

El ARN del SARS-CoV-2 de transcripción inversa puede integrarse en el genoma de células humanas cultivadas y puede expresarse en tejidos derivados de pacientes

Liguo Zhang , Alexsia Richards ,
Stephen H. Hughes ,
Jaenisch

M. Inmaculada Barrasa ,
Richard A. Young y Rudolf

[Ver todos los autores y afiliaciones](#)

PNAS 25 de mayo de

2021 118 (21) e2105968118; <https://doi.org/10.1073/pnas.2105968118>

1. Contribuido por Rudolf Jaenisch, 19 de abril de 2021 (enviado para revisión el 29 de marzo de 2021; revisado por Anton Berns y Anna Marie Skalka)

- [Artículo](#)
- [Figuras y SI](#)
- [Información y métricas](#)
- [PDF](#)

Significado

Un problema no resuelto de la enfermedad del SARS-CoV-2 es que los pacientes a menudo permanecen positivos para el ARN viral detectado por PCR muchas semanas después de la infección inicial en ausencia de evidencia de replicación viral. Mostramos aquí que el ARN del SARS-CoV-2 puede transcribirse de forma inversa e integrarse en el genoma de la célula infectada y expresarse como transcripciones quiméricas que fusionan secuencias virales con celulares. Es importante destacar que tales transcripciones quiméricas se

detectan en tejidos derivados de pacientes. Nuestros datos sugieren que, en algunos tejidos de pacientes, la mayoría de todas las transcripciones virales se derivan de secuencias integradas. Nuestros datos brindan una idea de las consecuencias de las infecciones por SARS-CoV-2 que pueden ayudar a explicar por qué los pacientes pueden continuar produciendo ARN viral después de la recuperación.

Resumen

La detección prolongada del ARN del coronavirus 2 (SARS-CoV-2) del síndrome respiratorio agudo severo y la recurrencia de las pruebas de PCR positivas se han informado ampliamente en pacientes después de la recuperación de COVID-19, pero algunos de estos pacientes no parecen diseminar el virus infeccioso. Investigamos la posibilidad de que los ARN del SARS-CoV-2 se puedan transcribir de forma inversa e integrar en el ADN de las células humanas en cultivo y que la transcripción de las secuencias integradas podría explicar algunas de las pruebas de PCR positivas observadas en los pacientes. En apoyo de esta hipótesis, encontramos que las copias de ADN de las secuencias del SARS-CoV-2 pueden integrarse en el genoma de las células humanas infectadas. Encontramos duplicaciones de sitios de destino que flanquean las secuencias virales y las secuencias de reconocimiento de la endonucleasa LINE1 de consenso en los sitios de integración, consistentes con una LINE1 mediada por retrotransposón, mecanismo de retroposición y transcripción inversa cebado por la diana. También encontramos, en algunos tejidos derivados de pacientes, evidencia que sugiere que una gran fracción de las secuencias virales se transcribe a partir de copias de ADN integradas de secuencias virales, generando transcripciones quiméricas de virus-huésped. La integración y transcripción de secuencias virales puede contribuir así a la detección de ARN viral por PCR en pacientes después de la infección y la recuperación clínica. Debido a que hemos detectado solo secuencias subgenómicas derivadas principalmente del extremo 3' del genoma viral integrado en el ADN de la célula huésped, el virus infeccioso no se puede producir a partir de las secuencias integradas del SARS-CoV-2 subgenómico. evidencia que sugiere que una gran fracción de las secuencias virales se transcribe a partir de copias de ADN integradas de secuencias virales, generando transcripciones quiméricas

de virus-huésped. La integración y transcripción de secuencias virales puede contribuir así a la detección de ARN viral por PCR en pacientes después de la infección y la recuperación clínica. Debido a que hemos detectado solo secuencias subgenómicas derivadas principalmente del extremo 3' del genoma viral integrado en el ADN de la célula huésped, el virus infeccioso no se puede producir a partir de las secuencias integradas del SARS-CoV-2 subgenómico.

- [SARS-CoV-2](#)
 - [transcripción inversa](#)
 - [LÍNEA 1](#)
 - [integración genómica](#)
 - [ARN quiméricos](#)
-

Se han informado pruebas de PCR continuas o recurrentes positivas para el síndrome respiratorio agudo severo por coronavirus 2 (SARS-CoV-2) en muestras tomadas de pacientes semanas o meses después de la recuperación de una infección inicial ([1](#) [↓](#) [↓](#) [↓](#) [↓](#) [↓](#) [↓](#) [↓](#) [↓](#) [↓](#) [↓](#) [↓](#) [↓](#) [↓](#) [↓](#) - [17](#)). Aunque recientemente se ha informado de una reinfección genuina con SARS-CoV-2 después de la recuperación ([18](#)), los estudios basados en cohortes con sujetos mantenidos en estricta cuarentena después de recuperarse de COVID-19 sugirieron que al menos algunos casos "re-positivos" no fueron causada por reinfección ([19](#) , [20](#)). Además, no se aisló ni se propagó ningún virus competente para la replicación de estos pacientes con PCR positiva ([1](#) [↓](#) - [3](#) , [5](#) , [6](#) , [12](#) , [16](#)), y la causa de la producción prolongada y recurrente de ARN viral sigue siendo desconocida. El SARS-CoV-2 es un virus de ARN de cadena positiva. Al igual que otros beta-coronavirus (SARS-CoV-1 y coronavirus relacionado con el síndrome respiratorio de Oriente Medio), el SARS-CoV-2 emplea una ARN polimerasa dependiente de ARN para replicar su ARN genómico y transcribir ARN subgenómicos ([21](#) [↓](#) [↓](#) - [24](#)). Una posible explicación para la detección continua del ARN viral del SARS-CoV-2 en ausencia de reproducción del virus es que, en algunos casos, las copias de ADN de los ARN subgenómicos virales pueden integrarse en el ADN de la célula huésped mediante un mecanismo de transcripción inversa. La transcripción de las copias de ADN integradas podría ser responsable de las pruebas de PCR positivas mucho después de que se eliminó la infección inicial. De hecho, se han detectado secuencias de virus de ARN no retrovirales en los genomas de muchas especies de vertebrados ([25](#) , [26](#)), con varias integraciones que exhiben señales consistentes con la integración de copias de ADN de ARNm virales en la línea germinal a través de antiguos retrotransposones de elementos nucleares intercalados (LINE). (revisado en ref. [27](#)). Además, los virus de ARN no retrovirales, como el virus de la estomatitis vesicular o el virus

de la coriomeningitis linfocítica (LCMV), se pueden transcribir de forma inversa en copias de ADN mediante una transcriptasa inversa endógena (RT), y se ha demostrado que las copias de ADN de las secuencias virales se integran en el ADN del huésped. celdas ([28](#) ↓ - [30](#)). Además, se ha demostrado que los ARN celulares, por ejemplo, las transcripciones de *APP* humanas , se transcriben de forma inversa mediante RT endógena en neuronas con los fragmentos de *APP* resultantes integrados en el genoma y expresados ([31](#)). Los elementos LINE1 humanos (~17% del genoma humano), un tipo de retrotransposones autónomos, que son capaces de retro-transponerse a sí mismos y a otros elementos no autónomos como Alu, son una fuente de RT celular endógena ([32](#) ↓ - [34](#)). Se ha demostrado que los elementos LINE1 endógenos se expresan en tejidos humanos envejecidos ([35](#)) y la retrotransposición somática mediada por LINE1 es común en pacientes con cáncer ([36](#) , [37](#)). Además, la expresión de LINE1 endógeno y otros retrotransposones en las células hospedadoras comúnmente se regula al alza tras la infección viral, incluida la infección por SARS-CoV-2 ([38](#) ↓ - [40](#)).

En este estudio, mostramos que las secuencias de SARS-CoV-2 pueden integrarse en el genoma de la célula huésped mediante un mecanismo de retroposición mediado por LINE1. Proporcionamos evidencia de que las secuencias virales integradas se pueden transcribir y que, en algunas muestras de pacientes, la mayoría de las transcripciones virales parecen derivar de secuencias virales integradas.

Resultados

Integración de secuencias de SARS-CoV-2 en el ADN de células huésped en cultivo.

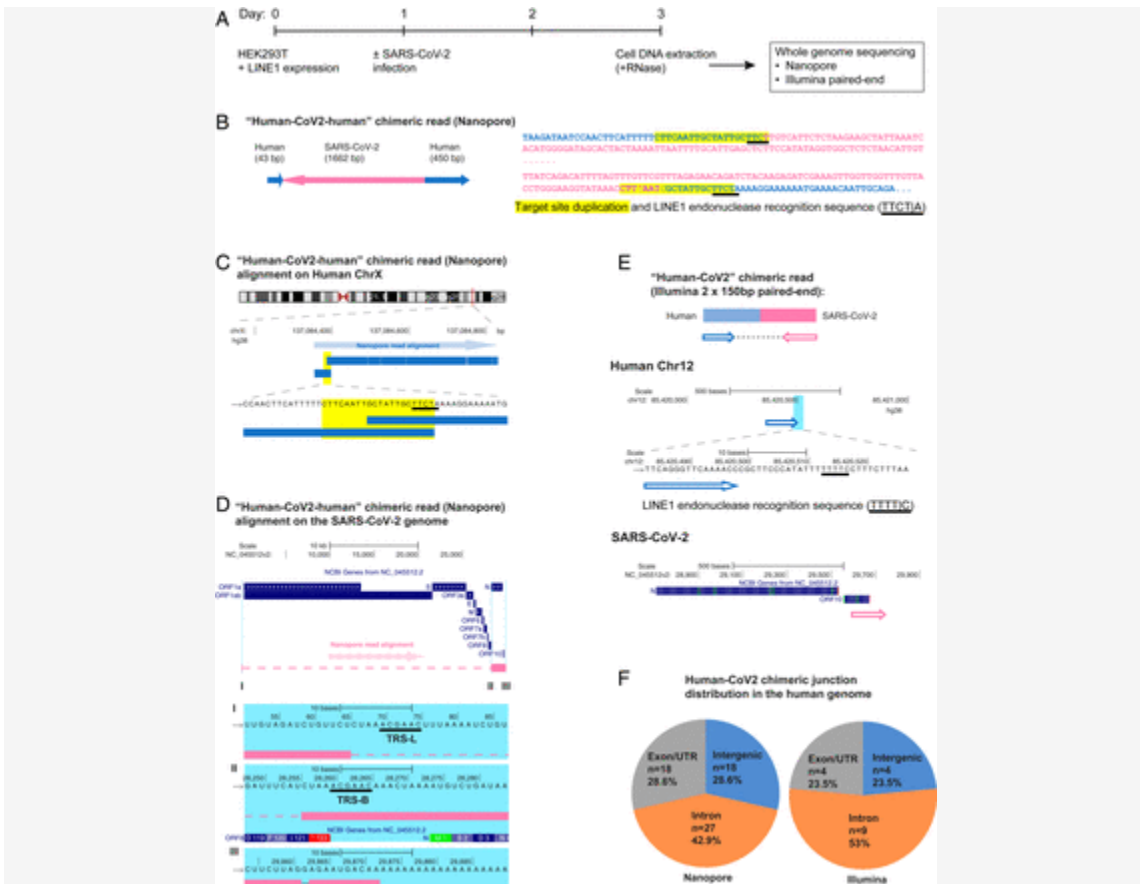
Utilizamos tres enfoques diferentes para detectar secuencias genómicas del SARS-CoV-2 integradas en el genoma de las células infectadas. Estos enfoques fueron la secuenciación de lectura larga de Nanopore, la secuenciación genómica completa de extremos emparejados de Illumina y la secuenciación de enriquecimiento del sitio de integración de ADN basada en la etiqueta Tn5. Los tres métodos proporcionaron evidencia de que las secuencias de SARS-CoV-2 pueden integrarse en el genoma de la célula huésped.

