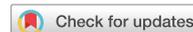


OPEN Las estimaciones del ruido del conjunto de genes destacan las vías críticas y predicen la gravedad de la enfermedad en H1N1, COVID-19 y la mortalidad en pacientes con sepsis



Tristan V. de Jong^{1,4}, Victor Guryev^{1,4} & Yuri M. Moshkin^{2,3,4}*

Encontrar nuevos biomarcadores para patologías humanas y predecir los resultados clínicos para los pacientes es un desafío. Esto se deriva de la respuesta heterogénea de los individuos a la enfermedad y se refleja en la variabilidad interindividual de las respuestas de expresión génica que oscurece el análisis diferencial de la expresión génica. Aquí, desarrollamos un enfoque alternativo que podría aplicarse para diseccionar los cambios moleculares asociados a la enfermedad. Definimos el ruido del conjunto de genes como una medida que representa una varianza para una colección de genes que codifican para miembros de vías biológicas conocidas o subunidades de complejos de proteínas anotadas y calculadas dentro de un individuo. El ruido del conjunto génico permite la identificación e interpretación holística del desequilibrio de la expresión génica a nivel de redes y sistemas genéticos. Al comparar los datos de expresión génica de pacientes con COVID-19, H1N1 y sepsis, identificamos alteraciones comunes en una serie de vías y complejos de proteínas relevantes para la patología de la sepsis. Entre otros, estos incluyen el complejo de cadena respiratoria mitocondrial I y los peroxisomas. Esto sugiere un efecto Warburg y el estrés oxidativo como características comunes de la respuesta inmune huésped-patógeno. Finalmente, demostramos que el ruido del conjunto de genes podría aplicarse con éxito para la predicción del resultado clínico, es decir, la mortalidad de los pacientes. Por lo tanto, concluimos que el ruido de conjunto de genes representa un enfoque prometedor para la investigación de los mecanismos moleculares de la patología a través de un prisma de alteraciones en la expresión coherente de los circuitos genéticos.

Tanto la neumonía viral como la bacteriana pueden conducir a una afección potencialmente mortal conocida como sepsis. Los casos más notables en la percepción pública incluyen infecciones virales pandémicas como la pandemia de gripe porcina de 2009 causada por el H1N1¹ y, más recientemente, la enfermedad por coronavirus de 2019 (COVID-19) causada por el coronavirus 2 del síndrome respiratorio agudo severo (SARS-CoV-2)². Al igual que con cualquier otra infección respiratoria aguda (IRAG) estacional, estas pandemias provocaron un aumento significativo de los pacientes con sepsis en las unidades de cuidados intensivos^{3,4}. La sepsis es una reacción compleja del huésped (humano) a una infección sistémica, viral o bacteriana, que a menudo resulta en shock

¹ Instituto Europeo de Investigación para la Biología del Envejecimiento, Universidad de Groningen, Centro Médico Universitario de Groningen, Groningen, Países Bajos. ² Centro Federal de Investigación, Instituto de Citología y Genética, SB RAS, Novosibirsk, Rusia. ³ Instituto de Biología Molecular y Celular, SB RAS, Novosibirsk, Rusia. ⁴ Gene Learning Association, Ginebra, Suiza. *correo electrónico: v.guryev@umcg.nl; yury.moshkin@gmail.com

séptico o la [muerte](#)⁵⁻⁷. Un problema tanto del tratamiento de la sepsis como de la predicción de los resultados clínicos de los pacientes se relaciona con la naturaleza altamente heterogénea de la sepsis⁸. A pesar de los avances recientes en la identificación de biomarcadores moleculares para la [sepsis](#)⁸⁻¹⁵, el tratamiento sigue siendo principalmente no curativo y los resultados clínicos se infieren principalmente a partir de signos clínicos ⁶.

Un enfoque canónico para la identificación de biomarcadores de enfermedades y sus posibles dianas terapéuticas se basa en el análisis de expresión génica diferencial (DGE) ya sea en los niveles de ARN o proteína. Esto se deriva de un modelo clásico de regulación génica de Jacob-Monod, que implica una respuesta específica de expresión génica (regulación ascendente o descendente) a una señal específica (véase la perspectiva reciente sobre los orígenes históricos del [modelo](#)¹⁶). Sin embargo, la expresión génica es un proceso estocástico y las respuestas celulares a las señales a menudo desencadenan una cascada de cambios en la expresión génica, lo que dificulta el descubrimiento de objetivos y biomarcadores específicos para la enfermedad.

La naturaleza estocástica de la expresión génica implica una variación natural en los números de copia de ARN y proteína ¹⁷. De acuerdo con la relación fluctuación-respuesta ^{18,19}, una cantidad de respuesta de expresión génica a una señal (fluctuación) es proporcional a su varianza (o al coeficiente biológico cuadrado de variación, *bcv*²) para los valores a escala logarítmica del número de copia de ^{ARN}²⁰. En consecuencia, la inferencia estadística de genes expresados diferencialmente estará sesgada hacia genes con alta varianza (*bcv*²) (Figura S1). Esto conduce a un conjunto de problemas intrínsecos con el análisis DGE. (1) Los genes con mayor variabilidad en la expresión responderán fuertemente a cualquier señal celular dirigida a ellos, aunque estos genes pueden no ser necesariamente causantes de un estado enfermo. Incluso en circunstancias normales, estos exhiben grandes fluctuaciones y, por lo tanto, están poco regulados. (2) Por el contrario, los genes con una baja variabilidad responderán solo modestamente. Estos genes están estrechamente regulados y cualquier fluctuación en su expresión podría ser causal de un estado enfermo.

Al llamar a genes significativamente modificados, para tener sentido biológico, estos genes se mapean a vías biológicas conocidas como GO o KEGG ^{21,22}, o subunidades de complejos de proteínas anotadas por CORUM u otras bases de datos de interacción²³. Por lo tanto, se requiere una segunda prueba estadística, a saber, el análisis de enriquecimiento de conjuntos de genes (GSEA). Sin embargo, esto no está exento de sus propias advertencias. La principal es que GSEA depende de la inferencia estadística de los límites de DGE y [DGE](#)^{24,25}. Como resultado, las interpretaciones biológicas de la DGE podrían verse afectadas drásticamente por las trampas derivadas de la relación fluctuación-respuesta, el umbral de DGE y la elección de un enfoque estadístico para GSEA.

Para eludir esto, razonamos que (1) los genes no funcionan de forma aislada, sino que actúan como conjuntos que representan vías biológicas y / o subunidades de complejos de proteínas. (2) La función normal de una vía biológica o complejo proteico requiere una expresión regulada (equilibrada) de todo el conjunto génico. (3) Cualquier alteración en la expresión de un conjunto de genes podría ser causal de una enfermedad o predictiva del resultado clínico. Para inferir las alteraciones en la expresión de los conjuntos génicos recurrimos a la estimación de sus varianzas (ruido del conjunto génico) a partir de los perfiles de expresión génica en sangre total de los individuos en condiciones normales y patológicas. A partir de la ley total de la varianza, el ruido del conjunto de genes ($\text{Var}[G]$) suma a partir de la varianza de las medias ($\text{E}[G]$) de los genes del conjunto y ($\text{Var}[\text{E}[G|g]]$), así como la expectativa de varianzas de los genes del conjunto ($\text{E}[\text{Var}[G|g]]$) (Figura S2). Por lo tanto, el ruido del conjunto de genes estima tanto: cambios en las estequiometrías de genes que codifican una vía biológica o subunidades complejas de proteínas, como cambios en la variabilidad media de la expresión génica para los genes en el conjunto.

A partir de los perfiles de expresión sanguínea total de pacientes en tratamiento de cuidados intensivos, estimamos cómo el ruido del conjunto de genes corresponde a un estado patológico, como sepsis, por ejemplo, neumonía adquirida en la comunidad / hospital (CAP), neumonía viral H1N1 (H1N1) y enfermedad COVID-19. A partir de este análisis, identificamos una serie de vías para las cuales el ruido del conjunto de genes se asoció positivamente con un estado de salud / enfermedad individual o riesgo de mortalidad cuando se trató como una variable ordinal (< fase temprana de H1N1 < fase H1N1 tardía, sepsis < sana (CAP, otros) sobrevivió a < sepsis (CAP, otros) pacientes fallecidos y covid-19 menos grave < grave). Finalmente, identificamos vías y complejos donde la desregulación se asocia con el riesgo de mortalidad por sepsis (NAC, otros) y la gravedad de la enfermedad COVID-19 en función del ruido del conjunto de genes con alta precisión. Nuestros hallazgos también fueron corroborados con el análisis de red de correlación ponderada (WGCNA). Concluimos que el ruido del conjunto de genes proporciona una herramienta poderosa para el descubrimiento de biomarcadores sistémicos de enfermedades, vías farmacéuticamente dirigidas y la predicción de un resultado clínico de la enfermedad.

Resultados

Respuesta de expresión génica media y de varianza a la infección y sepsis. Se cree que la sepsis desencadena una gran cantidad de respuestas heterogéneas del huésped a una infección sistémica ^{6,8}. Razonamos que esta heterogeneidad podría reflejarse en la variabilidad de la expresión génica interindividual (desviación estándar, σ o varianza, σ^2). Considerando que (a) el número de copia de ARN es un Poisson mixto (*p. ej.* binomio negativo) variable aleatoria ²⁶ y que (b) las intensidades de la señal de hibridación de microarrays transformadas en logarítmica se correlacionan con los números de copia de ARN-seq transformados en logarítmica²⁷. Es fácil demostrar que la varianza de la expresión génica logarítmica se aproxima al coeficiente biológico de variación (bcv^2)²⁰. De la expansión de Taylor de primer orden para la varianza: $\sigma^2 \approx \text{registro}(X) \cdot \text{registro}(X)$ donde $\text{registro}(X)$ es la expresión génica logarítmica. La variable aleatoria mixta de Poisson, CVD donde $X = bcv^2 \cdot CVD$, también conocido como el parámetro de sobredispersión, es independiente de la expresión génica media (μ). Así, para μ

$\gg 1$ (para genes con un gran número medio de copias de ARN), $\sigma_Y^2 \approx bcv^2$. En otras palabras, al estimar las variabilidades de la expresión génica logarítmica entre individuos, ya sea a partir de intensidades de señal de microarrays o recuentos de ARN-seq, podemos inferir aproximadamente los coeficientes biológicos de las variaciones a partir de los números de copia de ARN de los genes.

Estimamos las dispersiones para las expresiones génicas de registro de sangre total en pacientes con sepsis (8826 genes) y pacientes infectados por H1N1 (7240 genes) a partir de los dos conjuntos de datos GSE65682 y GSE21802 respectivamente (para una descripción detallada de las cohortes ver estudios originales y Métodos)^{8,12,28}. Para los pacientes con sepsis también se tuvo en cuenta la edad, incluyéndola como variable aleatoria en el Modelo Aditivo Generalizado para Localización, Escala y Forma (GAMLSS)²⁹, ver sección "Métodos". En promedio, las dispersiones en las expresiones génicas logarítmicas en pacientes con sepsis y H1N1 fueron significativamente mayores en comparación con individuos sanos (Fig. 1A). A eso, para las patentes de sepsis / CAP, las dispersiones en las expresiones génicas logarítmicas fueron significativamente mayores para los pacientes fallecidos en comparación con los que sobrevivieron. Del mismo modo, para los pacientes con H1N1, las dispersiones en las expresiones génicas logarítmicas aumentaron aún más en la fase tardía de la infección (Fig. 1A). Para los pacientes con sepsis, en promedio, las dispersiones de las expresiones génicas logarítmicas fueron comparables entre los pacientes supervivientes y fallecidos para todos los genes analizados (Fig. 1A). Sin embargo, para los genes para los cuales las dispersiones cambiaron significativamente entre individuos sanos y pacientes con sepsis (Bonferroni ajustó $p \leq 0,05$), sus dispersiones en promedio fueron mayores entre los pacientes fallecidos en comparación con los que sobrevivieron ($p < 0,001$). Juntos, estos sugieren que la respuesta del huésped a la infección aumenta los coeficientes biológicos de las variaciones en el número de copias de ARN de los genes (como $\sigma^2 \approx bcv^2$) y corrobora la heterogeneidad en la patogénesis de la sepsis⁸ desde la perspectiva de la expresión génica.

Debido a la relación [fluctuación-respuesta](#)¹⁸, los cambios absolutos en las expresiones génicas logarítmicas medias ($|\Delta\mu|$) en respuesta a la infección (sepsis/CAP, H1N1) y sepsis se correlacionaron significativamente con las varianzas de las expresiones génicas logarítmicas (Fig. 1B). Curiosamente, también observamos correlaciones significativas entre los cambios absolutos en las variabilidades de la expresión génica interindividual ($|\Delta\sigma^2|$) y las varianzas de las expresiones génicas logarítmicas (Fig. 1C). En consecuencia, $|\Delta\mu|$ y $|\Delta\sigma^2|$ también se correlacionaron (Fig. 1D). A partir de esto, concluimos que el H1N1 y la sepsis resultan en una coordinación coordinada.

