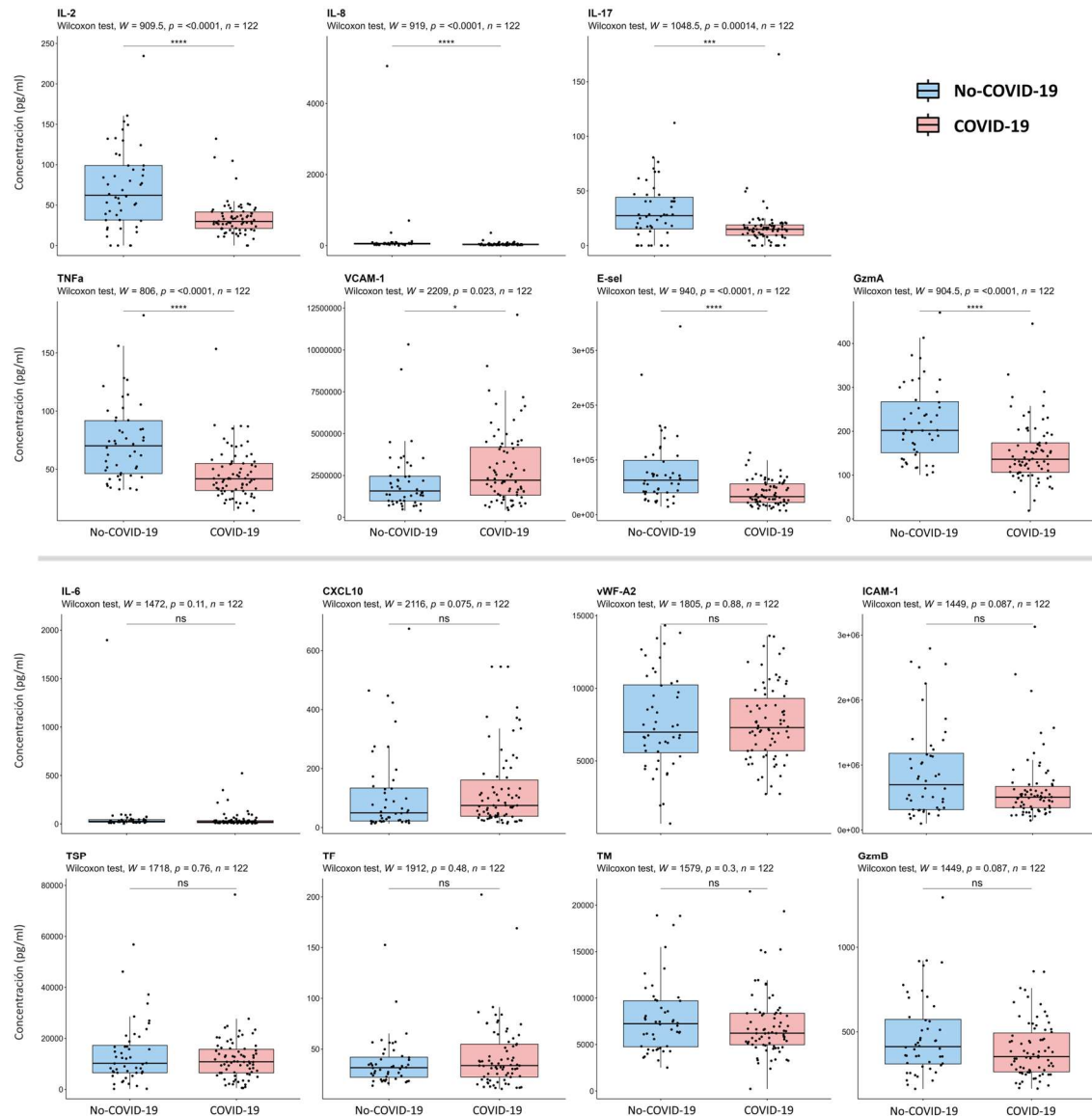


Figura 51. Comparación de los niveles de los biomarcadores en suero entre los pacientes con COVID-19 y pacientes no-COVID-19. Se utilizó la prueba de Mann-Whitney-Wilcoxon para las comparaciones entre grupos. *, **, ***, ****, ns: Diferencias significativas ($p < 0,05$; $0,01$; $0,001$; $0,0001$; no significativo, respectivamente). Se consideraron estadísticamente significativos los valores $p < 0,05$. vWF-A2: factor de von Willebrand A2; Gzm: granzima; TSP: trombospondina; ICAM-1: molécula de adhesión intercelular 1; IL: interleuquina; e-Sel: e-selectina; TNFa: factor de necrosis tumoral; CXCL10: ligando 10 de quimioquina con motivo C-X-C; VCAM-1: molécula de adhesión de células vasculares; TF: factor tisular; TM: trombomodulina.

Los niveles de CXCL10, VCAM-1, TF y TM mostraron un perfil similar en el mapa de calor (Figura 52), regulados al alza en algunos clústeres de pacientes COVID-19 y exhibiendo niveles bajos en el clado más homogéneo de los pacientes no-COVID-19. La visualización de la gravedad de los pacientes en el dendrograma superior de la figura 52 muestra una alta heterogeneidad en cuanto a la distribución de pacientes leves y graves.

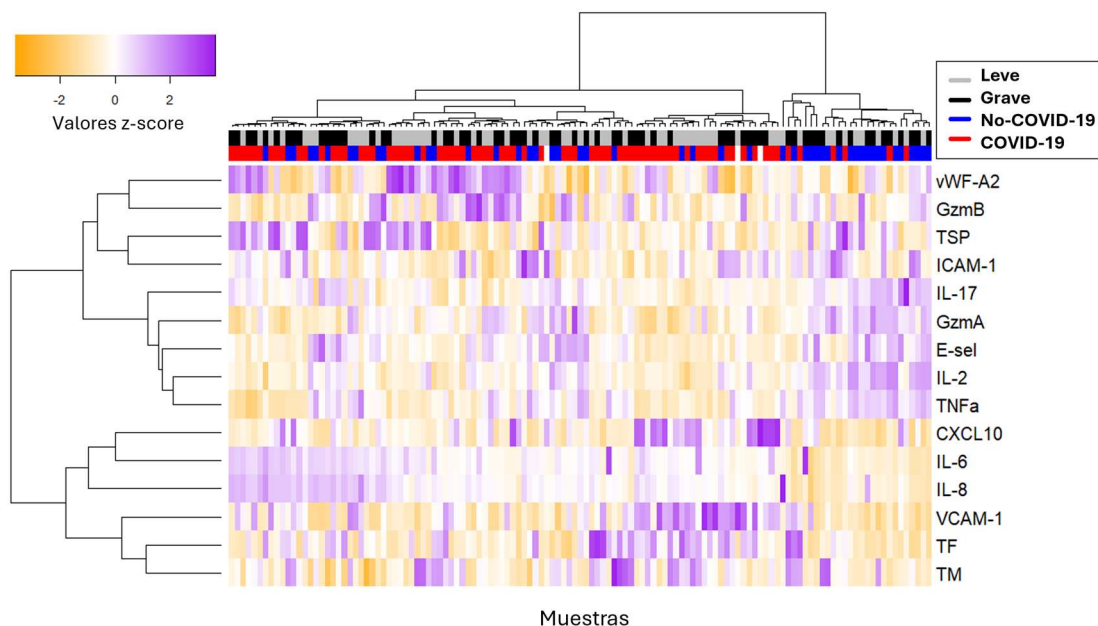


Figura 52. Comparación de los niveles de los biomarcadores en suero entre los pacientes COVID-19 y no-COVID-19. Mapa de calor o *heatmap* de las concentraciones de los parámetros en suero (estandarizadas mediante puntuación z o z-score) en suero de los pacientes. El dendrograma superior representa la distribución de los participantes, así como su gravedad, y el dendrograma de la izquierda representa la clusterización de los biomarcadores. vWF-A2: factor de von Willebrand A2; Gzm: granzima; TSP: trombospondina; ICAM-1: molécula de adhesión intercelular 1; IL: interleuquina; e-Sel: e-selectina; TNFa: factor de necrosis tumoral; CXCL10: ligando 10 de quimioquina con motivo C-X-C; VCAM-1: molécula de adhesión de células vasculares; TF: factor tisular; TM: trombomodulina.

2.3.3.2.2 Análisis del perfil de correlación entre los parámetros en suero y los compuestos orgánicos volátiles en aire expirado

Se analizó la correlación entre los niveles de todos los biomarcadores analizados (COVs expirados en aire y biomarcadores en suero) para los pacientes con COVID-19 y no-COVID-19 subgrupados por gravedad de la enfermedad (Figura 53).

Aparentemente, se observaron diferentes perfiles de correlación en pacientes con COVID-19 y no-COVID-19, así como en sus respectivos subgrupos leves y graves. El patrón de correlación de los pacientes no-COVID-19 (Figura 53 - A) mostró más simetría en cuanto a la gravedad que los pacientes con COVID-19 (Figura 53 - B). Este resultado indica que los pacientes con COVID-19 presentaron mayores diferencias en los perfiles de correlación según la gravedad.

Resultados

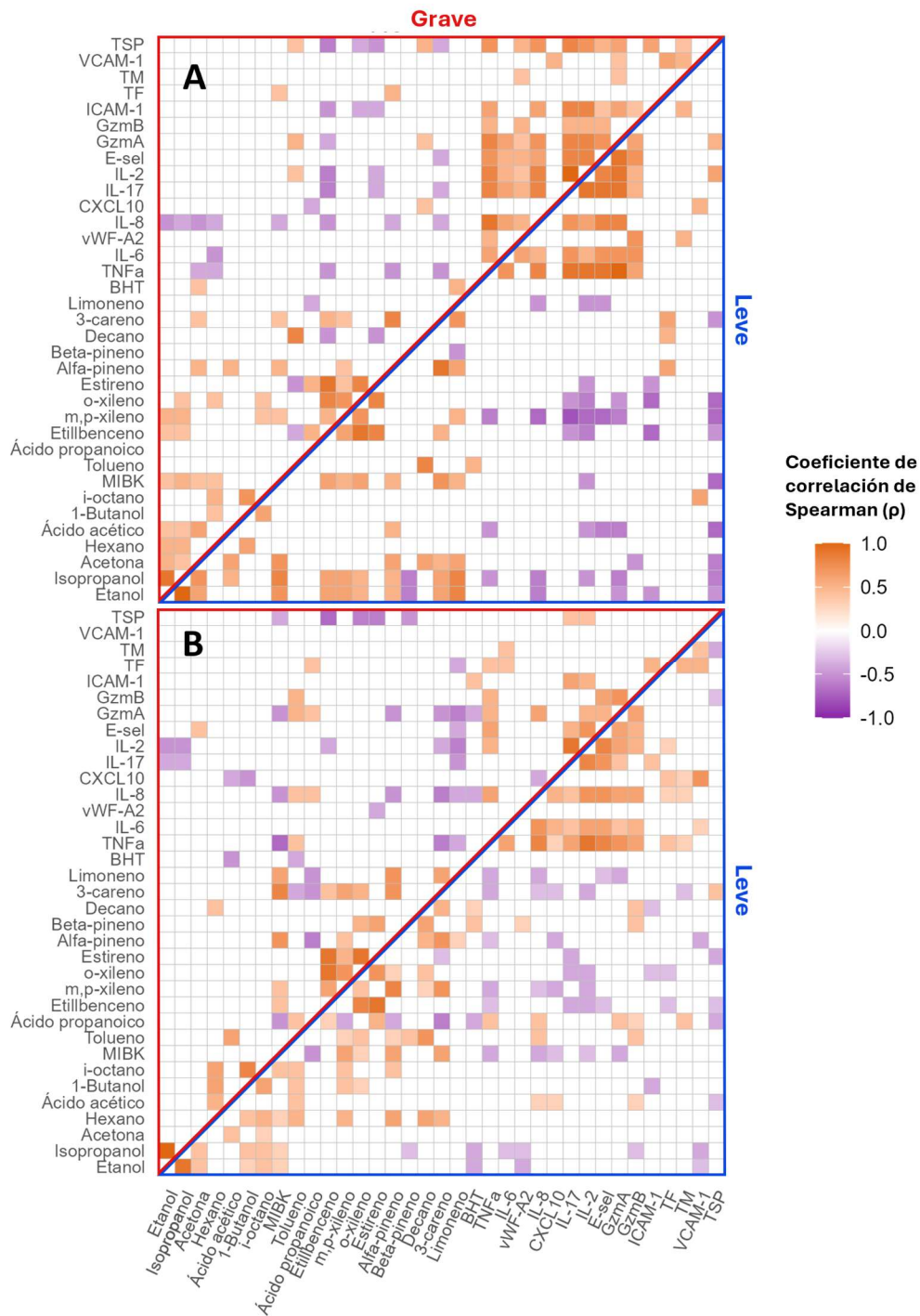


Figura 53. Heatmap que muestra la correlación de Spearman entre los COVs y biomarcadores de suero en los casos leves (triángulo superior izquierdo) y graves (triángulo inferior derecho), en **(A)** pacientes no-COVID-19 y **(B)** pacientes COVID-19. Los cuadrados blancos representan correlación no significativa ($p > 0,05$). vWF-A2: factor de von Willebrand A2; Gzm: granzima; TSP: trombospondina; ICAM-1: molécula de adhesión intercelular 1; IL: interleuquina; e-Sel: e-selectina; TNFa: factor de necrosis tumoral; CXCL10: ligando 10 de quimioquina con motivo C-X-C; VCAM-1: molécula de adhesión de células vasculares; TF: factor tisular; TM: trombomodulina; MIBK: metil isobutil cetona; BHT: butilhidroxitolueno.

2.3.3.3 Compuestos orgánicos volátiles y parámetros en suero como biomarcadores predictivos de gravedad de la enfermedad

2.3.3.3.1 Potencial de los compuestos orgánicos volátiles y parámetros en suero para discriminar la gravedad de los pacientes

Se calcularon curvas ROC para estudiar la capacidad discriminativa de la gravedad de enfermedad, para todos los biomarcadores estudiados y de manera individual para las cohortes de pacientes no-COVID-19 (Tabla 19) y COVID-19 (Tabla 20).

Tabla 19. Análisis del potencial discriminatorio en la clasificación de la gravedad de los pacientes no-COVID-19 basado en los COVs y biomarcadores en suero. Los puntos de corte o *cut-off* se calcularon mediante el método del Índice de Youden. También se calcularon la sensibilidad, especificidad, área bajo la curva (AUC, del inglés, *area under the curve*) y su error estándar e intervalo de confianza del 95 % (IC 95%). Se calculó el valor *p* para evaluar la diferencia en la capacidad predictiva entre la variable y el resultado binario, considerándose estadísticamente significativos los valores *p* inferiores a 0,05. * El valor del *cut-off* se expresa en ng/L para los COVs y en pg/mL para los biomarcadores en suero.

Compuesto	Cut-off*	AUC	Error estándar	IC 95 %	Valor <i>p</i>	Sensibilidad	Especificidad	Índice de Youden
Etanol	35,527	0,573	0,091	0,394 - 0,751	0,798	0,667	0,588	0,255
Isopropanol	71,912	0,590	0,097	0,401 - 0,780	0,849	0,833	0,471	0,304
Acetona	16,078	0,604	0,095	0,417 - 0,791	0,882	0,733	0,647	0,380
Hexano	5,401	0,655	0,088	0,483 - 0,827	0,962	0,700	0,647	0,347
Ácido acético	8,856	0,401	0,078	0,248 - 0,554	0,914	0,033	1,000	0,033
1-Butanol	1,178	0,573	0,085	0,407 - 0,739	0,816	0,833	0,412	0,245
<i>i</i> -octano	75,993	0,473	0,079	0,317 - 0,628	0,643	0,067	1,000	0,067
MIBK	3,709	0,712	0,083	0,549 - 0,874	0,992	0,933	0,529	0,463
Tolueno	679,306	0,598	0,092	0,417 - 0,779	0,867	0,833	0,471	0,304
Ácido propanoico	83,029	0,555	0,097	0,366 - 0,744	0,734	0,967	0,294	0,261
Etilbenceno	3925,157	0,692	0,085	0,526 - 0,859	0,986	0,833	0,588	0,422
<i>m,p</i> -xileno	279,817	0,571	0,097	0,381 - 0,760	0,789	0,833	0,412	0,245
<i>o</i> -xileno	175,145	0,676	0,085	0,509 - 0,844	0,978	0,633	0,824	0,457
Estireno	7329,303	0,637	0,089	0,463 - 0,812	0,940	0,633	0,824	0,457
Alfa-pineno	40,900	0,612	0,091	0,434 - 0,790	0,897	0,733	0,588	0,322
Beta-pineno	2,881	0,529	0,096	0,342 - 0,717	0,635	0,767	0,471	0,237
Decano	257,780	0,655	0,082	0,494 - 0,816	0,961	0,367	1,000	0,367
3-Careno	5,164	0,653	0,089	0,479 - 0,827	0,959	0,600	0,765	0,365
Limoneno	189,797	0,565	0,087	0,393 - 0,736	0,769	0,433	0,765	0,198
BHT	5,367	0,550	0,090	0,374 - 0,726	0,719	0,867	0,294	0,161
TNF- α	44,342	0,547	0,087	0,376 - 0,718	0,704	0,267	0,882	0,149
IL-6	27,909	0,557	0,100	0,362 - 0,752	0,741	0,633	0,588	0,222
vWF-A2	12190,161	0,486	0,096	0,298 - 0,675	0,443	0,933	0,235	0,169
IL-8	46,636	0,547	0,089	0,373 - 0,721	0,704	0,500	0,706	0,206
CXCL10	130,911	0,645	0,086	0,476 - 0,814	0,950	0,833	0,471	0,304
IL-17	25,716	0,459	0,092	0,278 - 0,640	0,684	0,600	0,588	0,188

Resultados

IL-2	71,949	0,502	0,091	0,324 - 0,680	0,496	0,500	0,647	0,147
E-Sel	76626,290	0,608	0,090	0,431 - 0,785	0,889	0,833	0,471	0,304
GzmA	210,474	0,490	0,093	0,309 - 0,672	0,548	0,500	0,647	0,147
GzmB	834,388	0,457	0,088	0,284 - 0,630	0,691	0,133	0,941	0,075
ICAM-1	442833,848	0,588	0,085	0,421 - 0,755	0,841	0,433	0,824	0,257
TF	37,560	0,549	0,089	0,374 - 0,724	0,296	0,400	0,824	0,224
TM	6224,759	0,612	0,086	0,443 - 0,781	0,107	0,733	0,529	0,263
VCAM-1	2461539,215	0,539	0,092	0,359 - 0,720	0,673	0,800	0,353	0,153
TSP	51491,250	0,443	0,093	0,262 - 0,625	0,266	1,000	0,059	0,059

MIBK: metil isobutil cetona; BHT: butilhidroxitolueno; TNF- α : factor de necrosis tumoral; IL: interleuquina; vWF-A2: factor de von Willebrand A2; CXCL10: ligando 10 de quimioquina con motivo C-X-C; e-Sel: e-selectina; Gzm: granzima; ICAM-1: molécula de adhesión intercelular 1; TF: factor tisular; TM: trombomodulina; VCAM-1: molécula de adhesión de células vasculares; TSP: trombospondina.

Tabla 20. Análisis del potencial discriminatorio en la clasificación de la gravedad de los pacientes COVID-19 basado en los COVs y biomarcadores en suero. Los puntos de corte o *cut-off* se calcularon mediante el método del Índice de Youden. También se calcularon la sensibilidad, especificidad, área bajo la curva (AUC, del inglés, *area under the curve*) y su error estándar e intervalo de confianza del 95 % (IC 95%). Se calculó el valor *p* para evaluar la diferencia en la capacidad predictiva entre la variable y el resultado binario, considerándose estadísticamente significativos los valores *p* inferiores a 0,05. * El valor del cut-off se expresa en ng/L para los COVs y en pg/mL para los biomarcadores de suero.

Compuesto	Cut-off*	AUC	Error estándar	IC 95 %	Valor <i>p</i>	Sensibilidad	Especificidad	Índice de Youden
Etanol	59,126	0,609	0,068	0,477 - 0,742	0,940	0,654	0,625	0,279
Isopropanol	44,457	0,622	0,066	0,492 - 0,752	0,958	0,538	0,729	0,268
Acetona	7,943	0,571	0,072	0,430 - 0,712	0,844	0,462	0,729	0,191
Hexano	6,100	0,515	0,070	0,378 - 0,652	0,417	0,654	0,500	0,154
Ácido acético	0,282	0,624	0,057	0,512 - 0,736	0,012	0,423	0,833	0,256
1-Butanol	3,512	0,579	0,066	0,451 - 0,708	0,103	0,231	0,958	0,189
<i>i</i> -octano	7,864	0,534	0,066	0,406 - 0,663	0,296	0,269	0,896	0,165
MIBK	5,626	0,444	0,070	0,308 - 0,581	0,786	0,538	0,500	0,038
Tolueno	263,840	0,764	0,057	0,652 - 0,877	<0,001	0,654	0,771	0,425
Ácido propanoico	3,337	0,595	0,065	0,466 - 0,723	0,080	0,692	0,563	0,255
Etilbenceno	5147,894	0,623	0,070	0,486 - 0,760	0,041	0,423	0,833	0,256
<i>m,p</i> -xileno	350,939	0,576	0,073	0,433 - 0,719	0,143	0,500	0,729	0,229
<i>o</i> -xileno	236,983	0,645	0,073	0,503 - 0,788	0,020	0,577	0,729	0,306
Estireno	10507,447	0,673	0,068	0,540 - 0,806	0,007	0,615	0,729	0,345
Alfa-pineno	100,922	0,512	0,073	0,369 - 0,656	0,433	0,308	0,792	0,099
Beta-pineno	2,948	0,605	0,070	0,468 - 0,742	0,068	0,654	0,625	0,279
Decano	1217,945	0,534	0,074	0,389 - 0,680	0,316	0,231	0,896	0,127
3-Careno	6,957	0,609	0,069	0,475 - 0,744	0,940	0,423	0,813	0,236
Limoneno	326,568	0,554	0,076	0,405 - 0,702	0,778	0,615	0,646	0,261
BHT	9,443	0,537	0,068	0,404 - 0,670	0,294	0,192	0,938	0,130
TNF- α	42,209	0,700	0,067	0,567 - 0,832	0,002	0,692	0,667	0,359
IL-6	22,181	0,804	0,052	0,702 - 0,905	<0,001	0,846	0,708	0,554

vWF-A2	4756,520	0,548	0,073	0,404 - 0,692	0,752	0,231	0,938	0,168
IL-8	30,445	0,780	0,053	0,675 - 0,884	<0,001	0,769	0,729	0,498
CXCL10	114,217	0,636	0,067	0,504 - 0,768	0,027	0,538	0,771	0,309
IL-17	16,926	0,569	0,077	0,418 - 0,719	0,168	0,538	0,750	0,288
IL-2	33,355	0,633	0,074	0,488 - 0,779	0,030	0,615	0,688	0,303
E-Sele	31677,957	0,616	0,068	0,483 - 0,750	0,051	0,731	0,583	0,314
GzmA	155,552	0,645	0,068	0,511 - 0,779	0,020	0,577	0,771	0,348
GzmB	345,574	0,599	0,072	0,458 - 0,739	0,083	0,731	0,583	0,314
ICAM-1	692569,232	0,613	0,075	0,466 - 0,760	0,056	0,423	0,854	0,277
TF	42,095	0,736	0,062	0,614 - 0,857	<0,001	0,615	0,833	0,449
TM	6112,634	0,747	0,062	0,625 - 0,869	<0,001	0,885	0,646	0,530
VCAM-1	2474977,371	0,688	0,065	0,561 - 0,815	0,004	0,692	0,667	0,359
TSP	13944,100	0,599	0,070	0,462 - 0,736	0,920	0,846	0,396	0,242

MIBK: metil isobutil cetona; BHT: butilhidroxitolueno; TNF- α : factor de necrosis tumoral; IL: interleuquina; vWF-A2: factor de von Willebrand A2; CXCL10: ligando 10 de quimioquina con motivo C-X-C; e-Sele: e-selectina; Gzm: granzima; ICAM-1: molécula de adhesión intercelular 1; TF: factor tisular; TM: trombomodulina; VCAM-1: molécula de adhesión de células vasculares; TSP: trombospondina.

Según el análisis de curvas ROC, se identificaron 14 compuestos con capacidad discriminativa significativa ($p < 0,05$) para predecir la gravedad de enfermedad en pacientes con COVID-19. Las curvas ROC de dichos compuestos se muestran en la figura 54.

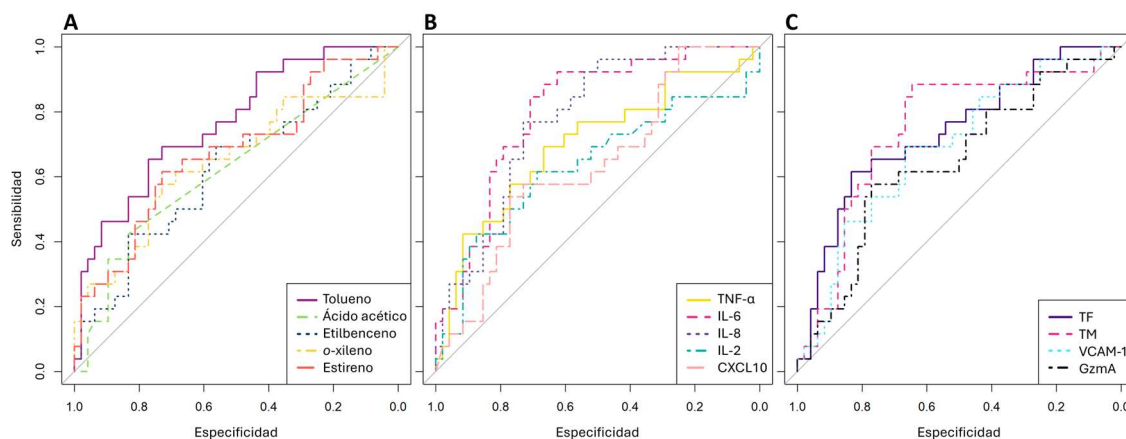


Figura 54. Análisis ROC (*Receiver Operating Characteristic*) que discrimina entre pacientes graves y leves pertenecientes al grupo de individuos con COVID-19. **(A)** Curvas ROC de los COVs que mostraron resultados estadísticamente significativos. **(B)** Curvas ROC de las citoquinas que mostraron resultados estadísticamente significativos. **(C)** Curvas ROC del resto de los biomarcadores en suero que mostraron resultados estadísticamente significativos. Se consideraron estadísticamente significativos los valores $p < 0,05$. TNF- α : factor de necrosis tumoral; IL: interleuquina; CXCL10: ligando 10 de quimioquina con motivo C-X-C; TF: factor tisular; TM: trombomodulina; VCAM-1: molécula de adhesión de células vasculares; Gzm: granzima.

Los parámetros IL-6, IL-8, TM, y tolueno mostraron valores $p < 0,001$, con valores AUC cerca de 0,80, indicando un alto rendimiento de las variables individuales. En particular, IL-6 y TM produjeron un Índice de Youden superior a 0.5, indicando un alto poder

Resultados

discriminativo y un buen equilibrio entre la sensibilidad y la especificidad. El análisis de las curvas ROC para IL-6, IL-8, TM y tolueno demostró una buena sensibilidad (70,8 %, 76,9 %, 88,5 % y 65,4 %, respectivamente) para predecir la gravedad de los pacientes con COVID-19. El TF, etilbenceno y ácido acético fueron los biomarcadores que presentaron mayor rendimiento en términos de especificidad, seguidos por el tolueno, CXCL10 y GzmA.

2.3.3.3.2 Desarrollo de modelos de predicción de gravedad para pacientes con COVID-19 y no-COVID-19

Con el objetivo de profundizar más en la caracterización de los perfiles de COVs y parámetros séricos, se realizó un análisis multivariante que combinó la información de todos los analitos estudiados. Se desarrollaron cinco modelos de clasificación de gravedad de enfermedad basados en algoritmos de ML (RL, RF, SVM, KNN y ANN), creados individualmente para los datos de COVID-19 y no-COVID-19. Estos modelos se desarrollaron con el set de datos de entrenamiento, es decir, con un 80 % del conjunto de datos original. Los detalles sobre la configuración de los parámetros principales de los modelos optimizados se especifican en la tabla 21.

Tabla 21. Parámetros del modelo de mejor ajuste para los algoritmos de ML desarrollados.

Modelo	Ajuste de parámetros del modelo	
	No-COVID-19	COVID-19
Regresión logística	Modelo completo (entrenado con todas las variables)	
<i>Random Forest</i>	Entrenado desarrollando 500 árboles 5 variables probadas en cada división	
<i>Support Vector Machines</i>	Función <i>kernel</i> de base radial Hiperparámetro C = 15	
<i>K-Nearest Neighbors</i>	k = 25	k = 9
	Validación cruzada de 10 pasos utilizando la exactitud para seleccionar el modelo óptimo	
Redes Neuronales Artificiales	5 capas ocultas <i>decay</i> = 1e-04 como parámetro de regularización para evitar sobreajuste	1 capa oculta Validación cruzada de 10 pasos utilizando la exactitud para seleccionar el modelo óptimo

El modelo desarrollado mediante RL se entrenó con todos los biomarcadores incluidos en este estudio. Se compararon los impactos relativos de las diferentes variables en los modelos no-COVID-19 (Figura 55 – A) y COVID-19 (Figura 55 – B). Los *forest plots* muestran los *odds ratios* (razones de probabilidad) de las variables incluidas en los modelos de RL para predecir la gravedad de la enfermedad. La mayoría de ellas mostraron

odds ratios no significativos (sus IC 95% incluyeron el 1), lo que sugiere que no tienen un efecto estadísticamente relevante en la probabilidad de gravedad en estos modelos. Los IC amplios de las variables indican una menor precisión en sus estimaciones. Solamente el etilbenceno, TM y vWF-A2 mostraron *odds ratios* significativos en el modelo de predicción de gravedad en los pacientes con COVID-19. En el caso del etilbenceno, su aumento se asoció con un incremento en las probabilidades de padecer una enfermedad grave. Por el contrario, el aumento de los niveles de TM y vW-FA2 se asoció con una disminución significativa en las probabilidades de gravedad.