Para aumentar la probabilidad de detectar eventos de integración raros, transfectamos células HEK293T con plásmidos de expresión LINE1 antes de la infección con SARS-CoV-2 y ADN aislado de las células 2 d después de la infección ([Apéndice S1, Fig. S1 A](#)). Detectamos copias de ADN de secuencias de nucleocápside (NC) del SARS-CoV-2 en las células infectadas mediante PCR ([Apéndice S1, Fig. S1 B](#)) y clonamos el gen NC completo ([Apéndice S1, Fig. S1 D](#)) de la célula de fragmento grande ADN genómico que se había purificado en gel ([Apéndice S1, Fig. S1 C](#)). La secuencia de ADN viral (NC) fue confirmada por

secuenciación de Sanger ([Dataset S1](#)). Estos resultados sugieren que el ARN del SARS-CoV-2 puede transcribirse de forma inversa y el ADN resultante podría integrarse en el genoma de la célula huésped.

Para demostrar directamente que las secuencias de SARS-CoV-2 se integraron en el genoma de la célula huésped, se usó ADN aislado de células HEK293T que sobreexpresan LINE1 infectadas para la secuenciación de lectura larga de Nanopore ([Fig. 1 A](#)). [La Fig. 1 B - D](#) muestra un ejemplo de una secuencia de ARN subgenómico NC viral de longitud completa (1662 pb) integrada en el cromosoma X de la célula y flanqueada en ambos lados por secuencias de ADN del huésped. Es importante destacar que las secuencias flanqueantes incluían una repetición directa de 20 pb. Esta duplicación del sitio de destino es una firma de la reintegración mediada por LINE1 ([41](#) , [42](#)). Otro integrante viral que comprende una secuencia de ARN subgenómico NC parcial que fue flanqueado por una secuencia diana de ADN de la célula huésped duplicada se muestra en la [SI Apéndice , Fig S2. A - C](#) . En ambos casos, las secuencias flanqueantes contenían una secuencia de reconocimiento de consenso de la endonucleasa LINE1 ([43](#)). Estos resultados indican que las secuencias de SARS-CoV-2 pueden integrarse en los genomas de células humanas cultivadas mediante un mecanismo de retroposición mediado por LINE1. [La Tabla 1](#) resume todas las secuencias del huésped del SARS-CoV-2 vinculadas que se recuperaron. Se encontraron copias de ADN de porciones del genoma viral en casi todos los cromosomas humanos. Además de los dos ejemplos dados en la [Fig.1](#) y el [Apéndice SI, Fig. S2](#) , también recuperamos secuencias celulares para 61 integrantes para los cuales solo se recuperó una de las dos uniones huésped-viral ([Apéndice SI, Fig. S2 D - F](#) y [Tabla 1](#) ; lecturas de Nanopore que contienen las secuencias quiméricas resumidas en el [Dataset S2](#)). Es importante destacar que aproximadamente el 67% de las secuencias humanas flanqueantes incluían un consenso o una secuencia de reconocimiento de endonucleasa LINE1 variante (como TTTT / A) ([Apéndice SI, Fig. S2 D - F](#) y [Tabla 1](#)). Estas secuencias de reconocimiento de LINE1 estaban en las uniones quiméricas que estaban directamente vinculadas al extremo 3' (cola poli-A) de las secuencias virales, o dentro de una distancia de 8-27 pb de las uniones que estaban vinculadas al extremo 5' de secuencias virales, que se encuentra dentro de la posible duplicación del sitio objetivo. Ambos resultados son consistentes con un modelo en el que la retroposición mediada por LINE1 proporciona un mecanismo para integrar copias de ADN de fragmentos subgenómicos del SARS-CoV-2 en el ADN genómico del huésped. Aproximadamente el 71% de las secuencias virales estaban flanqueadas por intrones o secuencias celulares intergénicas y el 29% por exones ([Fig.1 F](#) y [Tabla 1](#)). Por lo tanto, la asociación de las secuencias virales con exones es mucho mayor de lo que cabría esperar para la integración aleatoria en el genoma [genoma humano: 1,1% de exones, 24% de intrones y 75% de ADN intergénico ([44](#))], lo que sugiere una integración preferencial en

sitios de destino asociados al exón. Si bien estudios anteriores no mostraron preferencia por la retroposición de LINE1 en exones ([45](#) , [46](#)), nuestro hallazgo sugiere que la retroposición de algunos otros ARN mediada por LINE1 puede ser diferente. Observamos que los límites vírico-celulares estaban frecuentemente cerca de las regiones no traducidas (UTR) 5' o 3' de los genes celulares, lo que sugiere que existe una preferencia por la integración cerca de los promotores o sitios poli (A) en nuestro sistema experimental.



• [Descargar figura](#)

• [Abrir en una pestaña nueva](#)

• [Descargar powerpoint](#)

Figura 1.

El ARN del SARS-CoV-2 se puede transcribir de forma inversa e integrar en el genoma de la célula huésped. (A) Flujo de trabajo experimental. (B) Secuencia quimérica de una lectura de secuenciación de Nanopore que muestra la

integración de una secuencia de ARN subgenómico de SARS-CoV-2 NC de longitud completa (magenta) y secuencias genómicas humanas (azul) que flanquean ambos lados de la secuencia viral integrada. Las características indicativas de la "transcripción inversa cebada por el objetivo" mediada por LINE1 incluyen la duplicación del sitio objetivo (resaltado en amarillo) y la secuencia de reconocimiento de la endonucleasa LINE1 (subrayada). Las secuencias que podrían mapearse en ambos genomas se muestran en púrpura con desajustes con las secuencias genómicas humanas en cursiva. Las flechas indican la orientación de la secuencia con respecto a los genomas humanos y del SARS-CoV-2 como se muestra en *C* y *D*. (*C*) Alineación del Nanopore leído en *B* con el genoma humano (cromosoma X) que muestra el sitio de integración. Las secuencias humanas en la región de unión muestran el sitio objetivo, que se duplicó cuando se integró el ADNc del SARS-CoV-2 (resaltado en amarillo) y la secuencia de reconocimiento de la endonucleasa LINE1 (subrayada). (*D*) Alineación del nanoporo leído en *B* con el genoma del SARS-CoV-2 que muestra que el ADN viral integrado es una copia del ARN subgenómico NC de longitud completa. Las regiones resaltadas en azul claro se amplían para mostrar las secuencias TRS-L (I) y TRS-B (II) (subrayadas, estas son las secuencias donde la polimerasa viral salta para generar el ARN subgenómico) y el final de la secuencia viral en el cola de poli (A) (III). Estas características de la secuencia viral (I-III) muestran que una copia de ADN del ARN subgenómico NC de longitud completa fue reintegrada. (*E*) Un par de lectura quimérico humano-viral de la secuenciación del genoma completo de extremos emparejados de Illumina. El par de lectura se muestra alineado con los genomas humano (azul) y SARS-CoV-2 (magenta). Las flechas indican las orientaciones de lectura relativas a los genomas humano y SARS-CoV-2. La región resaltada (azul claro) del mapeo de lectura humana se amplía para mostrar la secuencia de reconocimiento LINE1 (subrayada). (*F*) Distribuciones de las uniones quiméricas humano-CoV2 de la secuenciación de Nanopore (*izquierda*) e Illumina (*derecha*) con respecto a las características del genoma humano.

• [VER EN LÍNEA](#)

• [VER VENTANA EMERGENTE](#)

Tabla 1.

Resumen de las secuencias quiméricas de CoV2 humano obtenidas por secuenciación de ADN Nanopore de células HEK293T que sobreexpresan LINE1 infectadas

Para confirmar la integración de las secuencias de SARS-CoV-2 en el ADN genómico mediante otro método, sometimos el ADN aislado de células HEK293T transfectadas con LINE1 e infectadas con SARS-CoV-2 a la secuenciación del genoma completo de extremos emparejados de Illumina, utilizando un Tn5- método de construcción de bibliotecas basado en (Illumina Nextera) para evitar artefactos de ligadura. Las lecturas de ADN viral se concentraron en el extremo 3' del genoma del SARS-CoV-2 ([Apéndice SI, Fig. S3](#)). Recuperamos 17 integrantes virales (suma de dos réplicas), mediante el mapeo de secuencias de ADN quimérico humano-viral ([Fig. 1 E](#) y [Tabla 2](#) , secuencias quiméricas resumidas en el [conjunto de datos S3](#)); 7 (41%) de las uniones contenían un consenso o una secuencia de reconocimiento de LINE1 variante en las secuencias celulares cerca de la unión ([Fig. 1 E](#) y [Tabla 2](#)), consistente con un mecanismo de retroposición mediado por LINE1. De manera similar a los resultados obtenidos de la secuenciación de Nanopore, aproximadamente el 76% de las secuencias virales estaban flanqueadas por intrones o secuencias celulares intergénicas y el 24% por exones ([Fig. 1 F](#) y [Tabla 2](#)).