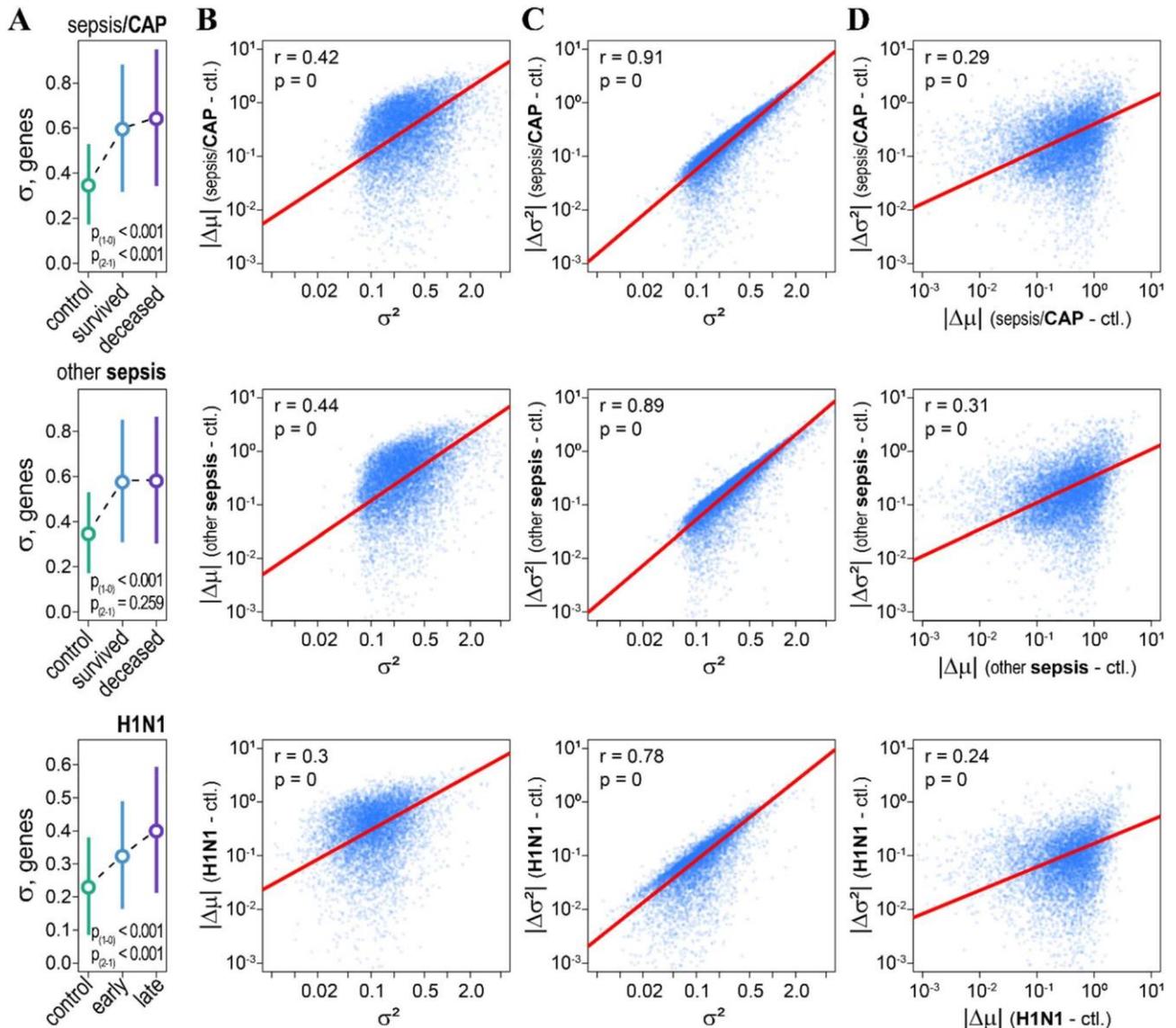


Figura 1. El H1N1 y la sepsis afectan coordinadamente la expresión génica media y la variabilidad de la expresión génica interindividual. **(A)** La variabilidad interindividual en la expresión génica en sangre total (σ) aumenta en los pacientes con sepsis / NAC (arriba), otra sepsis (medio) y H1N1 (abajo) en comparación con los individuos sanos. $p_{(1-0)}$ —valores p de las pruebas t que comparan las diferencias en la variabilidad de la expresión génica interindividual de individuos sanos (control) con pacientes infectados por H1N1 sobrevivientes (sepsis) y tempranos. $p_{(2-1)}$ —valores p de las pruebas t que comparan las diferencias de pacientes infectados por H1N1 sobrevivientes (sepsis) y tempranos con pacientes infectados por H1N1 fallecidos (sepsis) y tardíos infectados por H1N1. Los círculos y los bigotes indican medias y desviaciones estándar respectivamente. **(B)** Correlaciones entre las varianzas en la expresión génica en sangre total (σ^2) y los cambios absolutos en la expresión génica media ($|\Delta\mu|$) para individuos sanos (ctl.) y pacientes (sepsis, H1N1). Debido a la relación fluctuación-respuesta, la magnitud de la respuesta media de expresión génica depende de su varianza. Estimamos las variaciones comunes para los genes en pacientes sanos y con sepsis / CAP (arriba), pacientes sanos y otros pacientes con sepsis (medio) y pacientes sanos y con H1N1 (abajo). **(C)** Correlaciones entre las varianzas en la expresión génica de sangre total (σ^2) y los cambios absolutos en la variabilidad de la expresión génica interindividual ($|\Delta\sigma^2|$) para individuos de control (ctl.) y pacientes (sepsis, H1N1). **(D)** Correlaciones entre los cambios absolutos en la expresión génica media ($|\Delta\mu|$) y en la variabilidad de la expresión génica interindividual ($|\Delta\sigma^2|$).

cambios tanto en la media como en la heterogeneidad de la expresión de los genes y que las magnitudes de estos cambios dependen de los coeficientes biológicos de variación de los genes.

Respuesta de ruido de conjunto de genes a infecciones y sepsis. Tanto la media como la varianza se relacionan con estadísticas de población (interindividuales) que reflejan distintos aspectos de la regulación génica. Los cambios en los medios se ajustan a la visión clásica de DGE sobre la respuesta de los genes a la patología y otros procesos biológicos, mientras que los cambios en las varianzas producen una visión de la heterogeneidad de la respuesta génica. Sin embargo, como se señaló anteriormente (Fig. 1 y S1), la inferencia estadística de estos cambios está sesgada hacia una mayor significación para genes con un alto coeficiente biológico de variación.

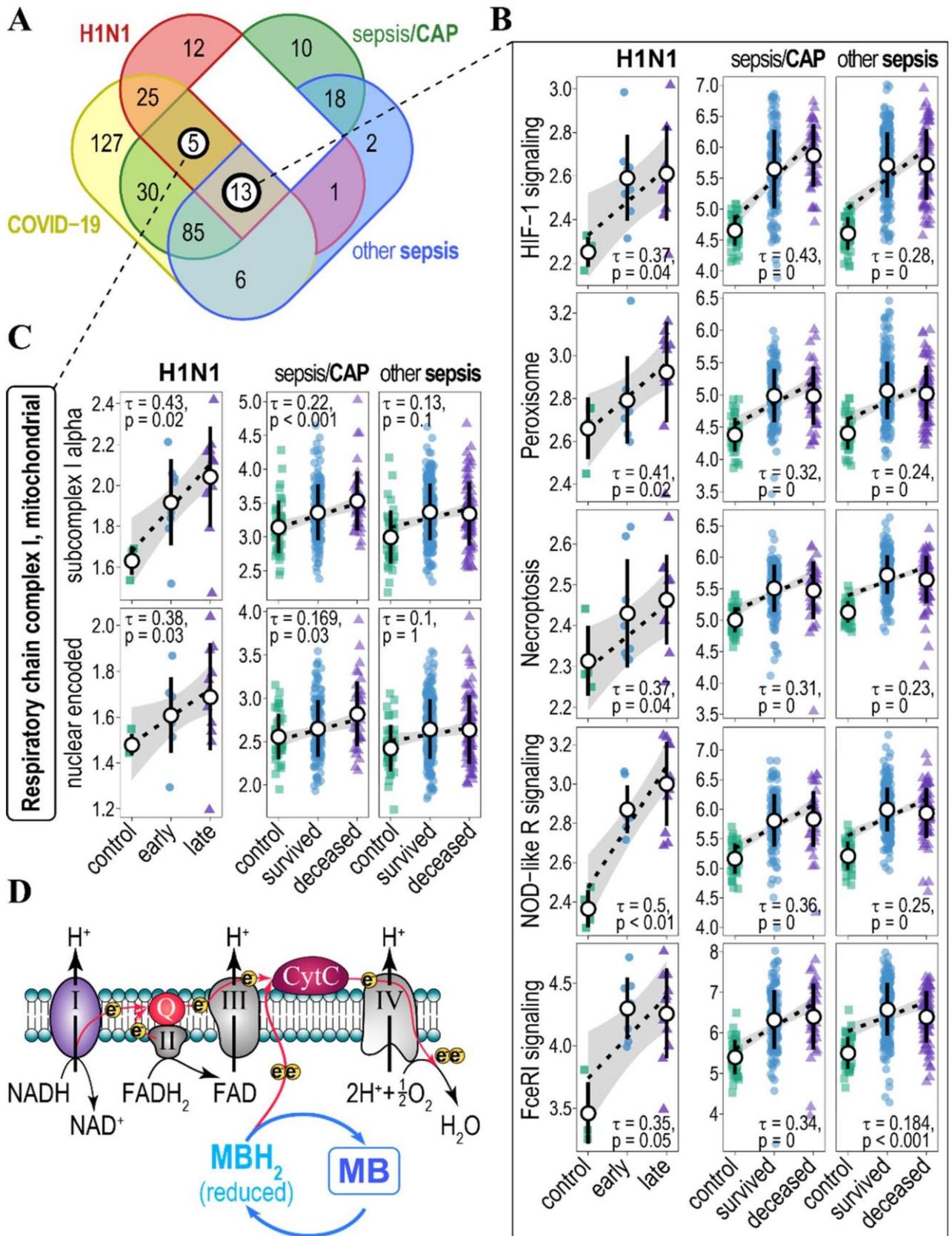


Figura 2. Asociación del ruido del conjunto de genes con la fase de infección por H1N1 y la mortalidad por sepsis. (A) Diagrama de Venn de vías biológicas/complejos de proteínas anotados por KEGG y CORUM para los cuales el ruido del conjunto de genes se asocia positivamente (aumenta) y significativamente con el estado de la enfermedad (leve / grave para el COVID-19, saludable / temprano / tardío para el H1N1 y

saludable / sobrevivido / fallecido para la sepsis / CAP y otros pacientes con sepsis). **(B)** Gráficos de ruido de conjunto de genes para genes involucrados en la señalización hif-1, peroxisoma, necroptosis, receptor similar a NOD y vías de señalización Fc epsilon RI. Las vías fueron anotadas por KEGG. Kendall tau, y FDR- (pacientes con H1N1) y Bonferroni- (pacientes con sepsis) ajustados p-valores están indicados. Se muestran líneas de tendencia de regresión basadas en rangos y bandas de confianza del 95% de la asociación del ruido del conjunto de genes con el estado de la enfermedad. Los círculos negros y los bigotes indican medias y desviaciones estándar. **(C)** Gráficos de ruido de conjunto de genes para genes que codifican subunidades anotadas por CORUM del complejo I de la cadena respiratoria mitocondrial (subcomplejo I alfa-panel superior y subunidades codificadas nuclearmente-panel inferior). Se muestran líneas de tendencia de regresión basadas en rangos y bandas de confianza del 95% de la asociación del ruido del conjunto de genes con el estado de la enfermedad. **(D)** El azul de metileno (MB) actúa como un donante de electrones alternativo a la cadena de transporte de electrones (flechas rojas) al desplazarse entre estados redox (MB-MBH2) y, por lo tanto, evitar el complejo de cadena respiratoria I. Se indica el complejo de cadena respiratoria I-IV y sus sustratos, Q-coenzima Q10, CytC-citocromo C. Los electrones se indican como círculos amarillos.

Aunque los cambios en el número de copias de ARN pueden servir en aplicaciones prácticas para el diagnóstico de la enfermedad y los resultados clínicos, la variabilidad interindividual no se puede utilizar para el diagnóstico. Al mismo tiempo, las fluctuaciones estocásticas en la expresión génica siguen siendo atractivas para la disección de nuevos mecanismos moleculares de la patología. Por lo tanto, esperamos que la estimación del ruido del conjunto de genes pueda proporcionar beneficios adicionales para el diagnóstico al cuantificar las fluctuaciones al tiempo que es informativa para el tratamiento personalizado.

Definimos el ruido de conjunto de genes como la varianza de los niveles de expresión normalizados transformados en logarítmica para una colección de genes $G = (g_1, \dots, g_i)$ que codifican para una proteína de una vía o subunidades de un complejo de proteínas. Para ello, mapeamos genes a las vías anotadas por KEGG y a los complejos proteicos anotados por CORUM^{22,23}. De la ley de la varianza total: $\text{Var}[G] = E[\text{Var}[G|g]] + \text{Var}[E[G|g]]$, el ruido del conjunto de genes depende de la variabilidad en la expresión de los genes en conjunto ($E[\text{Var}[G|g]]$) y de su estequiometría ($\text{Var}[E[G|g]]$) (Figura S1).