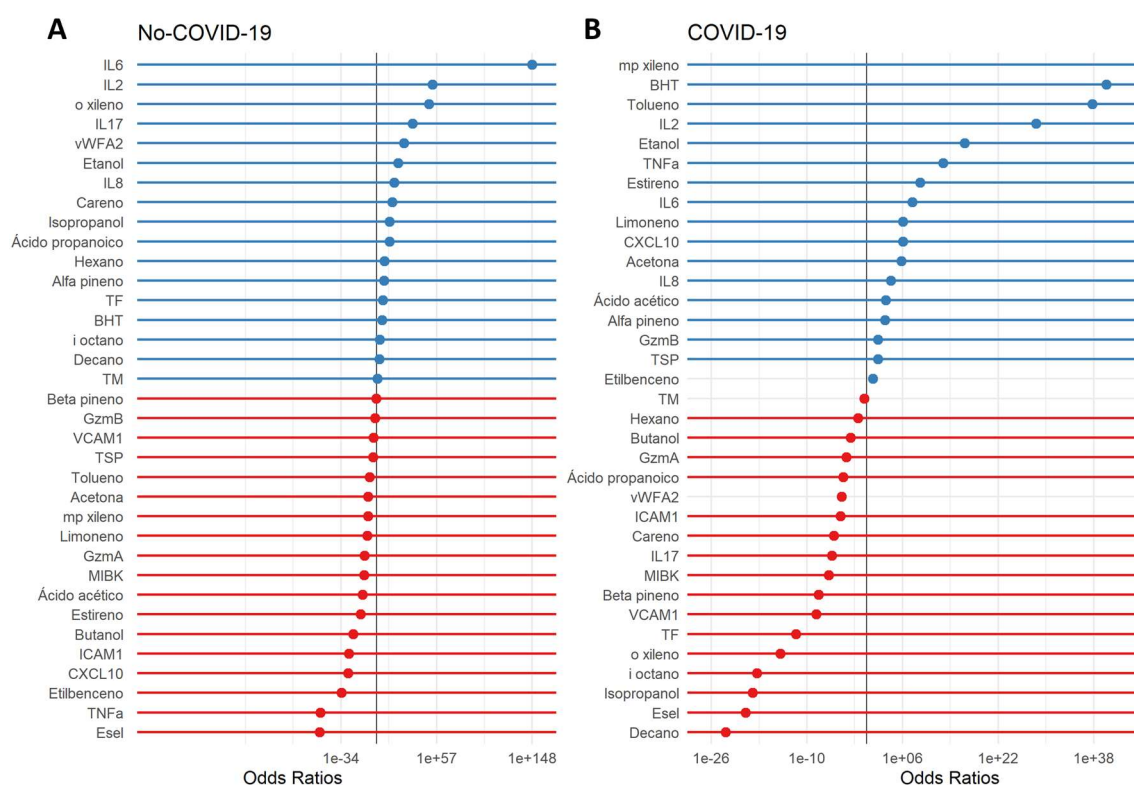


Figura 55. Forest plot que representa gráficamente las razones de probabilidades u *odds ratios* de las variables predictoras del modelo de regresión logística desarrollado para los datos de entrenamiento de los grupos (A) no-COVID-19 y (B) COVID-19, junto con sus intervalos de confianza, mostrando la magnitud y la significancia estadística de cada efecto. La línea vertical indica la neutralidad (sin efecto), *odds ratios* = 1. Los puntos desplazados hacia la derecha (> 1) indican un efecto positivo en la predicción de gravedad. Los puntos desplazados hacia la izquierda (< 1) indican un efecto negativo. La distancia del punto a la línea horizontal refleja la magnitud del efecto. Las barras de error representan el IC 95%. Si la barra del IC cruza el valor 1 en el eje horizontal, el efecto de esa variable no es estadísticamente significativo, ya que no se puede descartar que la variable tenga un efecto neutral. MIBK: metil isobutil cetona; BHT: butilhidroxitolueno; TNFA: factor de necrosis tumoral; IL: interleuquina; vWFA2: factor de von Willebrand A2; CXCL10: ligando 10 de quimioquina con motivo C-X-C; Esel: e-selectina; Gzm: granzima; ICAM1: molécula de adhesión intercelular 1; TF: factor tisular; TM: trombomodulina; VCAM1: molécula de adhesión de células vasculares; TSP: trombospondina.

Mediante la representación del *forest plot*, los logaritmos de probabilidades (*log-odds*) se transformaron a probabilidades relativas (*odds*), lo cual puede llevar a valores

Resultados

extremos que pueden quedar fuera del rango representado, como es el caso del *m,p*-xileno en el modelo COVID-19.

En general, los IC para todas las variables son extremadamente amplios, lo que indica que la estimación es muy imprecisa y que el modelo no proporciona información útil para esas variables.

Respecto a los modelos desarrollados con el algoritmo KNN, se empleó el método de validación cruzada de 10 pasos utilizando el valor de la exactitud para seleccionar el modelo óptimo. Dicha optimización resultó en un modelo que utiliza $k = 9$ vecinos más cercanos para hacer las predicciones en el caso del modelo COVID-19 (Figura 56 – A), y $k = 25$ vecinos más cercanos en el modelo no-COVID-19 (Figura 56 – B).

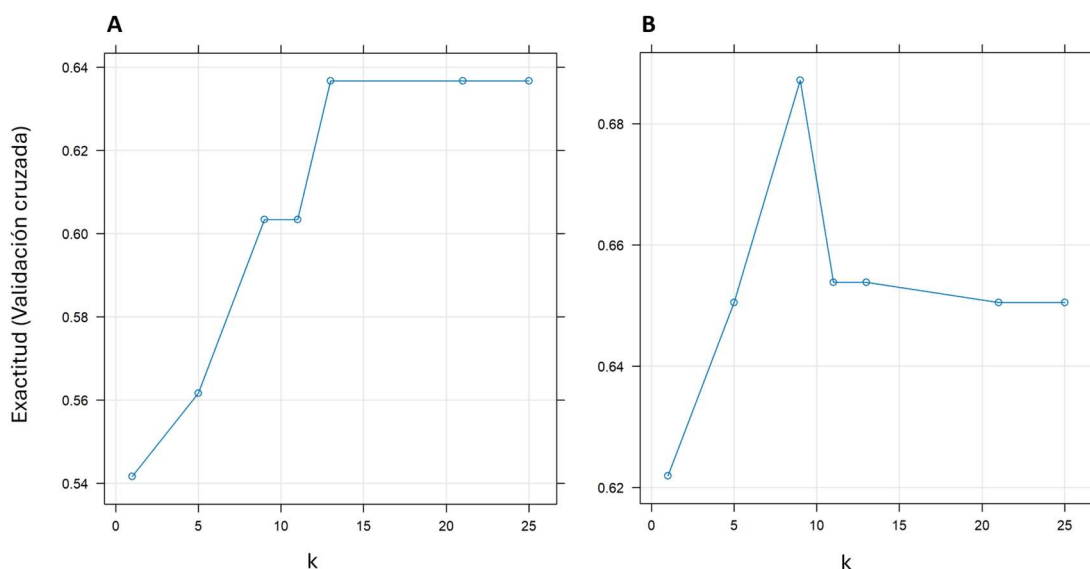


Figura 56. Gráficos que evalúan la exactitud de los modelos desarrollados mediante el algoritmo KNN para los datos de entrenamiento de pacientes (A) no-COVID-19 y (B) COVID-19 en base al número de vecinos utilizado para las predicciones (k). Esta evaluación se ha llevado a cabo mediante un proceso automatizado con validación cruzada de 10 pasos.

En cuanto al desarrollo de los modelos basados en ANN, se realizó la exploración de los mejores hiperparámetros para el ajuste de los datos mediante validación cruzada de 10 pasos. En el caso del grupo no-COVID-19, el modelo ANN óptimo (que presentó mayor exactitud) se construyó utilizando 5 capas ocultas con *weight decay* o penalización de pesos = 1×10^{-4} (Figura 57) como parámetro de regularización para evitar sobreajuste y mejorar la generalización del modelo.

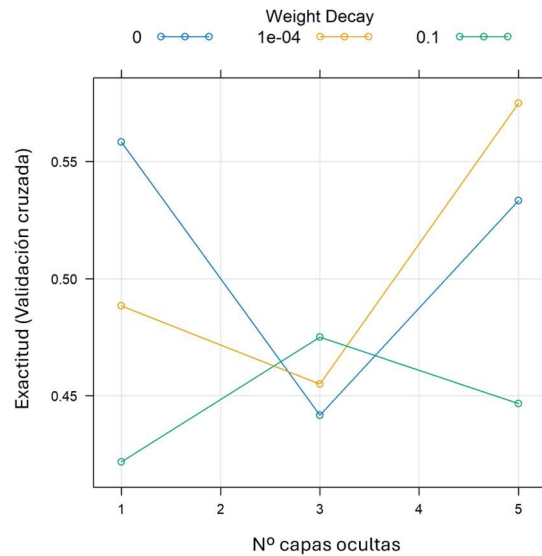


Figura 57. Gráfico que evalúa la exactitud de los modelos desarrollados mediante el algoritmo ANN para los datos de entrenamiento de pacientes no-COVID-19, en base al número de capas ocultas utilizado para las predicciones (eje x) y la penalización de pesos o *weight decay*. Este parámetro es una forma de regularización, que penaliza los pesos del modelo durante el entrenamiento, evitando el uso de pesos extremadamente grandes que impidan la correcta generalización del modelo. Esta evaluación se ha llevado a cabo mediante un proceso automatizado con validación cruzada de 10 pasos.

Por otro lado, el algoritmo desarrollado mediante ANN utilizando los datos de entrenamiento de los pacientes COVID-19, dio como resultado un modelo óptimo constituido por una capa oculta. En la figura 58 se muestra la estructura de la red neuronal con los biomarcadores como variables de entrada, la capa oculta y la capa de salida que muestra las posibles predicciones de gravedad.

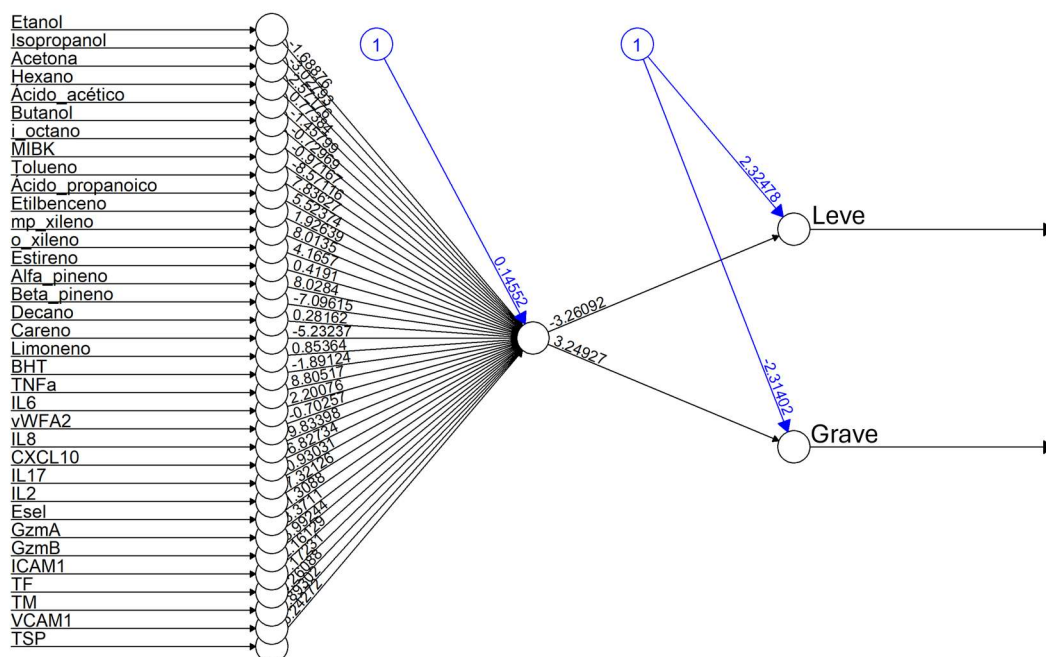


Figura 58. ANN desarrollada con los datos de entrenamiento de los pacientes COVID-19. Los círculos representan los nodos de la red neuronal, que incluye los nodos de entrada (todas las variables incluidas en el

Resultados

modelo), la capa oculta (centro del gráfico) y la capa de salida con las posibles predicciones de gravedad. Los números representan los pesos de los valores transferidos entre nodos. Los nodos azules presentes en la capa oculta y la capa de salida son los *bias node* o nodos de sesgo. Son una característica inherente de las redes neuronales para aumentar la flexibilidad del modelo y mejorar el ajuste de los datos. Específicamente, permite que la red neuronal ajuste los valores de activación incluso cuando todas las entradas tienen un valor de 0. MIBK: metil isobutil cetona; BHT: butilhidroxitolueno; TNFa: factor de necrosis tumoral; IL: interleuquina; vWFA2: factor de von Willebrand A2; CXCL10: ligando 10 de quimioquina con motivo C-X-C; Esel: e-selectina; Gzm: granzima; ICAM1: molécula de adhesión intercelular 1; TF: factor tisular; TM: trombomodulina; VCAM1: molécula de adhesión de células vasculares; TSP: trombospondina.

Una vez desarrollados los modelos de predicción mediante RL, KNN, ANN y SVM, se validó su comportamiento utilizando el set de datos de prueba (el 20 % restante de los datos originales, no utilizado en el entrenamiento) para determinar si los modelos eran capaces de realizar un buen trabajo de predicción para futuros nuevos datos. En la figura 59 se muestran las matrices de confusión, que comparan los resultados de las predicciones generadas por los modelos frente a las etiquetas de gravedad reales para esos mismos pacientes.

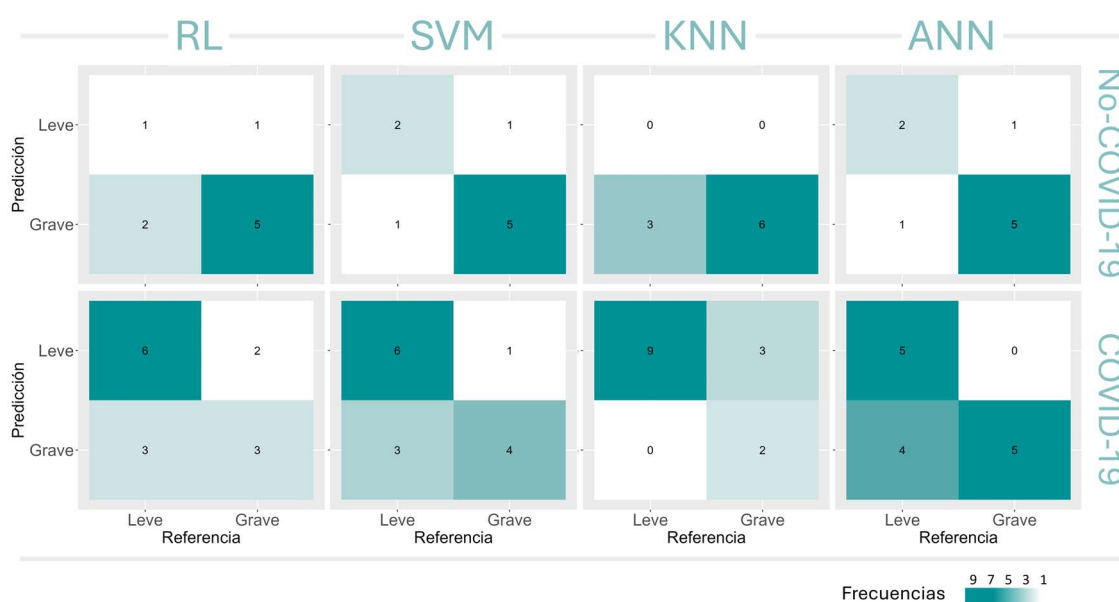


Figura 59. Matrices de confusión que representan los resultados de las predicciones realizadas por cada modelo (regresión logística (RL), *support vector machines* (SVM), *k-nearest neighbors* (KNN) y *artificial neural networks* (ANN) para no-COVID-19 y COVID-19 y las etiquetas reales (referencia) de gravedad de los pacientes del set de datos de prueba.

Se pudo observar la presencia de números elevados de FPs y FNs en los resultados de predicción de todos los modelos desarrollados mediante RL, SVM, KNN y ANN tanto para el grupo de pacientes no-COVID-19 como para el grupo de pacientes COVID-19. En base a estas predicciones, se calcularon las métricas que describen el comportamiento de los modelos en el set de datos de prueba (Tabla 22). Estas métricas reflejaron la

capacidad de los modelos para predecir el pronóstico de nuevos pacientes diferentes de los utilizados en el entrenamiento de los modelos.

Tabla 22. Resultados de validación de los algoritmos de ML estudiados para la predicción de gravedad en pacientes COVID-19 y no-COVID-19. Estos resultados involucran todos los COVs y parámetros séricos estudiados. Se muestran las métricas de evaluación del comportamiento de los modelos en el set de datos de evaluación. Se muestran la exactitud (*accuracy, Acc*) y su intervalo de confianza del 95% (IC 95%), el valor *p* (relativo a si la exactitud es superior al *Non Information Rate* (NIR), que es el porcentaje de individuos que pertenecen a la clase más numerosa de las clases estudiadas en el set de datos de prueba), área bajo la curva ROC (AUROC, del inglés, *area under the receiver operating characteristic*), kappa, índice F1, sensibilidad, especificidad, valor predictivo positivo (VPP) y valor predictivo negativo (VPN) y la exactitud balanceada (que toma en consideración si las clases presentan una distribución asimétrica). Se consideraron estadísticamente significativos los valores $p < 0,05$.

Modelo	Exactitud (IC 95 %)	Valor <i>p</i> [Acc > NIR]	AUROC (IC 95 %)	Kappa	F1	Sensibilidad	Especificidad	VPP	VPN	Exactitud balanceada
No-COVID-19										
RL	0,667 (0,299 - 0,925)	0,583	0,583 (0,218 - 0,949)	0,182	0,769	0,833	0,333	0,714	0,500	0,583
RF	0,889 (0,518 - 0,997)	0,143	0,944 (0,791 - 1,000)	0,727	0,923	1,000	0,667	0,857	1,000	0,833
SVM	0,778 (0,400 - 0,972)	0,377	0,833 (0,541 - 1,000)	0,500	0,833	0,833	0,667	0,833	0,667	0,750
KNN	0,667 (0,299 - 0,925)	0,650	0,750 (0,531 - 0,969)	0,000	0,800	1,000	0,000	0,667		0,500
ANN	0,778 (0,400 - 0,972)	0,377	0,833 (0,507 - 1,000)	0,500	0,833	0,833	0,667	0,833	0,667	0,750
COVID-19										
LR	0,643 (0,351 - 0,872)	0,619	0,511 (0,148 - 0,875)	0,255	0,545	0,600	0,667	0,500	0,750	0,633
RF	0,929 (0,661 - 0,998)	0,018	0,956 (0,853 - 1,000)	0,851	0,909	1,000	0,889	0,833	1,000	0,944
SVM	0,714 (0,419 - 0,916)	0,401	0,778 (0,512 - 1,000)	0,429	0,667	0,800	0,667	0,571	0,857	0,733
KNN	0,786 (0,492 - 0,953)	0,204	0,844 (0,648 - 1,000)	0,462	0,571	0,400	1,000	1,000	0,750	0,700
ANN	0,714 (0,419 - 0,916)	0,401	0,889 (0,705 - 1,000)	0,472	0,715	1,000	0,556	0,556	1,000	0,778
RL: Regresión logística; RF: <i>Random Forest</i> ; SVM: <i>Support Vector Machines</i> ; KNN: <i>K-Nearest Neighbors</i> ; ANN: Redes neuronales artificiales, del inglés, <i>Artificial Neural Networks</i> .										

Los datos mostrados en la tabla 22 incluyen los resultados de testar los algoritmos desarrollados mediante RF. Como se mostraba en la tabla 21, estos RF son el resultado del entrenamiento mediante la combinación de 500 árboles y del uso de 5 variables en cada división.

En cuanto al grupo de los pacientes con COVID-19, la validación del modelo RF mostró un 92,9 % de exactitud, 94,4 % de exactitud balanceada, kappa de 0,851, F1 de 0,909, 100% de sensibilidad y 88,9 de especificidad (Tabla 22). La precisión del modelo RF desarrollado con la cohorte de pacientes COVID-19 fue significativamente mayor ($p = 0,018$) que el *Non Information Rate* (NIR, calculado como el porcentaje de pacientes que pertenecen a la clase más numerosa del conjunto de datos de prueba).

En el caso de los pacientes no-COVID-19, el algoritmo con mejor rendimiento también fue el creado mediante RF. Este modelo clasificó los casos del set de prueba según su predicción de gravedad con un 88,9 % de exactitud, un 83,3 % de exactitud

Resultados

balanceada, kappa de 0,727, F1 de 0,923, un 100% de sensibilidad y un 66,7 % de especificidad.

En general, los modelos desarrollados mediante RF ofrecieron los mejores rendimientos para diferenciar pacientes leves y graves, sobre todo en cuanto a la exactitud, sensibilidad y especificidad tanto en el modelo no-COVID-19 como en el modelo COVID-19. Además, el resultado de exactitud del modelo COVID-19 mostró un valor p significativo.

Considerando estos hallazgos, se decidió profundizar y explicar de manera más exhaustiva el comportamiento de los modelos construidos con RF, así como sus resultados de validación y su ajuste al set de datos de prueba.

2.3.3.3 Explicación de los modelos *Random Forest* para la predicción de gravedad en pacientes con COVID-19 y no-COVID-19

Los resultados del modelo RF no-COVID-19 se muestran a continuación. El gráfico MDG (Figura 60) mostró los predictores más importantes en la clasificación en grupos de gravedad de los pacientes del set de datos de entrenamiento.

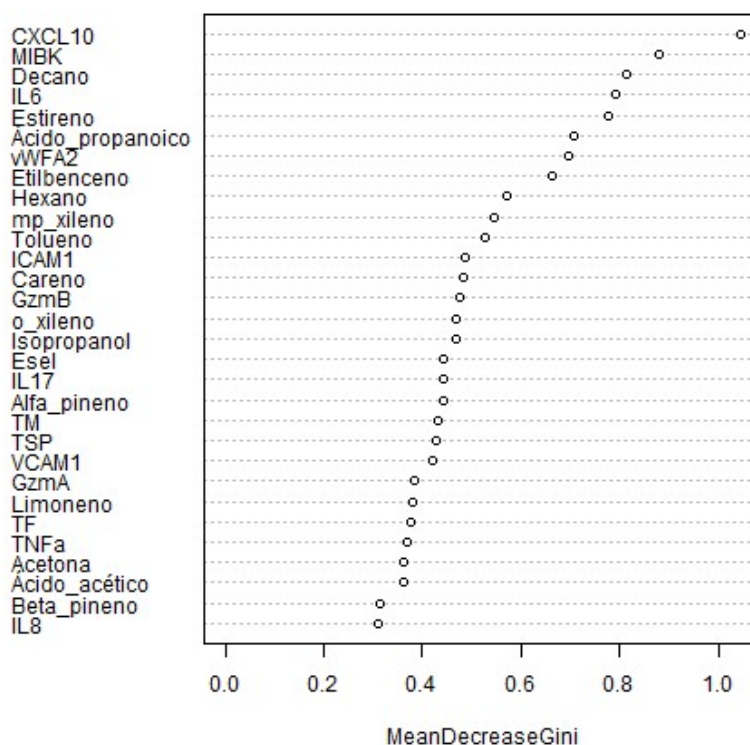


Figura 60. Resultados del modelo RF para el conjunto de datos no-COVID-19. El gráfico del coeficiente de Gini de disminución media (MDG, del inglés, *Mean Decrease Gini*) clasifica la importancia de la variable respecto a la gravedad de la enfermedad en este modelo de clasificación. MIBK: metil isobutil cetona; TNFa: factor de necrosis tumoral; IL: interleuquina; vWFA2: factor de von Willebrand A2; CXCL10: ligando 10 de quimioquina con motivo C-X-C; Esel: e-selectina; Gzm: granzima; ICAM1: molécula de adhesión intercelular 1; TF: factor tisular; TM: trombomodulina; VCAM1: molécula de adhesión de células vasculares; TSP: trombospondina.

Los parámetros con mayor ponderación fueron CXCL10, MIBK, decano, IL-6, estireno, ácido propanoico, vWF-A2, etilbenceno, hexano, *m,p*-xileno y tolueno para clasificar a los pacientes no-COVID-19 en los grupos de gravedad.

El modelo RF desarrollado en los datos de entrenamiento se validó en los datos de prueba. La matriz de confusión muestra la comparación entre los resultados de las predicciones y las etiquetas de gravedad reales (Figura 61), lo que arroja una exactitud del 88,9 % con un FP en la predicción de los pacientes nuevos.



Figura 61. Resultados del modelo RF para el conjunto de datos no-COVID-19. La matriz de confusión muestra el rendimiento del modelo en el conjunto de datos de prueba, comparando las predicciones del modelo con la etiqueta de gravedad real.

La figura 62 ilustra las 15 variables principales y su valor SHAP en el modelo de predicción de gravedad en pacientes no-COVID-19. MIBK, CXCL10, etilbenceno, vWF-A2 y GzmB fueron los compuestos más influyentes en la predicción de gravedad para el conjunto de datos de prueba. Concentraciones bajas de estas variables mostraron valores SHAP más altos. Esto puede interpretarse como que la probabilidad de que un paciente desarrolle una enfermedad no-COVID-19 grave aumenta a medida que disminuyen los niveles de estos biomarcadores. Por otro lado, TF, alfa-pineno y TM mostraron la tendencia opuesta.

Resultados

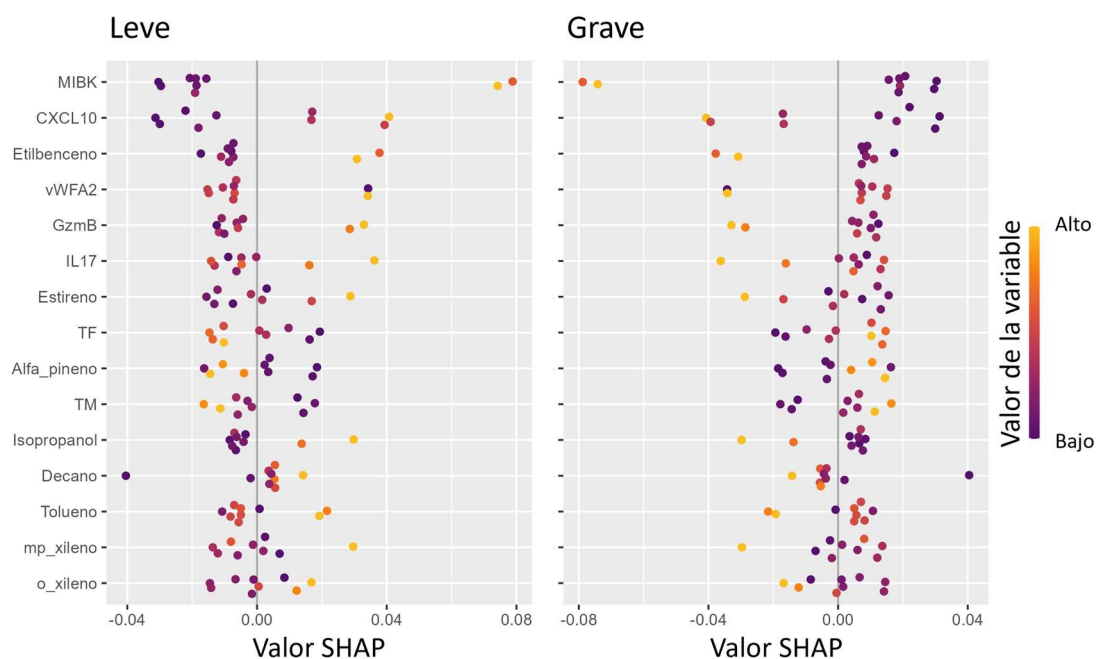


Figura 62. Resultados del modelo RF para el conjunto de datos no-COVID-19. El gráfico resumen de las explicaciones aditivas de Shapley (SHAP, del inglés, *Shapley Additive exPlanations*) cuantifica la importancia de las variables en el conjunto de datos de prueba. Los valores de SHAP indican la magnitud del cambio en las probabilidades logarítmicas (es decir, explican la probabilidad de que un paciente grave o leve). Los puntos coloreados con un gradiente violeta-naranja indican el valor original de la concentración para cada compuesto y paciente. MIBK: metil isobutil cetona; IL: interleuquina; vWFA2: factor de von Willebrand A2; CXCL10: ligando 10 de quimioquina con motivo C-X-C; Gzm: granzima; TF: factor tisular; TM: trombomodulina.

Para aclarar la predicción del modelo no-COVID-19 en pacientes individuales, se realizó un análisis LIME para un ejemplo leve (Figura 63) y un ejemplo grave (Figura 64). La predicción de la gravedad fue correcta para ambos casos. La probabilidad de predicción del paciente grave fue muy elevada, del 0,92. Es decir, el modelo predijo con un 92 % de confianza que el caso pertenece a la clase de pacientes graves. El ajuste o concordancia de la explicación fue elevado (0,92 en el paciente leve y 0,95 en el paciente grave), lo que demuestra un buen ajuste del modelo, robustez y alta confiabilidad de los resultados de probabilidad en estos ejemplos.

Los niveles de ácido propanoico inferiores a 13,2 ng/L, junto con los niveles de ICAM-1 inferiores a 1156 pg/ml, tuvieron la mayor influencia en la probabilidad de dar un pronóstico leve para el caso mostrado en la figura 63. VCAM-1 y 1-butanol tuvieron efecto en la otra dirección.

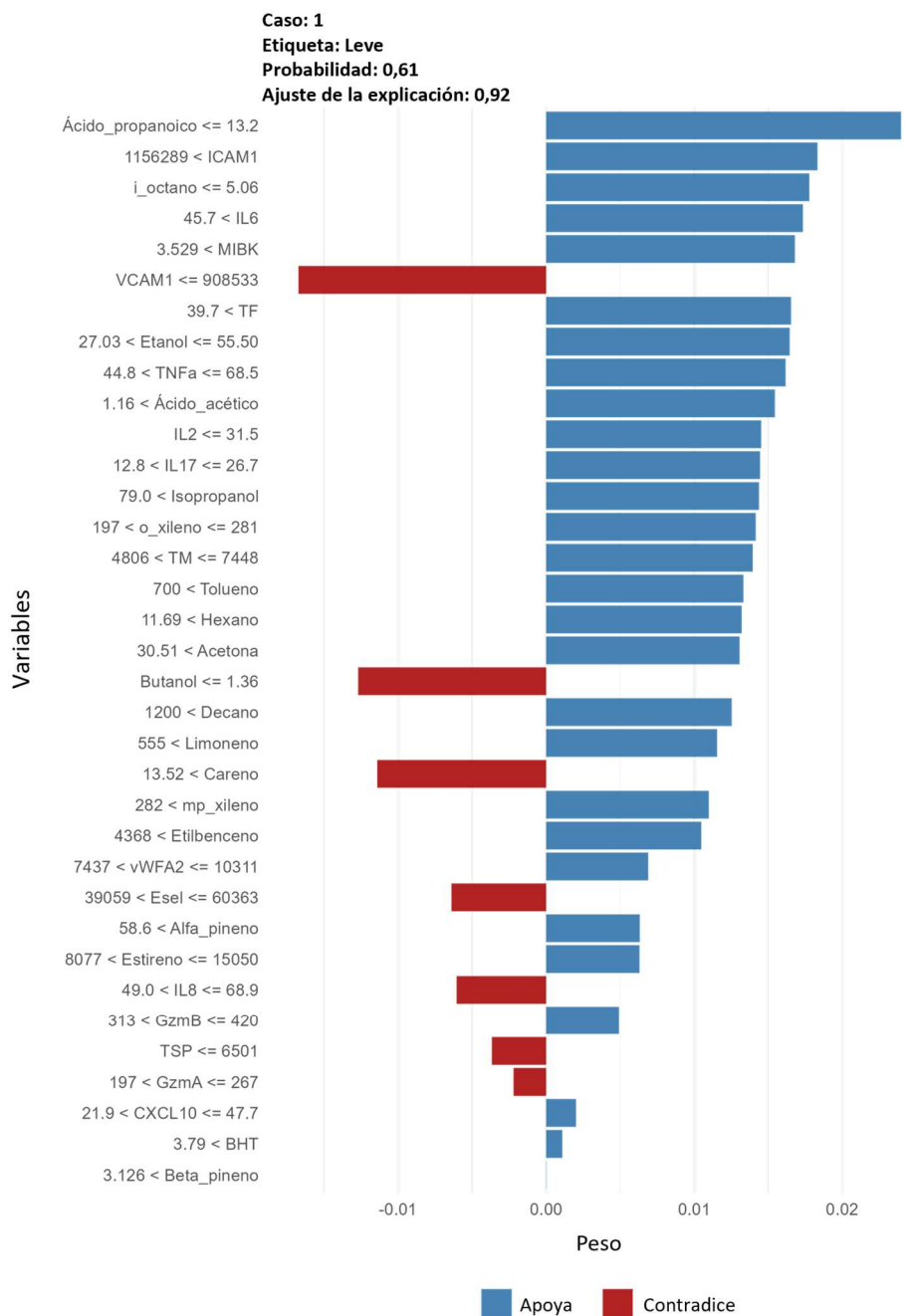


Figura 63. Resultados del modelo RF para el conjunto de datos no-COVID-19. Las explicaciones locales interpretables independientes del modelo (LIME, del inglés, *Local Interpretable Model-agnostic Explanations*) muestran predicciones individualizadas para un paciente leve como ejemplo. Muestra la probabilidad de ser clasificado como leve y la explicación adecuada para este paciente específico. La longitud de la barra representa en qué medida cada variable apoya (azul) o contradice (rojo) la clasificación como leve. Unidades de los COVs: ng/L. Unidades de los parámetros séricos: pg/ml. MIBK: metil isobutil cetona; BHT: butilhidroxitolueno; TNFa: factor de necrosis tumoral; IL: interleuquina; vWFA2: factor de von Willebrand A2; CXCL10: ligando 10 de quimioquina con motivo C-X-C; Esel: e-selectina; Gzm: granzima; ICAM1: molécula de adhesión intercelular 1; TF: factor tisular; TM: trombomodulina; VCAM1: molécula de adhesión de células vasculares; TSP: trombospondina.