- [VER EN LÍNEA](#)

- [VER VENTANA EMERGENTE](#)

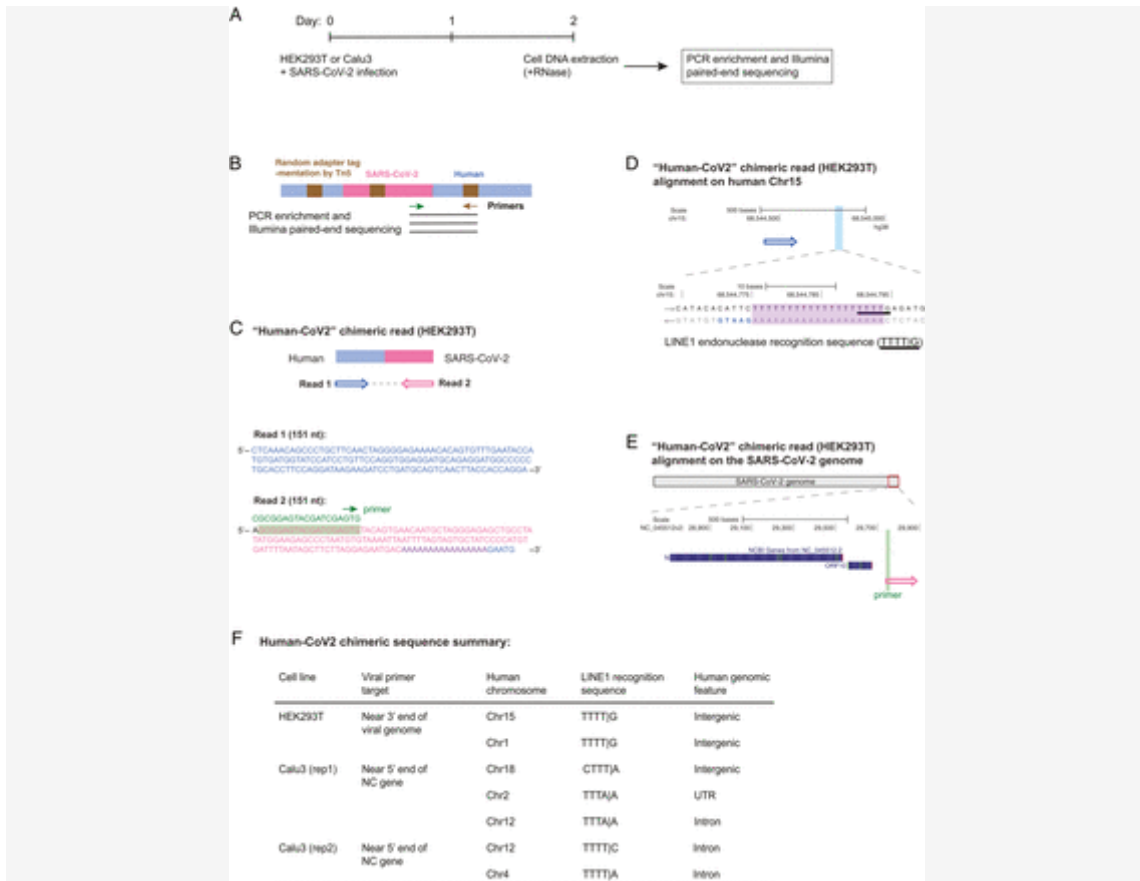
Tabla 2.

Resumen de las secuencias quiméricas de CoV2 humano obtenidas mediante la secuenciación de ADN del genoma completo de extremos emparejados de Illumina de células HEK293T infectadas que sobreexpresan LINE1

Aproximadamente el 32% de las secuencias de SARS-CoV-2 (6/21 eventos de integración en Nanopore, 4/10 en los datos de Illumina) se integraron en LINE, elementos nucleares intercalados cortos o elementos repetidos terminales largos sin evidencia de un sitio de reconocimiento LINE1, lo que sugiere que puede haber un mecanismo de integración / transcripción inversa alternativo, posiblemente similar al informado para las células infectadas de forma aguda con LCMV, que dio como resultado secuencias integradas de LCMV fusionadas con secuencias de partículas de tipo A intracisterno (IAP) ([29](#)).

Para evaluar si la integración genómica de las secuencias de SARS-CoV-2 también podría ocurrir en células infectadas que no sobreexpresaban RT, aislamos ADN de células HEK293T y Calu3 infectadas con virus que no fueron transfectadas con un plásmido de expresión de RT ([Fig. 2 A](#)). La secuenciación de enriquecimiento del sitio de integración de ADN mediada por la etiqueta Tn5 ([47](#) , [48](#)) ([Fig. 2 B](#) y [Apéndice SI, Fig. S4 A](#)) detectó un total de siete

secuencias de SARS-CoV-2 fusionadas con secuencias celulares en estas células (suma de tres infecciones independientes de dos líneas celulares), todas las cuales mostraron secuencias de reconocimiento LINE1 cercanas a las uniones de secuencias humano-SARS-CoV-2 ([Fig.2 C.-Apéndice F](#) y [SI, Fig. S4 B - D](#) , secuencias quiméricas resumidas en el [conjunto de datos S4](#)).



[• Descargar figura](#)

[• Abrir en una pestaña nueva](#)

[• Descargar powerpoint](#)

Figura 2.

Evidencia de la integración del ADNc del SARS-CoV-2 en células cultivadas que no sobreexpresan una transcriptasa inversa. (A) Flujo de trabajo experimental. (B) Diseño experimental para el método de secuenciación de enriquecimiento mediado por etiquetado de Tn5 utilizado para mapear los sitios de integración en el genoma de la célula huésped. (C) Un par de lectura

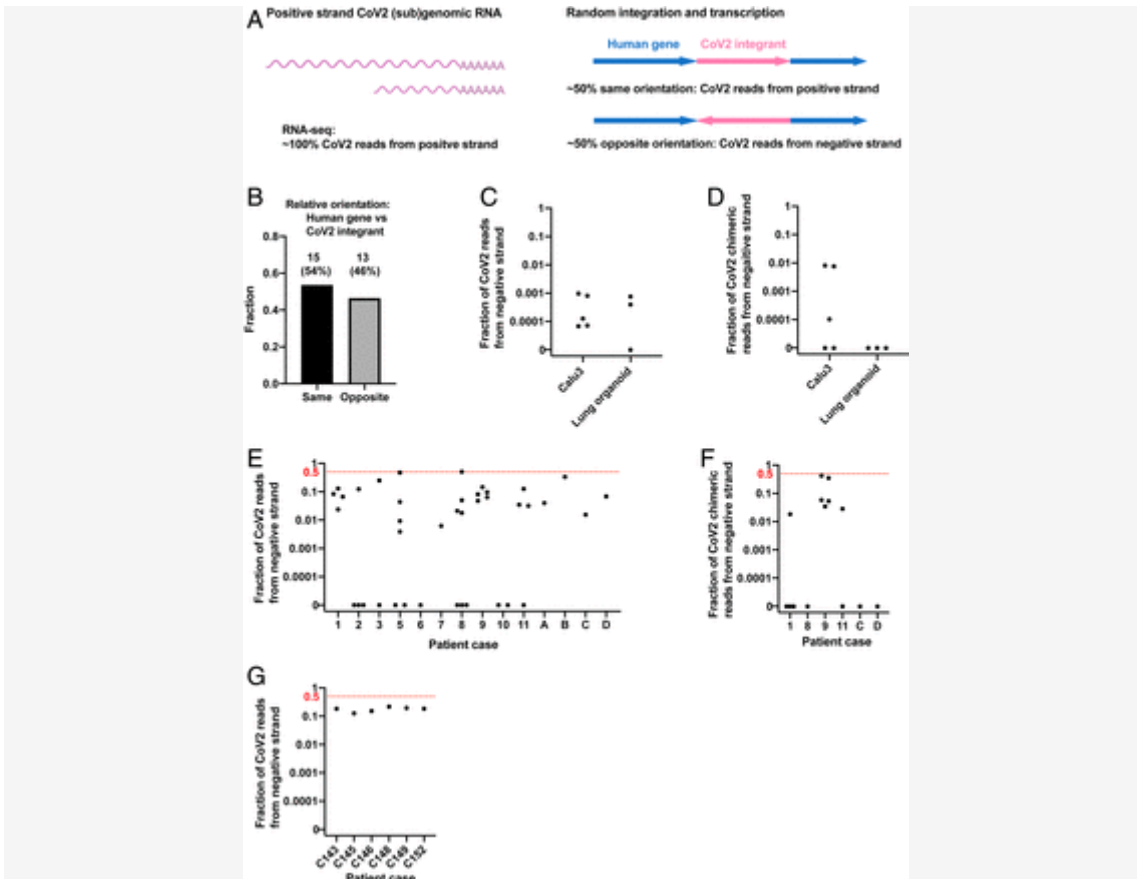
quimérico humano-viral que apoya la integración viral. Las lecturas están alineadas con las secuencias genómicas humana (azul) y SARS-CoV-2 (magenta). Las flechas indican las orientaciones de lectura en relación con el SARS-CoV-2 genomas humano y como se muestra en *D* y *E*. La secuencia del cebador viral utilizado para el enriquecimiento se muestra con un resaltado verde en la lectura (correspondiente a la flecha verde ilustrada en *B*). Las secuencias que podrían mapearse en ambos genomas se muestran en violeta. (*D*) Alineación del par leído en *C* con el genoma humano (cromosoma 15, flecha azul). La región resaltada (azul claro) de la secuencia humana se agranda para mostrar la secuencia de reconocimiento LINE1 (subrayada) con una secuencia poli-dT de 19 bases (resaltado púrpura) que podría ser recodido por la cola viral poli-A para "diana- transcripción inversa preparada ". Se capturó una secuencia humana adicional de 5 pb (GAATG, azul) en la lectura 2 (*C*), lo que respalda un sitio de integración genuino. (*E*) Alineación del par de lectura en *C* con el genoma del SARS-CoV-2 (magenta). La secuencia del cebador viral se muestra resaltada en verde. (*F*) Resumen de siete secuencias quiméricas humano-virales identificadas por el método de secuenciación de enriquecimiento en las dos líneas celulares que muestran los cromosomas humanos integrados, secuencias de reconocimiento LINE1 cercanas a la unión quimérica y características genómicas humanas en la unión de lectura.