Por lo tanto, la estimación simple de las varianzas ($\text{Var}[G]$) de los conjuntos de genes para cada individuo podría reflejar alteraciones en la función de las vías biológicas y los complejos de proteínas en el nivel de estequiometría y ruido génico. Correlacionamos $\text{Var}[G]$ para conjuntos con estados de enfermedad por H1N1 y sepsis. Para la infección viral por H1N1, el estado de la enfermedad se puede clasificar claramente: no infectado (sano) < fase temprana < fase tardía de la infección, por lo que representa una variable ordinal²⁸. Para los pacientes con sepsis, asumimos que una condición de los pacientes fallecidos era peor que la de los que sobrevivieron. Consideramos que los < sanos sobrevivieron < fallecidos también pueden representarse como una variable de estado de la enfermedad ordinal. Circunstancialmente, esto está respaldado por distintos endotipos 8 de expresión génica en sangre y una mayor heterogeneidad de expresión génica (Fig. 1A). La correlación de rango de Kendall identificó una serie de vías y complejos de proteínas para los cuales el ruido del conjunto de genes se asoció positiva y significativamente con el estado de la enfermedad en pacientes con H1N1 ($\text{FDR} \leq 0,05$) y sepsis (ajustado por Bonferroni $p \leq 0,05$) (Fig. 2A). Ninguna de las vías o complejos genéticos se asoció negativamente con el estado de la enfermedad en los umbrales de significación especificados. Se utilizaron diferentes procedimientos de ajuste del valor p (FDR, menos conservador, y Bonferroni, más conservador) para los pacientes con H1N1 y sepsis debido a las grandes diferencias en los tamaños de la muestra (número de pacientes) entre estos conjuntos de datos.

De todos los conjuntos de genes, 13 de ellos demostraron ser consistentes y correlacionados con el aumento del estado de la enfermedad en el ruido del conjunto de genes en las tres condiciones de la enfermedad (Fig. 2A, B, Tabla S1A). Se sabe que la mayoría de estos conjuntos de genes (vías) están involucrados en la patología de la sepsis a través de múltiples piezas de evidencia experimental (Tabla 1), lo que corrobora el poder del análisis de ruido de conjuntos de genes. Sin embargo, el ruido del conjunto de genes produce nuevos conocimientos sobre los mecanismos moleculares de la sepsis (H1N1, CAP u otras causas de sepsis) al sugerir una mala regulación holística en la estequiometría y el ruido genético para estos conjuntos de genes.

También se identificaron 5 conjuntos de genes para los cuales el ruido del conjunto de genes se correlacionó positiva y significativamente con el estado de la enfermedad en pacientes con H1N1 y CAP/sepsis (Fig. 2A, Tabla S1B). Sin embargo, el ruido del conjunto de genes para estas vías también aumentó significativamente en pacientes con sepsis (prueba t, Bonferroni ajustó $p < 0,01$) a pesar de la correlación de rango insignificante. A eso, algunas de estas vías pueden estar implicadas en la patología de la sepsis (Tabla 1). Dos de estos conjuntos fueron representados por genes que codifican el complejo de cadena respiratoria mitocondrial I (Complejo I) (Fig. 2C). Desde el punto de vista del ruido del conjunto de genes, esto sugiere una estequiometría alterada y un ruido génico en la expresión de las subunidades del Complejo I que, como resultado, podría conducir a su ensamblaje y función inadecuados en pacientes con H1N1 y sepsis. Previamente se ha establecido que la actividad del Complejo I está disminuida y se correlaciona negativamente con el resultado de la enfermedad de sepsis⁴⁶. El Complejo I es el primer conjunto de enzimas de la cadena respiratoria y es el punto de entrada para la mayoría de los electrones en la cadena de transporte de electrones⁴⁹. Curiosamente, en el caso de la inhibición o desregulación del Complejo I, el azul de metileno (MB) puede eludirlo actuando como un mediador redox alternativo en la cadena de transporte de electrones, restaurando así la respiración mitocondrial^{50,51} (Fig. 2D). La MB también se considera una terapéutica prometedora en el tratamiento del shock séptico^{52,53}. Por lo tanto, el

ruido del conjunto de genes podría proporcionar un atajo explicativo simple pero poderoso desde la expresión de miles de genes hasta la función de los conjuntos de genes que conducen a posibles objetivos farmacéuticos.

Predicción del resultado clínico para pacientes con sepsis a partir del ruido del conjunto de genes. El tratamiento de la sepsis es un desafío y las tasas de mortalidad entre los pacientes con sepsis son altas. Sin embargo, la predicción del resultado clínico también es un desafío debido a la heterogeneidad tanto en la patología⁸ como en la expresión génica (Fig. 1A). Recientemente, el consorcio Molecular Diagnosis and Risk Stratification of Sepsis (MARS) identificó el endotipo de expresión génica Mars1 que se asoció significativamente con la mortalidad aguda (28 días), sin embargo, para otros endotipos,

| Vías/Complejos | Papel en la patología de la sepsis | Referencias |
|--|--|-------------|
| KEGG: Vía de señalización HIF-1 | Reprogramación metabólica de células inmunes innatas durante las fases hiperinflamatoria e inmunotolerante de la sepsis | 30,31 |
| KEGG: Peroxisoma | El reciclaje defectuoso de peroxisomas altera la homeostasis redox celular y conduce a una respuesta exagerada al estrés oxidativo a la endotoxina (infección) y la sepsis | 32 |
| KEGG: Necroptosis | La necroptosis está implicada en enfermedades pulmonares y lesiones orgánicas asociadas a la sepsis | 33,34 |
| KEGG: Vía de señalización del receptor similar a NOD | La activación de la señalización de receptores tipo Toll y NOD protege a los ratones de la letalidad asociada a la sepsis polimicrobiana | 35 |
| KEGG: Vía de señalización Fc epsilon RI | Los receptores Fc se unen a anticuerpos unidos a patógenos invasores y su regulación ascendente puede servir como un biomarcador potencial para la sepsis. Los ratones deficientes para el gen FCER1G que codifica la subunidad γ de Fc epsilon RI muestran una mayor resistencia a la sepsis | 36,37 |
| KEGG: Autofagia—otros | La autofagia es un proceso protector adaptativo que elimina proteínas, orgánulos y patógenos dañados. Se cree que es un objetivo prometedor en el tratamiento de la sepsis | 38 |
| KEGG: Biosíntesis de aminoácidos | La sepsis resulta en trastornos significativos en el metabolismo de los aminoácidos | 39 |
| KEGG: Vía de señalización del glucagón | Los niveles de glucagón se asocian negativamente con el resultado clínico en pacientes con sepsis | 40 |
| KEGG: Metabolismo del propanoato (propionato) | La acidemia propiónica causada por el metabolismo alterado de los propionatos a menudo resulta en sepsis y muerte | 41 |
| KEGG: Ritmo circadiano | Hay evidencia acumulada de asociación entre la desalineación circadiana y la gravedad de las respuestas inflamatorias en la sepsis | 42 |
| KEGG: Sinapsis dopaminérgica | La dopamina media las comunicaciones neuroinmunes y la dopaminérgica está implicada en la inflamación y la sepsis | 43,44 |
| KEGG: Esclerosis lateral amiotrófica (ELA) | Los pacientes con ELA a menudo desarrollan insuficiencia pulmonar y tienen un mayor riesgo de sepsis | 45 |
| CORUM: Complejo de cadena respiratoria I, mitocondrial | La disfunción mitocondrial que resulta en una actividad reducida del complejo I de la cadena respiratoria y bajos niveles de ATP es una marca completa para la sepsis | 46 |
| KEGG: Diferenciación de osteoclastos | La expresión media de los genes de diferenciación de osteoclastos está regulada al alza en el shock séptico humano | 47 |
| KEGG: Unión estrecha | La sepsis interrumpe la barrera intestinal que conduce a un síndrome de disfunción de múltiples órganos y altera la expresión de proteínas de unión estrecha | 48 |

Tabla 1. Papel en la sepsis de las vías por las que el ruido del conjunto de genes se asocia positivamente (aumenta) con los estados de enfermedad (sanos < tempranos / sobrevividos < tardíos / fallecidos).

Marte²⁻⁴ discriminó pobremente entre la supervivencia y la mortalidad de los [pacientes](#)⁸. Por lo tanto, nos preguntamos si el resultado clínico (mortalidad) podría predecirse a partir del ruido del conjunto de genes.

Con este fin, entrenamos modelos de árbol de regresión impulsados por gradiente logístico binario utilizando la supervivencia y la mortalidad aguda como variable de respuesta binaria para el resultado clínico y la edad de los pacientes y el ruido del conjunto de genes sanguíneos como características del modelo. Los modelos fueron entrenados con XGBoost⁵⁴. Los pacientes con sepsis se dividieron en cohortes de descubrimiento (263 pacientes: 105 sepsis/CAP y 158 pacientes con sepsis) y validación (216 pacientes: 78 sepsis/CAP y 138 pacientes con sepsis) siguiendo la anotación GSE65682⁸. La cohorte de descubrimiento estuvo representada por los pacientes ingresados por sepsis (incluidos los pacientes sin NAC) en las unidades de cuidados intensivos (UCI) en los Países Bajos entre 2011 y 2012¹², y la validación, por los pacientes con sepsis debido a la NAC ingresados en las UCI en el Reino Unido entre 2006 y 2014⁵⁵. Dentro de las cohortes, las tasas de mortalidad fueron del 26,2% para los pacientes con sepsis (23,8% para la sepsis/NAC y 27,8% para otros pacientes con sepsis) dentro de la cohorte de descubrimiento, y del 20,8% para los pacientes con sepsis (19,2% para la sepsis/NAC y 21,7% para otros pacientes con sepsis) en la cohorte de validación.

En general, el desequilibrio de clases, el ruido debido a la heterogeneidad interindividual y la alta dimensionalidad de las características del modelo se encuentran entre los principales problemas del aprendizaje automático⁵⁶. En parte, el ruido del conjunto de genes conduce a una reducción en la variabilidad interindividual (Figura S3) y en la dimensionalidad, ya que las características del modelo no están representadas por genes individuales, sino por colecciones de genes. No obstante, redujimos aún más el número de características de ruido del conjunto de genes en los modelos mediante la selección de características de la prueba *t*. Para esto, comparamos el ruido del conjunto de genes entre pacientes sobrevivientes y fallecidos en las cohortes de descubrimiento. Los puntos de corte del valor *p* para las características del modelo se seleccionaron en función de la maximización de la precisión del entrenamiento de los modelos (consulte la sección "Métodos"). Parámetros de hiperajuste de XGBoost: tasa de aprendizaje, complejidad, profundidad, etc. se optimizaron en

función de la validación cruzada. Para evitar el sobreajuste, utilizamos la detención de época temprana, que se estimó a partir del pliegue de prueba de la cohorte de descubrimiento (consulte la sección "Métodos"). Debido al desequilibrio de clases, se utilizaron AUC (área bajo las curvas de características operativas del receptor (ROC)) para evaluar el rendimiento del modelo. Las cohortes de validación se ocultaron de la selección de características y la capacitación.

La Figura 3A, B muestra las puntuaciones del modelo y las curvas ROC para el modelo, prediciendo la mortalidad / supervivencia de los pacientes con sepsis en las cohortes de descubrimiento y validación. Las AUC para las cohortes de descubrimiento y validación fueron de 0,871 y 0,707 respectivamente, lo que sugiere una precisión razonable del modelo. Sin embargo, a partir de las puntuaciones del modelo y la evaluación de la especificidad / sensibilidad del modelo, parece que el modelo está sesgado hacia la predicción de la clase principal (sobrevivió) (Fig. 3A, Cuadro 2 y Cuadro S2A). Por lo tanto, las precisiones equilibradas de predicción de clases (bACC = Especificidad / 2 + Sensibilidad / 2) fueron de 0,799 y 0,701 para las cohortes de descubrimiento y validación respectivamente. No obstante, la probabilidad de supervivencia para los pacientes que se predijo que tenían un alto riesgo de mortalidad fue significativamente menor que la probabilidad de supervivencia para los pacientes que se predijo que tenían un bajo riesgo de mortalidad tanto en las cohortes de descubrimiento como en las de validación. Para eso, nuestro modelo predice mejor los riesgos de mortalidad en comparación con el endotipo Mars1 inferido a partir de la expresión génica logarítmica del aprendizaje no supervisado⁸ (Fig. 3C). Potencialmente, esto podría deberse a una menor variabilidad interindividual del ruido del conjunto de genes en comparación con la expresión génica logarítmica (Figura S3).

En un intento por aumentar la precisión de la predicción, entrenamos modelos de árbol potenciado por gradiente separado para sepsis/CAP (Fig. 4) y sepsis (Fig. 5) pacientes. De hecho, en ambos casos la exactitud de la predicción de la clase menor.