Resultados

En el caso grave (Figura 64), los niveles de vWF-A2 entre 5394 y 7437 pg/ml, junto con concentraciones de *m,p*-xileno superiores a 143 ng/L, fueron los factores más influyentes en la predicción de gravedad.

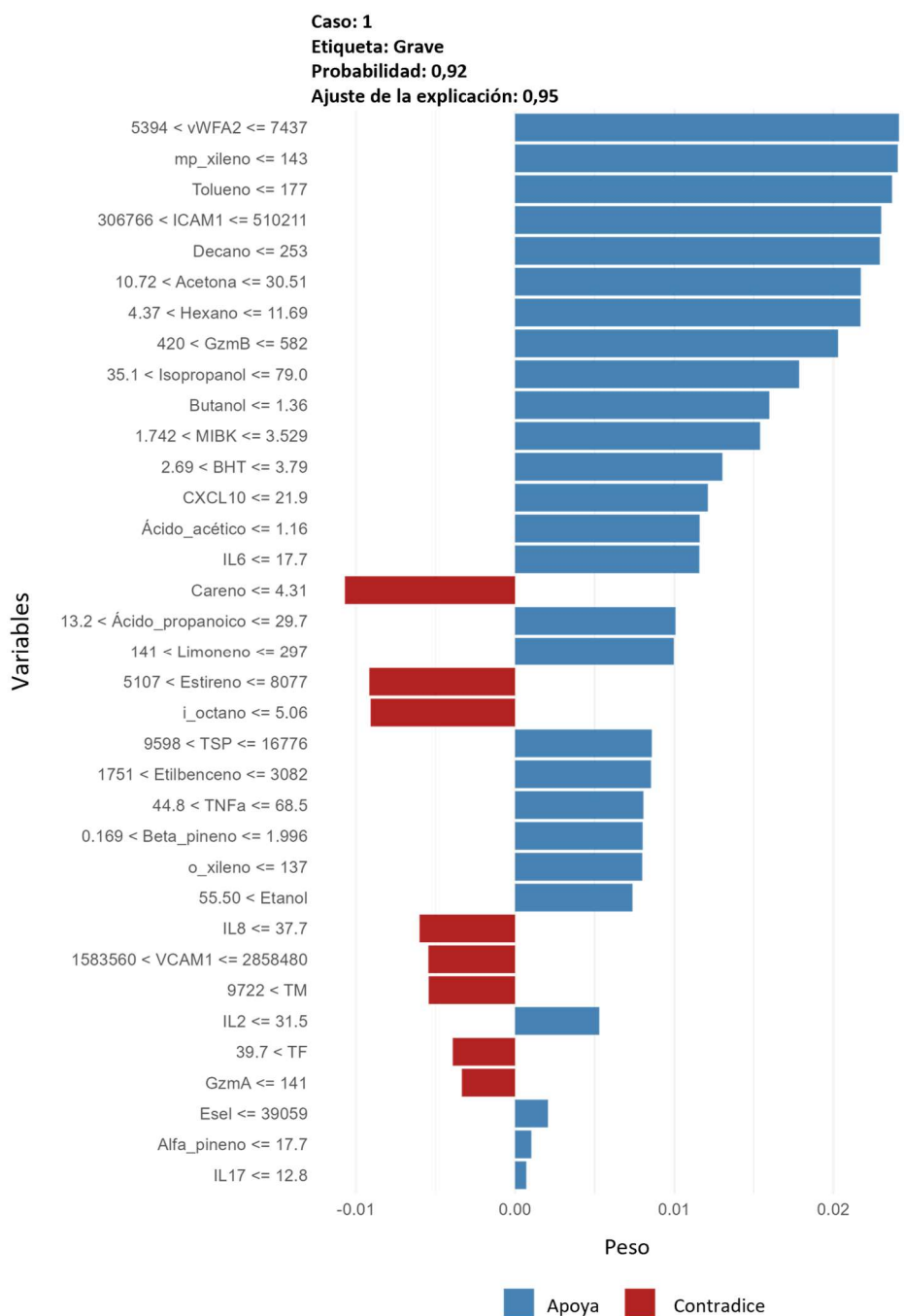


Figura 64. Resultados del modelo RF para el conjunto de datos no-COVID-19. Las explicaciones locales interpretables independientes del modelo (LIME, del inglés, *Local Interpretable Model-agnostic Explanations*) muestran predicciones individualizadas para un paciente grave como ejemplo. Muestra la probabilidad de ser clasificado como grave y la explicación adecuada para este paciente específico. La longitud de la barra representa en qué medida cada variable apoya (azul) o contradice (rojo) la clasificación como grave. Unidades de los COVs: ng/L. Unidades de los parámetros séricos: pg/ml. MIBK: metil isobutil cetona; BHT:

butilhidroxitolueno; TNFa: factor de necrosis tumoral; IL: interleuquina; vWFA2: factor de von Willebrand A2; CXCL10: ligando 10 de quimioquina con motivo C-X-C; Esel: e-selectina; Gzm: granzima; ICAM1: molécula de adhesión intercelular 1; TF: factor tisular; TM: trombomodulina; VCAM1: molécula de adhesión de células vasculares; TSP: trombospondina.

Por otro lado, se muestran los resultados del modelo RF relativo a la cohorte de pacientes COVID-19. El análisis MDG (Figura 65) mostró que IL-6, IL-8, tolueno, TM, TF, estireno, TNF- α , o-xileno, ICAM-1 y etilbenceno fueron los predictores más importantes de la gravedad de enfermedad en el desarrollo del modelo COVID-19.

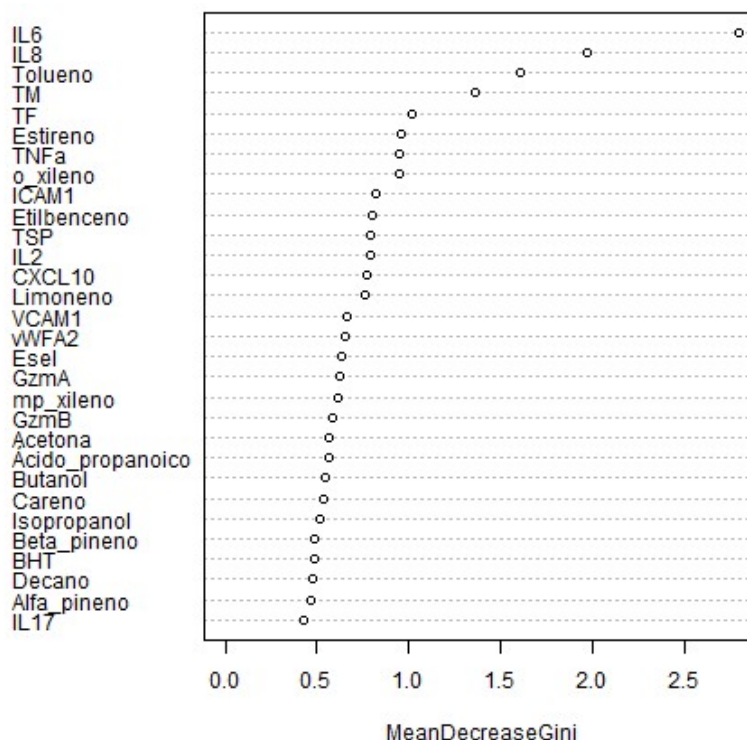


Figura 65. Resultados del modelo RF para el conjunto de datos COVID-19. El gráfico del coeficiente de Gini de disminución media (MDG, del inglés, *Mean Decrease Gini*) clasifica la importancia de la variable respecto a la gravedad de la enfermedad en este modelo de clasificación. BHT: butilhidroxitolueno; TNFa: factor de necrosis tumoral; IL: interleuquina; vWFA2: factor de von Willebrand A2; CXCL10: ligando 10 de quimioquina con motivo C-X-C; Esel: e-selectina; Gzm: granzima; ICAM1: molécula de adhesión intercelular 1; TF: factor tisular; TM: trombomodulina; VCAM1: molécula de adhesión de células vasculares; TSP: trombospondina.

La matriz de confusión (Figura 66) que se obtuvo de la comparación entre los resultados de las predicciones del modelo y las etiquetas de gravedad reales mostró un FP, lo que resultó en una precisión del 92,9 % ($p = 0,018$) cuando se expuso el modelo al conjunto de datos de prueba.

Resultados



Figura 66. Resultados del modelo RF para el conjunto de datos COVID-19. La matriz de confusión muestra el rendimiento del modelo en el conjunto de datos de prueba, comparando las predicciones del modelo con la etiqueta de gravedad real.

El análisis del rendimiento del modelo en los datos de prueba a través del método SHAP (Figura 67) señaló como variables más importantes las mismas 5 variables que resultaron de mayor importancia en la ejecución del modelo en los datos de entrenamiento (Figura 65). IL-6, IL-8, tolueno, TM y TF mostraron los valores SHAP más altos para la predicción de pacientes graves, seguidos de IL-2, ICAM-1, TNF- α , *m,p*-xileno e isopropanol. Cabe destacar que, a excepción del isopropanol y la TSP, todos los valores SHAP aumentaron a medida que aumentan los niveles de la variable. Por lo tanto, individuos con niveles elevados de estos compuestos presentaron mayores probabilidades de ser diagnosticados como pacientes graves.

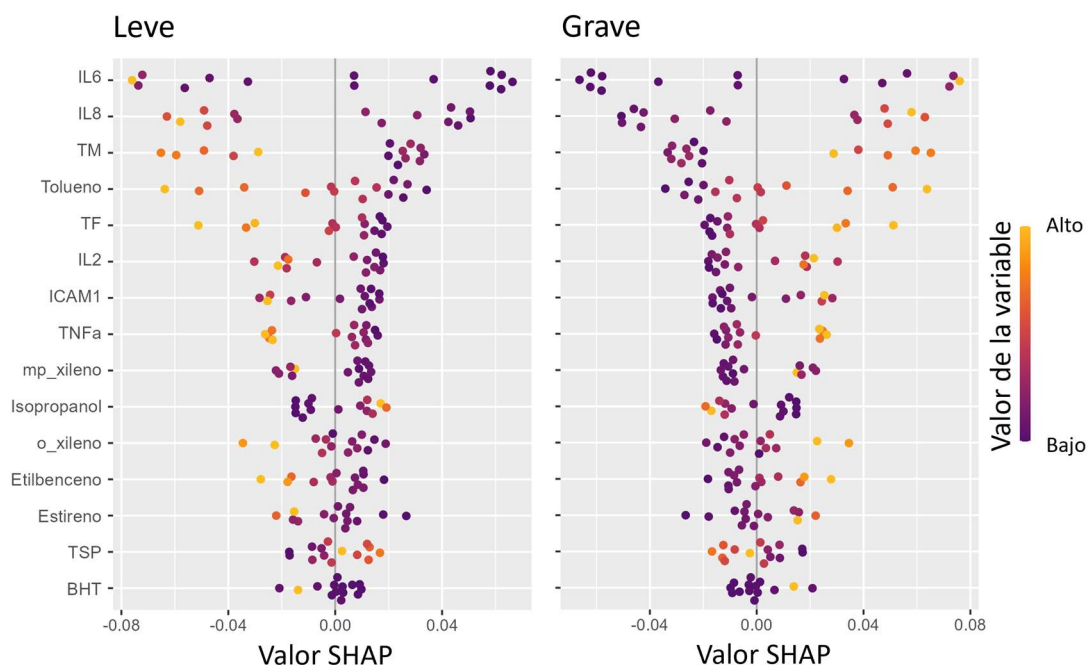


Figura 67. Resultados del modelo RF para el conjunto de datos COVID-19. El gráfico resumen de las explicaciones aditivas de Shapley (SHAP, del inglés, *Shapley Additive exPlanations*) cuantifica la importancia de las variables en el conjunto de datos de prueba. Los valores de SHAP indican la magnitud del cambio en las probabilidades logarítmicas (es decir, explican la probabilidad de que un paciente grave o leve). Los puntos coloreados con un gradiente violeta-naranja indican el valor original de la concentración para cada compuesto

y paciente. BHT: butilhidroxitolueno; TNFa: factor de necrosis tumoral; IL: interleuquina; ICAM1: molécula de adhesión intercelular 1; TF: factor tisular; TM: trombomodulina; TSP: trombospondina.

En cuanto al análisis LIME, se muestran ejemplos de un caso leve y un caso grave en las figuras 68 y 69, respectivamente. Se obtuvo un 0,96 para el ajuste de la explicación en el paciente leve y un 0,93 en el paciente grave, lo que mostró un alto grado de confianza de las predicciones de gravedad.

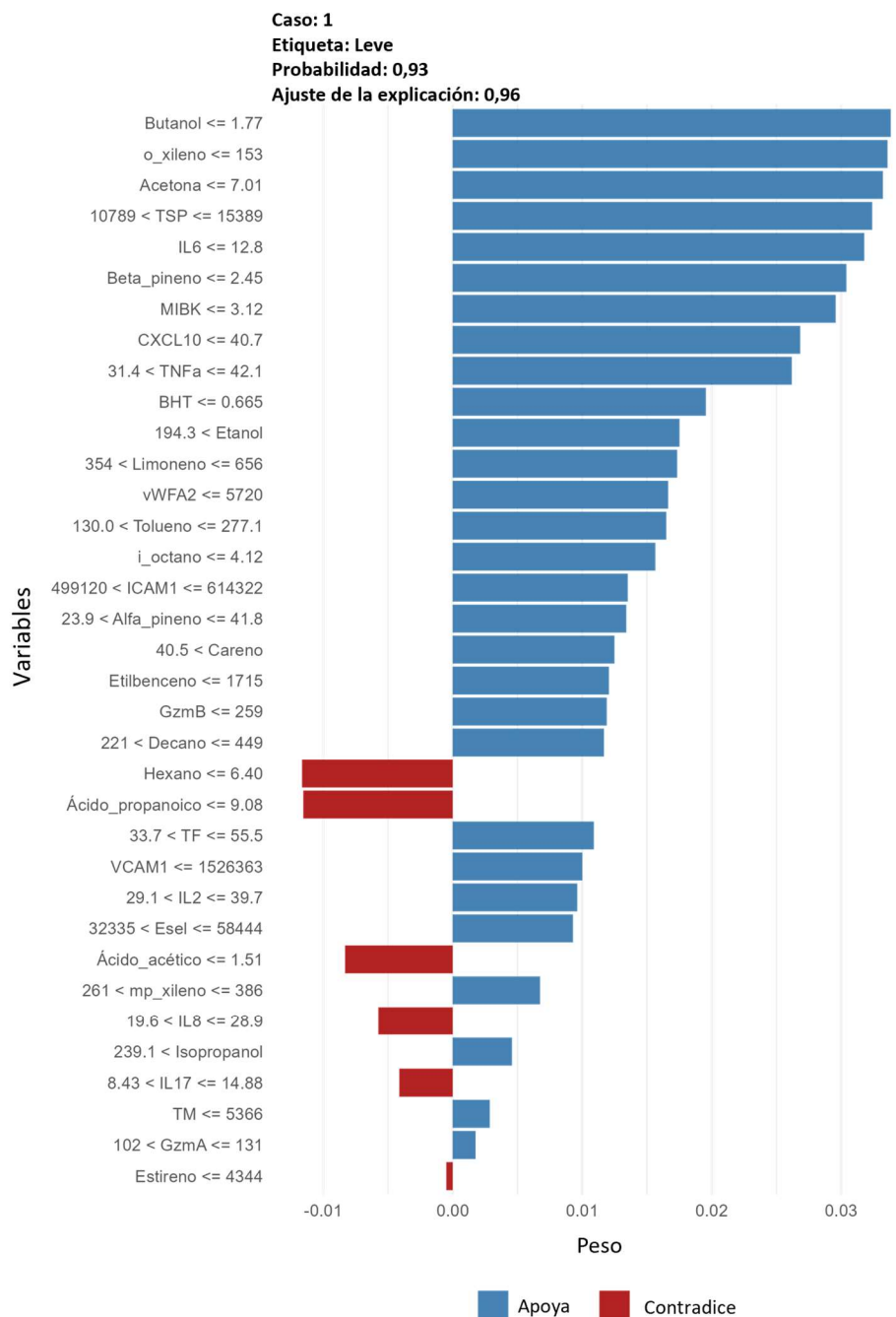


Figura 68. Resultados del modelo RF para el conjunto de datos COVID-19. Las explicaciones locales interpretables independientes del modelo (LIME, del inglés, *Local Interpretable Model-agnostic Explanations*) muestran predicciones individualizadas para un paciente leve como ejemplo. Muestra la probabilidad de ser

Resultados

clasificado como leve y la explicación adecuada para este paciente específico. La longitud de la barra representa en qué medida cada variable apoya (azul) o contradice (rojo) la clasificación como leve. Unidades de los COVs: ng/L. Unidades de los parámetros séricos: pg/ml. MIBK: metil isobutil cetona; BHT: butilhidroxitolueno; TNFa: factor de necrosis tumoral; IL: interleuquina; vWFA2: factor de von Willebrand A2; CXCL10: ligando 10 de quimioquina con motivo C-X-C; Esel: e-selectina; Gzm: granzima; ICAM1: molécula de adhesión intercelular 1; TF: factor tisular; TM: trombospondulina; VCAM1: molécula de adhesión de células vasculares; TSP: trombospondina.

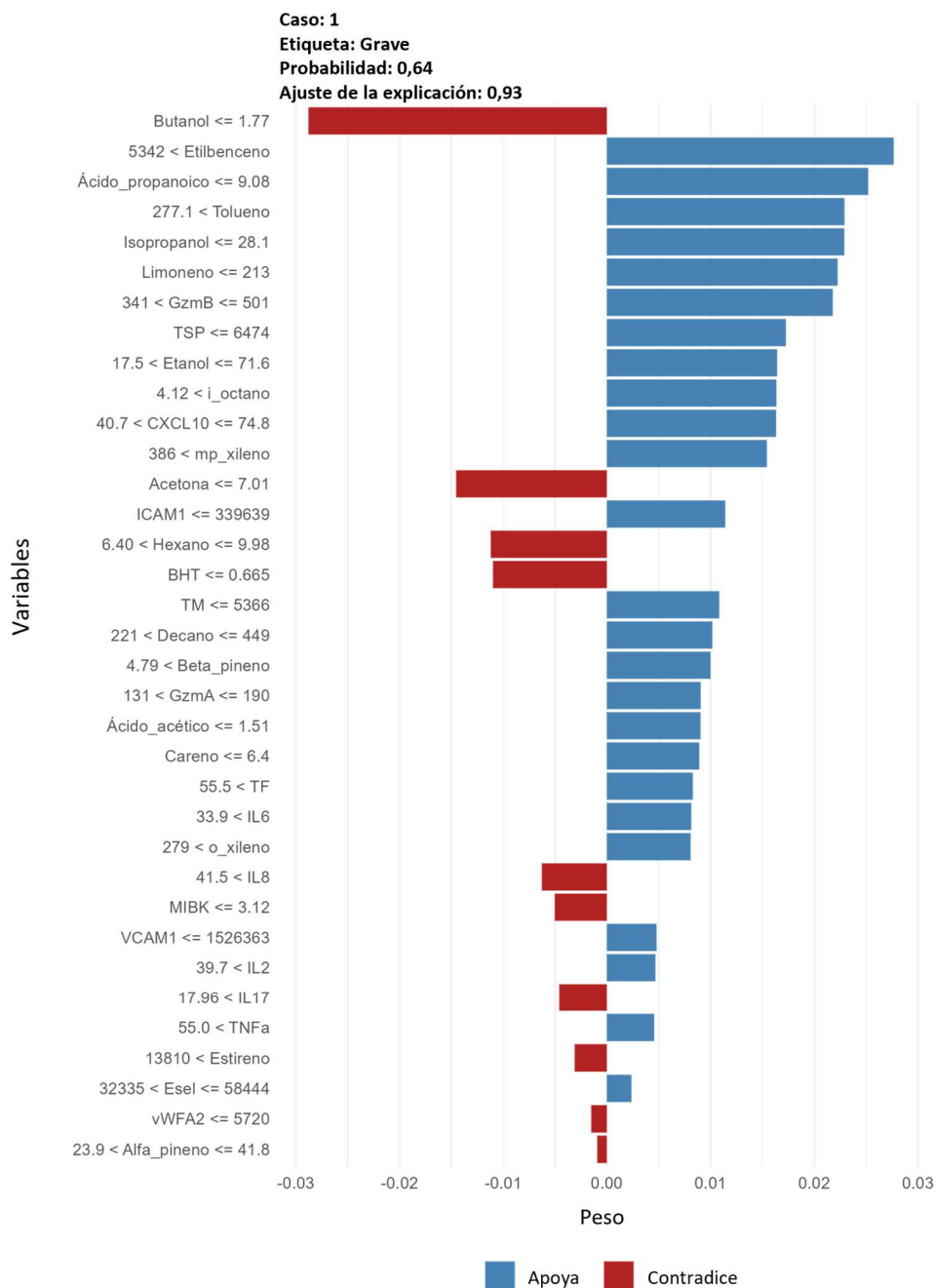


Figura 69. Resultados del modelo RF para el conjunto de datos COVID-19. Las explicaciones locales interpretables independientes del modelo (LIME, del inglés, *Local Interpretable Model-agnostic Explanations*) muestran predicciones individualizadas para un paciente grave como ejemplo. Muestra la probabilidad de ser clasificado como grave y la explicación adecuada para este paciente específico. La longitud de la barra representa en qué medida cada variable apoya (azul) o contradice (rojo) la clasificación como grave. Unidades de los COVs: ng/L. Unidades de los parámetros séricos: pg/ml. MIBK: metil isobutil cetona; BHT:

butilhidroxitolueno; TNFa: factor de necrosis tumoral; IL: interleuquina; vWFA2: factor de von Willebrand A2; CXCL10: ligando 10 de quimioquina con motivo C-X-C; Esel: e-selectina; Gzm: granzima; ICAM1: molécula de adhesión intercelular 1; TF: factor tisular; TM: trombomodulina; VCAM1: molécula de adhesión de células vasculares; TSP: trombospondina.

Para el caso leve, 1-butanol inferior a 1,77 ng/L y o-xileno inferior a 153 ng/L tuvieron la mayor influencia positiva sobre la probabilidad de clasificarlo como un paciente leve (Figura 68). En consecuencia, 1-butanol inferior a 1,77 ng/L contradijo la probabilidad de diagnosticar el caso como grave en el ejemplo de la figura 69. Asimismo, el etilbenceno, el ácido propanoico y el tolueno respaldaron en gran medida la predicción grave en ese paciente.

Finalmente, se incluyeron en el análisis las variables demográficas. En el caso del desarrollo del modelo COVID-19, el algoritmo resultante mantuvo las tres principales variables (IL-6, IL-8 y tolueno) como las más significativas, lo que refuerza su importancia en el pronóstico (Anexo 1 – Figura S2). Por el contrario, la edad surgió como la variable más significativa si se añadían estas variables en el modelo desarrollado con la cohorte de pacientes no-COVID-19 (Anexo 1 – Figura S3).

2.3.4 DISCUSIÓN

Nuestros hallazgos sugieren que la combinación del análisis de COVs en aire expirado junto con la detección de biomarcadores séricos es prometedora para el desarrollo de herramientas de pronóstico integrales que puedan mejorar el manejo de las enfermedades respiratorias. La ventaja principal de esta propuesta es la posibilidad de tomar muestras de forma no invasiva, lo que proporciona un beneficio práctico en entornos clínicos.

Los resultados obtenidos demuestran la existencia de perfiles de COVs distintos en pacientes COVID-19, no-COVID-19 y donantes sanos, algunos de los cuales confirman estudios previos, validando así nuestra cohorte. Por ejemplo, se han demostrado valores elevados de hexano y el etilbenceno en muestras de pacientes con COVID-19 en comparación con individuos sanos (311, 323). Por otro lado, se encontraron niveles más bajos de tolueno en pacientes COVID-19 que en sanos (327) y los niveles de decano fueron más elevados en pacientes no-COVID-19 comparados con los pacientes COVID-19 (320). Además de estos COVs, demostramos la existencia de diferencias entre los grupos en los niveles de terpenos: alfa-pineno, beta-pineno, 3-careno y limoneno. Previamente se ha sugerido que estos compuestos están alterados en condiciones de disfunción pulmonar no infecciosa (333). Por otro lado, se encontraron niveles más elevados de etanol en pacientes con COVID-19. Dicha regulación positiva fue previamente estudiada, y se sugirió que probablemente está asociada con los efectos del COVID-19 en el metabolismo del acetaldehído, que se produce principalmente en el hígado, lo que da lugar a un aumento de las concentraciones sanguíneas de este alcohol que se excreta a través de los pulmones (303). También encontramos diferencias entre los grupos en los niveles de MIBK, *m,p*-xileno, *o*-xileno, estireno y ácido propanoico, que no se habían informado previamente. Por otro lado, los inmunomoduladores séricos también mostraron perfiles distintos en pacientes con COVID-19 y no-COVID-19. En conjunto, los resultados obtenidos, que demostraron perfiles de COVs y parámetros séricos distintos para ambas cohortes de pacientes, nos impulsaron a realizar el estudio de predicción de gravedad por separado para pacientes con COVID-19 y no-COVID-19.

Numerosos estudios se han centrado en el estudio del perfil de citoquinas de los individuos para predecir el riesgo de complicaciones graves en infecciones respiratorias. La mayoría de ellos utilizan IL-6 (374, 377-379), IL-8 (377, 383), IL-2 (380, 381) y TFN- α (376, 380, 384) como biomarcadores pronósticos de COVID-19. En consonancia con sus resultados, las curvas ROC realizadas con los datos de pacientes COVID-19 para esos

parámetros fueron estadísticamente significativas en nuestro estudio, así como para CXCL10, TF, TM, VCAM-1 y GzmA, mostrando una buena capacidad para diferenciar los pacientes leves y graves. También encontramos que IL-6, IL-8 y TFN- α fueron las variables más importantes en los algoritmos de decisión desarrollados, especialmente para COVID-19. En cuanto al modelo no-COVID-19, CXCL10 es el parámetro sérico más importante junto con IL-6 en los datos de entrenamiento. Otros investigadores también han reconocido CXCL10 como un biomarcador que predice la gravedad de varias enfermedades (55). vWF-A2, IL-17 y GzmB también resultaron importantes en la visualización con SHAP del ajuste de datos de prueba al modelo de predicción de gravedad en pacientes no-COVID-19. En este sentido, un estudio reveló una disminución significativa de GzmB en pacientes con COVID-19 grave en comparación con casos leves (56), lo que en nuestro caso se aplica a individuos con infecciones no-COVID-19.

En cuanto a los COVs, la mayoría de los hidrocarburos aromáticos estudiados (tolueno, etilbenceno, o-xileno y estireno) mostraron curvas ROC significativas, demostrando su capacidad predictiva de la gravedad de enfermedad. El tolueno presentó el valor de AUC más alto para distinguir la gravedad en pacientes COVID-19. Como se esperaba, esos compuestos fueron variables importantes en el entrenamiento del algoritmo RF, así como para el ajuste de los datos de prueba en el modelo SHAP. Como se muestra en el gráfico SHAP de pacientes COVID-19, los niveles elevados de tolueno, o-xileno, etilbenceno y estireno respaldan el diagnóstico grave de los pacientes. Dado que generalmente se considera que los compuestos aromáticos provienen de fuentes exógenas, este resultado puede reflejar que estos COVs se habían absorbido de manera anómala en los tejidos pulmonares como una posible consecuencia de la alteración de la función pulmonar, de manera similar a lo que sucede en otras afecciones pulmonares como la EPOC, el asma o el cáncer (277, 300, 564).

Con respecto al origen endógeno de los ácidos orgánicos, se forman principalmente durante la fermentación bacteriana de carbohidratos y son especialmente abundantes en el intestino humano (289). Investigadores han sugerido que la disbiosis de la microbiota intestinal es una posible causa de la producción diferencial de ácido propanoico (280), que también podría afectar a la gravedad de la infección respiratoria (589). En nuestro estudio, el ácido propanoico fue muy relevante en el entrenamiento de los modelos para ambos grupos de pacientes, siendo la característica de mayor apoyo para clasificar no-COVID-19 leve y la tercera en COVID-19 grave.

Discusión

Por otro lado, MIBK demostró una importancia significativa en el diagnóstico de infecciones respiratorias, distinguiendo entre COVID-19 y no-COVID-19. Además, tuvo una alta relevancia en el modelo de RF desarrollado con la cohorte no-COVID-19, siendo el COV más importante para la ejecución del algoritmo tanto en los conjuntos de datos de entrenamiento como de prueba. Los niveles bajos de MIBK ayudaron al algoritmo RF del grupo no-COVID-19 a identificar a los individuos como pacientes graves. En los ejemplos explicados mediante LIME, MIBK tiene más peso en los casos leves que en los graves. A este respecto, Lima *et al.* propusieron que los niveles elevados de cetonas son resultado de una β -oxidación de ácidos grasos y un metabolismo proteico elevados (590), procesos que están alterados en los pacientes con COVID-19 (591). Esto respalda nuestro resultado de sobreproducción de MIBK en COVID-19 en comparación con pacientes no-COVID-19 y HD. Asimismo, un estudio reciente identificó la acetona como un COV predictivo de la gravedad de la enfermedad y los resultados de muerte (313). Nuestros resultados refuerzan esos hallazgos, dado que la acetona fue uno de los predictores de gravedad con mayor peso en nuestras descripciones LIME de COVID-19 leve y no-COVID-19 grave.

En lo que respecta al decano, también fue un compuesto crucial para el algoritmo de predicción de gravedad del grupo no-COVID-19. Kamal *et al.* concluyeron que las concentraciones de alcanos de cadena larga en el aliento exhalado se correlacionaban positivamente con las concentraciones nasales de citoquinas antivirales en pacientes con infecciones pulmonares (286). Nuestros resultados son consistentes con ellos, demostrando una correlación positiva del decano con CXCL10 en pacientes no-COVID-19. Anteriormente, se ha demostrado que los alcanos de cadena larga se producen debido al estrés oxidativo en humanos (284) y se reconoce comúnmente que la infección viral induce estrés oxidativo a través de varios mecanismos (592). Por el contrario, nuestros resultados SHAP para no-COVID-19 indican una mayor probabilidad de casos graves con niveles más bajos de decano. Posiblemente, este hecho se debe a la amplia variedad de enfermedades dentro del grupo no-COVID-19.

Cabe destacar que el 1-butanol surgió como la variable más significativa influyendo en las clasificaciones de gravedad en los ejemplos explicados con LIME de pacientes COVID-19. Por el contrario, esta variable no tuvo el mismo nivel de importancia durante el desarrollo del algoritmo. Esta observación subraya la importancia crítica de validar los modelos con datos nuevos que no se emplearon en el entrenamiento del algoritmo. Además, nuestros hallazgos destacan la relevancia de conservar variables que pueden no parecer útiles en análisis individuales.

La identificación temprana de pacientes con enfermedades respiratorias que tienen un alto riesgo de evolucionar hacia un estado grave es crucial para optimizar la toma de decisiones clínicas y la asignación de recursos médicos. Además del desarrollo de varias puntuaciones o índices de pronóstico (593), la investigación metabolómica ha avanzado notablemente en COVID-19 en comparación con otras enfermedades respiratorias (360). Sin embargo, los hallazgos y herramientas desarrollados para el diagnóstico y pronóstico de COVID-19 pueden servir como un excelente punto de partida para estudiar su aplicación en otras infecciones respiratorias.