Expresión de transcripciones quiméricas vírico-celulares en células cultivadas infectadas y tejidos derivados de pacientes.

Para investigar la posibilidad de que las secuencias de SARS-CoV-2 integradas en el genoma puedan expresarse, analizamos los datos publicados de RNA-seq de células infectadas con SARS-CoV-2 en busca de evidencia de transcripciones quiméricas (49). El examen de estos conjuntos de datos (50 ↓ ↓ ↓ ↓ - 55) (*Apéndice SI, Fig. S5*) reveló una serie de lecturas quiméricas humano-virales (*Apéndice SI, Fig. S6 A y B*). Estos ocurrieron en múltiples tipos de muestras, incluidas células cultivadas y organoides de tejidos de pulmón / corazón / cerebro / estómago (*Apéndice SI, Fig. S6 B*). La abundancia de lecturas quiméricas se correlacionó positivamente con el nivel de ARN viral en todos los tipos de muestras (*Apéndice SI, Fig. S6 B*). Las lecturas quiméricas generalmente representaron el 0.004–0.14% del total de lecturas de SARS-CoV-2 en las muestras. La mayoría de las uniones quiméricas se asignaron a la secuencia del gen NC del SARS-CoV-2 (*Apéndice SI, Fig. S6 C y D*). Esto concuerda con el hallazgo de que el ARN NC es el ARN subgenómico del SARS-CoV-2 más abundante (56), lo que lo convierte en el objetivo más probable para la transcripción inversa y la integración. Sin embargo, datos recientes mostraron que hasta el 1% de las lecturas de secuencia de ARN de células infectadas con SARS-CoV-2 pueden ser artefactos quiméricos como resultado del cambio de RT entre plantillas de ARN, lo que puede ocurrir durante el paso

de síntesis de ADNc en la preparación de una biblioteca de RNA-seq (57). Por tanto, debido a que hay una mezcla de ARNm del hospedador y ARNm viral de hebra positiva en las células infectadas, la identificación de transcripciones de ARN celular-viral quimérico genuino se ve comprometida por la generación de quimeras de artefactos en los ensayos.

Razonamos que la orientación de una copia de ADN integrada del ARN del SARS-CoV-2 debería ser aleatoria con respecto a la orientación del gen del huésped objetivo, prediciendo que aproximadamente la mitad de los ADN virales que se integraron en un gen huésped expresado deberían estar en un orientación opuesta a la dirección de la transcripción del gen de la célula huésped (Fig. 3 A). Como se predijo, aproximadamente el 50% de los integrantes virales en los genes humanos estaban en la orientación opuesta en relación con el gen del huésped en nuestro conjunto de datos Nanopore (integración en genes humanos con secuencias de reconocimiento LINE1, Fig.3 B). Por tanto, para las transcripciones quiméricas derivadas de secuencias virales integradas, se esperaría que aproximadamente el 50% de las transcripciones quiméricas contengan secuencias virales de cadena negativa unidas a secuencias de ARN del huésped de cadena positiva. Por lo tanto, determinamos la fracción de transcripciones quiméricas virales y humano-virales en células cultivadas / organoides infectados y en tejidos derivados de pacientes que contienen secuencias de ARN viral de hebra negativa.



• [Descargar figura](#)

• [Abrir en una pestaña nueva](#)

• [Descargar powerpoint](#)

Fig. 3.

Las lecturas de secuencia de ARN viral de hebra negativa sugieren que se expresan secuencias integradas de SARS-CoV-2. (A) Esquema que predice fracciones de lecturas de secuencia de ARN de SARS-CoV-2 de cadena positiva o negativa que se derivan de ARN (sub) genómicos virales o de transcripciones de secuencias virales integradas. Las flechas (*derecha*) muestran la orientación de una hebra positiva integrada para el SARS-CoV-2 (magenta) en relación con la orientación del gen celular del huésped (azul). (B) Fracciones de secuencias de SARS-CoV-2 integradas en genes humanos con la misma ($n = 15$) u opuesta ($n = 13$) orientación de la hebra viral positiva con respecto a la hebra positiva del gen humano. Se identificaron un total de 28 eventos de integración en genes humanos con secuencias de reconocimiento de endonucleasa LINE1 a partir de nuestra secuenciación de ADN Nanopore de células HEK293T que sobreexpresan LINE1 infectadas ([Fig. 1A](#)). (C) Fracción de lecturas virales totales que se derivan de ARN viral de cadena negativa en células u organoides con infección aguda (consulte el [Apéndice SI, Tabla S1](#) para obtener más detalles). (D) Fracción de lecturas quiméricas virales humanas que contienen secuencias virales derivadas de ARN viral de cadena negativa en células u organoides con infección aguda (consulte el [Apéndice SI, Tabla S1](#) para obtener detalles). (E) Fracción de lecturas virales totales que se derivan de ARN viral de hebra negativa en datos de secuencia de ARN de pacientes publicados (muestras de FFPE de autopsia, GSE150316, muestras sin lecturas virales o de baja calidad de varamiento de biblioteca no incluidas; consulte el [Apéndice SI, Tabla S2](#) para detalles; resultados de reanálisis consistentes con la publicación original). (F) Fracción de lecturas quiméricas virales humanas que contienen secuencias virales derivadas de ARN viral de hebra negativa en datos de secuencia de ARN de pacientes publicados (muestras de autopsia FFPE, GSE150316; consulte el [Apéndice SI, Tabla S2](#) para obtener detalles). (G) Fracción de lecturas virales totales que se derivan de ARN viral de hebra negativa en datos de secuencia de ARN de pacientes publicados (muestras BALF, GSE145926; ver [Apéndice SI, Tabla S3](#) para más detalles). Las líneas rojas discontinuas en E - G indican el nivel en el que el 50% de todas las lecturas

virales (*E* y *G*) o secuencias virales en lecturas quiméricas humano-virales (*F*) procedían de ARN virales de cadena negativa, un nivel esperado si todos los Las secuencias virales se derivaron de secuencias integradas.

La replicación del ARN del SARS-CoV2 requiere la síntesis de ARN viral de cadena negativa, que sirve como molde para la replicación del ARN genómico viral y la transcripción del ARN subgenómico de cadena positiva viral ([21](#)). Para evaluar la prevalencia de ARN viral de cadena negativa en células infectadas de forma aguda, determinamos la proporción de ARN de cadena positiva total a negativa. Entre el 0 y el 0,1% de las lecturas virales totales se derivaron de ARN de cadena negativa en células Calu3 u organoides pulmonares infectados de forma aguda [nuestros datos y datos publicados ([50](#) , [58](#))] ([Fig. 3 C](#) y [Apéndice SI, Tabla S1](#)), similar a lo que se ha informado en muestras clínicas tomadas poco después de la infección ([59](#)). Estos resultados argumentan que el nivel de ARN viral de cadena negativa es al menos 1000 veces menor que el del ARN viral de cadena positiva en células infectadas de forma aguda, debido al menos en parte a una producción masiva de ARN subgenómico de cadena positiva durante la replicación viral. . Esto reduce en gran medida la probabilidad de que el cambio de molde aleatorio durante el paso de transcripción inversa en la construcción de la biblioteca de secuencias de ARN genere una gran fracción de las lecturas quiméricas de artefactos que contendrían ARN de cadena negativa viral fusionado con secuencias de ARN de cadena positiva celular. Determinamos que entre el 0 y el 1% de las lecturas quiméricas virales humanas contenían secuencias virales de cadena negativa en las células / organoides con infección aguda ([Fig. 3 D](#) y [Apéndice SI, Tabla S1](#)), consistente con una pequeña fracción de lecturas virales derivadas de secuencias integradas de SARS-CoV-2.

A diferencia de los resultados obtenidos con células Calu3 u organoides pulmonares infectados de forma aguda, hasta el 51% de todas las lecturas virales y hasta el 42,5% de las lecturas quiméricas virales humanas se derivaron de la cadena negativa del ARN del SARS-CoV-2 en algunos tejidos derivados del paciente [datos publicados ([60](#) , [61](#)), antecedentes clínicos del paciente disponibles en las publicaciones originales] ([Fig. 3 E - G](#) y [Apéndice SI, Tablas S2 y S3](#)). El análisis unicelular de las células del líquido de lavado broncoalveolar pulmonar (BALF) de pacientes con COVID grave [datos publicados ([61](#))] mostró que hasta el 40% de todas las lecturas virales se derivaron del ARN del SARS-CoV-2 de cadena negativa ([Apéndice SI, Figura S7](#)). Las fracciones de ARN de cadena negativa en los tejidos de algunos pacientes eran órdenes de magnitud más altas que las de las células u organoides con infección aguda ([Fig. 3 C - G](#)). En muestras de autopsia fijadas (fijadas en formalina, incluidas en parafina [FFPE]), en 4 de 14 pacientes ([Fig. 3 E](#) y [Apéndice SI, Tabla S2](#)), y en muestras BALF, en 4 de 6 pacientes ([Fig. 3 G](#) y [SI Apéndice, Tabla S3](#)), al menos ~ 20% de las lecturas virales se derivaron

de ARN viral de cadena negativa. A diferencia de las células infectadas de forma aguda ([Fig.3 C y D](#) y [Apéndice SI, Tabla S1](#)), hubo poca o ninguna evidencia de reproducción del virus en estas muestras de autopsia ([60](#)). Como se resume en el [Apéndice SI, Tabla S2](#) , hubo secuencias virales de cadena negativa en una gran fracción de las lecturas quiméricas virales humanas (hasta ~ 40%) en las muestras de un paciente. Diferentes muestras derivadas del mismo paciente revelaron una fracción igualmente alta de lecturas de ARN humano de hebra viral negativa. Varias otras muestras de pacientes revelaron una fracción más baja de quimeras de ARN de cadena viral negativa-ARN humano, que, sin embargo, fueron significativamente más altas que las que se encontraron en las células con infección aguda ([Fig.3 D y F](#) y [Apéndice SI, Tabla S1 y S2](#)). Debido a que la capacidad para identificar lecturas quiméricas virales-humanas usando RNA-seq de lectura corta es limitada, nuestro análisis no pudo mostrar un número significativo de lecturas quiméricas en muestras de pacientes BALF ([Apéndice SI, Tabla S3](#)). En resumen, nuestros datos sugieren que en algunos tejidos derivados de pacientes, donde el número total de células con secuencia positiva de SARS-CoV-2 puede ser pequeño, una gran fracción de las transcripciones virales podrían haberse transcrito a partir de secuencias de SARS-CoV-2. integrado en el genoma del huésped.