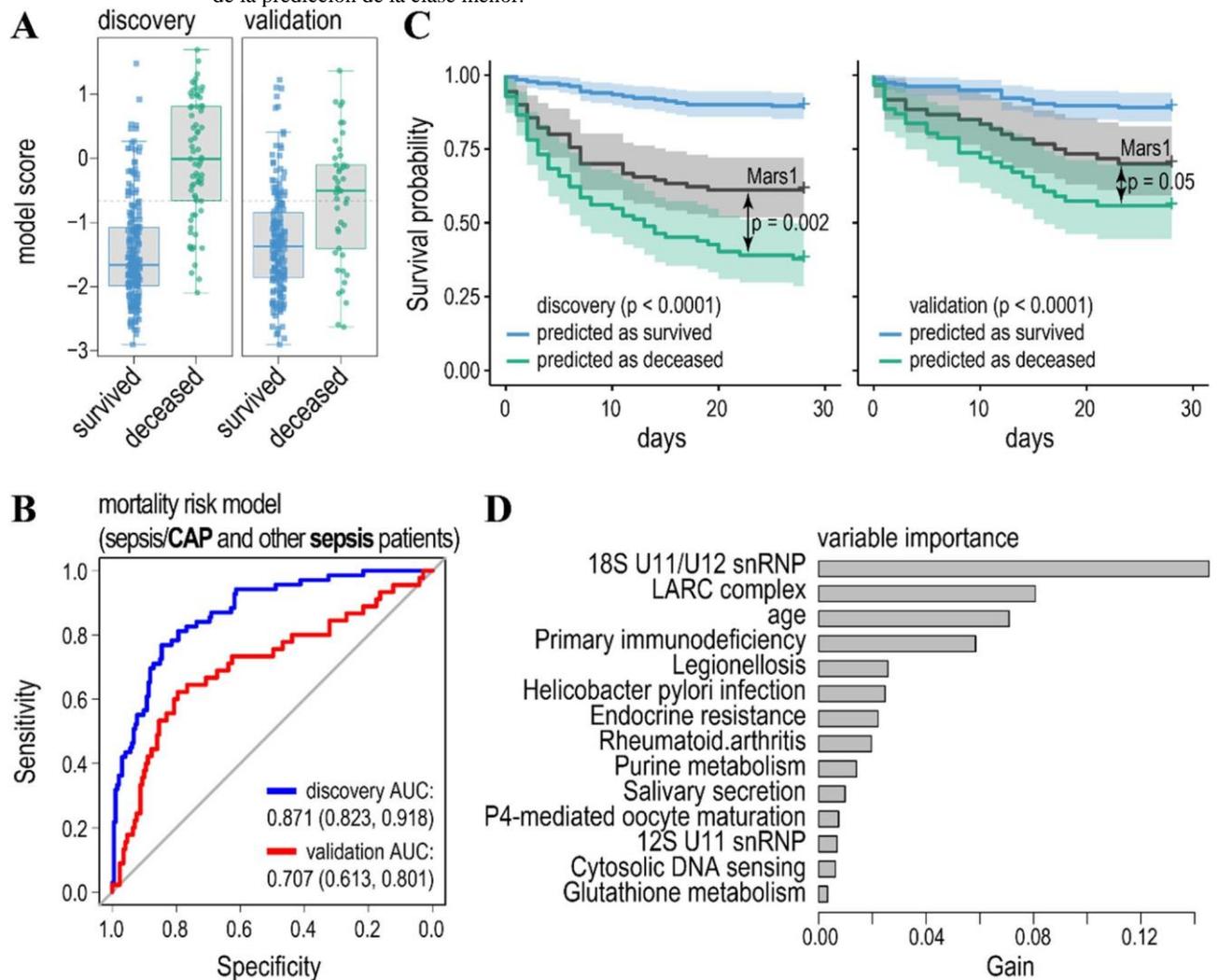


Figura 3. El modelo que predice la mortalidad/supervivencia de los pacientes con sepsis (agrupados como sepsis/CAP y otra sepsis). **(A)** Diagramas de caja de las puntuaciones del modelo que predicen la mortalidad / supervivencia en las cohortes de descubrimiento (izquierda) y validación (derecha). El modelo fue entrenado sobre los mismos datos que la cohorte de descubrimiento publicada por el árbol de regresión potenciado por gradiente y validado en una cohorte independiente⁸. Las líneas discontinuas indican los niveles umbral de

clasificación. El umbral se calculó maximizando un producto de la especificidad y sensibilidad de la predicción del modelo en la cohorte de descubrimiento. En las Tablas 2 y S2 se dan más detalles sobre la precisión del modelo. **(B)** Curvas de características operativas del receptor (ROC) para el modelo que predice la mortalidad (punto final: supervivencia o muerte dentro de los 28 días posteriores al tratamiento) en pacientes con sepsis (agrupados como sepsis / CAP y otra sepsis) (línea azul: cohorte de descubrimiento, línea roja: cohorte de validación). Las características fueron seleccionadas por la prueba t que compara el ruido del conjunto de genes entre los pacientes sobrevivientes y fallecidos en la cohorte de descubrimiento para lograr la máxima precisión de predicción para la cohorte de descubrimiento. Se indican los valores para el área bajo la curva ROC (AUC). **(C)** Probabilidad de supervivencia para los pacientes que se predice que tienen un riesgo de mortalidad bajo (línea azul) y alto (línea verde) para las cohortes de descubrimiento (panel izquierdo) y validación (panel derecho). *Los valores de P* indican diferencias significativas en los peligros para las clases predichas (supervivencia/mortalidad) según el modelo de riesgos proporcionales de Cox. Líneas negras: la probabilidad de supervivencia de los pacientes con endotipo 8 de Mars1 se comparó con la clase de fallecidos predicha para las cohortes de descubrimiento y validación. **(D)** La importancia variable del modelo clasifica las características de ruido del conjunto de genes de acuerdo con su contribución relativa (ganancia).

(pacientes fallecidos) aumentó (Tabla 2 y S2) tanto en las cohortes de descubrimiento como de validación. Asimismo, los AUC para las cohortes de validación también fueron mayores en comparación con el modelo que predice la mortalidad para ambas cohortes de pacientes con sepsis (comparar Fig. 4B y 5B con Fig. 3B). Para eso, las diferencias en las AUC entre las cohortes de descubrimiento y validación fueron menores para los modelos que predijeron la mortalidad por separado para los pacientes con sepsis / CAP y sepsis como para el modelo entrenado en ambos tipos de pacientes. Esto fue especialmente evidente para el modelo que predijo la mortalidad de los pacientes con sepsis/NAC (Fig. 4B). Por lo tanto, conocer la causa de la sepsis mejora la precisión de la predicción de los modelos.

Finalmente, debe tenerse en cuenta que tanto la selección de características como los árboles de regresión potenciados por gradiente permiten clasificar la importancia de las características del modelo (Figs. 3D, 4D, 5D y Tabla S3). En primer lugar, resultó que la edad de un paciente no contribuye notablemente a la predicción de la mortalidad en pacientes con sepsis / NAC y ocupa un lugar bajo en la predicción de la mortalidad de los pacientes con sepsis. En segundo lugar, los conjuntos genéticos de alto rango (vías) podrían inmediatamente

| Métrica | Sepsis | | Sepsis/CAP | | Otra sepsis | | COVID-19 | |
|------------------------------------|--------|--------|------------|--------|-------------|--------|----------|--------|
| | Disco | Válido | Disco | Válido | Disco | Válido | Disco | Válido |
| Ruido del conjunto genético | | | | | | | | |
| Bacc | 0.799 | 0.701 | 0.802 | 0.798 | 0.779 | 0.761 | 0.963 | 0.95 |
| Sensibilidad | 0.754 | 0.6 | 0.88 | 0.867 | 0.75 | 0.8 | 0.974 | 1 |
| Especificidad | 0.845 | 0.801 | 0.725 | 0.73 | 0.807 | 0.722 | 0.951 | 0.9 |
| WGCNA eigengenes | | | | | | | | |
| Bacc | 0.805 | 0.707 | 0.8 | 0.806 | 0.82 | 0.734 | 0.963 | 0.9 |
| Sensibilidad | 0.799 | 0.614 | 0.8 | 0.867 | 0.798 | 0.769 | 0.974 | 0.9 |
| Especificidad | 0.812 | 0.8 | 0.8 | 0.746 | 0.841 | 0.7 | 0.951 | 0.9 |

Tabla 2. Precisión de predicción de los modelos para los pacientes con sepsis y COVID-19 en función del conjunto de genes ruido y variables explicativas de WGCNA eigengenes para las cohortes de descubrimiento (Disc.) y validación (Valid.).

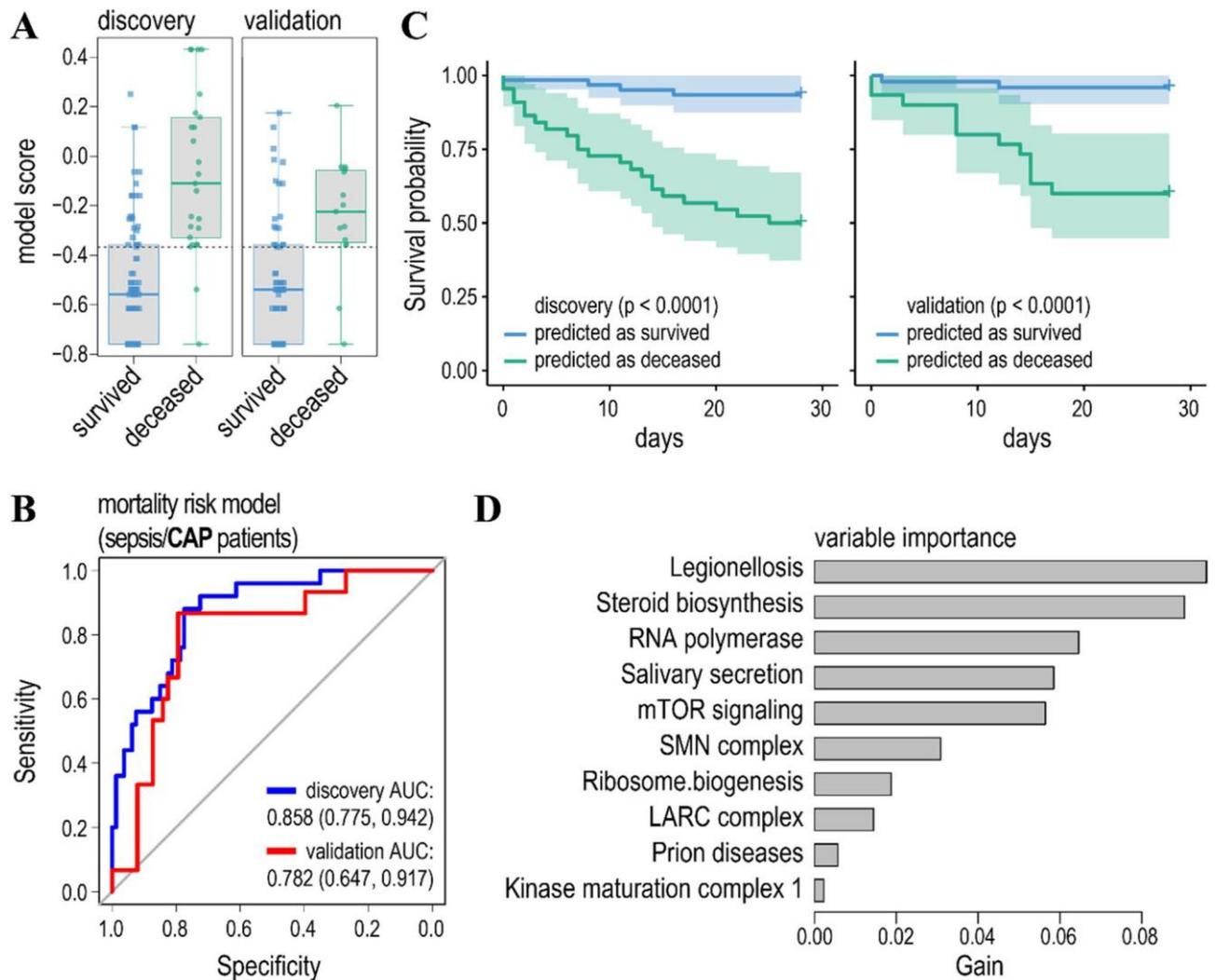


Figura 4. El modelo que predice la mortalidad/supervivencia de los pacientes con sepsis/NAC. **(A)** Diagramas de caja de las puntuaciones del modelo que predicen la mortalidad / supervivencia en las cohortes de descubrimiento (izquierda) y validación (derecha). **(B)** Curvas ROC para el modelo que predice la mortalidad en pacientes con sepsis/NAC en las cohortes de descubrimiento (línea azul) y validación (línea roja). Las cohortes se dividieron como en el estudio anterior⁸. **(C)** Probabilidad de supervivencia para los pacientes que se predice que tienen un riesgo de mortalidad bajo (línea azul) y alto (línea verde) para las cohortes de descubrimiento (panel izquierdo) y validación (panel derecho). **(D)** Contribución relativa de las características de ruido del conjunto de genes al modelo.