Por esta razón, investigamos la capacidad de una combinación de biomarcadores para la identificación temprana de pacientes con alto riesgo de desarrollar una enfermedad grave, tanto para los grupos de pacientes COVID-19 como para pacientes no-COVID-19 a través de modelos de inteligencia artificial. Los algoritmos de ML, como el modelo RF utilizado en este estudio, son particularmente valiosos para analizar datos clínicos extensos y hacer predicciones en base a ellos (594).

Sin embargo, los modelos RF son algoritmos de caja negra, lo que dificulta la comprensión del proceso que llevan a cabo para la toma de decisiones. Con el objetivo de mejorar su interpretabilidad, se emplearon los métodos SHAP y LIME para explicar las predicciones globales e individualizadas, respectivamente. El desarrollo de modelos transparentes y robustos tiene gran importancia para la clasificación de pacientes, el seguimiento de la enfermedad, la optimización del tratamiento y la asignación eficiente de recursos.

La limitación principal de este estudio fue el pequeño tamaño de la muestra y la heterogeneidad diagnóstica en el grupo no-COVID-19, lo que puede dificultar la generalización de los resultados. El elevado número de características analizadas plantea un riesgo debido al sobreajuste, lo que conduce a un buen rendimiento en los datos de entrenamiento que se ve disminuido cuando se testan los algoritmos en nuevos datos. Cabe señalar que, a pesar de los sólidos resultados de las pruebas, este estudio es exploratorio. Para lograr resultados fiables y generalizables, se necesitan conjuntos de datos clínicos más amplios y pruebas de muestras independientes.

En conclusión, este estudio demuestra por primera vez el valor pronóstico del perfil de COVs cuando se combina con la huella de biomarcadores relacionados con el sistema inmunológico en infecciones respiratorias. Proponemos este nuevo punto de vista, desarrollando algoritmos de decisión que pueden mejorar la predicción de la gravedad. Sin

Discusión

embargo, para lograr un rendimiento predictivo superior y generalizable, se requiere validar nuestros hallazgos desarrollando modelos que incluyan datos clínicos a gran escala.

Conclusiones

3 CONCLUSIONES

Los resultados obtenidos en este trabajo nos conducen a las siguientes conclusiones:

1. La GT muestra actividad antibacteriana contra cepas de *S. aureus*, tanto sensibles a meticilina como resistentes a ella y con resistencia intermedia a vancomicina.
2. La GT muestra sinergia con diferentes antibióticos empleados en clínica frente a cepas de *S. aureus*, como linezolid, ácido fusídico y vancomicina, permitiendo reducir la concentración a la que son eficaces para las cepas SAMS y SARM/SARIV *in vitro*.
3. La combinación de GT con vancomicina disminuye sus CPMs individuales y además muestra un potente efecto sinérgico *in vitro*, produciendo un efecto antibiótico significativo con concentraciones de vancomicina reducidas a la mitad o incluso un cuarto de su CMI para la cepa SARM/SARIV.
4. El modelo *in vivo* de infección por SARM/SARIV en *C. elegans* demuestra el efecto potenciador de la GT cuando se combina con vancomicina a unas concentraciones que además resultan mínimamente tóxicas para los nemátodos.
5. NRPS provenientes de cepas concretas de *A. fumigatus*, *A. Fischeri*, *A. lentutlus*, *A. novofumigatus*, *A. udawae*, *A. turcosus* y *A. thermomutatus* presentan alta homología de secuencia con la proteína gliP de *A. fumigatus* Af293, la cual resulta fundamental para la síntesis de GT.
6. Cepas concretas de *A. fumigatus*, *A. fischeri*, *A. turcosus*, *A. pseudofischeri* y *T. virens* son capaces de producir GT y bmGT *in vitro*.
7. La cepa *A. fumigatus* B5233 produce mayores concentraciones tanto de GT como de bmGT en ambos tipos de cultivo (Czapek-Dox y RPMI 1640), comparada con la producción de *A. fumigatus* 1631562 y el resto de las cepas.
8. Un elevado porcentaje de los pacientes diagnosticados como AI negativa incluidos en este estudio han recibido terapias farmacológicas mediante profilaxis y/o tratamiento con antifúngicos.
9. La utilidad de la bmGT como biomarcador resulta superior al GM y BDG en términos de especificidad y VPN si se elimina el sesgo del GM.
10. Existe una vinculación significativa entre los resultados de la primera muestra de bmGT analizada en los sueros de pacientes y el diagnóstico de AI en la población estudiada.

Conclusiones

11. La combinación de bmGT, GM, IL-8 y CMV como biomarcadores de AI supone una mejora sustancial en la discriminación de pacientes con AI probable respecto al GM en solitario.
12. El desarrollo de árboles de decisión para la predicción de AI probable pone en valor la gran utilidad de la detección temprana del valor de IL-8, que demuestra ser un factor clave en conjunto con GM y PCR.
13. Los pacientes con COVID-19, no-COVID-19 y donantes sanos presentan perfiles significativamente distintos de COVs expirados en aire e inmunomoduladores en suero.
14. IL-6, IL-8, IL-2, TFN- α , CXCL10, TF, TM, VCAM-1, GzmA, tolueno, ácido acético, etilbenceno, o-xileno y estireno demuestran ser efectivos predictores de la gravedad de los pacientes de manera individual.
15. La combinación de parámetros en suero y COVs para el pronóstico de gravedad mediante modelos de ML ofrece una elevada exactitud, sensibilidad y especificidad, destacando el mejor rendimiento del algoritmo RF comparado con RL, KNN, SVM y ANN tanto para los pacientes COVID-19 como no-COVID-19.
16. Los compuestos CXCL10 y MIBK son los más relevantes en la predicción de gravedad de los pacientes no-COVID-19.
17. Los parámetros IL-6, IL-8, tolueno y TM son los más relevantes en la predicción de gravedad de los pacientes COVID-19.

4 CONCLUSIONS

The results obtained through this work lead us to the following conclusions:

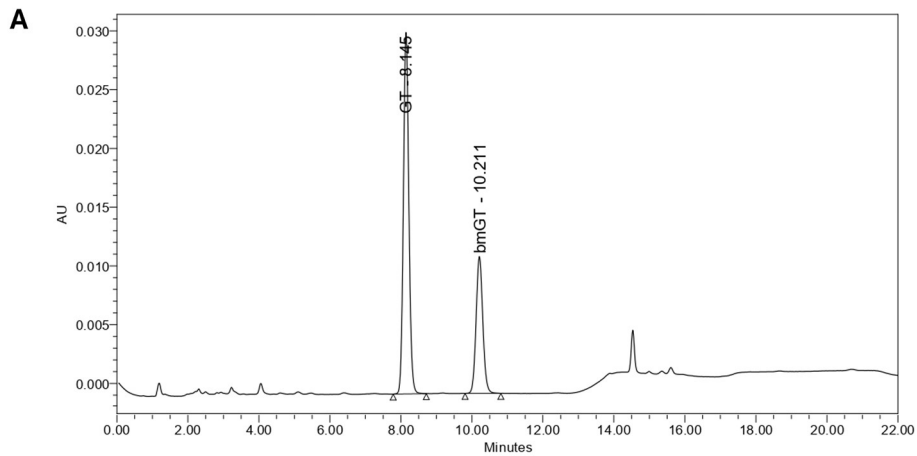
1. GT shows antibacterial activity against both methicillin-sensitive and methicillin-resistant *S. aureus* strains, including those with intermediate resistance to vancomycin.
2. GT exhibits synergy with various antibiotics commonly used in clinical practice, such as linezolid, fusidic acid, and vancomycin, reducing their effective concentrations against both MSSA and MRSA/SARIV strains *in vitro*.
3. The combination of GT with vancomycin lowers their individual CPMs and displays a potent synergistic effect *in vitro*, resulting in a significant antibiotic effect with vancomycin concentrations reduced to half or even a quarter of their MIC for the MRSA/SARIV strain.
4. *In vivo* studies using a *C. elegans* model of MRSA/SARIV infection demonstrate the potentiating effect of GT when combined with vancomycin at concentrations that remain minimally toxic to nematodes.
5. NRPS from specific *A. fumigatus* strains, such as *A. fischeri*, *A. lentutlus*, *A. novofumigatus*, *A. udawae*, *A. turcosus* and *A. thermomutatus* show high sequence homology with the gliP protein of *A. fumigatus* Af293, which is crucial for GT synthesis.
6. Specific strains of *A. fumigatus*, *A. fischeri*, *A. turcosus*, *A. pseudofischeri* and *T. virens* are capable of producing GT and bmGT *in vitro*.
7. The *A. fumigatus* B5233 strain produces higher concentrations of GT and bmGT in both Czapek-Dox and RPMI 1640 cultures, compared to *A. fumigatus* 1631562 and the rest of the strains.
8. A significant percentage of patients diagnosed as AI-negative included in this study had received pharmacological therapies through prophylaxis and/or antifungal treatments.
9. The utility of bmGT as a biomarker is superior to GM and BDG in terms of specificity and NPV if the GM bias is eliminated.
10. There is a significant association between the results of the first bmGT sample analysed in patients' serum and the diagnosis of AI in the studied population.
11. The combination of bmGT, GM, IL-8 and CMV as AI biomarkers greatly enhances the discrimination of patients with probable AI compared to GM alone.

Conclusions

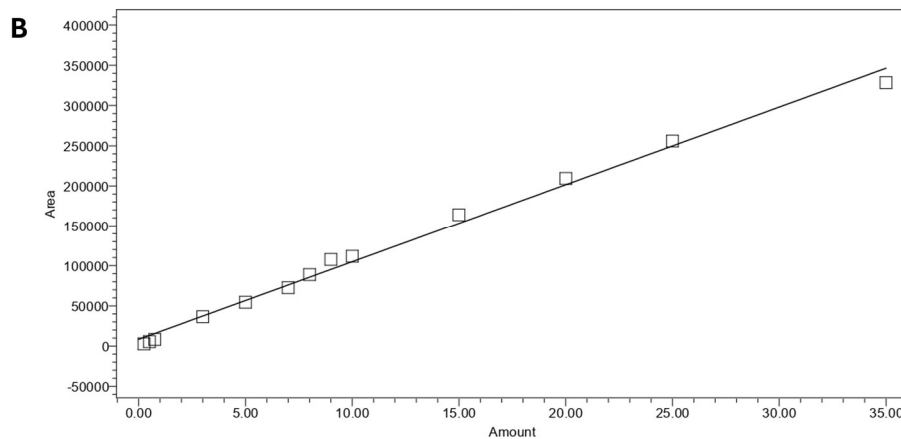
12. The development of decision trees for predicting probable AI highlights the importance of early IL-8 detection, which proves to be a key factor in combination with GM and PCR.
13. COVID-19 patients, non-COVID-19 patients, and healthy donors exhibit significantly different profiles of expired VOCs and serum immunomodulators.
14. Individually, IL-6, IL-8, IL-2, TFN- α , CXCL10, TF, TM, VCAM-1, GzmA, toluene, acetic acid, ethylbenzene, *o*-xylene and styrene are effective predictors of patients severity.
15. The combination of serum parameters and VOCs for severity prediction using ML models offers high accuracy, sensitivity and specificity. The RF algorithm demonstrates better performance compared to RL, KNN, SVM and ANN for both COVID-19 and non-COVID-19 patients.
16. CXCL10 and MIBK are the most relevant compounds for predicting the severity of non-COVID-19 patients.
17. IL-6, IL-8, toluene and TM are the most relevant parameters for predicting the severity of COVID-19 patients.

Anexos

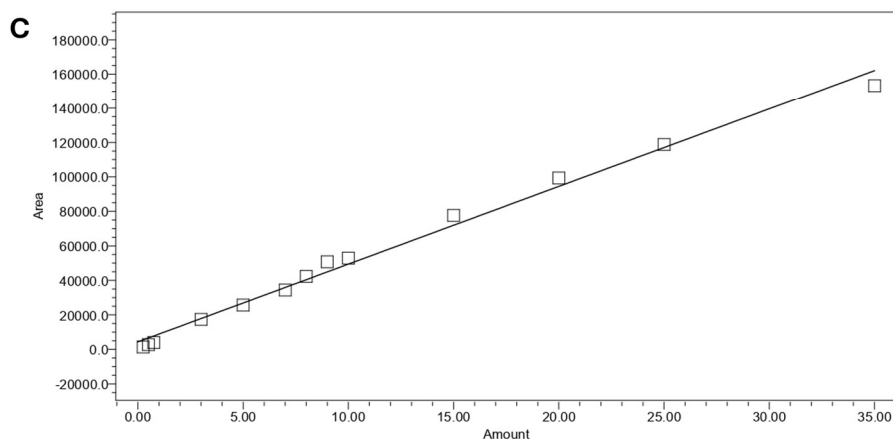
5 ANEXOS
4.1 Anexo 1



Peak Name	RT	Area	% Area	Height	Amount	Units
1 GT	8.145	328372	68.17	30792	35.000	ug/ml
2 bmGT	10.211	153324	31.83	11641	35.000	ug/ml



Peak Name: GT; RT: 8.268; Fit Type: Linear (1st Order); Cal Curve Id: 6532; R: 0.996230; R²: 0.992473;
 — Weighting: None; Equation: $Y = 9.66e+03 X + 8.38e+03$; Normalized Intercept/Slope: 0.049223; RSD(E): 8.306042



Peak Name: bmGT; RT: 10.396; Fit Type: Linear (1st Order); Cal Curve Id: 6533; R: 0.995654; R²: 0.991328; Weighting: None; Equation: $Y = 4.51e+03 X + 4.35e+03$; Normalized Intercept/Slope: 0.054811; RSD(E): 8.845175

Figura S1. (A) Cromatograma de los estándares de GT y bmGT analizados mediante HPLC. (B) Curva de calibración de la GT. (C) Curva de calibración de la bmGT.

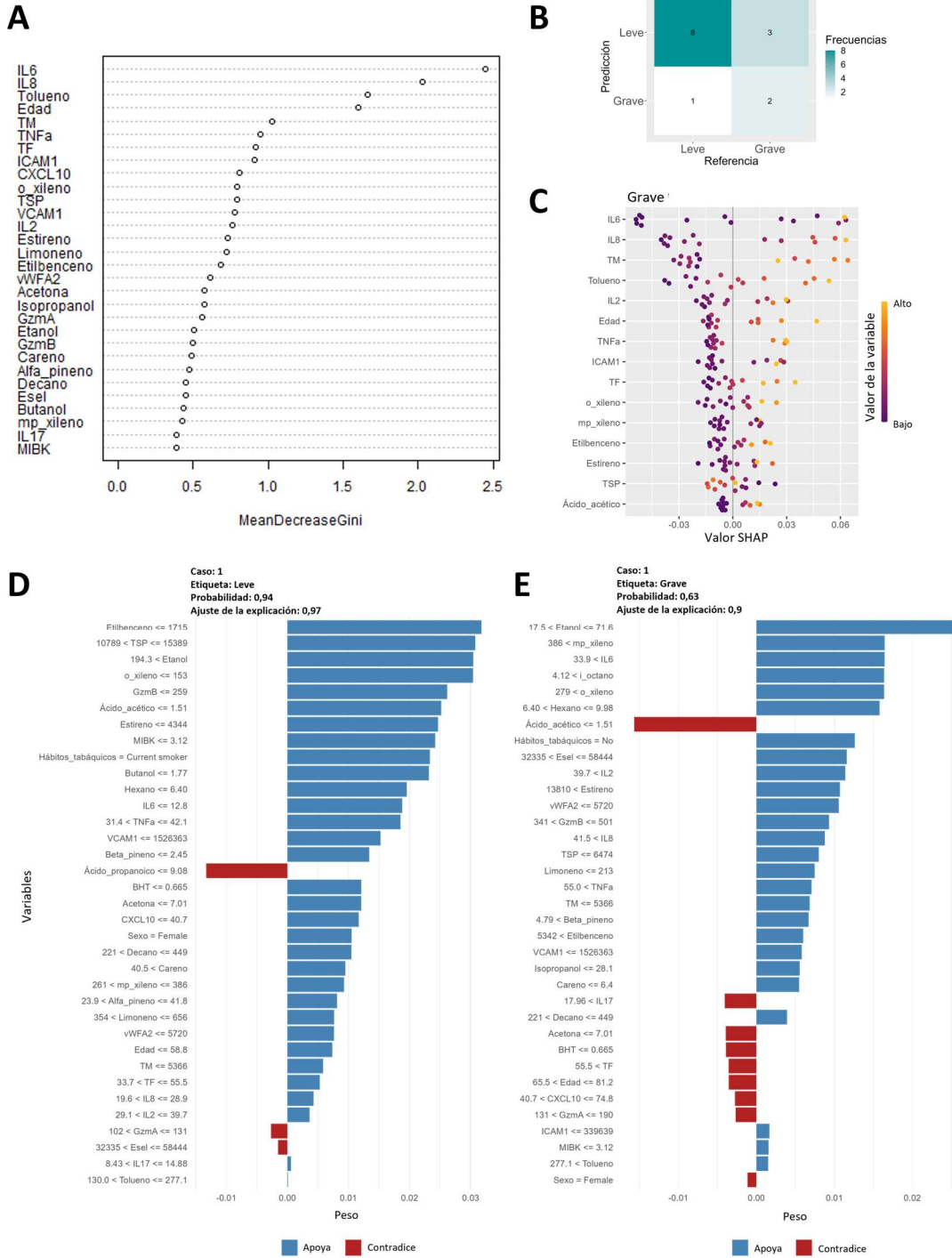


Figura S2. Resultados del modelo RF para el conjunto de datos COVID-19, añadiendo las variables demográficas como predictoras. **(A)** El gráfico del coeficiente de Gini de disminución media (MDG, del inglés, *Mean Decrease Gini*) clasifica la importancia de la variable respecto a la gravedad de la enfermedad en este modelo de clasificación. **(B)** La matriz de confusión muestra el rendimiento del modelo en el conjunto de datos de prueba, comparando las predicciones del modelo con la etiqueta de gravedad real. **(C)** El gráfico resumen de las explicaciones aditivas de Shapley (SHAP, del inglés, *Shapley Additive exPlanations*) cuantifica la importancia de las variables en el conjunto de datos de prueba. Los valores de SHAP indican la magnitud del cambio en las probabilidades logarítmicas (es decir, explican la probabilidad de que un paciente grave o leve). Los puntos coloreados con un gradiente violeta-naranja indican el valor original de la concentración para cada compuesto y paciente. **(D, E)** Las explicaciones locales interpretables independientes del modelo (LIME, del inglés, *Local Interpretable Model-agnostic Explanations*) muestran predicciones individualizadas para dos

ejemplos: (D) un paciente leve y (E) un paciente grave. Muestra la probabilidad de ser clasificado como leve (D) o grave (E), y la explicación adecuada para cada paciente específico. La longitud de la barra representa en qué medida cada variable apoya (azul) o contradice (rojo) la clasificación como leve (D) o grave (E). Unidades de los COVs: ng/L. Unidades de los parámetros séricos: pg/ml. MIBK: metil isobutil cetona; BHT: butilhidroxitolueno; TNFa: factor de necrosis tumoral; IL: interleuquina; vWFA2: factor de von Willebrand A2; CXCL10: ligando 10 de quimioquina con motivo C-X-C; Esel: e-selectina; Gzm: granzima; ICAM1: molécula de adhesión intercelular 1; TF: factor tisular; TM: trombosmodulina; VCAM1: molécula de adhesión de células vasculares; TSP: trombospondina.

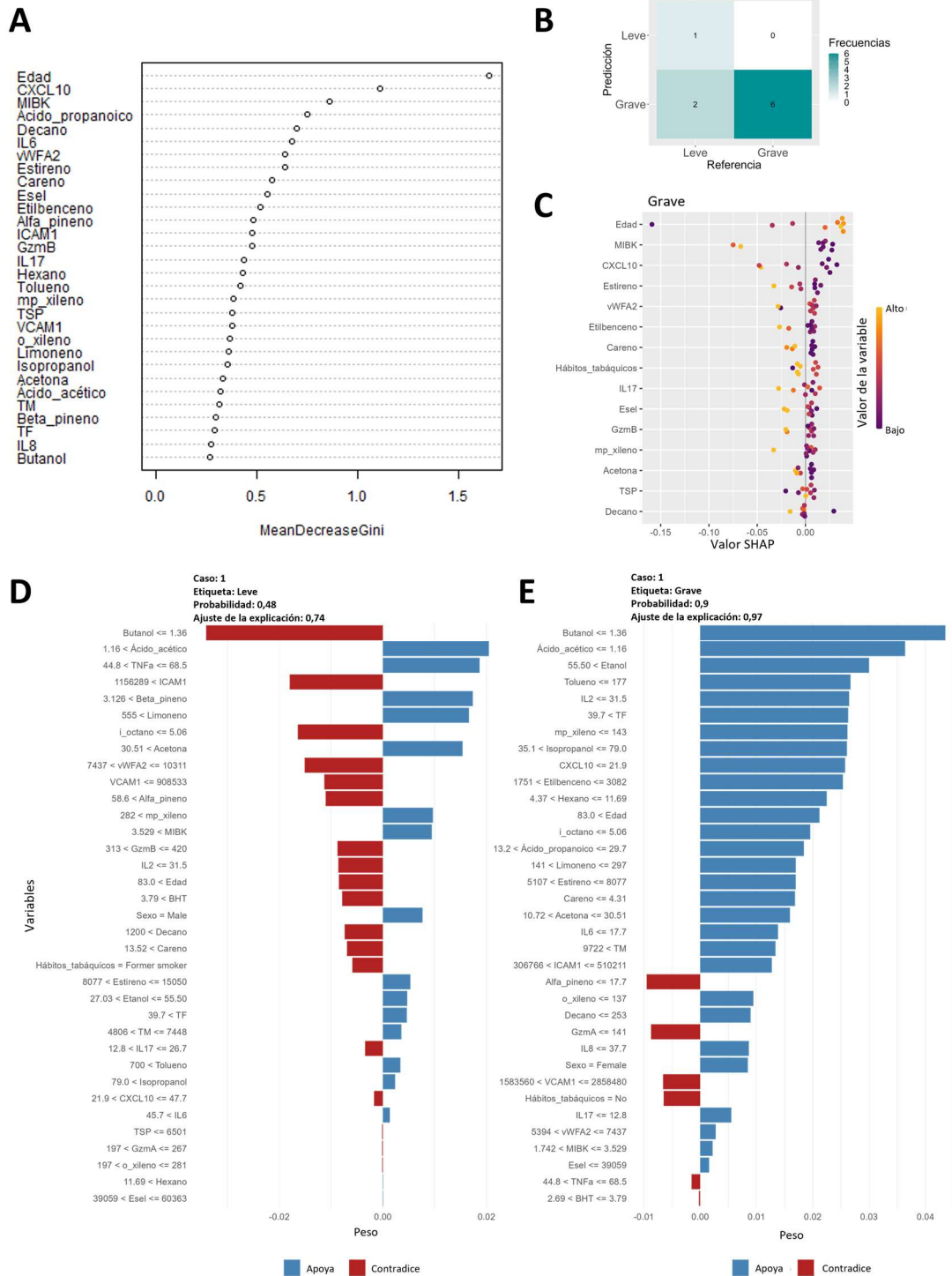


Figura S3. Resultados del modelo RF para el conjunto de datos no-COVID-19, añadiendo las variables demográficas como predictoras. (A) El gráfico del coeficiente de Gini de disminución media (MDG, del inglés,

Anexo 1

Mean Decrease Gini) clasifica la importancia de la variable respecto a la gravedad de la enfermedad en este modelo de clasificación. **(B)** La matriz de confusión muestra el rendimiento del modelo en el conjunto de datos de prueba, comparando las predicciones del modelo con la etiqueta de gravedad real. **(C)** El gráfico resumen de las explicaciones aditivas de Shapley (SHAP, del inglés, *Shapley Additive exPlanations*) cuantifica la importancia de las variables en el conjunto de datos de prueba. Los valores de SHAP indican la magnitud del cambio en las probabilidades logarítmicas (es decir, explican la probabilidad de que un paciente grave o leve). Los puntos coloreados con un gradiente violeta-naranja indican el valor original de la concentración para cada compuesto y paciente. **(D, E)** Las explicaciones locales interpretables independientes del modelo (LIME, del inglés, *Local Interpretable Model-agnostic Explanations*) muestran predicciones individualizadas para dos ejemplos: (D) un paciente leve y (E) un paciente grave. Muestra la probabilidad de ser clasificado como leve (D) o grave (E), y la explicación adecuada para cada paciente específico. La longitud de la barra representa en qué medida cada variable apoya (azul) o contradice (rojo) la clasificación como leve (D) o grave (E). Unidades de los COVs: ng/L. Unidades de los parámetros séricos: pg/ml. MIBK: metil isobutil cetona; BHT: butilhidroxitolueno; TNFa: factor de necrosis tumoral; IL: interleuquina; vWFA2: factor de von Willebrand A2; CXCL10: ligando 10 de quimioquina con motivo C-X-C; Esel: e-selectina; Gzm: granzima; ICAM1: molécula de adhesión intercelular 1; TF: factor tisular; TM: trombospondulina; VCAM1: molécula de adhesión de células vasculares; TSP: trombospondina.

4.2 Anexo 2

Tabla S1. Paquetes y funciones de R utilizadas.

Paquetes	Funciones
base	table, summary, writeLines, round, split, merge, paste, expand.grid, scale, max.col, mean, sqrt, replace, subset, duplicated, list, gsub, tapply, sapply, which.max, levels
biotools	boxM
C50	C5.0
caret	createDataPartition, train, confusionMatrix, varImp, nearZeroVar, trainControl
cdparcoord	demog
class	knn
corrplot	corrplot
DescTools	rev
devtools	Install
dichromat	colorRampPalette
dplyr	group_by, summarise, anti_join, select, mutate, filter, count
e1071	svm
factoextra	fviz_pca_ind, fviz_pca_var, fviz_pca_biplot, get_pca_var, fviz_cos2, fviz_contrib
FactoMineR	PCA
FSA	dunnTest
GGally	ggpairs
ggcorrplot	cor, cor_pmat, ggcorrplot
ggplot2	ggplot, ggsave, theme, labs, aes, coord_polar, geom_segment, geom_bar, annotate, geom_text, geom_density
ggpubr	ggboxplot, stat_pvalue_manual, ggarrange
ggstatsplot	ggbetweenstats, ggbarstats
ggVennDiagram	ggVennDiagram
gmodels	CrossTable, ci
gplots	heatmap.2, colorpanel, Venn
graphics	hist
graphics	par
grDevices	png, dev.off, colorRampPalette
kernelshap	kernelshap
kernlab	ksvm
knitr	kable
lime	as_classifier, lime, explain, plot_features
lplyr	pull
lubridate	floor
MASS	stepAIC
MVN	mvn
neuralnet	neuralnet, compute
nnet	nnet, multinom
nortest	lillie.test
pheatmap	pheatmap
plotly	last_plot

Anexo 2

plotrix	plot3d, pie3D
pROC	roc, plot.roc, coords, var, auc, ci.auc
purrr	map
randomForest	randomForest, varImpPlot
rcompanion	epsilonSquared
readxl	read.xlsx
reshape2	melt
reshape2	melt
ROCR	prediction, performance
rpart	rpart
rpart.plot	prp
rstatix	dunn_test, wilcox_test, anova_test, welch_anova_test, t_test, get_summary_stats, add_significance, wilcox_effsize
schoolmath	is.odd
shapviz	shapviz, sv_importance
stats	kruskal.test, heatmap, ks.test, qqnorm, sd, chisq.test, pairwise_prop_test, pairwise_t_test, aov, TukeyHSD, t.test, glm, predict, coef, reshape, cor_test, shapiro.test, wilcox.test, fitted, pt, fisher.test, p.adjust
stringr	str_replace, str_replace_all, str_split
tidyft	inner_join
tidyr	%>%, pivot_wider
tidyverse	
utils	capture.output, combn
verification	roc.area
viridis	viridis
webr	PieDonut
wordcloud	wordcloud
writexl	write_xlsx

Bibliografía

6 BIBLIOGRAFÍA

1. Niederman MS, Torres A. Respiratory infections. *European respiratory review : an official journal of the European Respiratory Society*. 2022;31(166):220150.
2. World Health Organization. (2019). *International Classification of Diseases for Mortality and Morbidity Statistics (11th ed.)*. <https://icd.who.int/browse11/l-m/en>.
3. Liu. PVDaC. Chapter 93 Infections of the Respiratory System. In: S. B, editor. *Medical Microbiology 4th edition*. University of Texas Medical Branch at Galveston. 1996.
4. Long Y, Zheng Y, Li C, Guo Z, Li P, Zhang F, et al. Respiratory pathogenic microbial infections: a narrative review. *International journal of medical sciences*. 2024;21(5):826-36.
5. Carroll KC, Adams LL. Lower Respiratory Tract Infections. *Microbiology spectrum*. 2016;4(4).
6. Azoulay E, Russell L, Van de Louw A, Metaxa V, Bauer P, Povoia P, et al. Diagnosis of severe respiratory infections in immunocompromised patients. *Intensive care medicine*. 2020;46(2):298-314.
7. Blasi F, Torres A. Respiratory infections management: Still a challenge. *Pulmonary pharmacology & therapeutics*. 2015;32:117-8.
8. Ritchie, H., Roser, M., & Ortiz-Ospina, E. (2021). Death rate from respiratory infections. *Our World in Data*. https://ourworldindata.org/grapher/death-rate-from-respiratory-infections?time=earliest..2021&country=~OWID_WRL.
9. Organización Mundial de la Salud. "The Top 10 Causes of Death." WHO, 2020, <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>.
10. Organización Mundial de la Salud. "Mortality and Global Health Estimates." WHO, 2021, <https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates>.
11. Global, regional, and national incidence and mortality burden of non-COVID-19 lower respiratory infections and aetiologies, 1990-2021: a systematic analysis from the Global Burden of Disease Study 2021. *The Lancet Infectious diseases*. 2024.
12. Fitzpatrick AM, Busse WW. Perspectives in Respiratory Infections and the Lung. *The journal of allergy and clinical immunology In practice*. 2022;10(3):694-6.
13. Velavan TP, Meyer CG. The COVID-19 epidemic. *Tropical medicine & international health : TM & IH*. 2020;25(3):278-80.
14. Gandhi L, Maisnam D, Rathore D, Chauhan P, Bonagiri A, Venkataramana M. Respiratory illness virus infections with special emphasis on COVID-19. *European journal of medical research*. 2022;27(1):236.
15. Parums DV. Editorial: The World Health Organization (WHO) Fungal Priority Pathogens List in Response to Emerging Fungal Pathogens During the COVID-19 Pandemic. *Medical science monitor : international medical journal of experimental and clinical research*. 2022;28:e939088.
16. Sati H, Beardsley J, Alastruey-Izquierdo A, Alffenaar J-W, Morrissey O, Beyer P, et al. WHO fungal priority pathogens list to guide research development and public health. 2022.
17. Jenks JD, Cornely OA, Chen SC, Thompson GR, 3rd, Hoenigl M. Breakthrough invasive fungal infections: Who is at risk? *Mycoses*. 2020;63(10):1021-32.
18. Kousha M, Tadi R, Soubani AO. Pulmonary aspergillosis: a clinical review. *European respiratory review : an official journal of the European Respiratory Society*. 2011;20(121):156-74.
19. Moldoveanu B, Gearhart AM, Jalil BA, Saad M, Guardiola JJ. Pulmonary Aspergillosis: Spectrum of Disease. *The American journal of the medical sciences*. 2021;361(4):411-9.
20. Pappas PG, Chen SC, Donnelly JP. The Evidence Supporting the Revised EORTC/MSGERC Definitions for Invasive Fungal Infections. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2021;72(Suppl 2):S77-s8.