Discusión

Presentamos aquí evidencia de que las secuencias de SARS-CoV-2 pueden transcribirse de forma inversa e integrarse en el ADN de las células humanas infectadas en cultivo. Para dos de los integrantes, recuperamos lecturas quiméricas "humano-viral-humano" que abarcan una repetición del sitio objetivo directo (20 o 13 pb), y un sitio de reconocimiento de consenso de la endonucleasa LINE1 estaba presente en ambos extremos del ADN del huésped que flanqueaba las secuencias virales. Estos y otros datos son consistentes con un mecanismo de integración de retroposición y transcripción inversa cebado por diana ([41](#) , [42](#)) y sugieren que LINE1 RT endógeno puede estar involucrado en la transcripción inversa e integración de secuencias de SARS-CoV-2 en los genomas de células infectadas.

Aproximadamente el 30% de los integrantes virales analizados en células cultivadas carecían de un sitio de reconocimiento de endonucleasa LINE1 cercano reconocible. Por tanto, también es posible que la integración pueda producirse mediante otro mecanismo. De hecho, existe evidencia de que los ADNc quiméricos pueden producirse en células infectadas de forma aguda con LCMV mediante la elección de copias con elementos IAP endógenos durante la transcripción inversa. Se espera que este mecanismo cree un ADNc quimérico complementario tanto de LCMV como de IAP. En algunos casos, los ADNc quiméricos resultantes se integraron sin la generación de una duplicación del sitio objetivo ([29](#)). Un estudio reciente también ha sugerido que la interacción

entre las secuencias de coronavirus y el retrotransposón endógeno podría ser un mecanismo de integración viral potencial ([40](#)).

Será importante, en los estudios de seguimiento, demostrar la presencia de secuencias de SARS-CoV-2 integradas en el genoma del huésped en los tejidos del paciente. Sin embargo, esto será un desafío técnico porque se espera que solo una pequeña fracción de las células en los tejidos del paciente sean positivas para las secuencias virales ([61](#)). De acuerdo con esta noción, se ha estimado que solo entre 1 de cada 1,000 y 1 de cada 100,000 células de ratón infectadas con LCMV, ya sea en cultivo o en el animal, portaban copias de ADN viral integradas en el genoma ([30](#)). Además, solo una fracción de los pacientes puede portar secuencias del SARS-CoV-2 integradas en el ADN de algunas células. Sin embargo, con más de 140 millones de humanos infectados con SARS-CoV-2 en todo el mundo (en abril de 2021), incluso un evento raro podría tener una relevancia clínica significativa. También es un desafío estimar la frecuencia de eventos de reintegración en ensayos de cultivo celular, ya que las células infectadas generalmente mueren y se pierden antes de la recolección de muestras. Por la misma razón, no se espera expansión clonal de células integradas en experimentos de infección aguda. Además, la posibilidad de integración en el mismo locus genómico en diferentes pacientes / tejidos puede ser baja, debido a un proceso de integración aleatorio.

La presencia de ARN quimérico de virus-hospedador en las células no puede tomarse por sí sola como una fuerte evidencia de la transcripción de secuencias virales integradas porque el cambio de molde puede ocurrir durante el paso de transcripción inversa de la preparación de la biblioteca de ADNc. Sin embargo, descubrimos que solo una fracción muy pequeña (0-1%) de lecturas quiméricas de células infectadas de forma aguda contenían secuencias de ARN viral de hebra negativa, mientras que, en las bibliotecas de ARN-seq preparadas a partir de algunos pacientes, la fracción de lecturas virales totales y la fracción de lecturas quiméricas virales humanas que se derivaron de ARN de cadena negativa del SARS-CoV-2 fue sustancialmente mayor. Para los eventos de integración mediados por retrotransposón, la orientación del ARN del SARS-CoV-2 de transcripción inversa debe ser aleatoria con respecto a la orientación de un gen huésped. Por tanto, para los ARN quiméricos derivados de secuencias virales integradas, aproximadamente la mitad de las lecturas quiméricas unirán secuencias de ARN del hospedador de cadena positiva a secuencias virales de cadena negativa. En algunas muestras de pacientes, las lecturas virales de cadena negativa representaron entre el 40 y el 50% de las secuencias totales de ARN viral y una fracción similar de las lecturas quiméricas contenían secuencias de ARN viral de cadena negativa, lo que sugiere que la mayoría, si no todos, de los ARN virales en estas muestras se derivaron de secuencias virales integradas.

Es importante señalar que, debido a que hemos detectado solo secuencias subgenómicas derivadas principalmente del extremo 3' del genoma viral integrado en el ADN de la célula huésped, el virus infeccioso no puede producirse a partir de tales secuencias subgenómicas integradas del SARS-CoV-2. La posibilidad de que las secuencias de SARS-CoV-2 puedan integrarse en el genoma humano y expresarse en forma de ARN quiméricos plantea varias preguntas para estudios futuros. ¿Las secuencias integradas del SARS-CoV-2 expresan antígenos virales en los pacientes y podrían influir en el curso clínico de la enfermedad? La evidencia clínica disponible sugiere que, como máximo, solo una pequeña fracción de las células en los tejidos del paciente expresan proteínas virales a un nivel que es detectable por inmunohistoquímica. Sin embargo, [62](#), [63](#)). La presencia de secuencias de LCMV integradas en los genomas de células infectadas de forma aguda en ratones llevó a los autores a especular que la expresión de tales secuencias "representa potencialmente una forma de vacuna de ADN producida naturalmente" ([30](#)). No se sabe cuántas células presentadoras de antígeno se necesitan para provocar una respuesta de antígeno, pero la expresión de LINE1 desreprimida, inducida por infección viral o por exposición a citocinas ([38](#) ↓ - [40](#)), puede estimular la integración del SARS-CoV-2 en el genoma de las células infectadas en los pacientes. De manera más general, nuestros resultados sugieren que la integración del ADN viral en las células somáticas puede representar una consecuencia de una infección natural que podría desempeñar un papel en los efectos de otros virus de ARN que causan enfermedades comunes, como el dengue, el Zika o el virus de la influenza.

Nuestros resultados también pueden ser relevantes para los ensayos clínicos actuales de terapias antivirales ([64](#)). Si la integración y expresión del ARN viral son bastante comunes, la dependencia de pruebas de PCR extremadamente sensibles para determinar el efecto de los tratamientos sobre la replicación viral y la carga viral puede no siempre reflejar la capacidad del tratamiento para suprimir completamente la replicación viral porque los ensayos de PCR pueden detectar virus. transcripciones que se derivan de secuencias de ADN viral que se han integrado de manera estable en el genoma en lugar de virus infecciosos.

Materiales y métodos

Cultivo celular y transfección de plásmidos.

Se obtuvieron células HEK293T de ATCC (CRL-3216) y se cultivaron en DMEM complementado con FBS inactivado por calor al 10% (HyClone; SH30396.03) y L- glutamina 2 mM (MP Biomedicals; IC10180683) siguiendo el método de ATCC. Se obtuvieron células Calu3 de ATCC (HTB-55) y se cultivaron en EMEM (ATCC; 30-2003) suplementado con FBS inactivado por calor al 10% (HyClone; SH30396.03) siguiendo el método de ATCC.

Los plásmidos para la expresión humana de LINE1, pBS-L1PA1-CH-mneo (CMV-LINE-1), fue un regalo de Astrid Roy-Engel, Tulane University Health Sciences Center, Nueva Orleans, LA (Addgene plásmido # 51288; <http://addgene.org/51288> ; RRID: Addgene_51288) ([65](#)); EF06R (5'UTR-LINE-1) fue un regalo de Eline Luning Prak, Universidad de Pennsylvania, Filadelfia, PA (Addgene plásmido # 42940; <http://addgene.org/42940> ; RRID: Addgene_42940) ([66](#)). La transfección se realizó con Lipofectamine 3000 (Invitrogen; L3000001) siguiendo el protocolo del fabricante.

Infección por SARS-CoV-2.

El SARS-CoV-2 USA-WA1 / 2020 (GenBank: [MN985325.1](https://www.ncbi.nlm.nih.gov/nuclseq/MN985325.1)) se obtuvo de BEI Resources y se expandió y tituló en células Vero. Las células se infectaron en DMEM más FBS al 2% durante 48 h usando una multiplicidad de infección (MOI) de 0,5 para la infección de células HEK293T y una MOI de 1 o 2 para células Calu3. Todo el procesamiento de muestras y la recolección con virus infecciosos se realizaron en las instalaciones de BSL3 en el Instituto Ragon.

Extracción de ácidos nucleicos y ensayo de PCR.