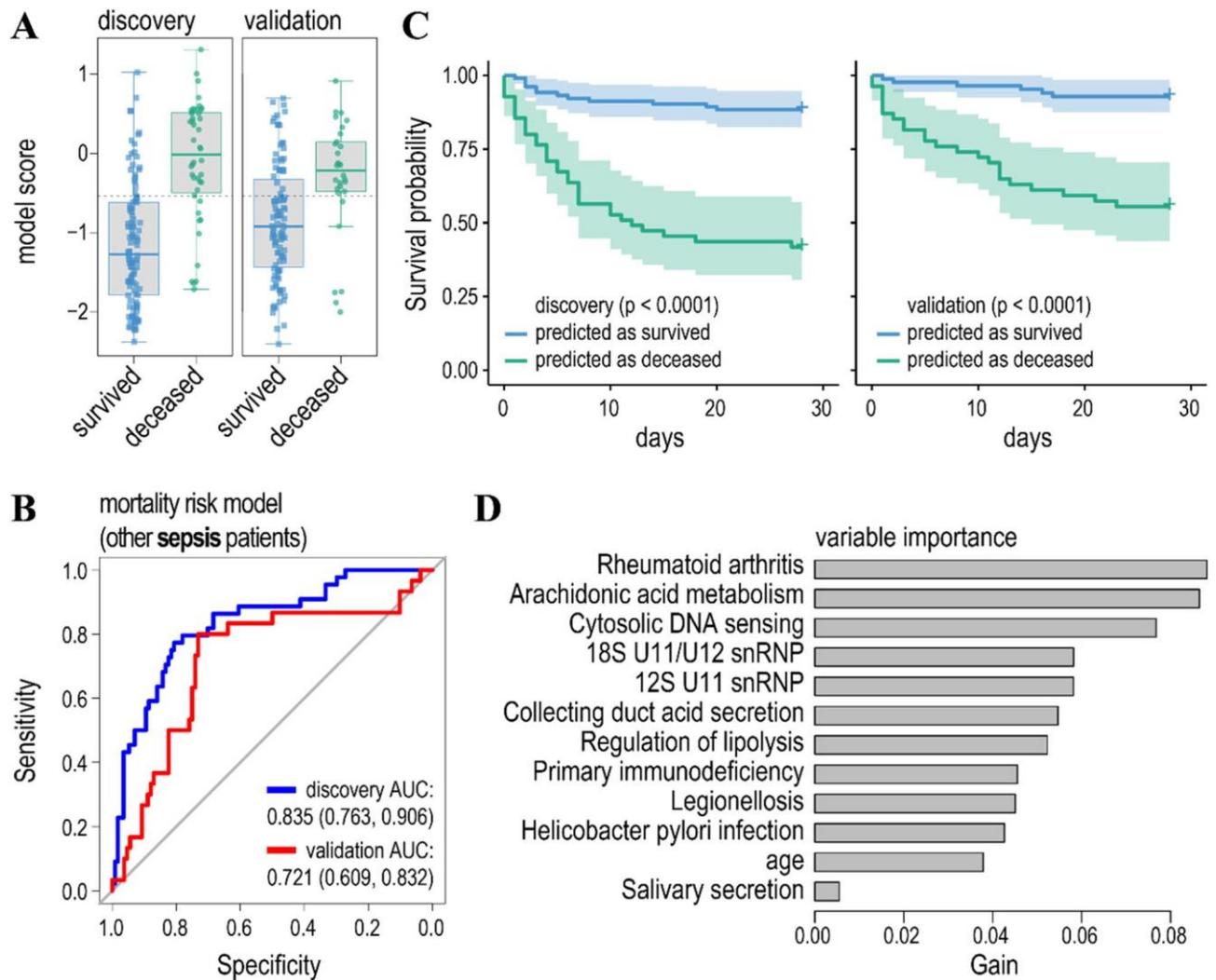


Figura 5. El modelo que predice la mortalidad/supervivencia de otros pacientes con sepsis. **(A)** Diagramas de caja de las puntuaciones del modelo. **(B)** Curvas ROC para el modelo de predicción de mortalidad en pacientes con sepsis. Las cohortes se dividieron como en el estudio anterior⁸. **(C)** Probabilidad de supervivencia para los pacientes que se predice que tienen un riesgo de mortalidad bajo (línea azul) y alto (línea verde). **(D)** Contribución relativa de las características de ruido del conjunto de genes al modelo.

estar asociado con la respuesta del huésped a la infección y, por lo tanto, a la patología de la sepsis. Estos incluyen legionelosis (una vía responsable de la neumonía atípica causada por la bacteria *Legionella*), señalización de células epiteliales en la infección por *Helicobacter pylori* y la leishmaniasis, y los desequilibrios en estas vías causados por las infecciones correspondientes o la activación inmune podrían conducir a la sepsis^{57–59}. A eso, el ruido del conjunto de genes en las vías inmunes, como la artritis reumatoide y la inmunodeficiencia primaria, contribuyen a la predicción del resultado clínico en pacientes con sepsis (Fig. 3D, 5D y Tabla S3).

Análisis de red de correlación ponderada de pacientes con sepsis. Como alternativa al ruido del conjunto de genes, se consideró el análisis de red de correlación ponderada (WGCNA) para predecir los resultados clínicos de los pacientes con sepsis. WGCNA agrupa genes en módulos distintos basados en la coexpresión adyacente y superposición topológica, y proyecta módulos al primer componente principal: eigengene₆₀. La identificación de los módulos de coexpresión para los pacientes con CAP y sepsis reveló sólo una asociación lineal débil de WGCNA eigengenes con el riesgo de mortalidad. Los valores de eigengene se compararon entre pacientes con CAP y sepsis supervivientes y fallecidos con pruebas t seguidas de un ajuste FDR de los valores p (Figura S5). A continuación, entrenamos los modelos de clasificación XGBoost para predecir la mortalidad de los pacientes con sepsis utilizando WGCNA eigengenes como variables explicativas. Curiosamente, las precisiones de clasificación resultantes de estos modelos fueron comparables a los modelos basados en las características de ruido del conjunto de genes (Tabla 2, Figura S6, Tabla S2). Esto sugiere que

losógenos propios de WGCNA contribuyen al riesgo de mortalidad de forma no lineal, ya que los conjuntos de árboles de decisión XGBoost son capaces de incorporar asociaciones no monótonas en el modelo. Al igual que con cualquier tipo de análisis DGE, los módulos WGCNA requieren GSEA para la interpretación biológica. Anotamos los módulos que se clasifican por encima del conjunto de genes no asignados (módulo 0) por la importancia variable para cada modelo utilizando anotaciones de genes KEGG y CORUM. Una prueba hipergeométrica estándar reveló una serie de enriquecimientos significativos para las anotaciones KEGG/CORUM, incluido el complejo de cadena respiratoria mitocondrial I (Complejo I) para algunos de los módulos WGCNA, incluidos (Figuras S7, S8, S9). Sin embargo, la mayor parte de los módulos WGCNA carecían de enriquecimientos KEGG/CORUM significativos que apuntaran a las limitaciones de la interpretabilidad biológica de WGCNA. El ruido del conjunto de genes anula esta limitación, ya que opera directamente a nivel de conjuntos biológicos conocidos (vías, subunidades complejas de proteínas, etc.) en lugar de genes individuales o propios.

El ruido del conjunto de genes y WGCNA predicen la gravedad de la enfermedad COVID-19 y descubren vías para la orientación farmacéutica.

Teniendo en cuenta el impacto de la pandemia de COVID-19 y la disponibilidad de datos de secuenciación de ARN leucocitario para los pacientes con COVID-1961, replicamos el ruido del conjunto de genes y el análisis WGCNA para obtener más información sobre la respuesta inmune al virus SARS-CoV-2. Los 100 pacientes con COVID-19 se asignaron en dos grupos de igual tamaño: graves y menos graves (leves), según los "días libres de hospitalización en el día 45" (HFD-45). El grupo grave corresponde a los pacientes con $HFD-45 < 26$ y los menos graves a $HFD-45 \geq 26$. Para cada grupo, evaluamos los coeficientes biológicos de las variaciones en los recuentos de ARN de los genes (bv^2) utilizando el modelo de binomio negativo (NB). Para ello, aplicamos GAMLSS (modelo aditivo generalizado para localización, escala y forma) para evaluar el impacto estadístico de la gravedad de la enfermedad COVID-19 (efecto fijo) en ambos parámetros de la distribución NB (μ —expresión media y bv^2). La edad y el sexo se incorporaron al modelo de μ como efectos aleatorios. Para los detalles de GAMLSS y su aplicación a la inferencia estadística de genes bv^2 ver ^{20,29}. Al igual que el H1N1 y la sepsis/CAP, la variabilidad interindividual de la expresión génica de leucocitos aumentó en los pacientes graves con COVID-19 (comparar la Fig. 1A con 6A). Esto indica que tanto las infecciones virales como las bacterianas causan una respuesta inmune heterogénea que se manifiesta en el aumento de bv^2 del número de copias de ARN.

El análisis del ruido del conjunto de genes reveló un impacto significativo de la enfermedad COVID-19 en muchos conjuntos anotados KEGG y CORUM (Figs. 2A, 6B, Figura S10). Además, COVID-19, H1N1 y sepsis ejercen un impacto común en el ruido del conjunto de genes del complejo de cadena respiratoria I, señalización HIF-1, peroxisoma, necroptosis, receptor similar a NOD, vías de señalización Fc epsilon RI, etc. (Fig. 2A) sugiriendo perturbaciones similares en la estequiometría/ruido genético de estos conjuntos de genes causados por una respuesta inmune. Esto, a su vez, indica que la focalización farmacéutica de estos conjuntos podría conducir al desarrollo de tratamientos comunes de primera línea en respuesta a pandemias presentes y futuras. Curiosamente, el complejo de cadena respiratoria I también se identificó como asociado con el estado de la enfermedad COVID-19 (leve / grave) en el análisis de los módulos WGCNA (Figura S11). Por lo tanto, ambos enfoques que toman rutas de análisis completamente diferentes convergieron en el mismo objetivo (Complejo I) que corrobora su papel en la enfermedad COVID-19.

Finalmente, entrenamos los modelos de árbol de regresión potenciado por gradiente para clasificar la gravedad de la enfermedad COVID-19 utilizando el ruido del conjunto de genes o los propios WGCNA como variables explicativas. Con este fin, dividimos aleatoriamente los datos en cohortes de descubrimiento del 80% (80 pacientes) y del 20% (20 pacientes) para replicar el diseño del estudio multiómico COVID-19 ⁶¹ y entrenamos los modelos XGBoost. El modelo basado en el ruido del conjunto de genes clasificó con precisión a los pacientes con COVID-19 leves y graves en las cohortes de descubrimiento y validación (Fig. 6C, Figura S12A, Tabla 2 y Tabla S4). Como tal, la precisión global del modelo (Sensibilidad, Especificidad, AUC, etc.) fue comparable a los modelos basados en los perfiles de transcriptoma, proteoma y metaboloma multiómico ⁶¹. Clasificación de la importancia variable de ruido del conjunto génico para el modelo (Fig. 6D, Tabla S5) reveló varios reguladores de la inflamación: complejo corepresor TLE1, complejo CtBP y complejo TNF α /NF- κ B 1062,63. Además, la importancia variable de los conjuntos de genes de citocromo c oxidasa y metabolismo del nitrógeno también se clasificó alto (Fig. 6D, cuadro S5). El metabolismo del nitrógeno está desregulado en los pacientes con COVID-19 ⁶⁴, mientras que las alteraciones en el ruido del conjunto del citocromo C junto con el complejo de la cadena respiratoria I apuntan además a la disfunción de la respiración mitocondrial.

La precisión del modelo entrenado con el conjunto de variables explicativas WGCNA fue comparable al modelo basado en el ruido del conjunto de genes (Figura S12A, B, Tabla 2 y Tabla S4). La clasificación de la importancia variable WGCNA eigengenes seguida del análisis de enriquecimiento KEGG/CORUM confirmó aún más una asociación del complejo de la cadena respiratoria I con la gravedad de la enfermedad COVID-19 (Figura S12C, D).

En conjunto, concluimos que el ruido del conjunto de genes proporciona nuevos enfoques y conocimientos para el descubrimiento de biomarcadores, la predicción del resultado clínico y los mecanismos moleculares de la patología desde el punto de vista de los desequilibrios en la estequiometría y el ruido genético en la expresión de conjuntos de genes biológicamente interpretables.

Discusión

Aquí intentamos una disección de los mecanismos moleculares de la patología humana, ejemplificada por la infección por SARS-CoV-2 y H1N1, y la sepsis, a través de un prisma de ruido de conjunto de genes. A

diferencia de la DGE clásica, el ruido de conjunto de genes permite la identificación de desequilibrios en la expresión de circuitos genéticos completos, en lugar de genes individuales a nivel de estequiometría y ruido genético. Este enfoque ofrece una alternativa, pero no mutuamente excluyente a la interpretación de DGE de una base molecular de la enfermedad y ambos tienen sus propias fortalezas y debilidades. Al mismo tiempo, la identificación de los mismos objetivos por ambos enfoques solo fortalece la confianza en los hallazgos, como se ejemplifica aquí mediante la comparación de los resultados del ruido del conjunto de genes y el análisis WGCNA.

Observamos en la introducción que debido a una fluctuación-respuesta una inferencia estadística de DGE podría estar sesgada hacia genes con alta variabilidad interindividual, *es decir*, genes "ruidosos"^{18,19} (Figura S1). Sin embargo, lo mismo se aplica al ruido del conjunto de genes (Figura S4). Esto impone un cierto problema a la interpretación tanto de la DGE como del ruido del conjunto de genes. Por un lado, se puede sugerir que las grandes desviaciones en la expresión de genes y conjuntos, que son naturalmente propensas a altas fluctuaciones, podrían no ser causantes de una enfermedad, ya que un organismo ya está adaptado a tales variaciones. Por otro lado, estos genes/conjuntos en sí mismos podrían desempeñar un importante papel adaptativo⁶⁵ y su sobrerrespuesta podría conducir a la enfermedad. Por el momento parece difícil llegar a una resolución entre estas dos posibilidades, pero deben ser consideradas, específicamente en la identificación de dianas farmacéuticas: genes o conjuntos de genes (vías, complejos proteicos).