21. Ortiz SC, Pennington K, Thomson DD, Bertuzzi M. Novel Insights into *Aspergillus fumigatus* Pathogenesis and Host Response from State-of-the-Art Imaging of Host-Pathogen Interactions during Infection. *Journal of fungi (Basel, Switzerland)*. 2022;8(3).
22. Yang F-F, Yu S-J, Du W-N, Wang H-M, Yao X-X, Xue D-D, et al. Global morbidity and mortality of lower respiratory infections: A population -based study. *Respiratory Medicine*. 2022;205:107042.
23. Khan K, Aslam MA, Zahra FT, Basheer H, Bilal M, Sumrin A. Potential biomarkers for the diagnosis of respiratory tract infection and lungs cancer. *Cellular and molecular biology (Noisy-le-Grand, France)*. 2017;63(11):46-52.
24. Zou XH, Cao B. [Advances in pathogen diagnosis of respiratory infection diseases in 2021]. *Zhonghua jie he he hu xi za zhi = Zhonghua jiehe he huxi zazhi = Chinese journal of tuberculosis and respiratory diseases*. 2022;45(1):78-82.
25. Lynch T, Bialy L, Kellner JD, Osmond MH, Klassen TP, Durec T, et al. A systematic review on the diagnosis of pediatric bacterial pneumonia: when gold is bronze. *PLoS one*. 2010;5(8):e11989.
26. Marimón JM, Navarro-Marí JM. [Rapid diagnostic test for respiratory infections]. *Enfermedades infecciosas y microbiología clínica*. 2017;35(2):108-15.
27. Póvoa P, Coelho L, Bos LDJ. New biomarkers for respiratory infections. *Current opinion in pulmonary medicine*. 2020;26(3):232-40.
28. Charles PG. Early diagnosis of lower respiratory tract infections (point-of-care tests). *Current opinion in pulmonary medicine*. 2008;14(3):176-82.
29. Bonner AB, Monroe KW, Talley LI, Klasner AE, Kimberlin DW. Impact of the rapid diagnosis of influenza on physician decision-making and patient management in the pediatric emergency department: results of a randomized, prospective, controlled trial. *Pediatrics*. 2003;112(2):363-7.
30. Riedel S, Beekmann SE, Heilmann KP, Richter SS, Garcia-de-Lomas J, Ferech M, et al. Antimicrobial use in Europe and antimicrobial resistance in *Streptococcus pneumoniae*. *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology*. 2007;26(7):485-90.
31. Ho J, Ip M. Antibiotic-Resistant Community-Acquired Bacterial Pneumonia. *Infectious disease clinics of North America*. 2019;33(4):1087-103.
32. Kontoyiannis DP, Bodey GP. Invasive aspergillosis in 2002: an update. *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology*. 2002;21(3):161-72.
33. Upton A, Kirby KA, Carpenter P, Boeckh M, Marr KA. Invasive Aspergillosis following Hematopoietic Cell Transplantation: Outcomes and Prognostic Factors Associated with Mortality. *Clinical Infectious Diseases*. 2007;44(4):531-40.
34. Latgé JP, Chamilos G. *Aspergillus fumigatus* and Aspergillosis in 2019. *Clinical microbiology reviews*. 2019;33(1).
35. Arias M, Santiago L, Vidal-Garcia M, Redrado S, Lanuza P, Comas L, et al. Preparations for Invasion: Modulation of Host Lung Immunity During Pulmonary Aspergillosis by Gliotoxin and Other Fungal Secondary Metabolites. *Frontiers in Immunology*. 2018;9:2549.
36. Boyer J, Feys S, Zsifkovits I, Hoenigl M, Egger M. Treatment of Invasive Aspergillosis: How It's Going, Where It's Heading. *Mycopathologia*. 2023;188(5):667-81.
37. Wisplinghoff H, Bischoff T, Tallent SM, Seifert H, Wenzel RP, Edmond MB. Nosocomial Bloodstream Infections in US Hospitals: Analysis of 24,179 Cases from a Prospective Nationwide Surveillance Study. *Clinical Infectious Diseases*. 2004;39(3):309-17.
38. De Lucca AJ, Boué SM, Carter-Wientjes CH, Bland JM, Bhatnagar D, Cleveland TE. Volatile profiles of toxigenic and non-toxigenic *Aspergillus flavus* using SPME for solid

- phase extraction. *Annals of agricultural and environmental medicine* : AAEM. 2010;17(2):301-8.
39. Weiner LM, Webb AK, Limbago B, Dudeck MA, Patel J, Kallen AJ, et al. Antimicrobial-Resistant Pathogens Associated With Healthcare-Associated Infections: Summary of Data Reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2011-2014. *Infection control and hospital epidemiology*. 2016;37(11):1288-301.
40. Ramírez P, Garnacho-Montero J. Aspergilosis invasiva en el paciente crítico. *Revista Iberoamericana de Micología*. 2018;35(4):210-6.
41. Kimura S-i. Invasive Aspergillosis in Hematological Patients. *Medical Mycology Journal*. 2016;57(2):J77-J88.
42. Garnacho-Montero J, Olaechea P, Alvarez-Lerma F, Alvarez-Rocha L, Blanquer J, Galván B, et al. Epidemiology, diagnosis and treatment of fungal respiratory infections in the critically ill patient. *Revista española de quimioterapia : publicación oficial de la Sociedad Española de Quimioterapia*. 2013;26(2):173-88.
43. Baddley JW, Stephens JM, Ji X, Gao X, Schlamm HT, Tarallo M. Aspergillosis in Intensive Care Unit (ICU) patients: epidemiology and economic outcomes. *BMC infectious diseases*. 2013;13(1):29.
44. Nicolle MC, Benet T, Thiebaut A, Bienvenu AL, Voirin N, Duclos A, et al. Invasive aspergillosis in patients with hematologic malignancies: incidence and description of 127 cases enrolled in a single institution prospective survey from 2004 to 2009. *Haematologica*. 2011;96(11):1685-91.
45. Neofytos D, Horn D, Anaissie E, Steinbach W, Olyaei A, Fishman J, et al. Epidemiology and Outcome of Invasive Fungal Infection in Adult Hematopoietic Stem Cell Transplant Recipients: Analysis of Multicenter Prospective Antifungal Therapy (PATH) Alliance Registry. *Clinical Infectious Diseases*. 2009;48(3):265-73.
46. Pfaller MA, Diekema DJ. Epidemiology of Invasive Candidiasis: a Persistent Public Health Problem. *Clinical Microbiology Reviews*. 2007;20(1):133-63.
47. Garnacho-Montero J, Amaya-Villar R, Ortiz-Leyba C, León C, Álvarez-Lerma F, Nolla-Salas J, et al. Isolation of *Aspergillus* spp. from the respiratory tract in critically ill patients: risk factors, clinical presentation and outcome. *Critical Care*. 2005;9(3):R191-R.
48. Meersseman W, Lagrou K, Maertens J, Wilmer A, Hermans G, Vanderschueren S, et al. Galactomannan in Bronchoalveolar Lavage Fluid. *American Journal of Respiratory and Critical Care Medicine*. 2008;177(1):27-34.
49. Graf K, Khani SM, Ott E, Mattner F, Gastmeier P, Sohr D, et al. Five-years surveillance of invasive aspergillosis in a university hospital. *BMC Infectious Diseases*. 2011;11(1):163-.
50. Rabagliati R. [Update in the diagnostic and therapeutic approach of invasive aspergillosis in adult population]. *Revista chilena de infectología : organo oficial de la Sociedad Chilena de Infectología*. 2018;35(5):531-44.
51. Solano C, Vázquez L. [Invasive aspergillosis in the patient with oncohematologic disease]. *Revista iberoamericana de micología*. 2018;35(4):198-205.
52. Raposo Puglia D, Raposo Puglia J, García-Cabrera E, Morales F, Camacho-Vega JC, Vilches-Arenas Á. Risk Factors and Environmental Preventive Actions for Aspergillosis in Patients with Hematological Malignancies. *Clinics and practice*. 2024;14(1):280-92.
53. Ascoglu S, Rex JH, de Pauw B, Bennett JE, Bille J, Crokaert F, et al. Defining opportunistic invasive fungal infections in immunocompromised patients with cancer and hematopoietic stem cell transplants: an international consensus. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2002;34(1):7-14.
54. Garcia-Vidal C, Carratalà J. [Pathogenesis of invasive fungal infections]. *Enfermedades infecciosas y microbiología clínica*. 2012;30(3):151-8.

Bibliografía

55. Sanna M, Caocci G, Ledda A, Orrù F, Fozza C, Deias P, et al. Glucose-6-phosphate dehydrogenase deficiency and risk of invasive fungal disease in patients with acute myeloid leukemia. *Leukemia & lymphoma*. 2017;58(11):2558-64.
56. Vallejo Llamas JC, Ruiz-Camps I. [Invasive fungal infection in haematology patients]. *Enfermedades infecciosas y microbiología clínica*. 2012;30(9):572-9.
57. Pemán J, Salavert M. [General epidemiology of invasive fungal disease]. *Enfermedades infecciosas y microbiología clínica*. 2012;30(2):90-8.
58. Kontoyiannis DP, Chamilos G, Lewis RE, Giralt S, Cortes J, Raad, II, et al. Increased bone marrow iron stores is an independent risk factor for invasive aspergillosis in patients with high-risk hematologic malignancies and recipients of allogeneic hematopoietic stem cell transplantation. *Cancer*. 2007;110(6):1303-6.
59. Garcia-Vidal C, Barba P, Arnan M, Moreno A, Ruiz-Camps I, Gudíol C, et al. Invasive aspergillosis complicating pandemic influenza A (H1N1) infection in severely immunocompromised patients. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2011;53(6):e16-9.
60. Ajmal S, Mahmood M, Abu Saleh O, Larson J, Sohail MR. Invasive fungal infections associated with prior respiratory viral infections in immunocompromised hosts. *Infection*. 2018;46(4):555-8.
61. Yılmaz G, Coşkun B, Elhan A, Azap A, Akan H. D-index: A New Scoring System in Febrile Neutropenic Patients for Predicting Invasive Fungal Infections. *Turkish journal of haematology : official journal of Turkish Society of Haematology*. 2016;33(2):102-6.
62. Cunha C, Giovannini G, Pierini A, Bell AS, Sorci G, Riuzzi F, et al. Genetically-determined hyperfunction of the S100B/RAGE axis is a risk factor for aspergillosis in stem cell transplant recipients. *PloS one*. 2011;6(11):e27962.
63. Ruiz Camps I. Aspergillosis: Beyond the oncohematological patient. *Enfermedades infecciosas y microbiología clínica (English ed)*. 2020;38(1):1-3.
64. Ghez D, Calleja A, Protin C, Baron M, Ledoux MP, Damaj G, et al. Early-onset invasive aspergillosis and other fungal infections in patients treated with ibrutinib. *Blood*. 2018;131(17):1955-9.
65. Nedel WL, Kontoyiannis DP, Pasqualotto AC. Aspergillosis in patients treated with monoclonal antibodies. *Revista iberoamericana de micología*. 2009;26(3):175-83.
66. Lamoth F, Calandra T. Pulmonary aspergillosis: diagnosis and treatment. *European Respiratory Review*. 2022;31(166):220114.
67. De Pauw B, Walsh TJ, Donnelly JP, Stevens DA, Edwards JE, Calandra T, et al. Revised definitions of invasive fungal disease from the European Organization for Research and Treatment of Cancer/Invasive Fungal Infections Cooperative Group and the National Institute of Allergy and Infectious Diseases Mycoses Study Group (EORTC/MSG) Consensus Group. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2008;46(12):1813-21.
68. Tsitsikas DA, Morin A, Araf S, Murtagh B, Johnson G, Vinnicombe S, et al. Impact of the revised (2008) EORTC/MSG definitions for invasive fungal disease on the rates of diagnosis of invasive aspergillosis. *Medical Mycology*. 2012;50(5):538-42.
69. Ostrosky-Zeichner L, Alexander BD, Kett DH, Vazquez J, Pappas PG, Saeki F, et al. Multicenter Clinical Evaluation of the (1→3) β-D-Glucan Assay as an Aid to Diagnosis of Fungal Infections in Humans. *Clinical Infectious Diseases*. 2005;41(5):654-9.
70. Bassetti M, Azoulay E, Kullberg B-J, Ruhnke M, Shoham S, Vazquez J, et al. EORTC/MSGERC Definitions of Invasive Fungal Diseases: Summary of Activities of the Intensive Care Unit Working Group. *Clinical Infectious Diseases*. 2021;72(Supplement_2):S121-S7.
71. Donnelly JP, Chen SC, Kauffman CA, Steinbach WJ, Baddley JW, Verweij PE, et al. Revision and Update of the Consensus Definitions of Invasive Fungal Disease From the European Organization for Research and Treatment of Cancer and the Mycoses Study

- Group Education and Research Consortium. *Clinical Infectious Diseases*. 2019;71(6):1367-76.
72. Patterson TF, Thompson GR, Denning DW, Fishman JA, Hadley S, Herbrecht R, et al. Practice Guidelines for the Diagnosis and Management of Aspergillosis: 2016 Update by the Infectious Diseases Society of America. *Clinical Infectious Diseases*. 2016;63(4):e1-e60.
73. Kosmidis C, Denning DW. Opportunistic and Systemic Fungi. *Infectious Diseases: Elsevier*; 2017. p. 1681-709.e3.
74. Ledoux MP, Guffroy B, Nivoix Y, Simand C, Herbrecht R. Invasive Pulmonary Aspergillosis. *Seminars in respiratory and critical care medicine*. 2020;41(1):80-98.
75. Lamberink H, Wagemakers A, Sigaloff KCE, van Houdt R, de Jonge NA, van Dijk K. The impact of the updated EORTC/MSG criteria on the classification of hematological patients with suspected invasive pulmonary aspergillosis. *Clinical Microbiology and Infection*. 2022.
76. Alhambra A, Cuétara MS, Ortiz MC, Moreno JM, del Palacio A, Pontón J, et al. False positive galactomannan results in adult hematological patients treated with piperacillin-tazobactam. *Revista Iberoamericana de Micología*. 2007;24(2):106-12.
77. Rybak MJ. The efficacy and safety of daptomycin: first in a new class of antibiotics for Gram-positive bacteria. *Clinical Microbiology and Infection*. 2006;12:24-32.
78. Fortún J, Martín-Dávila P, Alvarez ME, Norman F, Sánchez-Sousa A, Gajate L, et al. False-Positive Results of Aspergillus Galactomannan Antigenemia in Liver Transplant Recipients. *Transplantation*. 2009;87(2).
79. Mennink-Kersten MA, Verweij PE. Non-culture-based diagnostics for opportunistic fungi. *Infectious disease clinics of North America*. 2006;20(3):711-27, viii.
80. Park HS, Jun SC, Han KH, Hong SB, Yu JH. Diversity, Application, and Synthetic Biology of Industrially Important Aspergillus Fungi. *Advances in applied microbiology*. 2017;100:161-202.
81. Perrone G, Susca A, Cozzi G, Ehrlich K, Varga J, Frisvad JC, et al. Biodiversity of Aspergillus species in some important agricultural products. *Studies in mycology*. 2007;59:53-66.
82. Dennis Kunkel. *Science Photo Library*. 2004.
83. Hilvering B, Speirs J, Ent CK, Beekman J. Allergic bronchopulmonary aspergillosis and other fungal diseases. 2011. p. 97-114.
84. Rokas A. Aspergillus. *Current biology : CB*. 2013;23(5):R187-8.
85. Latgé JP. Aspergillus fumigatus and aspergillosis. *Clinical microbiology reviews*. 1999;12(2):310-50.
86. Lamoth F. Aspergillus fumigatus-Related Species in Clinical Practice. *Frontiers in microbiology*. 2016;7:683.
87. Samson RA, Visagie CM, Houbraeken J, Hong SB, Hubka V, Klaassen CH, et al. Phylogeny, identification and nomenclature of the genus Aspergillus. *Studies in mycology*. 2014;78:141-73.
88. Alastruey-Izquierdo A, Alcazar-Fuoli L, Cuenca-Estrella M. Antifungal Susceptibility Profile of Cryptic Species of Aspergillus. *Mycopathologia*. 2014;178(5):427-33.
89. Arias RS, Orner VA, Martinez-Castillo J, Sobolev VS. Aspergillus Section Flavi, Need for a Robust Taxonomy. *Microbiology resource announcements*. 2021;10(48):e0078421.
90. Frisvad JC, Hubka V, Ezekiel CN, Hong SB, Nováková A, Chen AJ, et al. Taxonomy of Aspergillus section Flavi and their production of aflatoxins, ochratoxins and other mycotoxins. *Studies in mycology*. 2019;93:1-63.
91. Lass-Flörl C, Dietl AM, Kontoyiannis DP, Brock M. Aspergillus terreus Species Complex. *Clinical microbiology reviews*. 2021;34(4):e0031120.
92. Denis B, Guiguet M, de Castro N, Mechali F, Revest M, Melica G, et al. Relevance of EORTC Criteria for the Diagnosis of Invasive Aspergillosis in HIV-Infected Patients, and

- Survival Trends Over a 20-Year Period in France. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2015;61(8):1273-80.
93. Pagano L, Caira M, Candoni A, Offidani M, Martino B, Specchia G, et al. Invasive aspergillosis in patients with acute myeloid leukemia: a SEIFEM-2008 registry study. *Haematologica*. 2010;95(4):644-50.
 94. Steinbach WJ, Marr KA, Anaissie EJ, Azie N, Quan S-P, Meier-Kriesche H-U, et al. Clinical epidemiology of 960 patients with invasive aspergillosis from the PATH Alliance registry. *Journal of Infection*. 2012;65(5):453-64.
 95. Bouza E, Guinea J, Peláez T, Pérez-Molina J, Alcalá L, Muñoz P. Workload Due to *Aspergillus fumigatus* and Significance of the Organism in the Microbiology Laboratory of a General Hospital. *Journal of Clinical Microbiology*. 2005;43(5):2075-9.
 96. Álvarez Lerma F, Olaechea Astigarraga P, Palomar Martínez M, Rodríguez Carvajal M, Machado Casas JF, Jiménez Quintana MM, et al. Infecciones respiratorias por *Aspergillus* spp. en pacientes críticos ingresados en unidades de cuidados intensivos. *Medicina Intensiva*. 2015;39(3):149-59.
 97. Ullmann AJ, Aguado JM, Arikan-Akdagli S, Denning DW, Groll AH, Lagrou K, et al. Diagnosis and management of *Aspergillus* diseases: executive summary of the 2017 ESCMID-ECMM-ERS guideline. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*. 2018;24 Suppl 1:e1-e38.
 98. Ruhnke M, Cornely OA, Schmidt-Hieber M, Alakel N, Boell B, Buchheidt D, et al. Treatment of invasive fungal diseases in cancer patients-Revised 2019 Recommendations of the Infectious Diseases Working Party (AGIHO) of the German Society of Hematology and Oncology (DGHO). *Mycoses*. 2020;63(7):653-82.
 99. Campoy S, Adrio JL. Antifungals. *Biochemical pharmacology*. 2017;133:86-96.
 100. Grela E, Zdybicka-Barabas A, Pawlikowska-Pawlega B, Cytrynska M, Wlodarczyk M, Grudzinski W, et al. Modes of the antibiotic activity of amphotericin B against *Candida albicans*. *Scientific reports*. 2019;9(1):17029.
 101. Samson RA, Peterson SW, Frisvad JC, Varga J. New species in *Aspergillus* section *Terrei*. *Studies in mycology*. 2011;69(1):39-55.
 102. Schoberle TJ, Nguyen-Coleman CK, Herold J, Yang A, Weirauch M, Hughes TR, et al. A Novel C2H2 Transcription Factor that Regulates *gliA* Expression Interdependently with *GliZ* in *Aspergillus fumigatus*. *PLoS Genetics*. 102014.
 103. Calvo AM, Wilson RA, Bok JW, Keller NP. Relationship between Secondary Metabolism and Fungal Development. *Microbiology and Molecular Biology Reviews*. 662002. p. 447-59.
 104. Coleman JJ, Ghosh S, Okoli I, Mylonakis E. Antifungal activity of microbial secondary metabolites. *PloS one*. 2011;6(9):e25321.
 105. Theobald S, Vesth TC, Andersen MR. Genus level analysis of PKS-NRPS and NRPS-PKS hybrids reveals their origin in *Aspergilli*. *BMC genomics*. 2019;20(1):847.
 106. Gibbons JG, Rokas A. The function and evolution of the *Aspergillus* genome. *Trends in microbiology*. 2013;21(1):14-22.
 107. Yaegashi J, Oakley BR, Wang CC. Recent advances in genome mining of secondary metabolite biosynthetic gene clusters and the development of heterologous expression systems in *Aspergillus nidulans*. *Journal of industrial microbiology & biotechnology*. 2014;41(2):433-42.
 108. Fedorova ND, Khaldi N, Joardar VS, Maiti R, Amedeo P, Anderson MJ, et al. Genomic islands in the pathogenic filamentous fungus *Aspergillus fumigatus*. *PLoS genetics*. 2008;4(4):e1000046.
 109. Cramer RA, Jr., Stajich JE, Yamanaka Y, Dietrich FS, Steinbach WJ, Perfect JR. Phylogenomic analysis of non-ribosomal peptide synthetases in the genus *Aspergillus*. *Gene*. 2006;383:24-32.

110. von Döhren H. A survey of nonribosomal peptide synthetase (NRPS) genes in *Aspergillus nidulans*. *Fungal genetics and biology : FG & B*. 2009;46 Suppl 1:S45-52.
111. Frisvad JC, Rank C, Nielsen KF, Larsen TO. Metabolomics of *Aspergillus fumigatus*. *Medical mycology*. 2009;47 Suppl 1:S53-71.
112. Gardiner DM, Waring P, Howlett BJ. The epipolythiodioxopiperazine (ETP) class of fungal toxins: distribution, mode of action, functions and biosynthesis. *Microbiology*. 2005;151(Pt 4):1021-32.
113. Weindling R, Emerson O. The Isolation of a toxic substance from the culture filtrate of trichoderma. 1936.
114. White EC. Bactericidal filtrates from a mold culture. *Science (New York, NY)*. 1940;92(2380).
115. J. P. El regreso al futuro 10 años después. *Revista Española de Quimioterapia*. 2007;20(2):241-4.
116. Johnson JR, Bruce WF, Dutcher JD, Miller LL. Gliotoxin, the Antibiotic Principle of *Gliocladium fimbriatum*. II. General Chemical Behavior and Crystalline Derivatives. *Journal of the American Chemical Society*. 1944;66(4):614-6.
117. Beecham A. F. FJaMA. The structure and absolute configuration of gliotoxin and sporidesmin. *Tetrahedron Letters*. 1966:3131-8.
118. Gardiner DM, Howlett BJ. Bioinformatic and expression analysis of the putative gliotoxin biosynthetic gene cluster of *Aspergillus fumigatus*. *FEMS Microbiology Letters*. 2005;248(2):241-8.
119. Bok JW, Chung D, Balajee SA, Marr KA, Andes D, Nielsen KF, et al. GliZ, a transcriptional regulator of gliotoxin biosynthesis, contributes to *Aspergillus fumigatus* virulence. *Infection and immunity*. 2006;74(12):6761-8.
120. Balibar CJ, Walsh CT. GliP, a multimodular nonribosomal peptide synthetase in *Aspergillus fumigatus*, makes the diketopiperazine scaffold of gliotoxin. *Biochemistry*. 2006;45(50):15029-38.
121. Sugui JA, Pardo J, Chang YC, Müllbacher A, Zarembek KA, Galvez EM, et al. Role of *laeA* in the Regulation of *alb1*, *gliP*, Conidial Morphology, and Virulence in *Aspergillus fumigatus*. *Eukaryotic cell*. 2007;6(9):1552-61.
122. Dolan SK, O'Keeffe G, Jones GW, Doyle S. Resistance is not futile: gliotoxin biosynthesis, functionality and utility. *Trends in microbiology*. 2015;23(7):419-28.
123. Chang SL, Chiang YM, Yeh HH, Wu TK, Wang CC. Reconstitution of the early steps of gliotoxin biosynthesis in *Aspergillus nidulans* reveals the role of the monooxygenase GliC. *Bioorganic & medicinal chemistry letters*. 2013;23(7):2155-7.
124. Davis CC, Stephen; Schrettl, Markus; Singh, Ishwar; Stephens, John C.; Barry, Sarah M.; Kavanagh, Kevin; Challis, Gregory L.; Brougham, Dermot; Doyle, Sean. The Role of Glutathione S-Transferase GliG in Gliotoxin Biosynthesis in *Aspergillus fumigatus*. *Chemistry & Biology*. 2011;18(4):542-52.
125. Scharf DH, Chankhamjon P, Scherlach K, Heinekamp T, Willing K, Brakhage AA, et al. Epidithiodiketopiperazine biosynthesis: a four-enzyme cascade converts glutathione conjugates into transannular disulfide bridges. *Angewandte Chemie (International ed in English)*. 2013;52(42):11092-5.
126. Schrettl M, Carberry S, Kavanagh K, Haas H, Jones GW, O'Brien J, et al. Self-protection against gliotoxin--a component of the gliotoxin biosynthetic cluster, GliT, completely protects *Aspergillus fumigatus* against exogenous gliotoxin. *PLoS pathogens*. 2010;6(6):e1000952.
127. Scharf DH, Remme N, Heinekamp T, Hortschansky P, Brakhage AA, Hertweck C. Transannular disulfide formation in gliotoxin biosynthesis and its role in self-resistance of the human pathogen *Aspergillus fumigatus*. *Journal of the American Chemical Society*. 2010;132(29):10136-41.

128. Ries LNA, Pardeshi L, Dong Z, Tan K, Steenwyk JL, Colabardini AC, et al. The *Aspergillus fumigatus* transcription factor RglT is important for gliotoxin biosynthesis and self-protection, and virulence. *PLoS pathogens*. 2020;16(7):e1008645.
129. Dolan SK, Owens RA, O'Keeffe G, Hammel S, Fitzpatrick DA, Jones GW, et al. Regulation of nonribosomal peptide synthesis: bis-thiomethylation attenuates gliotoxin biosynthesis in *Aspergillus fumigatus*. *Chemistry & biology*. 2014;21(8):999-1012.
130. Scharf DH, Groll M, Habel A, Heinekamp T, Hertweck C, Brakhage AA, et al. Flavoenzyme-catalyzed formation of disulfide bonds in natural products. *Angewandte Chemie (International ed in English)*. 2014;53(8):2221-4.
131. Owens RA, O'Keeffe G, Smith EB, Dolan SK, Hammel S, Sheridan KJ, et al. Interplay between Gliotoxin Resistance, Secretion, and the Methyl/Methionine Cycle in *Aspergillus fumigatus*. *Eukaryotic cell*. 2015;14(9):941-57.
132. Duell ER, Glaser M, Le Chapelain C, Antes I, Groll M, Huber EM. Sequential Inactivation of Gliotoxin by the S-Methyltransferase TmtA. *ACS chemical biology*. 2016;11(4):1082-9.
133. Cramer RA, Jr., Gamcsik MP, Brooking RM, Najvar LK, Kirkpatrick WR, Patterson TF, et al. Disruption of a nonribosomal peptide synthetase in *Aspergillus fumigatus* eliminates gliotoxin production. *Eukaryotic cell*. 2006;5(6):972-80.
134. Wang DN, Toyotome T, Muraosa Y, Watanabe A, Wuren T, Bunsupa S, et al. GliA in *Aspergillus fumigatus* is required for its tolerance to gliotoxin and affects the amount of extracellular and intracellular gliotoxin. *Medical mycology*. 2014;52(5):506-18.
135. Weindling R. *Trichoderma lignorum* as a parasite of other soil fungi. *Phytopathology*. 1932;22:837-45.
136. Anitha R, Murugesan K. Production of gliotoxin on natural substrates by *Trichoderma virens*. *Journal of basic microbiology*. 2005;45(1):12-9.
137. Shah DT, Larsen B. Clinical isolates of yeast produce a gliotoxin-like substance. *Mycopathologia*. 1991;116(3):203-8.
138. Glister G, Williams T. Production of gliotoxin by *Aspergillus fumigatus* mut. *helvola* Yuill. *Nature*. 1944;153(3891):651-.
139. Kupfahl C, Michalka A, Lass-Flörl C, Fischer G, Haase G, Ruppert T, et al. Gliotoxin production by clinical and environmental *Aspergillus fumigatus* strains. *International Journal of Medical Microbiology*. 2008;298(3-4):319-27.
140. Frisvad JC, Larsen TO. Extrolites of *Aspergillus fumigatus* and Other Pathogenic Species in *Aspergillus* Section *Fumigati*. *Frontiers in microbiology*. 2015;6:1485.
141. Larsen TO, Smedsgaard J, Nielsen KF, Hansen MA, Samson RA, Frisvad JC. Production of mycotoxins by *Aspergillus lentulus* and other medically important and closely related species in section *Fumigati*. *Medical mycology*. 2007;45(3):225-32.
142. Vidal-García M, Redrado S, Domingo MP, Marquina P, Colmenarejo C, Meis JF, et al. Production of the Invasive Aspergillosis Biomarker Bis(methylthio)gliotoxin Within the Genus *Aspergillus*: In Vitro and in Vivo Metabolite Quantification and Genomic Analysis. *Frontiers in microbiology*. 2018;9:1246.
143. Kwon-Chung KJ, Sugui JA. What do we know about the role of gliotoxin in the pathobiology of *Aspergillus fumigatus*? *Medical Mycology*. 2009;47(Suppl 1):S97-103.
144. Waring P, Beaver J. Gliotoxin and related epipolythiodioxopiperazines. *General Pharmacology: The Vascular System*. 1996;27(8):1311-6.
145. Sugui JA, Pardo J Fau - Chang YC, Chang Yc Fau - Zarembler KA, Zarembler Ka Fau - Nardone G, Nardone G Fau - Galvez EM, Galvez Em Fau - Müllbacher A, et al. Gliotoxin is a virulence factor of *Aspergillus fumigatus*: gliP deletion attenuates virulence in mice immunosuppressed with hydrocortisone. *Eukaryotic Cell*. 2007;6:1562-9.
146. Spikes S, Xu R, Nguyen CK, Chamilos G, Kontoyiannis Dimitrios P, Jacobson Raymond H, et al. Gliotoxin Production in *Aspergillus fumigatus* Contributes to Host-Specific Differences in Virulence. *The Journal of Infectious Diseases*. 2008;197(3):479-86.