La extracción de ADN celular se realizó mediante un método publicado ([31](#)). Para la purificación del ADN genómico, el ADN celular total se fraccionó en un gel de agarosa al 0,4% (p / vol) / 1 × TAE durante 1,5 h con un voltaje de 3 V / cm, con λ DNA-HindIII Digest (NEB; N3012S) como tamaño marcadores. Se cortaron fragmentos grandes (> 23,13 kb), se congelaron a -80 ° C y luego se trituraron con una punta de pipeta. Se agregaron tres volúmenes (vol / p) de tampón TE alto (Tris 10 mM - EDTA 10 mM, pH 8.0) y luego se agregó NaCl para dar una concentración final de 200 mM. La solución de gel se calentó a 70 ° C durante 15 min con mezcla constante y luego se extrajo con fenol: cloroformo: alcohol isoamílico (25: 24: 1, vol / vol / vol) (Life Technologies; 15593031) y cloroformo: alcohol isoamílico 24 : 1 (Sigma; C0549-1PT). El ADN se precipitó mediante la adición de acetato de sodio y alcohol isopropílico. Para muestras con baja concentración de ADN, glucógeno (Life Technologies;

La extracción de ARN se realizó con RNeasy Plus Micro Kit (Qiagen; 74034) siguiendo el protocolo del fabricante.

Para detectar copias de ADN de secuencias de SARS-CoV-2, elegimos cuatro conjuntos de cebadores de PCR dirigidos a genes NC que se utilizan en las pruebas COVID-19 [[Apéndice SI, Fig. S1 A](#) , fuente de cebadores de la Organización Mundial de la Salud ([67](#)), para que coincida con la versión del genoma de NC_045512.2]. Consulte el [Apéndice SI, Tabla S4](#) para conocer las secuencias de cebadores de PCR utilizadas en este estudio. La PCR se realizó utilizando AccuPrime Taq DNA Polymerase, de alta fidelidad (Life Technologies;

12346094). Los productos de PCR se procesaron en gel de agarosa al 1% o al 2% (peso / volumen) para mostrar amplificaciones.

Secuenciación y análisis de ADN de nanoporos.

Se utilizó un total de 1,6 µg de ADN extraído de células HEK293T transfectadas con el plásmido pBS-L1PA1-CH-mneo (CMV-LINE-1) e infectadas con SARS-CoV-2 para hacer una biblioteca de secuenciación con el kit SQK-LSK109 (Oxford Nanopore Technologies) y se secuenció en una celda de flujo R9 PromethION (FLO-PRO002) durante 3 días y 5 minutos. Los datos de secuenciación se denominaron de forma base utilizando Guppy 4.0.11 (Oxford Nanopore Technologies) utilizando el modelo de alta precisión.

Las lecturas de Nanopore se mapearon utilizando minimap2 ([68](#)) (versión 2.15) con los parámetros "-p 0.3 -ax map-ont" y un archivo fasta que contiene la secuencia del genoma humano de la versión 93 de ENSEMBL ([ftp://ftp.ensembl.org/pub /release-93/fasta/homo_sapiens/dna/Homo_sapiens.GRCh38.dna.primary_assembly.fasta](ftp://ftp.ensembl.org/pub/release-93/fasta/homo_sapiens/dna/Homo_sapiens.GRCh38.dna.primary_assembly.fasta)) concatenado a la secuencia SARS-CoV-2, GenBank ID: [MN988713.1](#), "Síndrome respiratorio agudo severo coronavirus 2 aislado SARS-CoV-2 / human / USA / IL-CDC-IL1 / 2020, genoma completo". Desde el archivo SAM, seleccionamos todas las secuencias que se asignaron al genoma viral y las dividimos en grupos según los cromosomas humanos a los que se asignaron. Explotamos las secuencias seleccionadas, usando blastn, contra una base de datos BLAST hecha con las secuencias humanas y de virus descritas anteriormente. Analizamos la salida de blast en un archivo de texto que contiene una fila por par de segmentos de alta puntuación (HSP) con un script de Perl personalizado. Además, filtramos ese archivo, para cada secuencia, seleccionando todas las HSP virales y las tres HSP humanas principales. Inspeccionamos esos archivos visualmente para identificar secuencias que contienen uniones humano-viral-humano o humano-viral. Para algunas secuencias, de más de 30 kb, inspeccionamos las 15 mejores HSP humanas. Adicionalmente, [69](#)) herramienta. Debido a errores en la secuenciación de Nanopore y / o llamadas de bases, existen "secuencias híbridas" de artefactos en un subconjunto de estas lecturas, a veces con cadenas de Watson y Crick del mismo fragmento de ADN presentes en la misma lectura. Por lo tanto, solo nos centramos en las secuencias quiméricas que muestran uniones virales-humanas claras y analizamos las características de retroposición mediadas por LINE1, como las duplicaciones de sitios diana y las secuencias de reconocimiento de endonucleasas LINE1, en busca de pruebas de integración.

Enriquecimiento del sitio de integración mediado por etiquetado Tn5.

Usamos un método basado en etiquetado para enriquecer los sitios de integración viral ([47](#) , [48](#)). Brevemente, usamos la transposasa Tn5 (Diagenode; C01070010) para etiquetar aleatoriamente el ADN celular con adaptadores (adaptador A, el sistema Illumina Nextera). La marcación se realizó utilizando 100 ng de ADN durante 10 min a 55 ° C, seguido de la eliminación de la transposasa Tn5 del ADN con SDS. Usamos un cebador inverso dirigido al extremo cercano a 5 ' del gen NC del SARS-CoV-2 (CCA AGA CGC AGT ATT ATT GGG TAA A) o un cebador directo dirigido al extremo cercano a 3 ' del genoma del SARS-CoV-2 (CTT GTG CAG AAT GAA TTC TCG TAA CT) para amplificar linealmente (PCR0, 45 ciclos) los fragmentos de ADN marcados que contienen secuencias virales. Tomamos el producto de PCR0 y amplificamos los fragmentos de ADN que contienen el adaptador y las secuencias virales (posibles sitios de integración) usando 15-20 ciclos de PCR1, con un cebador Nextera con código de barras (i5) (AAT GAT ACG GCG ACC ACC GAG ATC TAC ACN NNN NNN NTC GTC GGC AGC GTC, NNNNNNNN indica el código de barras) contra la secuencia del adaptador y un cebador viral. El cebador viral se diseñó para apuntar al extremo cercano a 5 ' del gen SARS-CoV-2 NC (GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GGCC GAC GTT GTT TTG ATC G , secuencia viral subrayada) o apuntar al extremo cercano a 3 ' del genoma SRAS-CoV-2 (GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G CGC GGA GTA CGA TCG AGT G , secuencia viral subrayada). El cebador viral también contenía una secuencia adaptadora para una posterior amplificación por PCR. Amplificamos el producto PCR1 mediante 15-20 ciclos de PCR2, utilizando un cebador corto (AAT GAT ACG GCG ACC ACC GA) contra la secuencia del cebador i5 Nextera y un cebador Nextera con código de barras (i7) (CAA GCA GAA GAC GGC ATA CGA GAT NNN NNN NNG TCT CGT GGG CTC GG, NNNNNNNN indica el código de barras) frente a la secuencia adaptadora introducida por el cebador viral en PCR1. El producto final de la amplificación PCR2 se fraccionó en gel de agarosa al 1,5% (Sage Science; HTC1510) con PippinHT (Sage Science; HTP0001) y se seleccionaron piezas de 500 a 1.000 pb para la secuenciación de extremos emparejados de Illumina. Los tres pasos de PCR (PCR0 – PCR2) se realizaron con KAPA HiFi HotStart ReadyMix (KAPA; KK2602).

Secuenciación y análisis de ADN de Illumina.

Construimos bibliotecas para la secuenciación del genoma completo de células HEK293T utilizando el kit Illumina DNA Prep basado en Tn5 (Illumina; 20018704). Las bibliotecas de secuenciación de genoma completo o las bibliotecas del enriquecimiento del sitio de integración mediado por Tn5 después del dimensionamiento (descrito anteriormente) se sometieron a secuenciación de Illumina. Se usó qPCR para medir las concentraciones de cada biblioteca usando el kit de cuantificación de la biblioteca de qPCR KAPA de acuerdo con el protocolo del fabricante. A continuación, se combinaron las bibliotecas a concentraciones equimolares, para cada carril, en función de las

concentraciones de qPCR. Las bibliotecas agrupadas se desnaturalizaron utilizando el protocolo de Illumina. Las bibliotecas desnaturalizadas se cargaron en una celda de flujo SP en un Illumina NovaSeq 6000 y se procesaron durante 2 x 150 ciclos de acuerdo con las instrucciones del fabricante. Los archivos Fastq se generaron y demultiplexaron con el software de conversión bcl2fastq (Illumina).

Para identificar las lecturas de ADN quimérico humano-SARS-CoV-2, las lecturas de secuenciación sin procesar se alinearon con STAR ([70](#)) (versión 2.7.1a) a un genoma humano más SARS-CoV-2 elaborado con un archivo fasta que contiene la versión de la secuencia del genoma humano hg38 sin cromosomas alternativos concatenados a la secuencia de SARS-CoV-2 de la secuencia de referencia NC_045512.2 del Centro Nacional de Información Biotecnológica (NCBI). Se utilizaron los siguientes parámetros STAR para llamar a lecturas quiméricas: `-alignIntronMax 1 \ -chimOutType Junctions SeparateSAMold WithinBAM HardClip \ -chimScoreJunctionNonGTAG 0 \ -alignSjstitchMismatchNmax -1-1 -1-1 \ -chimSegmentMin 25 \ -chimJunction \OverhangedMintype 25 \ -chimJunction \OverhangMintype`. Extrajimos lecturas virales del archivo BAM generado por samtools ([71](#)) (versión 1.11) usando el comando: `samtools view -b Aligned.sortedByCoord.out.bam NC_045512v2> NC_Aligned.sortedByCoord.out.bam`. Extrajimos lecturas quiméricas virales humanas usando los nombres de lectura del archivo Chimeric.out.junction generado por STAR para obtener las alineaciones de lectura del archivo Chimeric.out.sam generado por STAR por Picard (<http://broadinstitute.github.io/picard>), usando el comando: `java -jar picard.jar FilterSamReads I = Chimeric.out.sam O = hv-Chimeric.out.sam READ_LIST_FILE = hv-Chimeric.out.junction.ids FILTER = includeReadList`. Además, confirmamos cada una de las lecturas quiméricas y filtramos cualquier lectura poco convincente (demasiado corta o alineada con múltiples sitios del genoma humano) mediante inspección visual con UCSC BLAT ([69](#)) herramienta. También cargamos el archivo Aligned.sortedByCoord.out.bam generado por STAR o el archivo NC_Aligned.sortedByCoord.out.bam que contiene lecturas virales extraídas al genoma SARS-CoV-2 del navegador UCSC (NC_045512.2) para buscar lecturas quiméricas adicionales fueron omitidos por el método de llamada quimérico STAR. Para generar el archivo de cobertura del genoma, usamos bamCoverage de la suite deepTools ([72](#)) (versión 3.5.0) para convertir el archivo Aligned.sortedByCoord.out.bam generado por STAR en un archivo bigwig agrupado a 10 bp, usando el comando: `bamCoverage -b Aligned.sortedByCoord.out.bam -o Aligned.sortedByCoord.out.bw - binSize 10`.