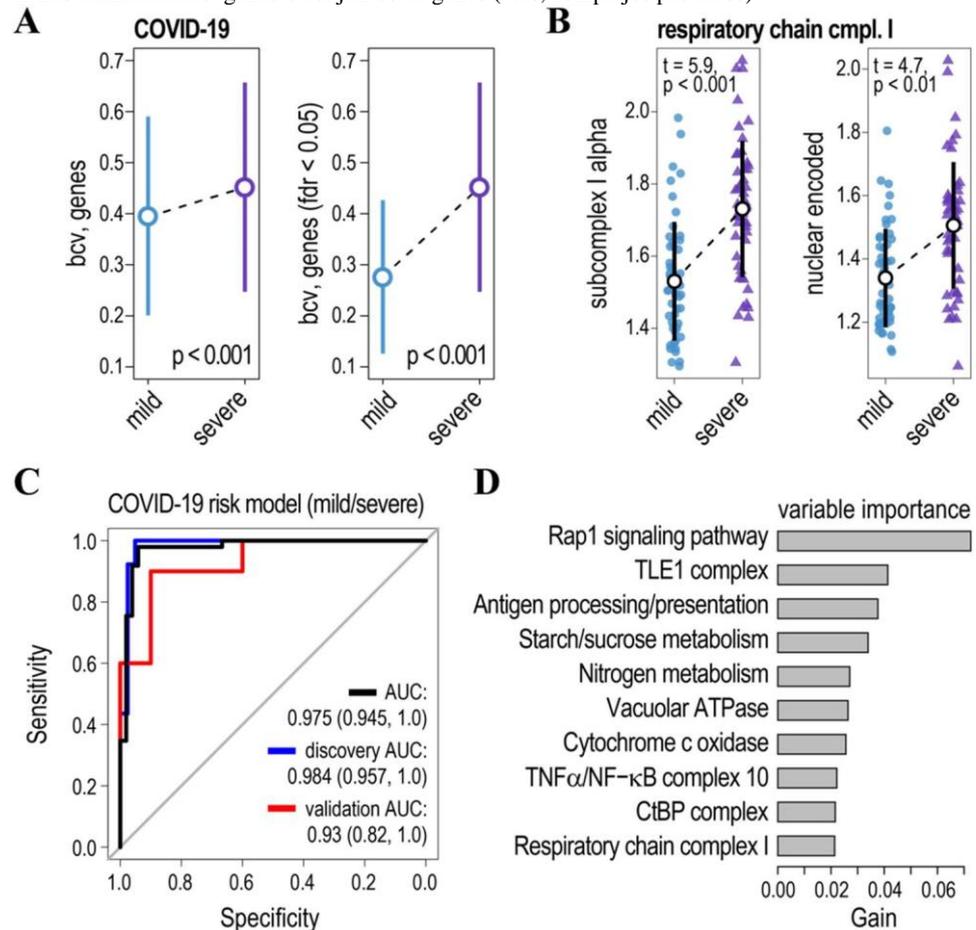


Figura 6. Asociación del ruido del conjunto de genes con el estado de la enfermedad COVID-19. **(A)** La variabilidad biológica interindividual en la expresión génica de leucocitos (bcv, coeficiente biológico de variación) aumenta en pacientes graves con COVID-19 en comparación con los leves. Panel izquierdo: estimaciones promedio del bcv para todos los genes expresados en los leucocitos de los pacientes, panel derecho: bcv para los genes con cambios significativos en la variabilidad biológica interindividual (tasa de descubrimiento falso, FDR < 0.05). *p*-valores de las pruebas t que comparan las diferencias en la variabilidad de la expresión génica interindividual para pacientes con COVID-19 leve y grave. Los círculos y los bigotes indican medias y desviaciones estándar respectivamente. **(B)** Gráficos de ruido de conjunto de genes para genes que codifican subunidades del complejo I de la cadena respiratoria mitocondrial (subcomplejo I alfa - panel izquierdo y subunidades codificadas nucleares - panel derecho). Los círculos negros y los bigotes indican medias y desviaciones estándar. Se muestran los valores t y p de las pruebas que comparan el ruido del conjunto de genes para pacientes con COVID-19 leve y grave. **(C)** Curvas ROC para el modelo basadas en el ruido del conjunto de genes para las cohortes de descubrimiento (línea azul) y validación (línea roja), y todas las muestras (línea negra). Para obtener más detalles sobre la precisión del modelo, consulte las Tablas 2 y S4.

(D) Contribución relativa de las características de ruido del conjunto de genes al modelo. Para más detalles, véase el cuadro S5.

En comparación con DGE, el ruido del conjunto de genes proporciona una interpretación holística de la mala regulación en la expresión génica bajo condiciones patológicas o de otro tipo. Como opera a nivel de conjuntos de genes, no requiere análisis de enriquecimiento de conjuntos de genes (GSEA), por lo que evita las posibles trampas de GSEA asociadas con el problema de corte de DGE^{24,25}. Al igual que con cualquier análisis de expresión génica, el ruido del conjunto de genes se basa en la calidad y la integridad de las vías y la anotación de los complejos de proteínas. Finalmente, observamos que la variabilidad interindividual del ruido del conjunto de genes es significativamente menor que la de la expresión génica individual (Figura S3). Esto, a su vez, podría mejorar la precisión de los modelos de resultados diagnósticos y clínicos. Aunque puede venir a expensas de que pocas características estén disponibles para la selección y el entrenamiento de modelos. Al mismo tiempo, en estudios futuros, se podría combinar tanto el DGE como el ruido del conjunto de genes.

En este estudio, aplicamos el concepto de ruido de conjunto génico al análisis de pacientes críticos con H1N1 y sepsis^{8,28}, así como a la enfermedad COVID-19. Observamos una respuesta génica a gran escala en dos dimensiones: en el nivel de expresión génica media y el nivel de varianza (variabilidad interindividual). Curiosamente, ambas respuestas fueron correlacionadas (Fig. 1D) y ambos dependían de la varianza génica, lo que sugiere que la fluctuación-respuesta podría impulsar cambios en estos dos parámetros de expresión génica de forma coordinada¹⁸. En todos los casos (COVID-19, H1N1 y sepsis), la variabilidad interindividual aumentó para la mayor parte de los genes. En consecuencia, solo identificamos vías o complejos genéticos para los cuales el ruido del conjunto de genes aumentó significativamente para los pacientes con H1N1 y sepsis en comparación con los individuos sanos, y para el COVID-19 en comparación entre los pacientes graves y menos graves. Es interesante observar que la desregulación de ciertas vías se ve comúnmente y es informativa para la predicción de la supervivencia del paciente en todas las enfermedades. Por lo tanto, la respuesta descoordinada a la neumonía atípica causada por la bacteria *Legionella*, así como la señalización de las células epiteliales en la infección por *Helicobacter pylori* y la leishmaniasis también podrían resaltar la deficiencia general en la respuesta inmune, lo que resulta en una mayor vulnerabilidad de los pacientes con sepsis. Nuestros resultados sugieren que la variabilidad de la expresión génica interindividual es un impulsor prominente del ruido del conjunto de genes en estos pacientes.

Debido a que las infecciones virales / bacterianas y la sepsis resultan en una respuesta abrumadora de expresión génica, es difícil identificar un conjunto razonablemente pequeño de genes o conjuntos de genes para la interpretación biológica. Por lo tanto, solo nos centramos en las vías (complejos de proteínas) para las cuales el ruido del conjunto de genes aumentó en todos los casos y las correlacionamos con un estado de enfermedad (Fig. 2A). A partir de esta intersección, inferimos 13 vías, la mayoría de las cuales han sido previamente implicadas en la sepsis (Tabla 1). Para eso, 5 vías (complejos de proteínas) mostraron una asociación significativa del ruido del conjunto de genes con la gravedad de covid-19, la fase de infección por H1N1 y el riesgo de mortalidad por sepsis (CAP, otros) (Fig. 2A). Potencialmente, estas vías podrían ser dirigidas para el tratamiento adyuvante de la sepsis. Especialmente, consideramos el complejo de cadena respiratoria mitocondrial I (Complejo I) (Fig. 2D) y peroxisoma prometedor para la orientación farmacéutica.

El aumento del ruido del conjunto de genes para el Complejo I implicaría una estequiometría alterada o un aumento del ruido de expresión génica para los genes que codifican subunidades del Complejo I o ambos. Como resultado, esto podría conducir a un ensamblaje inadecuado del Complejo I que resulte en la desregulación de la respiración mitocondrial y la manifestación del efecto Warburg. Curiosamente, el efecto Warburg es una de las señas de identidad de la sepsis y otras enfermedades inflamatorias⁶⁶. Asimismo, la disfunción mitocondrial también se evidenció a partir del reciente análisis multiómico de la enfermedad COVID-19 ya que se notó una disminución significativa en los niveles de citrato en los pacientes graves de COVID-19. La función deteriorada del Complejo I puede ser omitida por un mediador redox alternativo, como el azul de metileno^{50,51}. Para ello, el azul de metileno es un inhibidor selectivo de la vía del monofosfato de guanosina nítrico-guanosina cíclico (NO-cGMP)⁵³, y el aumento de los niveles de NO es un sello distintivo de la sepsis⁶⁷. Algunos estudios clínicos ya han indicado un papel beneficioso del azul de metileno en el tratamiento de la sepsis^{52,53}. Al igual que la respiración mitocondrial, los peroxisomas también juegan un papel importante en la patología de la sepsis, ya que la disfunción de los peroxisomas resulta en estrés oxidativo³². Una vez más, un aumento del ruido del conjunto de genes para la vía del peroxisoma indica un mecanismo potencial para dicha disfunción en pacientes con H1N1 y sepsis. Potencialmente la biogénesis del peroxisoma podría ser restaurada por el 4-fenilbutirato y existen varios estudios que indican su papel positivo en el tratamiento de la sepsis^{68,69}. Teniendo en cuenta las direcciones futuras, se podría proponer que la búsqueda de moduladores epigenéticos del ruido del conjunto de genes podría representar una nueva vía farmacéutica para los tratamientos adyuvantes de la sepsis.

Finalmente, exploramos la posibilidad de utilizar el ruido de conjunto de genes en la predicción de los resultados clínicos. Previamente se han identificado algunos biomarcadores prometedores y endotipos de expresión génica asociados con shock séptico y mortalidad a partir del análisis DGE^{8,15}. Sin embargo, como ya se mencionó, el ruido del conjunto de genes analiza la expresión génica desde un ángulo diferente, pero complementario, lo que permite la identificación de nuevas vías y biomarcadores para la sepsis y otras enfermedades. Para eso, los modelos que predicen la patología basados en el ruido del conjunto de genes podrían ser potencialmente más robustos, ya que la variabilidad interindividual para el ruido del conjunto de genes es menor que la de la expresión génica logarítmica (Figura S3). Además, los modelos de árbol de regresión potenciados por gradiente entrenados en pacientes con sepsis para predecir su mortalidad tuvieron una buena

precisión en la cohorte de validación (Fig. 3, Tabla 2). Estos superaron las predicciones basadas en el endotipo de expresión génica Mars1, que se demostró que se asocia con un mal pronóstico⁸, tanto en las cohortes de descubrimiento como de validación (Fig. 3C). Curiosamente, algunas características de ruido del conjunto de genes seleccionadas estadísticamente para los modelos que predicen la mortalidad en ambas cohortes de pacientes con sepsis podrían relacionarse inmediatamente con la respuesta del huésped a la infección. Por ejemplo, los aumentos en el ruido del conjunto de genes en la legionelosis, la señalización de las células epiteliales en la infección por *Helicobacter pylori* y las vías de leishmaniasis podrían servir potencialmente como biomarcadores de sepsis y su resultado.

También intentamos mejorar las precisiones de los modelos mediante el entrenamiento de perceptrones multicapa (MLP), un enfoque de aprendizaje profundo de vainilla, así como redes neuronales convolucionales (CNN) con Keras (<https://keras.io>). Sin embargo, las precisiones equilibradas de los modelos MLP y CNN para las cohortes de validación fueron más bajas que las de los árboles de regresión impulsados por gradiente (datos no mostrados), lo que podría deberse a un problema de muestra de aprendizaje de pequeño tamaño⁷⁰. No obstante, el aprendizaje profundo podría representar un enfoque poderoso para el análisis multiómico en el futuro, ya que recientemente se han desarrollado modelos CNN para la clasificación de tipos de tumores basados en ARN-seq⁷¹.

En conclusión, aquí mostramos el potencial del ruido del conjunto de genes en la interpretación biológica de una enfermedad, la identificación de vías farmacéuticamente dirigidas, nuevos biomarcadores y la predicción del resultado clínico. Vale la pena señalar que, si bien mostramos aplicaciones del ruido de conjunto de genes para la sepsis y COVID-19 utilizando complejos bioquímicos y vías KEGG, también se pueden emplear conjuntos de datos para otras enfermedades y tipos de conjuntos como la ontología génica, los elementos reguladores promotores comunes o los sitios miRNA. Creemos que el análisis de ruido de conjunto de genes podría aplicarse ampliamente junto con DGE para diseccionar los mecanismos moleculares de la patología en dos dimensiones complementarias: en la dimensión jacob-Monod de la regulación génica específica y una nueva dimensión de la regulación holística del circuito génico.

Métodos

Recursos y procesamiento de datos. El código informático y los modelos utilizados en este proyecto se depositan en el siguiente repositorio público: https://github.com/Vityay/GeneE_nsemb_leNoi_se.