147. Sugui JA, Rose SR, Nardone G, Swamydas M, Lee CR, Kwon-Chung KJ, et al. Host immune status-specific production of gliotoxin and bis-methyl-gliotoxin during invasive aspergillosis in mice. *Scientific reports*. 2017;7(1):10977.
148. Domingo MP, Colmenarejo C, Martínez-Lostao L, Müllbacher A, Jarne C, Revillo MJ, et al. Bis(methyl)gliotoxin proves to be a more stable and reliable marker for invasive aspergillosis than gliotoxin and suitable for use in diagnosis. *Diagnostic microbiology and infectious disease*. 2012;73(1):57-64.
149. Tsunawaki S, Yoshida LS, Nishida S, Kobayashi T, Shimoyama T. Fungal metabolite gliotoxin inhibits assembly of the human respiratory burst NADPH oxidase. *Infection and immunity*. 2004;72(6):3373-82.
150. Eichner RD, Al Salami M, Wood PR, Müllbacher A. The effect of gliotoxin upon macrophage function. *International journal of immunopharmacology*. 1986;8(7):789-97.
151. Müllbacher A, Eichner RD. Immunosuppression in vitro by a metabolite of a human pathogenic fungus. *Proceedings of the National Academy of Sciences of the United States of America*. 1984;81(12):3835-7.
152. Pahl HL, Krauss B, Schulze-Osthoff K, Decker T, Traenckner EB, Vogt M, et al. The immunosuppressive fungal metabolite gliotoxin specifically inhibits transcription factor NF-kappaB. *The Journal of experimental medicine*. 1996;183(4):1829-40.
153. Waring P, Eichner RD, Mullbacher A, Sjaarda A. Gliotoxin induces apoptosis in macrophages unrelated to its antiphagocytic properties. *The Journal of biological chemistry*. 1988;263(34):18493-9.
154. Stanzani M, Orciuolo E, Lewis R, Kontoyiannis DP, Martins SL, St John LS, et al. *Aspergillus fumigatus* suppresses the human cellular immune response via gliotoxin-mediated apoptosis of monocytes. *Blood*. 2005;105(6):2258-65.
155. Niide O, Suzuki Y, Yoshimaru T, Inoue T, Takayama T, Ra C. Fungal metabolite gliotoxin blocks mast cell activation by a calcium- and superoxide-dependent mechanism: implications for immunosuppressive activities. *Clinical immunology (Orlando, Fla)*. 2006;118(1):108-16.
156. Hubmann R, Hilgarth M, Schnabl S, Ponath E, Reiter M, Demirtas D, et al. Gliotoxin is a potent NOTCH2 transactivation inhibitor and efficiently induces apoptosis in chronic lymphocytic leukaemia (CLL) cells. *British journal of haematology*. 2013;160(5):618-29.
157. Piva TJ. Gliotoxin induces apoptosis in mouse L929 fibroblast cells. *Biochemistry and molecular biology international*. 1994;33(3):411-9.
158. Kweon YO, Paik YH, Schnabl B, Qian T, Lemasters JJ, Brenner DA. Gliotoxin-mediated apoptosis of activated human hepatic stellate cells. *Journal of hepatology*. 2003;39(1):38-46.
159. Nguyen VT, Lee JS, Qian ZJ, Li YX, Kim KN, Heo SJ, et al. Gliotoxin isolated from marine fungus *Aspergillus* sp. induces apoptosis of human cervical cancer and chondrosarcoma cells. *Marine drugs*. 2013;12(1):69-87.
160. Chen J, Wang C, Lan W, Huang C, Lin M, Wang Z, et al. Gliotoxin Inhibits Proliferation and Induces Apoptosis in Colorectal Cancer Cells. *Marine drugs*. 2015;13(10):6259-73.
161. Pardo J, Urban C, Galvez EM, Ekert PG, Muller U, Kwon-Chung J, et al. The mitochondrial protein Bak is pivotal for gliotoxin-induced apoptosis and a critical host factor of *Aspergillus fumigatus* virulence in mice. *The Journal of cell biology*. 2006;174(4):509-19.
162. Comas L, Polo E, Domingo MP, Hernández Y, Arias M, Esteban P, et al. Intracellular Delivery of Biologically-Active Fungal Metabolite Gliotoxin Using Magnetic Nanoparticles. *Materials (Basel, Switzerland)*. 2019;12(7).
163. Ráduly Z, Szabó L, Madar A, Pócsi I, Csernoch L. Toxicological and Medical Aspects of *Aspergillus*-Derived Mycotoxins Entering the Feed and Food Chain. *Frontiers in microbiology*. 2019;10:2908.

164. Gallo A, Giuberti G, Frisvad JC, Bertuzzi T, Nielsen KF. Review on Mycotoxin Issues in Ruminants: Occurrence in Forages, Effects of Mycotoxin Ingestion on Health Status and Animal Performance and Practical Strategies to Counteract Their Negative Effects. *Toxins*. 2015;7(8):3057-111.
165. Novak B, Rainer V, Sulyok M, Haltrich D, Schatzmayr G, Mayer E. Twenty-Eight Fungal Secondary Metabolites Detected in Pig Feed Samples: Their Occurrence, Relevance and Cytotoxic Effects In Vitro. *Toxins*. 2019;11(9).
166. Stanley SA, Grant SS, Kawate T, Iwase N, Shimizu M, Wivagg C, et al. Identification of novel inhibitors of *M. tuberculosis* growth using whole cell based high-throughput screening. *ACS chemical biology*. 2012;7(8):1377-84.
167. Tompsett R, Mc DW, Kidd JG. Tuberculostatic activity of blood and urine from animals given gliotoxin. *Journal of immunology (Baltimore, Md : 1950)*. 1950;65(1):59-63.
168. Liang WL, Le X, Li HJ, Yang XL, Chen JX, Xu J, et al. Exploring the chemodiversity and biological activities of the secondary metabolites from the marine fungus *Neosartorya pseudofischeri*. *Marine drugs*. 2014;12(11):5657-76.
169. Svahn KS, Göransson U, El-Seedi H, Bohlin L, Larsson DG, Olsen B, et al. Antimicrobial activity of filamentous fungi isolated from highly antibiotic-contaminated river sediment. *Infection ecology & epidemiology*. 2012;2.
170. Hatabu T, Hagiwara M, Taguchi N, Kiyozawa M, Suzuki M, Kano S, et al. Plasmodium falciparum: the fungal metabolite gliotoxin inhibits proteasome proteolytic activity and exerts a plasmodicidal effect on *P. falciparum*. *Experimental parasitology*. 2006;112(3):179-83.
171. Aljofan M, Sganga ML, Lo MK, Rootes CL, Porotto M, Meyer AG, et al. Antiviral activity of gliotoxin, gentian violet and brilliant green against Nipah and Hendra virus in vitro. *Virology Journal*. 2009;6:187.
172. McDougall JK. Antiviral action of gliotoxin. *Archiv fur die gesamte Virusforschung*. 1969;27(2):255-67.
173. Rightsel WA, Schneider HG, Sloan BJ, Graf PR, Miller FA, Bartz OR, et al. ANTIVIRAL ACTIVITY OF GLIOTOXIN AND GLIOTOXIN ACETATE. *Nature*. 1964;204:1333-4.
174. Larin NM, Copping MP, Herbst-Laier RH, Roberts B, Wenham RB. Antiviral activity of gliotoxin. *Chemotherapy*. 1965;10(1):12-23.
175. Jurado DN. Desarrollo de inmunobiosensores y sistemas de diagnóstico para aplicaciones biomédicas: análisis de la interacción ICAM-1/LFA-1 y detección de aspergilosis invasiva. CSIC - Instituto de Carboquímica (ICB). Universidad de Zaragoza; 2015.
176. Waring P, Newcombe N, Edel M, Lin QH, Jiang H, Sjaarda A, et al. Cellular uptake and release of the immunomodulating fungal toxin gliotoxin. *Toxicon*. 1994;32(4):491-504.
177. Amitani R, Taylor G, Elezis EN, Llewellyn-Jones C, Mitchell J, Kuze F, et al. Purification and characterization of factors produced by *Aspergillus fumigatus* which affect human ciliated respiratory epithelium. *Infection and Immunity*. 1995;63(9):3266-71.
178. Lewis RE, Wiederhold NP, Chi J, Han XY, Komanduri KV, Kontoyiannis DP, et al. Detection of gliotoxin in experimental and human aspergillosis. *Infection and immunity*. 2005;73(1):635-7.
179. Lewis RE, Wiederhold NP, Lionakis MS, Prince RA, Kontoyiannis DP. Frequency and Species Distribution of Gliotoxin-Producing *Aspergillus* Isolates Recovered from Patients at a Tertiary-Care Cancer Center. *Journal of Clinical Microbiology*. 432005. p. 6120-2.
180. Sulyok M, Krska R, Schuhmacher R. A liquid chromatography/tandem mass spectrometric multi-mycotoxin method for the quantification of 87 analytes and its application to semi-quantitative screening of moldy food samples. *Analytical and bioanalytical chemistry*. 2007;389(5):1505-23.
181. Vishwanath V, Sulyok M, Labuda R, Bicker W, Krska R. Simultaneous determination of 186 fungal and bacterial metabolites in indoor matrices by liquid

- chromatography/tandem mass spectrometry. *Analytical and bioanalytical chemistry*. 2009;395(5):1355-72.
182. Müllbacher A, Waring P, Tiwari-Palni U, Eichner RD. Structural relationship of epipolythiodioxopiperazines and their immunomodulating activity. *Molecular immunology*. 1986;23(2):231-5.
183. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet (London, England)*. 2020;395(10223):507-13.
184. Tan W, Zhao X, Ma X, Wang W, Niu P, Xu W, et al. A Novel Coronavirus Genome Identified in a Cluster of Pneumonia Cases - Wuhan, China 2019-2020. *China CDC weekly*. 2020;2(4):61-2.
185. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. Addendum: A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;588(7836):E6.
186. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature microbiology*. 2020;5(4):536-44.
187. Lam TT, Jia N, Zhang YW, Shum MH, Jiang JF, Zhu HC, et al. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature*. 2020;583(7815):282-5.
188. Chan JF, Yuan S, Kok KH, To KK, Chu H, Yang J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet (London, England)*. 2020;395(10223):514-23.
189. Rothe C, Schunk M, Sothmann P, Bretzel G, Froeschl G, Wallrauch C, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. *The New England journal of medicine*. 2020;382(10):970-1.
190. WHO. (2020). WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020.
191. Momtazmanesh S, Ochs HD, Uddin LQ, Perc M, Routes JM, Vieira DN, et al. All together to Fight COVID-19. *The American journal of tropical medicine and hygiene*. 2020;102(6):1181-3.
192. Gobierno de España (2020). Plan de desescalada.
193. Kyriakidis NC, López-Cortés A, González EV, Grimaldos AB, Prado EO. SARS-CoV-2 vaccines strategies: a comprehensive review of phase 3 candidates. *NPJ vaccines*. 2021;6(1):28.
194. Watson OJ, Barnsley G, Toor J, Hogan AB, Winskill P, Ghani AC. Global impact of the first year of COVID-19 vaccination: a mathematical modelling study. *The Lancet Infectious diseases*. 2022;22(9):1293-302.
195. Chen Z, Zheng W, Wu Q, Chen X, Peng C, Tian Y, et al. Global diversity of policy, coverage, and demand of COVID-19 vaccines: a descriptive study. *BMC medicine*. 2022;20(1):130.
196. WHO. (2024). WHO Coronavirus (COVID-19) Dashboard.
197. Lucas C, Wong P, Klein J, Castro TBR, Silva J, Sundaram M, et al. Longitudinal analyses reveal immunological misfiring in severe COVID-19. *Nature*. 2020;584(7821):463-9.
198. Medzhitov R. Origin and physiological roles of inflammation. *Nature*. 2008;454(7203):428-35.
199. Liu J, Cao X. Cellular and molecular regulation of innate inflammatory responses. *Cellular & molecular immunology*. 2016;13(6):711-21.
200. Chen L, Deng H, Cui H, Fang J, Zuo Z, Deng J, et al. Inflammatory responses and inflammation-associated diseases in organs. *Oncotarget*. 2018;9(6):7204-18.
201. Land WG. Role of DAMPs in respiratory virus-induced acute respiratory distress syndrome-with a preliminary reference to SARS-CoV-2 pneumonia. *Genes and immunity*. 2021;22(3):141-60.

202. Pober JS, Sessa WC. Evolving functions of endothelial cells in inflammation. *Nature reviews Immunology*. 2007;7(10):803-15.
203. Abdulkhaleq LA, Assi MA, Abdullah R, Zamri-Saad M, Taufiq-Yap YH, Hezmee MNM. The crucial roles of inflammatory mediators in inflammation: A review. *Veterinary world*. 2018;11(5):627-35.
204. Nathan C. Points of control in inflammation. *Nature*. 2002;420(6917):846-52.
205. Serhan CN, Savill J. Resolution of inflammation: the beginning programs the end. *Nature immunology*. 2005;6(12):1191-7.
206. Herold S, Mayer K, Lohmeyer J. Acute lung injury: how macrophages orchestrate resolution of inflammation and tissue repair. *Frontiers in Immunology*. 2011;2:65.
207. Wang J, Yang X, Li Y, Huang J-a, Jiang J, Su N. Specific cytokines in the inflammatory cytokine storm of patients with COVID-19-associated acute respiratory distress syndrome and extrapulmonary multiple-organ dysfunction. *Virology Journal*. 2021;18(1):117.
208. Sehgal T, Gupta N, Kohli S, Khurana A, Dass J, Diwan S, et al. A Prospective Study of Specialized Coagulation Parameters in Admitted COVID-19 Patients and Their Correlation With Acute Respiratory Distress Syndrome and Outcome. *Cureus*. 2021;13(8):e17463.
209. Junqueira C, Crespo Â, Ranjbar S, de Lacerda LB, Lewandowski M, Ingber J, et al. FcγR-mediated SARS-CoV-2 infection of monocytes activates inflammation. *Nature*. 2022;606(7914):576-84.
210. Bohn MK, Lippi G, Horvath A, Sethi S, Koch D, Ferrari M, et al. Molecular, serological, and biochemical diagnosis and monitoring of COVID-19: IFCC taskforce evaluation of the latest evidence. *Clinical chemistry and laboratory medicine*. 2020;58(7):1037-52.
211. Moore JB, June CH. Cytokine release syndrome in severe COVID-19. *Science*. 2020;368(6490):473-4.
212. Chen J, Subbarao K. The Immunobiology of SARS. *Annual Review of Immunology*. 2007;25(1):443-72.
213. Montazersaheb S, Hosseiniyan Khatibi SM, Hejazi MS, Tarhriz V, Farjami A, Ghasemian Sorbeni F, et al. COVID-19 infection: an overview on cytokine storm and related interventions. *Virology Journal*. 2022;19(1):92.
214. Fajgenbaum DC, June CH. Cytokine Storm. *The New England journal of medicine*. 2020;383(23):2255-73.
215. Tisoncik JR, Korth MJ, Simmons CP, Farrar J, Martin TR, Katze MG. Into the eye of the cytokine storm. *Microbiology and molecular biology reviews : MMBR*. 2012;76(1):16-32.
216. Bartee E, McFadden G. Cytokine synergy: an underappreciated contributor to innate anti-viral immunity. *Cytokine*. 2013;63(3):237-40.
217. O'Shea JJ, Gadina M, Siegel RM. 9 - Cytokines and Cytokine Receptors. In: Rich RR, Fleisher TA, Shearer WT, Schroeder HW, Frew AJ, Weyand CM, editors. *Clinical Immunology (Fifth Edition)*. London: Elsevier; 2019. p. 127-55.e1.
218. Griffith JW, Sokol CL, Luster AD. Chemokines and chemokine receptors: positioning cells for host defense and immunity. *Annual Review of Immunology*. 2014;32:659-702.
219. Xiong Y, Liu Y, Cao L, Wang D, Guo M, Jiang A, et al. Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients. *Emerging microbes & infections*. 2020;9(1):761-70.
220. Zhang C, Wu Z, Li JW, Zhao H, Wang GQ. Cytokine release syndrome in severe COVID-19: interleukin-6 receptor antagonist tocilizumab may be the key to reduce mortality. *International journal of antimicrobial agents*. 2020;55(5):105954.
221. Nieto-Torres JL, DeDiego ML, Verdiá-Báguena C, Jimenez-Guardeño JM, Regla-Nava JA, Fernandez-Delgado R, et al. Severe acute respiratory syndrome coronavirus envelope protein ion channel activity promotes virus fitness and pathogenesis. *PLoS pathogens*. 2014;10(5):e1004077.

222. Zhou Y, Fu B, Zheng X, Wang D, Zhao C, Qi Y, et al. Pathogenic T-cells and inflammatory monocytes incite inflammatory storms in severe COVID-19 patients. *National Science Review*. 2020;7(6):998-1002.
223. Hultström M, Fromell K, Larsson A, Persson B, Nilsson B, Quaggin SE, et al. Angiopoietin-2 Inhibition of Thrombomodulin-Mediated Anticoagulation-A Novel Mechanism That May Contribute to Hypercoagulation in Critically Ill COVID-19 Patients. *Biomedicines*. 2022;10(6).
224. Goshua G, Pine AB, Meizlish ML, Chang CH, Zhang H, Bahel P, et al. Endotheliopathy in COVID-19-associated coagulopathy: evidence from a single-centre, cross-sectional study. *The Lancet Haematology*. 2020;7(8):e575-e82.
225. Escher R, Breakey N, Lämmle B. Severe COVID-19 infection associated with endothelial activation. *Thrombosis research*. 2020;190:62.
226. Birnhuber A, Fließer E, Gorkiewicz G, Zacharias M, Seeliger B, David S, et al. Between inflammation and thrombosis: endothelial cells in COVID-19. *The European respiratory journal*. 2021;58(3).
227. Singh V, Kaur R, Kumari P, Pasricha C, Singh R. ICAM-1 and VCAM-1: Gatekeepers in various inflammatory and cardiovascular disorders. *Clinica chimica acta; international journal of clinical chemistry*. 2023;548:117487.
228. Kim IS, Lee SG, Shin SG, Jeong H, Sohn KM, Park KS, et al. Dysregulated thrombospondin 1 and miRNA-29a-3p in severe COVID-19. *Scientific reports*. 2022;12(1):21227.
229. Sachetto ATA, Mackman N. Tissue Factor and COVID-19: An Update. *Current drug targets*. 2022;23(17):1573-7.
230. Subramaniam S, Kothari H, Bosmann M. Tissue factor in COVID-19-associated coagulopathy. *Thrombosis research*. 2022;220:35-47.
231. Won T, Wood MK, Hughes DM, Talor MV, Ma Z, Schneider J, et al. Endothelial thrombomodulin downregulation caused by hypoxia contributes to severe infiltration and coagulopathy in COVID-19 patient lungs. *EBioMedicine*. 2022;75:103812.
232. Voskoboinik I, Whisstock JC, Trapani JA. Perforin and granzymes: function, dysfunction and human pathology. *Nature reviews Immunology*. 2015;15(6):388-400.
233. Bovenschen N, Kummer JA. Orphan granzymes find a home. *Immunological reviews*. 2010;235(1):117-27.
234. Andrade F. Non-cytotoxic antiviral activities of granzymes in the context of the immune antiviral state. *Immunological reviews*. 2010;235(1):128-46.
235. Richardson KC, Jung K, Pardo J, Turner CT, Granville DJ. Noncytotoxic Roles of Granzymes in Health and Disease. *Physiology (Bethesda, Md)*. 2022;37(6):323-48.
236. Arias M, Martínez-Lostao L, Santiago L, Ferrandez A, Granville DJ, Pardo J. The Untold Story of Granzymes in Oncoimmunology: Novel Opportunities with Old Acquaintances. *Trends in cancer*. 2017;3(6):407-22.
237. Metkar SS, Menea C, Pardo J, Wang B, Wallich R, Freudenberg M, et al. Human and mouse granzyme A induce a proinflammatory cytokine response. *Immunity*. 2008;29(5):720-33.
238. Santiago L, Castro M, Sanz-Pamplona R, Garzón M, Ramirez-Labrada A, Tapia E, et al. Extracellular Granzyme A Promotes Colorectal Cancer Development by Enhancing Gut Inflammation. *Cell reports*. 2020;32(1):107847.
239. Santiago L, Menea C, Arias M, Martin P, Jaime-Sánchez P, Metkar S, et al. Granzyme A Contributes to Inflammatory Arthritis in Mice Through Stimulation of Osteoclastogenesis. *Arthritis & rheumatology (Hoboken, NJ)*. 2017;69(2):320-34.
240. Bem RA, Bos AP, Bots M, Wolbink AM, van Ham SM, Medema JP, et al. Activation of the granzyme pathway in children with severe respiratory syncytial virus infection. *Pediatric research*. 2008;63(6):650-5.

241. Spaeny-Dekking EH, Hanna WL, Wolbink AM, Wever PC, Kummer JA, Swaak AJ, et al. Extracellular granzymes A and B in humans: detection of native species during CTL responses in vitro and in vivo. *Journal of immunology (Baltimore, Md : 1950)*. 1998;160(7):3610-6.
242. van Daalen KR, Reijneveld JF, Bovenschen N. Modulation of Inflammation by Extracellular Granzyme A. *Frontiers in Immunology*. 2020;11:931.
243. Jiang Y, Wei X, Guan J, Qin S, Wang Z, Lu H, et al. COVID-19 pneumonia: CD8(+) T and NK cells are decreased in number but compensatory increased in cytotoxic potential. *Clinical immunology (Orlando, Fla)*. 2020;218:108516.
244. Malengier-Devlies B, Filtjens J, Ahmadzadeh K, Boeckx B, Vandenhaute J, De Visscher A, et al. Severe COVID-19 patients display hyper-activated NK cells and NK cell-platelet aggregates. *Frontiers in Immunology*. 2022;13:861251.
245. Mazzoni A, Salvati L, Maggi L, Capone M, Vanni A, Spinicci M, et al. Impaired immune cell cytotoxicity in severe COVID-19 is IL-6 dependent. *The Journal of clinical investigation*. 2020;130(9):4694-703.
246. Li M, Guo W, Dong Y, Wang X, Dai D, Liu X, et al. Elevated Exhaustion Levels of NK and CD8(+) T Cells as Indicators for Progression and Prognosis of COVID-19 Disease. *Frontiers in Immunology*. 2020;11:580237.
247. Witkowski M, Tizian C, Ferreira-Gomes M, Niemeyer D, Jones TC, Heinrich F, et al. Untimely TGFβ responses in COVID-19 limit antiviral functions of NK cells. *Nature*. 2021;600(7888):295-301.
248. Zenarruzabeitia O, Astarloa-Pando G, Terren I, Orrantia A, Perez-Garay R, Seijas-Betolaza I, et al. T Cell Activation, Highly Armed Cytotoxic Cells and a Shift in Monocytes CD300 Receptors Expression Is Characteristic of Patients With Severe COVID-19. *Frontiers in Immunology*. 2021;12:655934.
249. Mitsuyama Y, Yamakawa K, Kayano K, Maruyama M, Umemura Y, Wada T, et al. Residual persistence of cytotoxicity lymphocytes and regulatory T cells in patients with severe coronavirus disease 2019 over a 1-year recovery process. *Acute medicine & surgery*. 2022;9(1):e803.
250. Zheng M, Gao Y, Wang G, Song G, Liu S, Sun D, et al. Functional exhaustion of antiviral lymphocytes in COVID-19 patients. *Cellular & molecular immunology*. 2020;17(5):533-5.
251. Ramljak D, Vukoja M, Curlin M, Vukojevic K, Barbaric M, Glamoclija U, et al. Early Response of CD8+ T Cells in COVID-19 Patients. *Journal of personalized medicine*. 2021;11(12).
252. Elemam NM, Hammoudeh S, Salameh L, Mahboub B, Alsafar H, Talaat IM, et al. Identifying Immunological and Clinical Predictors of COVID-19 Severity and Sequelae by Mathematical Modeling. *Frontiers in Immunology*. 2022;13:865845.
253. Kang CK, Han GC, Kim M, Kim G, Shin HM, Song KH, et al. Aberrant hyperactivation of cytotoxic T-cell as a potential determinant of COVID-19 severity. *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases*. 2020;97:313-21.
254. Schultze JL, Aschenbrenner AC. COVID-19 and the human innate immune system. *Cell*. 2021;184(7):1671-92.
255. Schrikker AC, de Vries WR, Zwart A, Luijendijk SC. The excretion of highly soluble gases by the lung in man. *Pflugers Archiv : European journal of physiology*. 1989;415(2):214-9.
256. Shirasu M, Touhara K. The scent of disease: volatile organic compounds of the human body related to disease and disorder. *The Journal of Biochemistry*. 2011;150(3):257-66.

257. Amann A, Costello Bde L, Miekisch W, Schubert J, Buszewski B, Pleil J, et al. The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva. *Journal of breath research*. 2014;8(3):034001.
258. Phillips M, Gleeson K, Hughes JM, Greenberg J, Cataneo RN, Baker L, et al. Volatile organic compounds in breath as markers of lung cancer: a cross-sectional study. *Lancet (London, England)*. 1999;353(9168):1930-3.
259. Gordon SM, Szidon JP, Krotoszynski BK, Gibbons RD, O'Neill HJ. Volatile organic compounds in exhaled air from patients with lung cancer. *Clinical chemistry*. 1985;31(8):1278-82.
260. Schnabel R, Fijten R, Smolinska A, Dallinga J, Boumans ML, Stobberingh E, et al. Analysis of volatile organic compounds in exhaled breath to diagnose ventilator-associated pneumonia. *Scientific reports*. 2015;5:17179.
261. Lichtenstein M, Turjerman S, Pinto JM, Barash O, Koren O. Pathophysiology of SARS-CoV-2 Infection in the Upper Respiratory Tract and Its Relation to Breath Volatile Organic Compounds. *mSystems*. 2021;6(4):e0010421.
262. Ratiu IA, Ligor T, Monedeiro F, Milanowski M, Rudnicka J, Buszewski B. Volatile Organic Compounds Emitted by Biological Matrices. In: Buszewski B, Baranowska I, editors. *Handbook of Bioanalytics*. Cham: Springer International Publishing; 2020. p. 1-17.
263. Ratiu IA, Bocos-Bintintan V, Turner M, Moll V-H, Thomas Lawrence Paul C. Discrimination of Chemical Profiles of Some Bacterial Species by Analyzing Culture Headspace Air Samples Using TD-GC/MS. *Current Analytical Chemistry*. 2014;10(4):488-97.
264. Ratiu IA, Ligor T, Bocos-Bintintan V, Al-Suod H, Kowalkowski T, Rafinska K, et al. The effect of growth medium on an *Escherichia coli* pathway mirrored into GC/MS profiles. *Journal of breath research*. 2017;11(3):036012.
265. Furuhashi T, Ishii R, Onishi H, Ota S. Elucidation of Biochemical Pathways Underlying VOCs Production in A549 Cells. *Frontiers in Molecular Biosciences*. 2020;7:116.
266. Silva CL, Perestrelo R, Silva P, Tomas H, Camara JS. Volatile metabolomic signature of human breast cancer cell lines. *Scientific reports*. 2017;7(1):43969.
267. Barash O, Zhang W, Halpern JM, Hua Q-L, Pan Y-Y, Kayal H, et al. Differentiation between genetic mutations of breast cancer by breath volatolomics. *Oncotarget*. 2015;6(42):44864.
268. Hubbard HF, Sobus JR, Pleil JD, Madden MC, Tabucchi S. Application of novel method to measure endogenous VOCs in exhaled breath condensate before and after exposure to diesel exhaust. *Journal of Chromatography B*. 2009;877(29):3652-8.
269. Remy R, Kemnitz N, Trefz P, Fuchs P, Bartels J, Klemenz AC, et al. Profiling of exhaled volatile organics in the screening scenario of a COVID-19 test center. *iScience*. 2022;25(10):105195.
270. Leopold JH, Philipp A, Bein T, Redel A, Gruber M, Schultz MJ, et al. Volatile organic compound profiles in outlet air from extracorporeal life-support devices differ from breath profiles in critically ill patients. *ERJ open research*. 2019;5(2).
271. Crowell SR, Smith JN, Creim JA, Faber W, Teeguarden JG. Physiologically based pharmacokinetic modeling of ethyl acetate and ethanol in rodents and humans. *Regulatory toxicology and pharmacology : RTP*. 2015;73(1):452-62.
272. Wishart DS, Guo A, Oler E, Wang F, Anjum A, Peters H, et al. HMDB 5.0: the Human Metabolome Database for 2022. *Nucleic acids research*. 2022;50(D1):D622-d31.
273. Buszewski B, Ligor T, Jezierski T, Wenda-Piesik A, Walczak M, Rudnicka J. Identification of volatile lung cancer markers by gas chromatography-mass spectrometry: comparison with discrimination by canines. *Analytical and bioanalytical chemistry*. 2012;404(1):141-6.

Bibliografía

274. Xu Zq, Broza YY, Ionsecu R, Tisch U, Ding L, Liu H, et al. A nanomaterial-based breath test for distinguishing gastric cancer from benign gastric conditions. *British journal of cancer*. 2013;108(4):941-50.
275. Kumar S, Mohan A, Guleria R. Biomarkers in cancer screening, research and detection: present and future: a review. *Biomarkers : biochemical indicators of exposure, response, and susceptibility to chemicals*. 2006;11(5):385-405.
276. Murray RK, Granner DK, Mayes PA, Rodwell VW. *Harper's illustrated biochemistry*. 2003.
277. Oxner M, Trang A, Mehta J, Forsyth C, Swanson B, Keshavarzian A, et al. The Versatility and Diagnostic Potential of VOC Profiling for Noninfectious Diseases. *BME frontiers*. 2023;4:0002.
278. Oyerinde AS, Selvaraju V, Babu JR, Geetha T. Potential Role of Oxidative Stress in the Production of Volatile Organic Compounds in Obesity. *Antioxidants (Basel, Switzerland)*. 2023;12(1).
279. Ahmed WM, Lawal O, Nijssen TM, Goodacre R, Fowler SJ. Exhaled Volatile Organic Compounds of Infection: A Systematic Review. *ACS infectious diseases*. 2017;3(10):695-710.
280. Belizário JE, Faintuch J, Malpartida MG. Breath Biopsy and Discovery of Exclusive Volatile Organic Compounds for Diagnosis of Infectious Diseases. *Frontiers in cellular and infection microbiology*. 2020;10:564194.
281. Kalapos MP. On the mammalian acetone metabolism: from chemistry to clinical implications. *Biochimica et Biophysica Acta - General Subjects*. 2003;1621(2):122-39.
282. U.S. Department of Health and Human Services. Hazardous Substances Data Bank (HSDB, online database). National Toxicology Information Program, National Library of Medicine, Bethesda, MD. 1993.
283. Sola-Martínez RA, Sanchez-Solis M, Lozano-Terol G, Gallego-Jara J, García-Marcos L, Cánovas Díaz M, et al. Relationship between lung function and exhaled volatile organic compounds in healthy infants. *Pediatric pulmonology*. 2022;57(5):1282-92.
284. Phillips M, Cataneo RN, Greenberg J, Grodman R, Salazar M. Breath markers of oxidative stress in patients with unstable angina. *Heart disease (Hagerstown, Md)*. 2003;5(2):95-9.
285. Muller-Decker K, Amtmann E, Sauer G. Inhibition of the phosphorylation of the regulatory non-structural protein of vesicular stomatitis virus by an antiviral xanthate compound. *The Journal of general virology*. 1987;68 (Pt 12):3045-56.
286. Kamal F, Kumar S, Edwards MR, Veselkov K, Belluomo I, Kebabze T, et al. Virus-induced Volatile Organic Compounds Are Detectable in Exhaled Breath during Pulmonary Infection. *American journal of respiratory and critical care medicine*. 2021;204(9):1075-85.
287. Peled N, Barash O, Tisch U, Ionescu R, Broza YY, Ilouze M, et al. Volatile fingerprints of cancer specific genetic mutations. *Nanomedicine : nanotechnology, biology, and medicine*. 2013;9(6):758-66.
288. Ratcliffe N, Wiczorek T, Drabinska N, Gould O, Osborne A, De Lacy Costello B. A mechanistic study and review of volatile products from peroxidation of unsaturated fatty acids: an aid to understanding the origins of volatile organic compounds from the human body. *Journal of breath research*. 2020;14(3):034001.
289. La JA, Lim S, Park HJ, Heo MJ, Sang BI, Oh MK, et al. Plasmonic-based colorimetric and spectroscopic discrimination of acetic and butyric acids produced by different types of *Escherichia coli* through the different assembly structures formation of gold nanoparticles. *Analytica chimica acta*. 2016;933:196-206.
290. Cazzola M, Segreti A, Capuano R, Bergamini A, Martinelli E, Calzetta L, et al. Analysis of exhaled breath fingerprints and volatile organic compounds in COPD. *COPD Research and Practice*. 2015;1(1):7.