RNA-Seq y análisis.

Para identificar lecturas quiméricas de SARS-CoV-2 humano, se descargaron datos de secuencia de ARN publicados de Gene Expression Omnibus (GEO) con

los números de acceso GSE147507 ([50](#)), GSE153277 ([51](#)), GSE156754 ([52](#)), GSE157852 ([53](#)), GSE153684 ([54](#)) y GSE154998 ([55](#)) (resumidos en el *Apéndice SI, Fig. S5 C*). Las lecturas de secuenciación sin procesar se alinearon con STAR ([70](#)) (versión 2.7.1a) al genoma y transcriptoma humano más SARS-CoV-2 elaborado con un archivo fasta que contiene la secuencia del genoma humano versión hg38 sin cromosomas alternativos concatenados a la secuencia SARS-CoV-2 de la secuencia de referencia NCBI NC_045512.2 y un archivo gtf que contiene las anotaciones de genes humanos de ENSEMBL versión GRCh38.97 concatenadas con las anotaciones de genes SARS-CoV-2 de NCBI (<http://hgdownload.soe.ucsc.edu/goldenPath/wuhCor1/bigZips/genes/>) . Los siguientes parámetros STAR ([56](#)) se utilizaron para llamar lecturas quiméricas a menos que se especifique lo contrario (*Apéndice SI, Fig. S5 C*): - Uniones chimOutType SeparateSAMold WithinBAM HardClip \ - chimScoreJunctionNonGTAG 0 \ -alignSjstitchMismatchNmax -1-1 -1-1 \ - chimSegmentMin 50 \ -chimJunctionOverhangMin 50.

Para el análisis de cadena de ARN-seq, generamos datos de ARN-seq utilizando ARN de células Calu3 infectadas con SARS-CoV-2. Se construyeron bibliotecas trenzadas con el kit Kapa mRNA HyperPrep (Roche; 08098115702). Las bibliotecas se sometieron a qPCR usando un kit de cuantificación de bibliotecas de qPCR de KAPA según el protocolo del fabricante. A continuación, se combinaron las bibliotecas a concentraciones equimolares, para cada carril, en función de las concentraciones de qPCR. Las bibliotecas agrupadas se desnaturalizaron utilizando el protocolo de Illumina. Las bibliotecas desnaturalizadas se cargaron en un HiSeq 2500 (Illumina) y se secuenciaron durante 120 ciclos desde un extremo de los fragmentos. Las llamadas base se realizaron utilizando el llamador de bases sin conexión (OLB) de Illumina y luego se demultiplexaron. Descargamos datos publicados de RNA-seq (bibliotecas trenzadas) de GEO con los números de acceso GSE147507 ([50](#)) (Calu3, *SI Apéndice, Tabla S1*), GSE148697 ([58](#)) (organoides pulmonares, *Apéndice SI, Tabla S1*) y GSE150316 ([60](#)) (tejidos FFPE del paciente, *Apéndice SI, Tabla S2*). Las lecturas de secuencia de ARN sin procesar se alinearon como se describe anteriormente, utilizando los parámetros - chimSegmentMin 30 \ -chimJunctionOverhangMin 30 para llamar a las lecturas quiméricas. Extrajimos lecturas virales totales y lecturas quiméricas virales humanas como se describió anteriormente. Convertimos los archivos BAM de lectura viral en archivos Bed usando la utilidad bamToBed en BEDTools ([73](#)). Luego contamos los números de lectura totales y varados en los archivos BED convertidos.

Los datos de secuencia de ARN de una sola célula publicados se descargaron de GEO con el número de acceso GSE145926 ([61](#)) (muestras de pacientes BALF, *Apéndice SI, Tabla S3*). Para el análisis masivo, dedup_hash (https://github.com/mvdbeek/dedup_hash) eliminó las lecturas duplicadas

con las mismas secuencias read1 (UMI) y read2 en archivos fastq sin procesar . Luego, el grupo de read2 se alineó como se describió anteriormente, utilizando los parámetros `-chimSegmentMin 30 \ -chimJunctionOverhangMin 30` para llamar a lecturas quiméricas. El mapeo de lectura se analizó como se describe anteriormente. Para el análisis de una sola célula, generamos un genoma personalizado por Cell Ranger (10 × Genomics Cell Ranger 3.0.2) ([74](#)) mkref, utilizando un archivo fasta que contiene la secuencia del genoma humano de la versión 93 de ENSEMBL (ftp://ftp.ensembl.org/pub/release-93/fasta/homo_sapiens/dna/Homo_sapiens.GRCh38.dna.primary_assembly.fasta) concatenado a la secuencia SARS-CoV-2, GenBank ID: [MN988713.1](#) y un archivo gtf que contiene anotaciones humanas y virales. El mapeo de lectura, la asignación de lecturas a los códigos de barras de las células y la eliminación de los duplicados de PCR se realizaron con el recuento de Cell Ranger (10 × Genomics Cell Ranger 4.0.0) ([74](#)), utilizando el genoma personalizado descrito anteriormente. Procesamos los recuentos utilizando Seurat (versión 3.2.2) ([75](#)). Eliminamos las células que tenían menos de 200 genes detectados o más del 20% de los recuentos de transcripciones derivadas de las mitocondrias. Para cada célula, contamos el número de lecturas que se asignan a la hebra viral positiva o negativa.

Disponibilidad de datos

Todos los datos que respaldan los hallazgos de este estudio están disponibles en el artículo y la información de respaldo. Todos los datos de secuenciación generados en este estudio se han depositado en Sequence Read Archive, <https://www.ncbi.nlm.nih.gov/sra> (número de acceso [PRJNA721333](#)). Todos los datos publicados analizados en este estudio se citan en este artículo con los métodos de adhesión proporcionados en [Materiales y métodos](#) .

Expresiones de gratitud

Agradecemos a los miembros de los laboratorios de RJ y RAY y a otros colegas del Whitehead Institute y del Massachusetts Institute of Technology (MIT) por los útiles debates y recursos. Agradecemos a Thomas Volkert y al personal del núcleo de genómica de Whitehead, y a Stuart Levine del centro BioMicro del MIT / Koch Institute por su apoyo en la secuenciación. Agradecemos a Lorenzo Bombardelli por compartir el protocolo y los consejos para la secuenciación de enriquecimiento de integración mediada por etiquetado de Tn5. Agradecemos a Jerold Chun, Inder Verma, Joseph Ecker y Daniel W. Bellott por la discusión y las sugerencias. Este trabajo fue apoyado por subvenciones del NIH a RJ (1U19AI131135-01; 5R01MH104610-21) y por una generosa donación de Dewpoint Therapeutics y de Jim Stone. SHH fue apoyado por el Programa de

Investigación Intramural del Centro de Investigación del Cáncer del Instituto Nacional del Cáncer.

Notas al pie

- [↵](#)¹ A quién puede dirigirse la correspondencia. Correo electrónico: jaenisch@wi.mit.edu .
 - Contribuciones de los autores: investigación diseñada por LZ, RAY y RJ; LZ y AR realizaron experimentos; Datos analizados LZ, AR, MIB, SHH, RAY y RJ; y LZ y RJ escribieron el artículo con aportes de todos los autores.
 - Revisores: AB, Instituto del Cáncer de los Países Bajos; y AMS, Fox Chase Cancer Center.
 - Declaración de intereses en competencia: RJ es asesor / cofundador de Fate Therapeutics, Fulcrum Therapeutics, Omega Therapeutics y Dewpoint Therapeutics. RAY es fundador y accionista de Syros Pharmaceuticals, Camp4 Therapeutics, Omega Therapeutics y Dewpoint Therapeutics.
-
- Este artículo contiene información de apoyo en línea en <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2105968118/-/DCSupplemental> .
 - Copyright © 2021 del autor (es). Publicado por PNAS.

Este artículo de acceso abierto se distribuye bajo [Creative Commons Attribution License 4.0 \(CC BY\)](#) .

Referencias

1. [↵](#)
 1. Agencia Coreana de Control y Prevención de Enfermedades , Hallazgos de la investigación y análisis de casos re-positivos . <https://www.kdca.go.kr/board/board.es?mid=a30402000000&bid=0030> . Consultado el 12 de junio de 2020 .
[Google Académico](#)
2. [↵](#)
 1. J. Bullard y col.

., Predicción del síndrome respiratorio agudo severo infeccioso coronavirus 2 a partir de muestras de diagnóstico . Clin. Infectar. Dis. **71** , 2663 - 2666 (2020).

[Google Académico](#)

3. [↵](#)

1. X. He et al

., Dinámica temporal en la diseminación viral y transmisibilidad de COVID-19 . Nat. Medicina. **26** , 672 - 675 (2020).