GSE65682 Affymetrix Genoma Humano U219 Array Se utilizaron perfiles de expresión génica en sangre entera para el análisis de pacientes con sepsis por neumonía adquirida en la comunidad y en el hospital (HAP/CAP) (aquí denominada sepsis/CAP) y pacientes con sepsis sin CAP (aquí denominada otra sepsis)^{8,12}. En resumen, la cohorte consistió en 42 individuos sanos (24 hombres, 18 mujeres), 183 pacientes con sepsis / CAP (111 hombres, 72 mujeres) y otros 296 pacientes con sepsis (161 hombres, 135 mujeres). La edad media de los pacientes con sepsis/CAP ($61,5 \pm 1,2$) y otra sepsis ($60,6 \pm 0,8$) no difirió significativamente ($t(350,37) = 0,59, p = 0,56$), sin embargo, los individuos sanos fueron significativamente más jóvenes en comparación con la sepsis/CAP ($t(54,0) = 4,7, p < 0,001$) y otras sepsis ($t(47,2) = 4,6, p < 0,001$) pacientes. De los 183 pacientes con sepsis/CAP, 40 murieron en 28 días y de otros 296 pacientes con sepsis, 74 murieron en 28 días. Por lo tanto, dividimos la sepsis / CAP y otros pacientes con sepsis en grupos de sobrevivientes y fallecidos, considerando estos dos estados como una variable de factor ordenado (ordinal) (sobreviviente < fallecido). En general, la sepsis/CAP y otros pacientes con sepsis fueron similares en el tratamiento y el resultado, y otros pacientes con sepsis se sospecharon para la CAP y se trataron con antibióticos (ver Scicluna et al.¹² para la descripción detallada de las cohortes). Para el análisis se utilizaron 521 muestras de 802 por las siguientes razones: (1) para comparar nuestros modelos de riesgo de mortalidad con los publicados anteriormente¹² en las mismas cohortes de descubrimiento y validación y (2) para los 281 restantes no había datos de supervivencia disponibles.

GSE21802 Illumina human-6 v2.0 expression bead-chip perfiles de expresión génica de sangre entera se utilizaron para el análisis de pacientes infectados por H1N1²⁸. La cohorte consistió en 4 individuos sanos y 19 pacientes con H1N1 (8 en la fase temprana y 11 en la fase tardía de la enfermedad). La fase temprana se definió como temprana, desde el inicio de los síntomas, del día 0 al día 8, y tardía, desde el día 9 y superior. Las estadísticas de las cohortes se dan en la Tabla 1 del estudio original²⁸, sin embargo, ni las asignaciones de sexo ni edad estaban disponibles para los pacientes de la anotación de la serie GSE21802. Utilizamos 521 de 802 por las siguientes razones: (1) para comparar nuestros modelos de riesgo de mortalidad con el estudio publicado previamente por Scicluna, et al. 2015 sobre las mismas cohortes de descubrimiento y validación y (2) para los 281 restantes no había datos de supervivencia disponibles. Para el análisis de la enfermedad COVID-19 se utilizaron perfiles de expresión génica de ARN-seq leucocitario GSE157103. La cohorte consistió en 100 pacientes con COVID-19 divididos en grupos menos graves (leves) y graves según los días libres en el hospital como se describe⁶¹.

Las intensidades de señal de microarrays GSE65682 fueron preprocesadas (corregidas en segundo plano y normalizadas por RMA) con el paquete de oligoconductores⁷². Los genes poco expresados y atípicos se identificaron en dimensiones altas utilizando el algoritmo de signos espaciales (*sign2*) del paquete *R de mvoulter* con un valor crítico para la detección de valores atípicos en 0,9. Los componentes principales robustos explicaron la varianza de 0,9573. Las intensidades de señal GSE21802 significativamente por encima del fondo fueron cuantiles normalizadas⁷⁴. Los genes fueron anotados con bioconductor *hgu219.db* e *illuminaHumanv2.db* paquetes de base de datos para GSE65682 (8826 genes) y GSE21802 (7240 genes) respectivamente. Los recuentos de GSE157103 RNA-seq se analizaron con GAMLSS como se describió anteriormente²⁰.

Análisis estadístico de la variabilidad de la expresión génica y el ruido del conjunto génico. El

análisis estadístico se realizó utilizando paquetes R y R/Bioconductor.

Para estimar la variabilidad de la expresión génica interindividual para pacientes sanos y con sepsis, se tuvo en cuenta la edad como un efecto aleatorio. Para ello, se utilizó el Modelo Aditivo Generalizado para Ubicación, Escala y Forma (GAMLSS)^{20,29}. En resumen, para las intensidades de microarrays transformadas en logarítmica normalmente distribuidas ($Y = \log(X)$, $Y \sim N(\mu_Y, \sigma_Y)$), GAMLSS permite el modelado de ambos parámetros de expresión génica (media y dispersión):

$$\mu_Y \sim D \beta_\mu + Z u_\mu, \log(\sigma_Y)$$

$$\sim D \beta_\sigma + Z u_\sigma,$$

donde $\mu_Y = (\mu_{Y_1}, \dots, \mu_{Y_n})^T$ y $\sigma_Y = (\sigma_{Y_1}, \dots, \sigma_{Y_n})^T$ son los vectores de medias y dispersiones para

$\beta_\mu = (\beta_{\mu_1}, \dots, \beta_{\mu_k})^T$ and $\beta_\sigma = (\beta_{\sigma_1}, \dots, \beta_{\sigma_k})^T$ para el estado de la enfermedad (sano, sobrevivido, fallecido).

$Y = (Y_1, \dots, Y_n)$. $D = n \times p$ matriz de diseño de efectos fijos para el estado de la enfermedad (sano, sobrevivido, fallecido). T_{rr} — coeficientes de efecto fijo estimados para la media y la dispersión. $Z = n \times k$ matriz de diseño de efectos aleatorios para la edad (la edad se unió en 10 deciles). $u_\mu = (u_{\mu_1}, \dots, u_{\mu_k})^T$ y $u_\sigma = (u_{\sigma_1}, \dots, u_{\sigma_k})^T$ — coeficientes de efecto aleatorio estimados para media y dispersión, donde $u \sim N(0, \delta)$. Con GAMLSS también es sencillo probar la importancia de un efecto factorial en la media, la varianza o ambas con la prueba de razón de verosimilitud²⁰.

Las listas de conjuntos de genes se generaron mediante el mapeo de genes a las vías biológicas anotadas por KEGG o subunidades anotadas por CORUM de complejos de proteínas de mamíferos^{22,23}. Su ruido genético se estimó para cada individuo calculando las varianzas de las expresiones de genes transformadas en logaríto para cada conjunto (Fig. 1S). Las estimaciones del ruido del conjunto de genes se correlacionaron con los estados de la enfermedad (< fase temprana < fase tardía para el H1N1 y < sanos sobrevivieron < fallecidos por sepsis) mediante la correlación de rango de Kendall, tratando el estado de la enfermedad como una variable ordinal. Las tendencias lineales entre los estados de enfermedad y el ruido del conjunto de genes se estimaron mediante regresión basada en rangos⁷⁵.

Modelos de árbol de regresión potenciados por gradiente. Para predecir la mortalidad de los pacientes con sepsis, entrenamos modelos de árbol de regresión potenciados por gradiente con un sistema escalable de aumento de árboles XGBoost⁵⁴ utilizando la mortalidad dentro de los 28 días como variable de respuesta binaria, y el ruido del conjunto de genes y la edad como características independientes del modelo. Con este fin, dividimos a los individuos en cohortes de descubrimiento y validación siguiendo la misma partición que se anota en GSE65682. Luego, entrenamos 3 modelos: (1) un modelo que predice la mortalidad para todos los pacientes con sepsis, (2) un modelo que predice la mortalidad para los pacientes con sepsis / CAP y (3) un modelo que predice la mortalidad para otros pacientes con sepsis. Las características del modelo se preseleccionaron utilizando cohortes de descubrimiento mediante la prueba t que compara el ruido del conjunto de genes para pacientes sobrevivientes y fallecidos para maximizar la precisión del entrenamiento XGBoost en los conjuntos de datos de descubrimiento. Para los modelos de sepsis por todas las causas (1) y sepsis sin CAP (3), el punto de corte para las características del modelo se estableció en $p \leq 0,01$, y para el modelo de sepsis/CAP (2), en $p \leq 0,05$. Los parámetros de hiperajuste de XGBoost (tasa de aprendizaje (η), complejidad (γ), profundidad, etc.) se optimizaron mediante validación cruzada. Para evitar el sobreajuste, encontramos parámetros de detención de época temprana dividiendo aleatoriamente la cohorte de descubrimiento en dos pliegues iguales: entrenamiento y prueba. Luego, las cohortes de validación, que se ocultaron de la selección de características y el entrenamiento del modelo, se utilizaron para verificar la precisión de los modelos finales.

Recibido: 3 de diciembre de 2020; Aprobado: 7 mayo 2021

Published online: 24 May 2021

Referencias

1. Trifonov, V., Khiabani, H. & Rabadan, R. Dependencia geográfica, vigilancia y orígenes del virus de la influenza A (H1N1) de 2009. *N. Inglés J. Med.* **361**, 115–119 (2009).
2. Grupo de Estudio Coronaviridae del Comité Internacional de Taxonomía de Virus. La especie Coronavirus relacionado con el síndrome respiratorio agudo severo: clasificación de 2019-nCoV y denominación SARS-CoV-2. *NAT. Microbiol.* **5**, 536–544 (2020).
3. van Asten, L. *et al.* Estimación de la gravedad de las epidemias de influenza por infecciones respiratorias agudas graves (IRAG) en unidades de cuidados intensivos. *Crit. Atención* **22**, 351 (2018).
4. Alhazzani, W. *et al.* Campaña Sobrevivir a la Sepsis: pautas sobre el manejo de adultos gravemente enfermos con enfermedad por coronavirus 2019 (COVID-19). *Medicina de Cuidados Intensivos.* **46**, 854–887 (2020).
5. Liu, V. *et al.* Muertes hospitalarias en pacientes con sepsis de 2 cohortes independientes. *JAMA* **312**, 90 (2014).
6. Angus, D.C. & van der Poll, T. Sepsis severa y shock séptico. *N. Inglés J. Med.* **369**, 840–851 (2013).
7. Cohen, J. *et al.* Sepsis: una hoja de ruta para futuras investigaciones. *Lancet Infect. Dis.* **15**, 581–614 (2015).
8. Scicluna, B. P. *et al.* Clasificación de pacientes con sepsis según el endotipo genómico sanguíneo: un estudio de cohorte prospectivo. *Lancet Respir. Med.* **5**, 816–826 (2017).
9. Maslove, D.M. & Wong, H. R. Perfil de expresión génica en sepsis: tiempo, tejido y consideraciones traslacionales. *Tendencias Mol. Med.* **20**, 204–213 (2014).