291. Cho KS, Lim YR, Lee K, Lee J, Lee JH, Lee IS. Terpenes from Forests and Human Health. *Toxicological research*. 2017;33(2):97-106.
292. Masyita A, Mustika Sari R, Dwi Astuti A, Yasir B, Rahma Rumata N, Emran TB, et al. Terpenes and terpenoids as main bioactive compounds of essential oils, their roles in human health and potential application as natural food preservatives. *Food chemistry: X*. 2022;13:100217.
293. Haick H, Broza YY, Mochalski P, Ruzsanyi V, Amann A. Assessment, origin, and implementation of breath volatile cancer markers. *Chemical Society reviews*. 2014;43(5):1423-49.
294. Miekisch W, Schubert JK, Noeldge-Schomburg GF. Diagnostic potential of breath analysis--focus on volatile organic compounds. *Clinica chimica acta; international journal of clinical chemistry*. 2004;347(1-2):25-39.
295. Boots AW, van Berkel JJ, Dallinga JW, Smolinska A, Wouters EF, van Schooten FJ. The versatile use of exhaled volatile organic compounds in human health and disease. *Journal of breath research*. 2012;6(2):027108.
296. Bos LD, Sterk PJ, Schultz MJ. Volatile metabolites of pathogens: a systematic review. *PLoS pathogens*. 2013;9(5):e1003311.
297. Hussain I, Cher GLY, Abid MA, Abid MB. Role of gut microbiome in COVID-19: an insight into pathogenesis and therapeutic potential. *Frontiers in immunology*. 2021;12:765965.
298. Giron LB, Dweep H, Yin X, Wang H, Damra M, Goldman AR, et al. Plasma markers of disrupted gut permeability in severe COVID-19 patients. *Frontiers in immunology*. 2021;12:686240.
299. Berna AZ, Odom John AR. Breath Metabolites to Diagnose Infection. *Clinical chemistry*. 2021;68(1):43-51.
300. Traxler S, Bischoff AC, Saß R, Trefz P, Gierschner P, Brock B, et al. VOC breath profile in spontaneously breathing awake swine during Influenza A infection. *Scientific reports*. 2018;8(1).
301. Broza YY, Zhou X, Yuan M, Qu D, Zheng Y, Vishinkin R, et al. Disease Detection with Molecular Biomarkers: From Chemistry of Body Fluids to Nature-Inspired Chemical Sensors. *Chemical reviews*. 2019;119(22):11761-817.
302. van Keulen KE, Jansen ME, Schrauwen RWM, Kolkman JJ, Siersema PD. Volatile organic compounds in breath can serve as a non-invasive diagnostic biomarker for the detection of advanced adenomas and colorectal cancer. *Alimentary pharmacology & therapeutics*. 2020;51(3):334-46.
303. Wilson AD, Forse LB. Potential for Early Noninvasive COVID-19 Detection Using Electronic-Nose Technologies and Disease-Specific VOC Metabolic Biomarkers. *Sensors (Basel, Switzerland)*. 2023;23(6).
304. Liangou A, Tasoglou A, Huber HJ, Wistrom C, Brody K, Menon PG, et al. A method for the identification of COVID-19 biomarkers in human breath using Proton Transfer Reaction Time-of-Flight Mass Spectrometry. *EclinicalMedicine*. 2021;42:101207.
305. Laird S, Debenham L, Chandla D, Chan C, Daulton E, Taylor J, et al. Breath Analysis of COVID-19 Patients in a Tertiary UK Hospital by Optical Spectrometry: The E-Nose CoVal Study. *Biosensors*. 2023;13(2).
306. Tsou PH, Lin ZL, Pan YC, Yang HC, Chang CJ, Liang SK, et al. Exploring Volatile Organic Compounds in Breath for High-Accuracy Prediction of Lung Cancer. *Cancers*. 2021;13(6).
307. Solga SF, Alkhuraishe A, Cope K, Tabesh A, Clark JM, Torbenson M, et al. Breath biomarkers and non-alcoholic fatty liver disease: preliminary observations. *Biomarkers : biochemical indicators of exposure, response, and susceptibility to chemicals*. 2006;11(2):174-83.

308. Di Gilio A, Palmisani J, Picciariello A, Zambonin C, Aresta A, De Vietro N, et al. Identification of a characteristic VOCs pattern in the exhaled breath of post-COVID subjects: are metabolic alterations induced by the infection still detectable? *Journal of breath research*. 2023;17(4).
309. Subali AD, Wiyono L, Yusuf M, Zaky MFA. The potential of volatile organic compounds-based breath analysis for COVID-19 screening: a systematic review & meta-analysis. *Diagnostic microbiology and infectious disease*. 2022;102(2):115589.
310. Chen H, Qi X, Ma J, Zhang C, Feng H, Yao M. Breath-borne VOC Biomarkers for COVID-19. *Journal of breath research*. 2020;2020.06.21.20136523.
311. Myers R, Ruszkiewicz DM, Meister A, Bartolomeu C, Atkar-Khattra S, Thomas CLP, et al. Breath testing for SARS-CoV-2 infection. *EBioMedicine*. 2023;92:104584.
312. Gould O, Ratcliffe N, Krol E, de Lacy Costello B. Breath analysis for detection of viral infection, the current position of the field. *Journal of breath research*. 2020;14(4):041001.
313. Ruszkiewicz DM, Sanders D, O'Brien R, Hempel F, Reed MJ, Riepe AC, et al. Diagnosis of COVID-19 by analysis of breath with gas chromatography-ion mobility spectrometry - a feasibility study. *EClinicalMedicine*. 2020;29:100609.
314. Martinez-Lozano Sinues P, Meier L, Berchtold C, Ivanov M, Sievi N, Camen G, et al. Breath analysis in real time by mass spectrometry in chronic obstructive pulmonary disease. *Respiration; international review of thoracic diseases*. 2014;87(4):301-10.
315. Chen H, Qi X, Zhang L, Li X, Ma J, Zhang C, et al. COVID-19 screening using breath-borne volatile organic compounds. *Journal of breath research*. 2021;15(4).
316. Bajtarevic A, Ager C, Pienz M, Klieber M, Schwarz K, Ligor M, et al. Noninvasive detection of lung cancer by analysis of exhaled breath. *BMC cancer*. 2009;9:348.
317. Nurputra DK, Kusumaatmaja A, Hakim MS, Hidayat SN, Julian T, Sumanto B, et al. Fast and noninvasive electronic nose for sniffing out COVID-19 based on exhaled breath-print recognition. *NPJ digital medicine*. 2022;5(1):115.
318. Schleich FN, Zanella D, Stefanuto PH, Bessonov K, Smolinska A, Dallinga JW, et al. Exhaled Volatile Organic Compounds Are Able to Discriminate between Neutrophilic and Eosinophilic Asthma. *American journal of respiratory and critical care medicine*. 2019;200(4):444-53.
319. Smolinska A, Tedjo DI, Blanchet L, Bodelier A, Pierik MJ, Masclee AAM, et al. Volatile metabolites in breath strongly correlate with gut microbiome in CD patients. *Analytica chimica acta*. 2018;1025:1-11.
320. Sharma R, Zang W, Tabartehfarahani A, Lam A, Huang X, Sivakumar AD, et al. Portable Breath-Based Volatile Organic Compound Monitoring for the Detection of COVID-19 During the Circulation of the SARS-CoV-2 Delta Variant and the Transition to the SARS-CoV-2 Omicron Variant. *JAMA network open*. 2023;6(2):e230982.
321. Bos LD, Weda H, Wang Y, Knobel HH, Nijsen TM, Vink TJ, et al. Exhaled breath metabolomics as a noninvasive diagnostic tool for acute respiratory distress syndrome. *The European respiratory journal*. 2014;44(1):188-97.
322. Berna AZ, Akaho EH, Harris RM, Congdon M, Korn E, Neher S, et al. Reproducible Breath Metabolite Changes in Children with SARS-CoV-2 Infection. *ACS infectious diseases*. 2021;7(9):2596-603.
323. Xue C, Xu X, Liu Z, Zhang Y, Xu Y, Niu J, et al. Intelligent COVID-19 screening platform based on breath analysis. *Journal of breath research*. 2022;17(1).
324. Caldeira M, Perestrelo R, Barros AS, Bilelo MJ, Morete A, Camara JS, et al. Allergic asthma exhaled breath metabolome: a challenge for comprehensive two-dimensional gas chromatography. *Journal of chromatography A*. 2012;1254:87-97.
325. Chen X, Xu F, Wang Y, Pan Y, Lu D, Wang P, et al. A study of the volatile organic compounds exhaled by lung cancer cells in vitro for breath diagnosis. *Cancer*. 2007;110(4):835-44.

326. Gashimova E, Temerdashev A, Porkhanov V, Polyakov I, Perunov D, Azaryan A, et al. Investigation of different approaches for exhaled breath and tumor tissue analyses to identify lung cancer biomarkers. *Heliyon*. 2020;6(6):e04224.
327. Zhang P, Ren T, Chen H, Li Q, He M, Feng Y, et al. A feasibility study of Covid-19 detection using breath analysis by high-pressure photon ionization time-of-flight mass spectrometry. *Journal of breath research*. 2022;16(4).
328. Li ZT, Zeng PY, Chen ZM, Guan WJ, Wang T, Lin Y, et al. Exhaled Volatile Organic Compounds for Identifying Patients With Chronic Pulmonary Aspergillosis. *Frontiers in medicine*. 2021;8:720119.
329. Lim SH, Martino R, Anikst V, Xu Z, Mix S, Benjamin R, et al. Rapid Diagnosis of Tuberculosis from Analysis of Urine Volatile Organic Compounds. *ACS Sensors Journal*. 2016;1(7):852-6.
330. Smith D, Sovova K, Dryahina K, Dousova T, Drevinek P, Spanel P. Breath concentration of acetic acid vapour is elevated in patients with cystic fibrosis. *Journal of breath research*. 2016;10(2):021002.
331. Phillips CO, Syed Y, Parthala NM, Zwiggelaar R, Claypole TC, Lewis KE. Machine learning methods on exhaled volatile organic compounds for distinguishing COPD patients from healthy controls. *Journal of breath research*. 2012;6(3):036003.
332. Schaber CL, Katta N, Bollinger LB, Mwale M, Mlotha-Mitole R, Trehan I, et al. Breathprinting Reveals Malaria-Associated Biomarkers and Mosquito Attractants. *The Journal of infectious diseases*. 2018;217(10):1553-60.
333. Ogbodo JO, Arazu AV, Iguh TC, Onwodi NJ, Ezike TC. Volatile organic compounds: A proinflammatory activator in autoimmune diseases. *Frontiers in Immunology*. 2022;13:928379.
334. O'Hara ME, Fernandez Del Rio R, Holt A, Pemberton P, Shah T, Whitehouse T, et al. Limonene in exhaled breath is elevated in hepatic encephalopathy. *Journal of breath research*. 2016;10(4):046010.
335. van de Kant KD, van der Sande LJ, Jöbsis Q, van Schayck OC, Dompeling E. Clinical use of exhaled volatile organic compounds in pulmonary diseases: a systematic review. *Respiratory research*. 2012;13(1):117.
336. van Aerde KJ, Jansen R, Merkus PJ, van der Flier M. Breath Test: Clinical Application of Breath Analysis in Lower Respiratory Tract Infection Diagnosis. *The Pediatric infectious disease journal*. 2021;40(11):e434-e6.
337. Xie N-n, Zhang W-c, Chen J, Tian F-b, Song J-x. Clinical Characteristics, Diagnosis, and Therapeutics of COVID-19: A Review. *Current Medical Science*. 2023;43(6):1066-74.
338. Ghassemi N, Shoeibi A, Khodatars M, Heras J, Rahimi A, Zare A, et al. Automatic diagnosis of COVID-19 from CT images using CycleGAN and transfer learning. *Applied soft computing*. 2023;144:110511.
339. Meng Z, Guo S, Zhou Y, Li M, Wang M, Ying B. Applications of laboratory findings in the prevention, diagnosis, treatment, and monitoring of COVID-19. *Signal Transduction and Targeted Therapy*. 2021;6(1):316.
340. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet (London, England)*. 2020;395(10223):497-506.
341. Wiersinga WJ, Rhodes A, Cheng AC, Peacock SJ, Prescott HC. Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review. *Jama*. 2020;324(8):782-93.
342. Elfiky AA. Ribavirin, Remdesivir, Sofosbuvir, Galidesivir, and Tenofovir against SARS-CoV-2 RNA dependent RNA polymerase (RdRp): A molecular docking study. *Life sciences*. 2020;253:117592.

343. Shyu D, Dorroh J, Holtmeyer C, Ritter D, Upendran A, Kannan R, et al. Laboratory Tests for COVID-19: A Review of Peer-Reviewed Publications and Implications for Clinical Use. *Missouri medicine*. 2020;117(3):184-95.
344. Ascierto PA, Fox BA, Urba WJ, Anderson AC, Atkins MB, Borden EC, et al. Insights from immuno-oncology: the Society for Immunotherapy of Cancer Statement on access to IL-6-targeting therapies for COVID-19. *Journal for immunotherapy of cancer*. 2020;8(1).
345. Jacofsky D, Jacofsky EM, Jacofsky M. Understanding Antibody Testing for COVID-19. *The Journal of arthroplasty*. 2020;35(7S):S74-S81.
346. Dutta NK, Mazumdar K, Gordy JT. The Nucleocapsid Protein of SARS-CoV-2: a Target for Vaccine Development. *Journal of virology*. 2020;94(13).
347. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Velesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*. 2020;181(2):281-92.e6.
348. Augustine R, Das S, Hasan A, SA, Abdul Salam S, Augustine P, et al. Rapid Antibody-Based COVID-19 Mass Surveillance: Relevance, Challenges, and Prospects in a Pandemic and Post-Pandemic World. *Journal of clinical medicine*. 2020;9(10).
349. Prazuck T, Colin M, Giachè S, Gubavu C, Seve A, Rzepecki V, et al. Evaluation of performance of two SARS-CoV-2 Rapid IgM-IgG combined antibody tests on capillary whole blood samples from the fingertip. *PLoS one*. 2020;15(9):e0237694.
350. Weissleder R, Lee H, Ko J, Pittet MJ. COVID-19 diagnostics in context. *Science translational medicine*. 2020;12(546).
351. Ge H, Wang X, Yuan X, Xiao G, Wang C, Deng T, et al. The epidemiology and clinical information about COVID-19. *European journal of clinical microbiology & infectious diseases* : official publication of the European Society of Clinical Microbiology. 2020;39(6):1011-9.
352. Zhou M, Zhang X, Qu J. Coronavirus disease 2019 (COVID-19): a clinical update. *Frontiers of medicine*. 2020;14(2):126-35.
353. Bellou V, Tzoulaki I, van Smeden M, Moons KGM, Evangelou E, Belbasis L. Prognostic factors for adverse outcomes in patients with COVID-19: a field-wide systematic review and meta-analysis. *The European respiratory journal*. 2022;59(2).
354. World Health Organization. (2023). Clinical management of COVID-19: Living guideline (2nd ed.). World Health Organization. [Disponible en <https://www.who.int/publications/i/item/WHO-2019-nCoV-clinical-2023.2>].
355. National Institutes of Health. Coronavirus disease 2019 (COVID-19) treatment guidelines. Feb 2024 [Disponible en: <https://www.ncbi.nlm.nih.gov/books/NBK570371/>].
356. Bhimraj A, Morgan RL, Hirsch Shumaker A, et al. Infectious Diseases Society of America guidelines on the treatment and management of patients with COVID-19. Jun 2023 [Disponible en: <https://www.idsociety.org/practice-guideline/covid-19-guideline-treatment-and-management>].
357. Documento técnico. Manejo del COVID-19: atención hospitalaria. Ministerio de Sanidad, Gobierno de España, Junio 2023. https://www.sanidad.gob.es/areas/alertasEmergenciasSanitarias/alertasActuales/nCov/documentos/Protocolo_manejo_clinico_ah_COVID-19.pdf.
358. Centro de Coordinación de Alertas y Emergencias Sanitarias. (2021). Información científica-técnica: Enfermedad por coronavirus, COVID-19, Actualización 28 de octubre de 2021. Ministerio de Sanidad, España. [Disponible en https://www.sanidad.gob.es/areas/alertasEmergenciasSanitarias/alertasActuales/nCov/documentos/20211028_CLINICA.pdf].
359. Bivona G, Agnello L, Ciaccio M. Biomarkers for Prognosis and Treatment Response in COVID-19 Patients. *Annals of laboratory medicine*. 2021;41(6):540-8.
360. Hasan MR, Suleiman M, Pérez-López A. Metabolomics in the Diagnosis and Prognosis of COVID-19. *Frontiers in genetics*. 2021;12:721556.

361. Shi C, Wang L, Ye J, Gu Z, Wang S, Xia J, et al. Predictors of mortality in patients with coronavirus disease 2019: a systematic review and meta-analysis. *BMC infectious diseases*. 2021;21(1):663.
362. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature*. 2020;584(7821):430-6.
363. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet*. 2020;395(10229):1054-62.
364. Özceylan G, Altuntaş SB. Relationship between initial symptoms and prognosis of patients with Covid-19. *Atencion primaria*. 2022;54(1):102146.
365. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. *The New England journal of medicine*. 2020;382(18):1708-20.
366. Ruan Q, Yang K, Wang W, Jiang L, Song J. Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. *Intensive care medicine*. 2020;46(5):846-8.
367. Merad M, Martin JC. Pathological inflammation in patients with COVID-19: a key role for monocytes and macrophages. *Nature Reviews Immunology*. 2020;20(6):355-62.
368. Li K, Wu J, Wu F, Guo D, Chen L, Fang Z, et al. The Clinical and Chest CT Features Associated With Severe and Critical COVID-19 Pneumonia. *Investigative radiology*. 2020;55(6):327-31.
369. Oi Y, Ogawa F, Yamashiro T, Matsushita S, Oguri A, Utada S, et al. Prediction of prognosis in patients with severe COVID-19 pneumonia using CT score by emergency physicians: a single-center retrospective study. *Scientific reports*. 2023;13(1):4045.
370. Pujadas E, Chaudhry F, McBride R, Richter F, Zhao S, Wajnberg A, et al. SARS-CoV-2 viral load predicts COVID-19 mortality. *The Lancet Respiratory medicine*. 2020;8(9):e70.
371. Mehta P, McAuley DF, Brown M, Sanchez E, Tattersall RS, Manson JJ. COVID-19: consider cytokine storm syndromes and immunosuppression. *Lancet (London, England)*. 2020;395(10229):1033-4.
372. Channappanavar R, Perlman S. Pathogenic human coronavirus infections: causes and consequences of cytokine storm and immunopathology. *Seminars in immunopathology*. 2017;39(5):529-39.
373. Zhang L, Huang B, Xia H, Fan H, Zhu M, Zhu L, et al. Retrospective analysis of clinical features in 134 coronavirus disease 2019 cases. *Epidemiology and infection*. 2020;148:e199.
374. Han H, Ma Q, Li C, Liu R, Zhao L, Wang W, et al. Profiling serum cytokines in COVID-19 patients reveals IL-6 and IL-10 are disease severity predictors. *Emerging microbes & infections*. 2020;9(1):1123-30.
375. Qin C, Zhou L, Hu Z, Zhang S, Yang S, Tao Y, et al. Dysregulation of Immune Response in Patients With Coronavirus 2019 (COVID-19) in Wuhan, China. *Clinical Infectious Diseases*. 2020;71(15):762-8.
376. Del Valle DM, Kim-Schulze S, Huang HH, Beckmann ND, Nirenberg S, Wang B, et al. An inflammatory cytokine signature predicts COVID-19 severity and survival. *Nature medicine*. 2020;26(10):1636-43.
377. Kesmez Can F, Özkurt Z, Öztürk N, Sezen S. Effect of IL-6, IL-8/CXCL8, IP-10/CXCL 10 levels on the severity in COVID 19 infection. *International journal of clinical practice*. 2021;75(12):e14970.
378. Coomes EA, Haghbayan H. Interleukin-6 in Covid-19: A systematic review and meta-analysis. *Reviews in medical virology*. 2020;30(6):1-9.
379. Liu T, Zhang J, Yang Y, Ma H, Li Z, Zhang J, et al. The role of interleukin-6 in monitoring severe case of coronavirus disease 2019. *EMBO molecular medicine*. 2020;12(7):e12421.

380. Zanza C, Romenskaya T, Manetti AC, Franceschi F, La Russa R, Bertozzi G, et al. Cytokine Storm in COVID-19: Immunopathogenesis and Therapy. *Medicina (Kaunas, Lithuania)*. 2022;58(2).
381. Taghiloo S, Soltanshahi M, Aliyali M, Abedi S, Mehravaran H, Ajami A, et al. Cytokine profiling in Iranian patients with COVID-19; association with clinical severity. *Iranian journal of immunology : IJI*. 2021;18(1):54-64.
382. Liu Y, Zhang C, Huang F, Yang Y, Wang F, Yuan J, et al. Elevated plasma levels of selective cytokines in COVID-19 patients reflect viral load and lung injury. *National Science Review*. 2020;7(6):1003-11.
383. Li L, Li J, Gao M, Fan H, Wang Y, Xu X, et al. Interleukin-8 as a Biomarker for Disease Prognosis of Coronavirus Disease-2019 Patients. *Frontiers in Immunology*. 2020;11:602395.
384. Smail SW, Babaei E, Amin K, Abdulahad WH. Serum IL-23, IL-10, and TNF-alpha predict in-hospital mortality in COVID-19 patients. *Frontiers in Immunology*. 2023;14:1145840.
385. Coperchini F, Chiovato L, Ricci G, Croce L, Magri F, Rotondi M. The cytokine storm in COVID-19: Further advances in our understanding the role of specific chemokines involved. *Cytokine & Growth Factor Reviews*. 2021;58:82-91.
386. Ikeagwulonu RC, Ugwu NI, Ezeonu CT, Ikeagwulonu ZC, Uro-Chukwu HC, Asiegbu UV, et al. C-Reactive Protein and Covid-19 Severity: A Systematic Review. *West African journal of medicine*. 2021;Vol. 38(10):1011-23.
387. Tang N, Bai H, Chen X, Gong J, Li D, Sun Z. Anticoagulant treatment is associated with decreased mortality in severe coronavirus disease 2019 patients with coagulopathy. *Journal of thrombosis and haemostasis : JTH*. 2020;18(5):1094-9.
388. Soraya GV, Ulhaq ZS. Crucial laboratory parameters in COVID-19 diagnosis and prognosis: An updated meta-analysis. *Medicina clinica*. 2020;155(4):143-51.
389. Laatifi M, Douzi S, Ezzine H, Asry CE, Naya A, Bouklouze A, et al. Explanatory predictive model for COVID-19 severity risk employing machine learning, shapley addition, and LIME. *Scientific reports*. 2023;13(1):5481.
390. Liang W, Liang H, Ou L, Chen B, Chen A, Li C, et al. Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients With COVID-19. *JAMA internal medicine*. 2020;180(8):1081-9.
391. Baker KF, Hanrath AT, Schim van der Loeff I, Kay LJ, Back J, Duncan CJ. National Early Warning Score 2 (NEWS2) to identify inpatient COVID-19 deterioration: a retrospective analysis. *Clinical medicine (London, England)*. 2021;21(2):84-9.
392. Williams B. Evaluation of the utility of NEWS2 during the COVID-19 pandemic. *Clinical medicine (London, England)*. 2022;22(6):539-43.
393. Iijima Y, Okamoto T, Shirai T, Mitsumura T, Sakakibara R, Honda T, et al. MuLBSTA score is a useful tool for predicting COVID-19 disease behavior. *Journal of infection and chemotherapy : official journal of the Japan Society of Chemotherapy*. 2021;27(2):284-90.
394. Carriel J, Muñoz-Jaramillo R, Bolaños-Ladinez O, Heredia-Villacreses F, Menéndez-Sanchón J, Martín-Delgado J. [CURB-65 as a predictor of 30-day mortality in patients hospitalized with COVID-19 in Ecuador: COVID-EC study]. *Revista clinica espanola*. 2022;222(1):37-41.
395. Chang SC, Grunkemeier GL, Goldman JD, Wang M, McKelvey PA, Hadlock J, et al. A simplified pneumonia severity index (PSI) for clinical outcome prediction in COVID-19. *PloS one*. 2024;19(5):e0303899.
396. Esmaeili Tarki F, Afaghi S, Rahimi FS, Kiani A, Varahram M, Abedini A. Serial SOFA-score trends in ICU-admitted COVID-19 patients as predictor of 28-day mortality: A prospective cohort study. *Health science reports*. 2023;6(5):e11116.

397. Yang Z, Hu Q, Huang F, Xiong S, Sun Y. The prognostic value of the SOFA score in patients with COVID-19: A retrospective, observational study. *Medicine*. 2021;100(32):e26900.
398. Khari S, Salimi Akin Abadi A, Pazokian M, Yousefifard M. CURB-65, qSOFA, and SIRS Criteria in Predicting In-Hospital Mortality of Critically Ill COVID-19 Patients; a Prognostic Accuracy Study. *Archives of academic emergency medicine*. 2022;10(1):e36.
399. Alencar J, Marina Gómez Gómez L, Cortez AL, Possolo de Souza H, Levin AS, Salomão MC. Performance of NEWS, qSOFA, and SIRS Scores for Assessing Mortality, Early Bacterial Infection, and Admission to ICU in COVID-19 Patients in the Emergency Department. *Frontiers in medicine*. 2022;9:779516.
400. Anurag A, Preetam M. Validation of PSI/PORT, CURB-65 and SCAP scoring system in COVID-19 pneumonia for prediction of disease severity and 14-day mortality. *The clinical respiratory journal*. 2021;15(5):467-71.
401. Satué Gracia EM, Vila Córcoles Á. [COVID-19 prognostic indicators in Primary Care]. *Atencion primaria*. 2022;54(7):102308.
402. Dhimi S, Thompson D, El Akoum M, Bates DW, Bertollini R, Sheikh A. Data-enabled responses to pandemics: policy lessons from COVID-19. *Nature medicine*. 2022;28(11):2243-6.
403. Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. *Science*. 2015;349(6245):255-60.
404. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. *BMC Medical Research Methodology*. 2019;19(1):64-.
405. Peiffer-Smadja N, Dellièrè S, Rodriguez C, Birgand G, Lescure FX, Fourati S, et al. Machine learning in the clinical microbiology laboratory: has the time come for routine practice? *Clinical Microbiology and Infection*. 2020;26(10):1300-9.
406. Hoque J, Hossen, Sayeed MS, Ho CK, Tawsif K, Rahman MA, et al. A Survey on Cleaning Dirty Data Using Machine Learning Paradigm for Big Data Analytics. *Indonesian Journal of Electrical Engineering and Computer Science*. 2018;10:1234-43.
407. Chen H, Ding J, Lu W, Bhuyan J. "Data Quality for Big Data and Machine Learning" *Frontiers in Big Data* 2022.
408. Pyle D. *Data Preparation for Data Mining*: Morgan Kaufmann Publishers Inc.; 1999.
409. Jiawei Han MK, Jian Pei. *Minería de datos: conceptos y técnicas*. Tercera edición ed: Morgan Kaufmann; 2011.
410. Venkatesh KA, Mohanasundaram K, Pothiyachi V. 8 - Regression tasks for machine learning. In: Goswami T, Sinha GR, editors. *Statistical Modeling in Machine Learning*: Academic Press; 2023. p. 133-57.
411. Simeone O. *A Brief Introduction to Machine Learning for Engineers*. *Foundations and Trends® in Signal Processing*. 2018;12(3-4):200-431.
412. Capozzoli A, Cerquitelli T, Piscitelli MS. Chapter 11 - Enhancing energy efficiency in buildings through innovative data analytics technologies. In: Dobre C, Xhafa F, editors. *Pervasive Computing*. Boston: Academic Press; 2016. p. 353-89.
413. Simeone O. *A Very Brief Introduction to Machine Learning With Applications to Communication Systems*. *IEEE Transactions on Cognitive Communications and Networking*. 2018;4(4):648-64.
414. Bishop CM. *Pattern Recognition and Machine Learning (Information Science and Statistics)*: Springer-Verlag; 2006.
415. Analytics Vidhya. (s.f.). *Logistic Regression*. [Available from: <https://www.analyticsvidhya.com/logistic-regression/>].
416. Singh, A. (2023, 24 de junio). *Python Train Model for Logistic Regression*. Vitalflux. [Available from: <https://vitalflux.com/python-train-model-logistic-regression/>].

Bibliografía

417. Voxco. (s.f.). Linear Regression vs Logistic Regression: Difference and Working. Voxco. [Available from: <https://www.voxco.com/blog/linear-regression-vs-logistic-regression-difference-and-working/>].
418. Manickam, D. (2020, 28 de febrero). An Introduction to Logistic Regression in Python with StatsModels and Scikit-Learn. LevelUp. [Available from: <https://levelup.gitconnected.com/an-introduction-to-logistic-regression-in-python-with-statsmodels-and-scikit-learn-1a1fb5ce1c13>].
419. Haga T, Okuno T, editors. Selection of variables in multiple regression analysis. Proceedings of the Third Japan — USSR Symposium on Probability Theory; 1976 1976//; Berlin, Heidelberg: Springer Berlin Heidelberg.
420. Hocking RR. A Biometrics Invited Paper. The Analysis and Selection of Variables in Linear Regression. Biometrics. 1976;32(1):1-49.
421. Statistics Solutions. (s.f.). Stepwise Regression: What is it and should you use it?. [Available from: <https://www.statisticssolutions.com/stepwise-regression-what-is-it-and-should-you-use-it/>].
422. Rokach L, Maimon O. Data mining with decision trees. Theory and applications. 2008.
423. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. Boca Raton, Florida: Chapman & Hall, CRC Press; 1984.
424. Quinlan JR. Induction of decision trees. Machine Learning. 1986;1(1):81-106.
425. Salzberg SL. C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993. Machine Learning. 1994;16(3):235-40.
426. Wu X, Kumar V, Quinlan R, Ghosh J, Yang Q, Motoda H, et al. Top 10 algorithms in data mining. Knowledge and Information Systems. 2007;14.
427. Lantz B. Machine Learning with R: Expert techniques for predictive modeling, 3rd Edition: Packt Publishing; 2019.
428. Fix E, Hodges JL. Discriminatory Analysis: Nonparametric Discrimination: Consistency Properties: USAF School of Aviation Medicine; 1951.
429. Cover T, Hart P. Nearest neighbor pattern classification. IEEE Transactions on Information Theory. 1967;13(1):21-7.
430. Walczak S, Cerpa N. Artificial Neural Networks. In: Meyers RA, editor. Encyclopedia of Physical Science and Technology (Third Edition). New York: Academic Press; 2003. p. 631-45.
431. UNIR. (n.d.). Redes neuronales artificiales: qué son, cómo funcionan y sus aplicaciones. UNIR. [Available from: <https://www.unir.net/ingenieria/revista/redes-neuronales-artificiales/>].
432. MIT News. (2017, April 14). Explained: Neural networks and deep learning. MIT News. [Available from: <https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414>].
433. Pérez Aguilar D, Risco R, Casaverde-Pacherrez L. Transfer learning en la clasificación binaria de imágenes térmicas. Ingenius. 2021:71-86.
434. IBM. (n.d.). Neural networks. IBM. [Available from: <https://www.ibm.com/topics/neural-networks>].
435. Devijver PAKJ. Pattern recognition : a statistical approach. Englewood Cliffs, N.J.: Prentice/Hall International; 1982.
436. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 2011.
437. Zhou Z-H. Ensemble Methods: Foundations and Algorithms. 1st ed: Chapman & Hall/CRC; 2012.
438. Opitz D, Maclin R. Popular Ensemble Methods: An Empirical Study. Journal of Artificial Intelligence Research. 1999;11:169-98.