10. McHugh, L. *et al.* Un ensayo de respuesta molecular del huésped para discriminar entre sepsis e inflamación sistémica negativa a la infección en pacientes críticamente enfermos: descubrimiento y validación en cohortes independientes. *PLoS Med.* **12**, e1001916 (2015).
11. Parlato, M. & Cavaillon, J.-M. *Biomarcadores de respuesta del huésped en el diagnóstico de sepsis: una visión general* 149–211 (Humana Press, 2015). https://doi.org/10.1007/978-1-4939-1776-1_15.
12. Scicluna, B. P. *et al.* Un biomarcador molecular para diagnosticar la neumonía adquirida en la comunidad al ingreso en la unidad de cuidados intensivos. *AM. J. Respir. Crit. Cuidado Médico.* **192**, 826–835 (2015).
13. Sweeney, T. E., Shidham, A., Wong, H. R. & Khatri, P. Un análisis multicohorte integral basado en el curso del tiempo de la sepsis y la inflamación estéril revela un sólido conjunto de genes de diagnóstico. *Sci. Trad. Med.* **7**, 287ra71 (2015).
14. Sweeney, T. E. & Wong, H. R. Estratificación del riesgo y pronóstico en la sepsis: ¿Qué hemos aprendido de los microarrays?. *Clin. Medicina de pecho.* **37**, 209–218 (2016).
15. Hu, Y., Cheng, L., Zhong, W., Chen, M. & Zhang, Q. Análisis bioinformático de perfiles de expresión génica para la predicción del riesgo en pacientes con shock séptico. *Med. Sci. Monit.* **25**, 9563–9571 (2019).
16. Peluffo, A. E. El "Programa genético": Detrás de la génesis de una metáfora influyente. *Genética* **200**, 685–696 (2015).
17. Raser, J.M. & O'Shea, E. K. Noise in gene expression: origins, consequences, and control. *Ciencia* **309**, 2010–2013 (2005).
18. Sato, K., Ito, Y., Yomo, T. & Kaneko, K. Sobre la relación entre fluctuación y respuesta en sistemas biológicos. *Proc. Natl. Acad. Sci.* **100**, 14086–14090 (2003).
19. Lehner, B. & Kaneko, K. Fluctuación y respuesta en biología. *Celda. Mol. Ciencia de la vida.* **68**, 1005–1010 (2011).
20. de Jong, T. V., Moshkin, Y.M. & Guryev, V. Variabilidad de la expresión génica: la otra dimensión en el análisis del transcriptoma. *Fisiol. Genómica* **51**, 145–158 (2019).
21. Ashburner, M. *et al.* Ontología génica: herramienta para la unificación de la biología. *NAT. Jineta.* **25**, 25–29 (2000).
22. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG como recurso de referencia para la anotación de genes y proteínas. *Nucl. Ácidos Res.* **44**, D457–D462 (2016).
23. Giurgiu, M. *et al.* CORUM: el recurso integral de los complejos de proteínas de mamíferos—2019. *Nucl. Ácidos Res.* **47**, D559–D563 (2019).
24. Pan, K.-H., Lih, C.-J. & Cohen, S. N. Effects of threshold choice on biological conclusions reached during analysis of gene expression by DNA microarrays. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 8961–8965 (2005).
25. Simillion, C., Liechti, R., Lischer, H. E. L., Ioannidis, V. & Bruggmann, R. Evitar las trampas del análisis de enriquecimiento de conjuntos de genes con SetRank. *BMC Bioinform.* **18**, 151 (2017).
26. McCarthy, D. J., Chen, Y. & Smyth, G. K. Análisis de expresión diferencial de experimentos multifactoriales de ARN-Seq con respecto a la variación biológica. *Nucl. Ácidos Res.* **40**, 4288–4297 (2012).
27. Zhao, S., Fung-Leung, W.-P., Bittner, A., Ngo, K. & Liu, X. Comparación de RNA-Seq y microarray en el perfil del transcriptoma de las células T activadas. *PLoS ONE* **9**, 78644 (2014).
28. Bermejo-Martin, J. F. *et al.* Deficiencia de inmunidad adaptativa del huésped en la influenza pandémica grave. *Crit. Cuidado* **14**, R167 (2010).
29. Rigby, R. A. & Stasinopoulos, D.M. Modelos aditivos generalizados para ubicación, escala y forma (con discusión). *J. R. Stat. Soc. Ser. C Appl. Estadísticas.* **54**, 507–554 (2005).
30. Fitzpatrick, S. F. Inmunometabolismo y sepsis: ¿un papel para HIF?. *Frente. Mol. Biosci.* **6**, 85 (2019).
31. Vanderhaeghen, T., Vandewalle, J. & Libert, C. Factores inducibles por hipoxia en la reprogramación metabólica durante la sepsis. *FEBS J.* **287**, 1478–1495 (2020).
32. Vasko, R. Peroxisomas y lesión renal. *Antioxidante. Señal Redox.* **25**, 217–231 (2016).
33. Choi, M. E., Price, D. R., Ryter, S. W. & Choi, A.M. K. Necroptosis: a crucial pathogenic mediator of human disease. *JCI insight* **4**, (2019).
34. Sauler, M., Bazan, I. S. & Lee, P. J. Muerte celular en el pulmón: el eje apoptosis-necroptosis. *Annu. Rev. Physiol.* **81**, 375–402 (2019).
35. Zhou, H. *et al.* La activación de la señalización TLR y NOD confiere al huésped una protección mediada por la inmunidad innata contra la infección microbiana. *Frente. Immunol.* **9**, 3082 (2018).
36. da PinheiroSilva, F. *et al.* CD16 promueve la sepsis por *Escherichia coli* a través de una vía inhibidora de la gamma FcR que previene la fagocitosis y facilita la inflamación. *NAT. Med.* **13**, 1368–1374 (2007).
37. Yang, Y.-X. & Li, L. Identificación de biomarcadores potenciales de sepsis mediante análisis bioinformático. *Exp. Ther. Med.* **13**, 1689–1696 (2017).
38. Yin, X., Xin, H., Mao, S., Wu, G. & Guo, L. El papel de la autofagia en la sepsis: protección y lesión de órganos. *Frente. Fisiol.* **10**, 1071 (2019).
39. Su, L. *et al.* Cambios dinámicos en los perfiles de concentración de aminoácidos en pacientes con sepsis. *PLoS ONE* **10**, e021933 (2015).
40. Jung, W. J. *et al.* Niveles de glucagón, gravedad de la enfermedad y resultado de sepsis grave. *Shock* **43**, 563–568 (2015).
41. Baumgartner, M. R. *et al.* Directrices propuestas para el diagnóstico y manejo de la acidemia metilmalónica y propiónica. *Orphanet J. Rare Dis.* **9**, 130 (2014).
42. Truong, K. K., Lam, M. T., Grandner, M. A., Sasso, C. S. & Malhotra, A. El tiempo importa: ritmo circadiano en la sepsis, enfermedad pulmonar obstructiva, apnea obstructiva del sueño y cáncer. *Ann. AM. Thorac. Soc.* **13**, 1144–1154 (2016).
43. Pacheco, R., Contreras, F. & Zouali, M. El sistema dopaminérgico en enfermedades autoinmunes. *Frente. Immunol.* **5**, 117 (2014).
44. Feketeova, E. *et al.* Control dopaminérgico de la inflamación y la glucemia en sepsis y diabetes. *Frente. Immunol.* **9**, 943 (2018).
45. Lee, C.W.-S., Chen, H.-J., Liang, J.-A. & Kao, C.-H. Riesgo de sepsis en pacientes con esclerosis lateral amiotrófica: un estudio de cohorte retrospectivo basado en la población en Taiwán. *BMJ Open* **7**, e013761 (2017).
46. Brealey, D. & Singer, M. Disfunción mitocondrial en la sepsis. *Curr. Infecta. Dis. Representante.* **5**, 365–371 (2003).
47. Mukhopadhyay, S. *et al.* El metanálisis transcriptómico revela una regulación al alza de la expresión génica funcional en la diferenciación de osteoclastos en el shock séptico humano. *PLoS ONE* **12**, e0171689 (2017).
48. Yoseph, B. P. *et al.* Mecanismos de disfunción de la barrera intestinal en la sepsis. *Shock* **46**, 52–59 (2016).

49. Sharma, L., Lu, J. & Bai, Y. Complejo respiratorio mitocondrial i: estructura, función e implicación en enfermedades humanas. *Curr. Med. Chem.* **16**, 1266–1277 (2009).
50. Tretter, L., Horvath, G., Hölgyesi, A., Essek, F. & Adam-Vizi, V. La generación mejorada de peróxido de hidrógeno acompaña los efectos bioenergéticos beneficiosos del azul de metileno en mitocondrias cerebrales aisladas. *Radic libre. Biol. Med.* **77**, 317–330 (2014).
51. Gureev, A. P., Shaforostova, E. A., Popov, V. N. & Starkov, A. A. El azul de metileno no evita el bloqueo de antimicina del Complejo III en las mitocondrias cerebrales de ratón. *FEBS Lett.* **593**, 499–503 (2019).
52. Kwok, E. S. H. & Howes, D. Uso de azul de metileno en sepsis: una revisión sistemática. *J. Cuidados Intensivos Med.* **21**, 359–363 (2006).
53. Jang, D. H., Nelson, L. S. & Hoffman, R. S. Azul de metileno para choque distributivo: un nuevo uso potencial de un antídoto viejo. *J. Med. Toxicol.* **9**, 242–249 (2013).
54. Chen, T. & Guestrin, C. XGBoost. en *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (ACM, 2016). <https://doi.org/10.1145/2939672.2939785>
55. Davenport, E. E. *et al.* Panorama genómico de la respuesta individual del huésped y los resultados en la sepsis: un estudio de cohorte prospectivo. *Lancet Respir. Med.* **4**, 259–271 (2016).
56. Valencia-Zapata, G. A., Mejia, D., Klimeck, G., Zentner, M. & Ersoy, O. A Statistical Approach to Increase Classification Accuracy in Supervised Learning Algorithms. (2017).
57. Demello, D., Kierol-Andrews, L. & Scalise, P. J. Sepsis grave y síndrome de dificultad respiratoria aguda por neumonía por legionela adquirida en la comunidad: informe de caso. *AM. J. Crit. Atención* **16**, 320 (2007).
58. Han, X. Y., Tarrand, J. J., Dickey, B. F. & Esteva, F. J. Helicobacteremia pylori con síndrome de sepsis. *J. Clin. Microbiol.* **48**, 4661–4663 (2010).
59. Endris, M. *et al.* Sepsis bacteriana en pacientes con leishmaniasis visceral en el noroeste de Etiopía. *Biomed Res. Int.* **2014**, págs. 1–7 (2014).
60. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559 (2008).
61. Overmyer, K. A. *et al.* Análisis multiómico a gran escala de la gravedad de la COVID-19. *Cell Syst.* **12**, 23–40.e7 (2021).
62. Shen, Y. *et al.* El estado bioenergético regula las respuestas inflamatorias innatas a través del co-represor transcripcional CtBP. *NAT. Comunidad.* **8**, 624 (2017).
63. Ramasamy, S. *et al.* El supresor tumoral Tle1 regula negativamente la inflamación in vivo y modula la vía inflamatoria NF- κ B. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 1871–1876 (2016).
64. Thomas, T. *et al.* La infección por COVID-19 produce alteraciones de la vía de la quinurenina y el metabolismo de los ácidos grasos que se correlacionan con los niveles de IL-6 y el estado renal. *medRxiv* 2020.05.14.20102491 (2020). <https://doi.org/10.1101/2020.05.14.20102491>
65. Blake, W. J. *et al.* Consecuencias fenotípicas del ruido transcripcional mediado por promotores. *Mol. Celda* **24**, 853–865 (2006).
66. Yang, L. *et al.* PKM2 regula el efecto Warburg y promueve la liberación de HMGB1 en la sepsis. *NAT. Comunidad.* **5**, 4436 (2014).
67. Fernandes, D. & Assreuy, J. Óxido nítrico y reactividad vascular en sepsis. *Shock* **30**, 10–13 (2008).
68. Wei, H., Kemp, S., McGuinness, M.C., Moser, A.B. & Smith, K. D. Inducción farmacológica de peroxisomas en trastornos de biogénesis de peroxisomas. *Ann. Neurol.* **47**, 286–296 (2000).
69. Chen, X. *et al.* La hemo oxigenasa-1 reduce el estrés del retículo endoplásmico inducido por sepsis y la lesión pulmonar aguda. *Mediadores Inflamamm.* **2018**, 1–10 (2018).
70. Galushkin, A. I. Métodos de resolución de problemas en la base lógica de la red neuronal. en *Neural Networks Theory* 341–376 (Springer, 2007). https://doi.org/10.1007/978-3-540-48125-6_18
71. Lyu, B. & Haque, A. Clasificación de tipos de tumores basada en el aprendizaje profundo utilizando datos de expresión génica. en *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics* 89–96 (ACM, 2018). <https://doi.org/10.1145/3233547.3233588>
72. Carvalho, B. S. & Irizarry, R. A. A framework for oligonucleotide microarray preprocessing. *Bioinformática* **26**, 2363–2367 (2010).
73. Filzmoser, P., Maronna, R. & Werner, M. Identificación de valores atípicos en altas dimensiones. *Informática. Estadísticas. Datos Anales.* **52**, 1694–1711 (2008).
74. Bolstad, B.M., Irizarry, R., Astrand, M. & Speed, T. P. Una comparación de métodos de normalización para datos de matriz de oligonucleótidos de alta densidad basada en varianza y sesgo. *Bioinformática* **19**, 185–193 (2003).
75. Kloke, J. D., McKean, J. W. & Rashid, M.M. Estimación basada en rangos e inferencias asociadas para modelos lineales con errores correlacionados con clústeres. *J. Am. Estadísticas. Assoc.* **104**, 384–390 (2009).

Agradecimientos

Agradecemos a Oleg Derkach que motivó este trabajo y al Dr. Laurent Schwartz, Olga y Errol Fontanellaz que señalaron el uso potencial del azul de metileno. Este trabajo ha sido apoyado por la subvención de la Fundación Rusa de Ciencias: 20-14-00055 a YMM y Gene Learning Association.

Contribuciones de los autores

Y.M.M. concibió y diseñó el estudio. T.D.J., V.G. e Y.M.M. realizaron el análisis de datos, la interpretación y el dibujo de cifras. Y.M.M. escribió el manuscrito. T.D.J., V.G. e Y.M.M. revisaron y editaron el manuscrito.

Intereses contrapuestos

Los autores no declaran intereses contrapuestos.

Información adicional

Información complementaria La versión en línea contiene material complementario disponible en <https://doi.org/10.1038/s41598-021-90192-9>.

La correspondencia y las solicitudes de materiales deben dirigirse a V.G. o Y.M.M.

La información sobre reimpresiones y permisos está disponible en www.nature.com/reprints.

Nota del editor Springer Nature se mantiene neutral con respecto a las reclamaciones jurisdiccionales en los mapas publicados y las afiliaciones institucionales.



Acceso abierto Este artículo está licenciado bajo una licencia Creative Commons Atribución 4.0 Internacional

Licencia, que permite el uso, intercambio, adaptación, distribución y reproducción en cualquier medio o formato, siempre que dé el crédito adecuado al autor (s) original (s) y la fuente, proporcione un enlace a la licencia Creative Commons e indique si se realizaron cambios. Las imágenes u otro material de terceros en este artículo se incluyen en la licencia Creative Commons del artículo, a menos que se indique lo contrario en una línea de crédito al material. Si el material no está incluido en la licencia Creative Commons del artículo y su uso previsto no está permitido por la regulación legal o excede el uso permitido, deberá obtener el permiso directamente del titular de los derechos de autor. Para ver una copia de esta licencia, visite <http://creativecommons.org/licenses/by/4.0/>.

© Autor(es) 2021