439. Rokach L. Ensemble-based classifiers. *Artificial Intelligence Review*. 2010;33(1):1-39.
440. Polikar R. Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*. 2006;6(3):21-45.
441. Pirayonesi SM, El-Diraby TE. Role of Data Analytics in Infrastructure Asset Management: Overcoming Data Size and Quality Problems. 2020;146(2):04020022.
442. Hastie T, Tibshirani R, Friedman J. *The elements of statistical learning : data mining, inference, and prediction : with 200 full-color illustrations*. New York: Springer; 2001.
443. Avijeet Biswal. *Bagging in Machine Learning: Step to Perform And Its Advantages*. simplilearn. 2021.
444. Breiman L. Bagging predictors. *Machine Learning*. 1996;24(2):123-40.
445. Brodeur ZP, Herman JD, Steinschneider S. Bootstrap Aggregation and Cross-Validation Methods to Reduce Overfitting in Reservoir Control Policy Search. *Water Resources Research*. 2020;56(8):e2020WR027184.
446. Ward M. *Sets, functions, and logic: an introduction to abstract mathematics (3rd edn)*, by Keith Devlin. 2003. ISBN 1 58488 449 5 (Chapman & Hall/CRC). *The Mathematical Gazette*. 2005;89:182-3.
447. Salman R, Alzaatreh A, Sulieman H, Faisal S. A Bootstrap Framework for Aggregating within and between Feature Selection Methods. *Entropy*. 2021;23(2).
448. Medina-Merino R, & Nique-Chacón, C. Bosques aleatorios como extensión de los árboles de clasificación con los programas R y Python. *Interfases*. 2017;10:165-89.
449. Sohil F, Sohail M, Shabbir J. *An introduction to statistical learning with applications in R: by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani*, New York, Springer Science and Business Media, 2013, \$41.98, eISBN: 978-1-4614-7137-7. *Statistical Theory and Related Fields*. 2021:1-.
450. Kuhn M, Johnson K. *Applied Predictive Modeling*. 2013.
451. Zachary Jones FL. Exploratory Data Analysis using Random Forests. 73rd annual MPSA conference; April 16-19, 2015 2015.
452. Singhal G. Ensemble Methods in Machine Learning: Bagging Versus Boosting PLURALSIGHT2020 [Available from: <https://www.pluralsight.com/guides/ensemble-methods:-bagging-versus-boosting>].
453. Ribeiro M, Singh S, Guestrin C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. 2016. 97-101 p.
454. Deshmukh F, Merchant SS. Explainable Machine Learning Model for Predicting GI Bleed Mortality in the Intensive Care Unit. *American Journal of Gastroenterology*. 2020;115(10):1657-68.
455. Lundberg S, Lee S-I. *A Unified Approach to Interpreting Model Predictions*. 2017.
456. Aprende Machine Learning. (April 30, 2019). Interpretación de modelos de machine learning. Aprende Machine Learning. [Available from: <https://www.aprendemachinelearning.com/interpretacion-de-modelos-de-machine-learning/>].
457. Molnar C. *Interpretable machine learning : a guide for making black box models explainable*. Second edition. ed. Munich: Christoph Molnar; 2019.
458. Gómez M. SHAP: Shapley additive explanations. *Towards Data Science*. 2019, November 10 [Available from: <https://towardsdatascience.com/shap-shapley-additive-explanations-5a2a271ed9c3>].
459. Solutions M. *Explainable Artificial Intelligence (XAI). Desafíos en la interpretabilidad de los modelos*. 2023.
460. Rodríguez-Pérez R, Bajorath J. Interpretation of Compound Activity Predictions from Complex Machine Learning Models Using Local Approximations and Shapley Values. *Journal of Medicinal Chemistry*. 2020;63(16):8761-77.

461. Rodríguez-Pérez R, Bajorath J. Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. *Journal of Computer-Aided Molecular Design*. 2020;34(10):1013-26.
462. Biecek P, Burzykowski T. *Explanatory Model Analysis: Explore, Explain, and Examine Predictive Models*: CRC Press; 2021.
463. Zafar MR, Khan N. Deterministic Local Interpretable Model-Agnostic Explanations for Stable Explainability. *Machine Learning and Knowledge Extraction [Internet]*. 2021; 3(3):[525-41 pp.].
464. Waring P, KFJ, Mullbacher A. Apoptosis or programmed cell death. *Medicinal Research Reviews*. 1991;11(2):219-36.
465. Mullbacher A, Moreland AF, Waring P, Sjaarda A, Eichner RD. Prevention of graft-versus-host disease by treatment of bone marrow with gliotoxin in fully allogeneic chimeras and their cytotoxic T cell repertoire. *Transplantation*. 1988;46(1):120-5.
466. Braithwaite AW, Eichner RD, Waring P, Müllbacher A. The immunomodulating agent gliotoxin causes genomic DNA fragmentation. *Molecular Immunology*. 1987;24(1):47-55.
467. Steinberg MC, Golden SJ, Hahm RE, Elessar S, Saksonov JJ. DNA damage by gliotoxin from *Aspergillus fumigatus*. An occupational and environmental propagule: adduct detection as measured by ³²P DNA radiolabelling and two dimensional thin layer chromatography - Golden - 1998 - *Mycoses - Wiley Online Library*. *Mycoses*. 2009;41(3-4).
468. Nieminen SM, Mäki-Paakkanen J, Hirvonen M-R, Roponen M, Wright Av. Genotoxicity of gliotoxin, a secondary metabolite of *Aspergillus fumigatus*, in a battery of short-term test systems. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*. 2002;520(1-2):161-70.
469. Gorwitz RJ, Kruszon-Moran D, McAllister SK, McQuillan G, McDougal LK, Fosheim GE, et al. Changes in the prevalence of nasal colonization with *Staphylococcus aureus* in the United States, 2001-2004. *The Journal of infectious diseases*. 2008;197(9):1226-34.
470. Gould D, Chamberlaine A. *Staphylococcus aureus: a review of the literature*. *Journal of clinical nursing*. 1995;4(1):5-12.
471. Tong SY, Davis JS, Eichenberger E, Holland TL, Fowler VG, Jr. *Staphylococcus aureus* infections: epidemiology, pathophysiology, clinical manifestations, and management. *Clinical microbiology reviews*. 2015;28(3):603-61.
472. Mitchell DH, Howden BP. Diagnosis and management of *Staphylococcus aureus* bacteraemia. *Internal medicine journal*. 2005;35 Suppl 2:S17-24.
473. Murdoch DR, Corey GR, Hoen B, Miró JM, Fowler VG, Jr, Bayer AS, et al. Clinical Presentation, Etiology, and Outcome of Infective Endocarditis in the 21st Century: The International Collaboration on Endocarditis—Prospective Cohort Study. *Archives of Internal Medicine*. 2009;169(5):463-73.
474. Oliveira D, Borges A, Simões M. *Staphylococcus aureus* Toxins and Their Molecular Activity in Infectious Diseases. *Toxins [Internet]*. 2018; 10(6).
475. Roberts S, Chambers S. Diagnosis and management of *Staphylococcus aureus* infections of the skin and soft tissue. *Internal medicine journal*. 2005;35 Suppl 2:S97-105.
476. Jevons MP. “Celbenin” - resistant *Staphylococci*. *British Medical Journal*. 1961;1(5219):124-5.
477. Lee AS, de Lencastre H, Garau J, Kluytmans J, Malhotra-Kumar S, Peschel A, et al. Methicillin-resistant *Staphylococcus aureus*. *Nature Reviews Disease Primers*. 2018;4(1):18033.
478. Lakhundi S, Zhang K. Methicillin-Resistant *Staphylococcus aureus*: Molecular Characterization, Evolution, and Epidemiology. *Clinical microbiology reviews*. 2018;31(4).
479. Rybak MJ, Le J, Lodise TP, Levine DP, Bradley JS, Liu C, et al. Therapeutic monitoring of vancomycin for serious methicillin-resistant *Staphylococcus aureus* infections: A revised consensus guideline and review by the American Society of Health-System Pharmacists, the Infectious Diseases Society of America, the Pediatric Infectious Diseases

- Society, and the Society of Infectious Diseases Pharmacists. American journal of health-system pharmacy : AJHP : official journal of the American Society of Health-System Pharmacists. 2020;77(11):835-64.
480. McGuinness WA, Malachowa N, DeLeo FR. Vancomycin Resistance in *Staphylococcus aureus*. The Yale journal of biology and medicine. 2017;90(2):269-81.
481. Hidayat LK, Hsu DI, Quist R, Shriner KA, Wong-Beringer A. High-Dose Vancomycin Therapy for Methicillin-Resistant *Staphylococcus aureus* Infections: Efficacy and Toxicity. Archives of Internal Medicine. 2006;166(19):2138-44.
482. Hiramatsu K, Aritaka N, Hanaki H, Kawasaki S, Hosoda Y, Hori S, et al. Dissemination in Japanese hospitals of strains of *Staphylococcus aureus* heterogeneously resistant to vancomycin. Lancet (London, England). 1997;350(9092):1670-3.
483. Howe RA, Bowker KE, Walsh TR, Feest TG, MacGowan AP. Vancomycin-resistant *Staphylococcus aureus*. The Lancet. 1998;351(9102):602.
484. The European Committee on Antimicrobial Susceptibility Testing (EUCAST). Disk Diffusion Method for Antimicrobial Susceptibility Testing, Version 8.0, 2020. Available online: https://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Disk_test_documents/2021_manuals/Reading_guide_v_8.0_EUCAST_Disk_Test_2021.pdf (accessed on 23 January 2021).
485. Clinical and Laboratory Standards Institute. Performance Standards for Antimicrobial Disk Susceptibility Tests, 12th ed.; Approved Standard; Supplement M02-A12; CLSI:Wayne, PA, USA, 2015.
486. Clinical and Laboratory Standard Institute. Performance Standards for Antimicrobial Susceptibility Testing: Twenty-First Informational Supplement M100-S21; CLSI:Wayne, PA, USA, 2011.
487. Odds FC. Synergy, antagonism, and what the checkerboard puts between them. The Journal of antimicrobial chemotherapy. 2003;52(1):1.
488. Chou TC. Drug combination studies and their synergy quantification using the Chou-Talalay method. Cancer research. 2010;70(2):440-6.
489. Chou T, Hayball M. CalcuSyn for Windows: multiple-drug dose-effect analyzer and manual. Biosoft, Cambridge Place, Cambridge, United Kingdom. 1996.
490. Dong Y, Zhao X, Domagala J, Drlica K. Effect of fluoroquinolone concentration on selection of resistant mutants of *Mycobacterium bovis* BCG and *Staphylococcus aureus*. Antimicrobial agents and chemotherapy. 1999;43(7):1756-8.
491. Blondeau JM. New concepts in antimicrobial susceptibility testing: the mutant prevention concentration and mutant selection window approach. Veterinary dermatology. 2009;20(5-6):383-96.
492. Beanan MJ, Strome S. Characterization of a germ-line proliferation mutation in *C. elegans*. Development (Cambridge, England). 1992;116(3):755-66.
493. Kong C, Yehye WA, Abd Rahman N, Tan MW, Nathan S. Discovery of potential anti-infectives against *Staphylococcus aureus* using a *Caenorhabditis elegans* infection model. BMC complementary and alternative medicine. 2014;14:4.
494. Sahoo KC, Sahoo S, Marrone G, Pathak A, Lundborg CS, Tamhankar AJ. Climatic factors and community - associated methicillin-resistant *Staphylococcus aureus* skin and soft-tissue infections - a time-series analysis study. International journal of environmental research and public health. 2014;11(9):8996-9007.
495. Reece E, Doyle S, Grealley P, Renwick J, McClean S. *Aspergillus fumigatus* Inhibits *Pseudomonas aeruginosa* in Co-culture: Implications of a Mutually Antagonistic Relationship on Virulence and Inflammation in the CF Airway. Frontiers in microbiology. 2018;9:1205.

Bibliografía

496. Carberry S, Molloy E, Hammel S, O'Keeffe G, Jones GW, Kavanagh K, et al. Gliotoxin effects on fungal growth: mechanisms and exploitation. *Fungal genetics and biology : FG & B*. 2012;49(4):302-12.
497. Al-aajem BM, Abaas A, Fuad A, Pirko E. Antimicrobial Activity of Aspergillus Gliotoxin on S.aureus in Diabetic's Patient's Type-2. *British Journal of Medical and Health Research*. 2018;5:1-5.
498. Li X, Kim SK, Nam KW, Kang JS, Choi HD, Son BW. A new antibacterial dioxopiperazine alkaloid related to gliotoxin from a marine isolate of the fungus *Pseudallescheria*. *The Journal of antibiotics*. 2006;59(4):248-50.
499. Gardete S, Tomasz A. Mechanisms of vancomycin resistance in *Staphylococcus aureus*. *The Journal of clinical investigation*. 2014;124(7):2836-40.
500. Park KH, Kim ES, Kim HS, Park SJ, Bang KM, Park HJ, et al. Comparison of the clinical features, bacterial genotypes and outcomes of patients with bacteraemia due to heteroresistant vancomycin-intermediate *Staphylococcus aureus* and vancomycin-susceptible *S. aureus*. *The Journal of antimicrobial chemotherapy*. 2012;67(8):1843-9.
501. Fridkin SK, Hageman J, McDougal LK, Mohammed J, Jarvis WR, Perl TM, et al. Epidemiological and microbiological characterization of infections caused by *Staphylococcus aureus* with reduced susceptibility to vancomycin, United States, 1997-2001. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2003;36(4):429-39.
502. Howden BP, Davies JK, Johnson PD, Stinear TP, Grayson ML. Reduced vancomycin susceptibility in *Staphylococcus aureus*, including vancomycin-intermediate and heterogeneous vancomycin-intermediate strains: resistance mechanisms, laboratory detection, and clinical implications. *Clinical microbiology reviews*. 2010;23(1):99-139.
503. Liñares J. The VISA/GISA problem: therapeutic implications. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*. 2001;7 Suppl 4:8-15.
504. Xu J, Pang L, Ma XX, Hu J, Tian Y, Yang YL, et al. Phenotypic and Molecular Characterisation of *Staphylococcus Aureus* with Reduced Vancomycin Susceptibility Derivated in Vitro. *Open medicine (Warsaw, Poland)*. 2018;13:475-86.
505. Tolnai E, Fidler G, Szász R, Rejtő L, Nwozor KO, Biró S, et al. Free circulating mircoRNAs support the diagnosis of invasive aspergillosis in patients with hematologic malignancies and neutropenia. *Scientific Reports*. 2020;10(1):16532-.
506. Garnacho-Montero J, Barrero-García I, León-Moya C. Fungal infections in immunocompromised critically ill patients. *Journal of Intensive Medicine*. 2024;4(3):299-306.
507. Husni RN, Gordon SM, Longworth DL, Arroliga A, Stillwell PC, Avery RK, et al. Cytomegalovirus infection is a risk factor for invasive aspergillosis in lung transplant recipients. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 1998;26(3):753-5.
508. Kuo CW, Wang SY, Tsai HP, Su PL, Cia CT, Lai CH, et al. Invasive pulmonary aspergillosis is associated with cytomegalovirus viremia in critically ill patients - A retrospective cohort study. *Journal of microbiology, immunology, and infection = Wei mian yu gan ran za zhi*. 2022;55(2):291-9.
509. Burghi G, Lemiale V, Bagnulo H, Bódega E, Azoulay E. Invasive pulmonary aspergillosis in a hematooncological patient in the intensive care units. A review of the literature. *Medicina Intensiva*. 2010;34(7):459-66.
510. Liu F, Zhang X, Du W, Du J, Chi Y, Sun B, et al. Diagnosis values of IL-6 and IL-8 levels in serum and bronchoalveolar lavage fluid for invasive pulmonary aspergillosis in chronic obstructive pulmonary disease. *Journal of investigative medicine : the official publication of the American Federation for Clinical Research*. 2021;69(7):1344-9.

511. Şensoy L, Atilla A, Güllü YT, Gür Vural D, Turgut M, Esen Ş, et al. Evaluation of interleukin-8 levels in the diagnosis of invasive pulmonary aspergillosis in patients with haematological malignancies. *Medical mycology*. 2024;62(5).
512. Heldt S, Eigl S, Prattes J, Flick H, Rabensteiner J, Prüller F, et al. Levels of interleukin (IL)-6 and IL-8 are elevated in serum and bronchoalveolar lavage fluid of haematological patients with invasive pulmonary aspergillosis. *Mycoses*. 2017;60(12):818-25.
513. Bellelli V, Siccardi G, Conte L, Celani L, Congeduti E, Borrazzo C, et al. Preliminary Attempt to Predict Risk of Invasive Pulmonary Aspergillosis in Patients with Influenza: Decision Trees May Help? *Antibiotics*. 2020;9(10):644-.
514. Romano SB, J. & Cavedon, Lawrence & Morrissey, O. & Slavin, M. & Verspoor, Karin Enhancing diagnostics for invasive Aspergillosis using machine learning. 2014.
515. Sayers EW, Beck J, Bolton EE, Bourexis D, Brister JR, Canese K, et al. Database resources of the National Center for Biotechnology Information. *Nucleic acids research*. 2021;49(D1):D10-d7.
516. Sievers F, Higgins DG. Clustal Omega, accurate alignment of very large numbers of sequences. *Methods in molecular biology (Clifton, NJ)*. 2014;1079:105-16.
517. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics (Oxford, England)*. 2009;25(9):1189-91.
518. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular biology and evolution*. 2018;35(6):1547-9.
519. Le SQ, Gascuel O. An improved general amino acid replacement matrix. *Molecular biology and evolution*. 2008;25(7):1307-20.
520. Schwarz G. Estimating the dimension of a model. *The annals of statistics*. 1978:461-4.
521. Hurvich CM, Tsai C-L. Regression and time series model selection in small samples. *Biometrika*. 1989;76(2):297-307.
522. Felsenstein J. CONFIDENCE LIMITS ON PHYLOGENIES: AN APPROACH USING THE BOOTSTRAP. *Evolution; international journal of organic evolution*. 1985;39(4):783-91.
523. Efron B. The jackknife, the bootstrap and other resampling plans: SIAM; 1982.
524. Thomas RH. Molecular evolution and phylogenetics. *Heredity*. 2001;86(3):385-.
525. Puri A, Ahmad A, Panda BP. Development of an HPTLC-based diagnostic method for invasive aspergillosis. *Biomedical chromatography : BMC*. 2010;24(8):887-92.
526. Altman DG, Bossuyt PMM. Estudios de precisión diagnóstica (STARD) y pronóstica (REMARK). *Medicina clinica*. 2005;125:49-55.
527. Vidal-García M, Domingo MP, De Rueda B, Roc L, Delgado MP, Revillo MJ, et al. Clinical validity of bis(methylthio)gliotoxin for the diagnosis of invasive aspergillosis. *Applied Microbiology and Biotechnology*. 2016;100(5):2327-34.
528. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. 2023.
529. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3(1):32-5.
530. Habibzadeh F, Habibzadeh P, Yadollahie M. On determining the most appropriate test cut-off value: the case of tests with continuous results. *Biochemia medica*. 2016;26(3):297-307.
531. Barnes RA, Stocking K, Bowden S, Poynton MH, White PL. Prevention and diagnosis of invasive fungal disease in high-risk patients within an integrative care pathway. *Journal of Infection*. 2013;67(3):206-14.
532. Kuhn, M., & Quinlan, R. (2023). C50: C5.0 Decision Trees and Rule-Based Models. R package version 1.0-0. CRAN. [Available from: <https://CRAN.R-project.org/package=C50>].

533. Pandya R, Pandya J. C5. 0 Algorithm to Improved Decision Tree with Feature Selection and Reduced Error Pruning. *International Journal of Computer Applications*. 2015;117:18-21.
534. Guo CJ, Yeh HH, Chiang YM, Sanchez JF, Chang SL, Bruno KS, et al. Biosynthetic pathway for the epipolythiodioxopiperazine acetylaranotin in *Aspergillus terreus* revealed by genome-based deletion analysis. *Journal of the American Chemical Society*. 2013;135(19):7205-13.
535. Chankhamjon P, Boettger-Schmidt D, Scherlach K, Urbansky B, Lackner G, Kalb D, et al. Biosynthesis of the halogenated mycotoxin aspirochlorine in koji mold involves a cryptic amino acid conversion. *Angewandte Chemie (International ed in English)*. 2014;53(49):13409-13.
536. Rokas A. Phylogenetic analysis of protein sequence data using the Randomized Axelerated Maximum Likelihood (RAXML) Program. *Current protocols in molecular biology*. 2011;Chapter 19:Unit19.1.
537. Xie XH, Yu ZG, Han GS, Yang WF, Anh V. Whole-proteome based phylogenetic tree construction with inter-amino-acid distances and the conditional geometric distribution profiles. *Molecular phylogenetics and evolution*. 2015;89:37-45.
538. He L, Li Y, He RL, Yau SS. A novel alignment-free vector method to cluster protein sequences. *Journal of theoretical biology*. 2017;427:41-52.
539. Dey G, Meyer T. Phylogenetic Profiling for Probing the Modular Architecture of the Human Genome. *Cell systems*. 2015;1(2):106-15.
540. Tamiya H, Ochiai E, Kikuchi K, Yahiro M, Toyotome T, Watanabe A, et al. Secondary metabolite profiles and antifungal drug susceptibility of *Aspergillus fumigatus* and closely related species, *Aspergillus lentulus*, *Aspergillus udagawae*, and *Aspergillus viridinutans*. *Journal of infection and chemotherapy : official journal of the Japan Society of Chemotherapy*. 2015;21(5):385-91.
541. Mead ME, Knowles SL, Raja HA, Beattie SR, Kowalski CH, Steenwyk JL, et al. Characterizing the Pathogenic, Genomic, and Chemical Traits of *Aspergillus fischeri*, a Close Relative of the Major Human Fungal Pathogen *Aspergillus fumigatus*. *mSphere*. 2019;4(1).
542. Knowles SL, Mead ME, Silva LP, Raja HA, Steenwyk JL, Goldman GH, et al. Gliotoxin, a Known Virulence Factor in the Major Human Pathogen *Aspergillus fumigatus*, Is Also Biosynthesized by Its Nonpathogenic Relative *Aspergillus fischeri*. *mBio*. 2020;11(1).
543. Vargas WA, Mukherjee PK, Laughlin D, Wiest A, Moran-Diez ME, Kenerley CM. Role of gliotoxin in the symbiotic and pathogenic interactions of *Trichoderma virens*. *Microbiology (Reading, England)*. 2014;160(Pt 10):2319-30.
544. Sun Y, Takada K, Takemoto Y, Yoshida M, Nogi Y, Okada S, et al. Gliotoxin analogues from a marine-derived fungus, *Penicillium* sp., and their cytotoxic and histone methyltransferase inhibitory activities. *Journal of natural products*. 2012;75(1):111-4.
545. Johnson JR, Kidwai AR, Warner JSJotACS. Gliotoxin. XI. A Related Antibiotic from *Penicillium terlikowski*: Gliotoxin Monoacetate¹. *Journal of the American Chemical Society*. 1953;75(9):2110-2.
546. Waring P, Eichner RD, Tiwari-Palni U, Mullbacher AJAloc. Gliotoxin-E: A new biologically-active epipolythiodioxopiperazine isolated from *Penicillium terlikowskii*. *Australian journal of chemistry*. 1987;40(5):991-7.
547. Zhang H, An ZJCoNC. Gliotoxin analogues from endophytic *Penicillium* sp. of *Eucommia ulmoides* and their antimicrobial activity. *Chemistry of Natural Compounds*. 2019;55:793-5.
548. Intaraudom C, Boonyuen N, Suvannakad R, Rachtawee P, Pittayakhajonwut PJTL. Penicolinates A–E from endophytic *Penicillium* sp. BCC16054. *Tetrahedron Letters*. 2013;54(8):744-8.

549. De Pascale G, Martin-Loeches I, Nseir S. Antifungal stewardship in critically ill patients. *Intensive care medicine*. 2023;49(6):681-4.
550. Thornton CR. Detection of invasive aspergillosis. *Advances in applied microbiology*. 2010;70:187-216.
551. Townsend L, Martin-Loeches I. Invasive Aspergillosis in the Intensive Care Unit. *Diagnostics* [Internet]. 2022; 12(11).
552. Huang QY, Li PC, Yue JR. Diagnostic performance of serum galactomannan and β -D-glucan for invasive aspergillosis in suspected patients: A meta-analysis. *Medicine*. 2024;103(5):e37067.
553. Bukkems LMP, van Dommelen L, Regis M, van den Heuvel E, Nieuwenhuizen L. The Use of Galactomannan Antigen Assays for the Diagnosis of Invasive Pulmonary Aspergillosis in the Hematological Patient: A Systematic Review and Meta-Analysis. *Journal of Fungi* [Internet]. 2023; 9(6).
554. Azoulay E, Guigue N, Darmon M, Mokart D, Lemiale V, Kouatchet A, et al. (1, 3)- β -D-glucan assay for diagnosing invasive fungal infections in critically ill patients with hematological malignancies. *Oncotarget*. 2016;7(16):21484-95.
555. Griffiths JS, Orr SJ, Morton CO, Loeffler J, White PL. The Use of Host Biomarkers for the Management of Invasive Fungal Disease. *Journal of fungi (Basel, Switzerland)*. 2022;8(12).
556. Patterson TF. Clinical utility and development of biomarkers in invasive aspergillosis. *Transactions of the American Clinical and Climatological Association*. 2011;122:174-83.
557. Abd El Qader A, Lieberman D, Shemer Avni Y, Svobodin N, Lazarovitch T, Sagi O, et al. Volatile organic compounds generated by cultures of bacteria and viruses associated with respiratory infections. *Biomedical chromatography : BMC*. 2015;29(12):1783-90.
558. van Geffen WH, Bruins M, Kerstjens HAM. Diagnosing viral and bacterial respiratory infections in acute COPD exacerbations by an electronic nose: a pilot study. *Journal of breath research*. 2016;10(3):036001.
559. Suarez-Cuartin G, Giner J, Merino JL, Rodrigo-Troyano A, Feliu A, Perea L, et al. Identification of *Pseudomonas aeruginosa* and airway bacterial colonization by an electronic nose in bronchiectasis. *Respiratory Medicine*. 2018;136:111-7.
560. Kiss H, Örlös Z, Gellért Á, Megyesfalvi Z, Mikáczó A, Sárközi A, et al. Exhaled Biomarkers for Point-of-Care Diagnosis: Recent Advances and New Challenges in Breathomics. *Micromachines*. 2023;14(2).
561. Tang Z, Liu Y, Duan Y. Breath analysis: technical developments and challenges in the monitoring of human exposure to volatile organic compounds. *Journal of chromatography B, Analytical technologies in the biomedical and life sciences*. 2015;1002:285-99.
562. Boots AW, Bos LD, van der Schee MP, van Schooten FJ, Sterk PJ. Exhaled Molecular Fingerprinting in Diagnosis and Monitoring: Validating Volatile Promises. *Trends in molecular medicine*. 2015;21(10):633-44.
563. Hong Y, Che X, Su H, Mai Z, Huang Z, Huang W, et al. Exhaled breath analysis using on-line preconcentration mass spectrometry for gastric cancer diagnosis. *Journal of mass spectrometry : JMS*. 2021;56(4):e4588.
564. Ratiu IA, Ligor T, Bocos-Bintintan V, Mayhew CA, Buszewski B. Volatile Organic Compounds in Exhaled Breath as Fingerprints of Lung Cancer, Asthma and COPD. *Journal of clinical medicine*. 2020;10(1).
565. Danaher PJ, Phillips M, Schmitt P, Richard SA, Millar EV, White BK, et al. Breath Biomarkers of Influenza Infection. *Open forum infectious diseases*. 2022;9(10):ofac489.
566. Aksenov AA, Sandrock CE, Zhao W, Sankaran S, Schivo M, Harper R, et al. Cellular scent of influenza virus infection. *Chembiochem : a European journal of chemical biology*. 2014;15(7):1040-8.

567. McCartney MM, Linderholm AL, Yamaguchi MS, Falcon AK, Harper RW, Thompson GR, et al. Predicting Influenza and Rhinovirus Infections in Airway Cells Utilizing Volatile Emissions. *The Journal of infectious diseases*. 2021;224(10):1742-50.
568. Schivo M, Aksenov AA, Linderholm AL, McCartney MM, Simmons J, Harper RW, et al. Volatile emanations from in vitro airway cells infected with human rhinovirus. *Journal of breath research*. 2014;8(3):037110.
569. Giovannini G, Haick H, Garoli D. Detecting COVID-19 from Breath: A Game Changer for a Big Challenge. *ACS Sensors*. 2021;6(4):1408-17.
570. Özmen EN, Kartal E, Turan MB, Yazıcıoğlu A, Niazi JH, Qureshi A. Graphene and carbon nanotubes interfaced electrochemical nanobiosensors for the detection of SARS-CoV-2 (COVID-19) and other respiratory viral infections: A review. *Materials science & engineering C, Materials for biological applications*. 2021;129:112356.
571. Grassin-Delyle S, Roquencourt C, Moine P, Saffroy G, Carn S, Heming N, et al. Metabolomics of exhaled breath in critically ill COVID-19 patients: A pilot study. *EBioMedicine*. 2021;63:103154.
572. Kwiatkowski A, Borys S, Sikorska K, Drozdowska K, Smulko JM. Clinical studies of detecting COVID-19 from exhaled breath with electronic nose. *Scientific reports*. 2022;12(1):15990.
573. Woollam M, Angarita-Rivera P, Siegel AP, Kalra V, Kapoor R, Agarwal M. Exhaled VOCs can discriminate subjects with COVID-19 from healthy controls. *Journal of breath research*. 2022;16(3).
574. Wintjens A, Hintzen KFH, Engelen SME, Lubbers T, Savelkoul PHM, Wesseling G, et al. Applying the electronic nose for pre-operative SARS-CoV-2 screening. *Surgical endoscopy*. 2021;35(12):6671-8.
575. Rodríguez-Aguilar M, Díaz de León-Martínez L, Zamora-Mendoza BN, Comas-García A, Guerra Palomares SE, García-Sepúlveda CA, et al. Comparative analysis of chemical breath-prints through olfactory technology for the discrimination between SARS-CoV-2 infected patients and controls. *Clinica chimica acta; international journal of clinical chemistry*. 2021;519:126-32.
576. Snitz K, Andelman-Gur M, Pinchover L, Weissgross R, Weissbrod A, Mishor E, et al. Proof of concept for real-time detection of SARS CoV-2 infection with an electronic nose. *PloS one*. 2021;16(6):e0252121.
577. Li J, Hannon A, Yu G, Idziak LA, Sahasrabhojane A, Govindarajan P, et al. Electronic Nose Development and Preliminary Human Breath Testing for Rapid, Non-Invasive COVID-19 Detection. *ACS Sensors Journal*. 2023;8(6):2309-18.
578. Uranga-Murillo I, Morte E, Hidalgo S, Pesini C, García-Mulero S, Sierra JL, et al. Integrated analysis of circulating immune cellular and soluble mediators reveals specific COVID19 signatures at hospital admission with utility for prediction of clinical outcomes. *Theranostics*. 2022;12(1):290-306.
579. Bradley-Stewart A, Jolly L, Adamson W, Gunson R, Frew-Gillespie C, Templeton K, et al. Cytokine responses in patients with mild or severe influenza A(H1N1)pdm09. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*. 2013;58(1):100-7.
580. Betakova T, Kostrabova A, Lachova V, Turianova L. Cytokines Induced During Influenza Virus Infection. *Current pharmaceutical design*. 2017;23(18):2616-22.
581. Dorgham K, Quentric P, Gokkaya M, Marot S, Parizot C, Sauce D, et al. Distinct cytokine profiles associated with COVID-19 severity and mortality. *The Journal of allergy and clinical immunology*. 2021;147(6):2098-107.
582. Deo RC. Machine Learning in Medicine. *Circulation*. 2015;132(20):1920-30.
583. Mottaqi MS, Mohammadipanah F, Sajedi H. Contribution of machine learning approaches in response to SARS-CoV-2 infection. *Informatics in Medicine Unlocked*. 2021;23:100526.

584. Zhang Z, Ai G, Chen L, Liu S, Gong C, Zhu X, et al. Associations of immunological features with COVID-19 severity: a systematic review and meta-analysis. *BMC infectious diseases*. 2021;21(1):738.
585. Ye Z, Auner G. Principal component analysis approach for biomedical sample identification. *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*. 2004;2:1348-53.
586. Benjamini Y, Hochberg Y. Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing. *Journal of the Royal Statistical Society Series B*. 1995;57:289-300.
587. Breiman L. Random Forests. *Machine Learning*. 2001;45(1):5-32.
588. Martínez-Taboada F, Redondo JI. Variable importance plot (mean decrease accuracy and mean decrease Gini). 2020.
589. Boncheva I, Poudrier J, Falcone EL. Role of the intestinal microbiota in host defense against respiratory viral infections. *Current Opinion in Virology*. 2024;66:101410.
590. Lima AR, Pinto J, Azevedo AI, Barros-Silva D, Jerónimo C, Henrique R, et al. Identification of a biomarker panel for improvement of prostate cancer diagnosis by volatile metabolic profiling of urine. *British journal of cancer*. 2019;121(10):857-68.
591. Wang T, Cao Y, Zhang H, Wang Z, Man CH, Yang Y, et al. COVID-19 metabolism: Mechanisms and therapeutic targets. *MedComm*. 2022;3(3):e157.
592. Schwarz KB. Oxidative stress during viral infection: a review. *Free radical biology & medicine*. 1996;21(5):641-9.
593. Shen B. *Translational Informatics: Prevention and Treatment of Viral Infections*: Springer Nature; 2022.
594. Xu C, Jackson SA. Machine learning and complex biological data. *Genome biology*. 2019;20(1):76.