[CrossRefPubMedGoogle Académico](#)

4. [↵](#)

1. N. Li ,
2. X. Wang ,
3. T. Lv

, Liberación prolongada de ARN del SARS-CoV-2: no es un fenómeno raro . J. Med. Virol. **92** , 2286 - 2287 (2020).

[Google Académico](#)

5. [↵](#)

1. MJ Mina ,
2. R. Parker ,
3. DB Larremore

, Repensar la sensibilidad de la prueba COVID-19: una estrategia de contención . N. Engl. J. Med. **383** , e120 (2020).

[CrossRefPubMedGoogle Académico](#)

6. [↵](#)

1. N. Sethuraman ,
2. SS Jeremías ,
3. A. Ryo

, Interpretación de pruebas de diagnóstico para SARS-CoV-2 . JAMA **323** , 2249 - 2251 (2020).

[CrossRefPubMedGoogle Académico](#)

7. [↵](#)

1. J.-R. Yang y col.

., Positividad persistente de ARN viral durante el período de recuperación de un paciente con infección por SARS-CoV-2 . J. Med. Virol. **92** , 1681 - 1683 (2020).

[Google Académico](#)

8. [↵](#)

1. J. An y col.

., Características clínicas de los pacientes con COVID-19 recuperados con prueba de ARN positiva re-detectable . Ana. Transl. Medicina. **8** , 1084 (2020).

[Google Académico](#)

9. [↵](#)

1. D. Chen y col.

., Recurrencia de ARN de SARS-CoV-2 positivo en COVID-19: Reporte de un caso . En t. J. Infect. Dis. **93** , 297 - 299 (2020).

[CrossRefPubMedGoogle Académico](#)

10. [↵](#)

1. L. Lan y col.

., Resultados positivos de la prueba de RT-PCR en pacientes recuperados de COVID-19 . JAMA **323** , 1502 - 1503 (2020).

[Google Académico](#)

11. [↵](#)

1. D. Loconsole y col.

., Recurrencia de COVID-19 después de la recuperación: Reporte de un caso de Italia . Infección **48** , 965 - 967 (2020).

[PubMedGoogle Académico](#)

12. [↵](#)

1. J. Lu y col.

., Caracterización clínica, inmunológica y virológica de los pacientes COVID-19 que dan positivo en la prueba del SRAS-CoV-2 mediante RT-PCR . EBioMedicine **59** , 102960 (2020).

[Google Académico](#)

13. [↵](#)

1. S. Luo ,
2. Y. Guo ,
3. X. Zhang ,
4. H. Xu

, Un estudio de seguimiento de pacientes recuperados con COVID-19 en Wuhan, China . En t. J. Infect. Dis. **99** , 408 - 409 (2020).

[CrossRefGoogle Académico](#)

14. [↵](#)

1. G. Ye y col.

., Características clínicas del síndrome respiratorio agudo severo reactivación del coronavirus 2 . J. Infect. **80** , e14 - e17 (2020).

[CrossRefPubMedGoogle Académico](#)

15. [↵](#)

1. R. Wölfel y col.

., Evaluación virológica de pacientes hospitalizados con COVID-2019 . Nature **581** , 465 - 469 (2020).

[CrossRefPubMedGoogle Académico](#)

16. [↵](#)

1. M. Cevik y col.

., SARS-CoV-2, SARS-CoV y dinámica de carga viral MERS-CoV, duración de la diseminación viral e infecciosidad: una revisión sistemática y un metanálisis . Lancet Microbe **2** , e13 - e22 (2021).

[Google Académico](#)

17. [↵](#)

1. AL Rasmussen ,
2. SV Popescu

., Transmisión del SARS-CoV-2 sin síntomas . Science **371** , 1206 - 1207 (2021).

[Resumen / Texto completo GRATISGoogle Académico](#)

18. [↵](#)

1. KK To y col.

., Reinfeción por COVID-19 por una cepa de coronavirus-2 SARS filogenéticamente distinta confirmada por secuenciación del genoma completo . Clin. Infectar. Dis. , doi: 10.1093 / cid / ciaa1275 (2020).

[CrossRefPubMedGoogle Académico](#)

19. [↵](#)

1. J. Huang y col.

., Recurrencia de la positividad de la PCR para SARS-CoV-2 en pacientes con COVID-19: una experiencia de un solo centro y posibles implicaciones . medRxiv [Preprint] (2020). <https://doi.org/10.1101/2020.05.06.20089573> (Consultado el 6 de junio de 2020).

[Google Académico](#)

20. [↵](#)

1. B. Yuan y col.

., Recurrencia de ARN viral SARS-CoV-2 positivo en pacientes con COVID-19 recuperados durante la observación de aislamiento médico . Sci. Rep. **10** , 11887 (2020).

[CrossRefPubMedGoogle Académico](#)

21. [↵](#)

1. P. V'Kovski ,
2. A. Kratzel ,
3. S. Steiner ,
4. H. Stalder ,
5. V. Thiel

, Biología y replicación del coronavirus: implicaciones para el SARS-CoV-2 . Nat. Rev. Microbiol. **19** , 155 - 170 (2021).

[Google Académico](#)

22. [↵](#)

1. L. Alanagreh ,
2. F. Alzoughool ,
3. M. Atoum

, La enfermedad por coronavirus humano COVID-19: su origen, características y conocimientos sobre fármacos potenciales y sus mecanismos . Patógenos **9** , 331 (2020).

[Google Académico](#)

23. [↵](#)

1. E. de Wit ,
2. N. van Doremalen ,
3. D. Falzarano ,
4. VJ Munster

, SARS y MERS: conocimientos recientes sobre coronavirus emergentes . Nat. Rev. Microbiol. **14** , 523 - 534 (2016).

[CrossRefPubMedGoogle Académico](#)

24. [↵](#)

1. AR Fehr ,
2. S. Perlman

, Coronavirus: una descripción general de su replicación y patogénesis . Methods Mol. Biol. **1282** , 1 - 23 (2015).

[CrossRefPubMedGoogle Académico](#)

25. [↵](#)

1. VA Belyi ,
2. AJ Levine ,
3. AM Skalka

, Herencia inesperada: múltiples integraciones de secuencias antiguas de bornavirus y ebolavirus / marburgvirus en genomas de vertebrados . PLoS Pathog. **6** , e1001030 (2010).

[CrossRefPubMedGoogle Académico](#)

26. [↵](#)

1. M. Horie y col.

., Elementos de virus de ARN endógenos no retrovirales en genomas de mamíferos . Nature **463** , 84 - 87 (2010).

[CrossRefPubMedGoogle Académico](#)

27. [↵](#)

1. M. Horie ,
2. K. Tomonaga

, Fósiles no retrovirales en genomas de vertebrados . Los virus **3** , 1836 - 1848 (2011).

[CrossRefPubMedGoogle Académico](#)

28. [↵](#)

1. A. Shimizu y col.

., Caracterización de ADN citoplasmático complementario a virus ARN no retrovirales en células humanas . Sci. Rep. **4** , 5074 (2014).

[CrossRefPubMedGoogle Académico](#)

29. [↵](#)

1. MB Geuking y col.

., La recombinación de retrotransposón y virus de ARN exógeno da como resultado la integración de ADNc no retroviral . Science **323** , 393 - 396 (2009).

[Resumen / Texto completo GRATISGoogle Académico](#)

30. [↵](#)

1. P. Klenerman ,
2. H. Hengartner ,
3. RM Zinkernagel

, Un virus de ARN no retroviral persiste en forma de ADN . Nature **390** , 298 - 301 (1997).

[CrossRefPubMedGoogle Académico](#)

31. [↵](#)

1. MH Lee y col.

., Recombinación del gen de la APP somática en la enfermedad de Alzheimer y neuronas normales . Nature **563** , 639 - 645 (2018).

[CrossRefPubMedGoogle Académico](#)

32. [↵](#)

1. CR Huang ,
2. KH Burns ,
3. JD Boeke

, Transposición activa en genomas . Annu. Rev. Genet. **46** , 651 - 675 (2012).

[CrossRefPubMedGoogle Académico](#)

33. [↵](#)

1. HH Kazazian Jr ,
2. JV Moran

, ADN móvil en salud y enfermedad . N. Engl. J. Med. **377** , 361 - 370 (2017).

[CrossRefPubMedGoogle Académico](#)

34. [↵](#)

1. Ataúd JM ,
2. H. Ventilador

, El descubrimiento de la transcriptasa inversa . Annu. Rev. Virol. **3** , 29 - 51 (2016).

[Google Académico](#)

35. [↵](#)

1. M. De Cecco y col.

, L1 impulsa el IFN en las células senescentes y promueve la inflamación asociada a la edad . Nature **566** , 73 - 78 (2019).

[Google Académico](#)

36. [↵](#)

1. B. Rodríguez-Martin y col .; Grupo de trabajo de variación estructural del PCAWG; Consorcio PCAWG

, El análisis de pan-cáncer de genomas completos identifica reordenamientos impulsores promovidos por la retrotransposición de LINE-1 . Nat. Genet. **52** , 306 - 319 (2020).

[Google Académico](#)

37. [↵](#)

1. EC Scott y col.

, Un retrotransposón L1 caliente evade la represión somática e inicia el cáncer colorrectal humano . Genome Res. **26** , 745 - 755 (2016).

[Resumen / Texto completo GRATISGoogle Académico](#)

38. [↵](#)

1. RB Jones y col.

, El ADN del elemento retrotransponible LINE-1 se acumula en las células infectadas por el VIH-1 . J. Virol. **87** , 13307 - 13320 (2013).

[Resumen / Texto completo GRATISGoogle Académico](#)

39. [↵](#)

1. MG Macchietto ,
2. RA Langlois ,
3. SS Shen

, Regulación positiva de la expresión de elementos transponibles inducida por virus en células huésped humanas y de ratón . Life Sci. Alliance **3** , e201900536 (2020).

[Resumen / Texto completo GRATISGoogle Académico](#)

40. [↵](#)

1. Y. Yin ,
2. XZ Liu ,
3. X. Él ,
4. LQ Zhou

, El coronavirus exógeno interactúa con el retrotransposón endógeno en las células humanas . Parte delantera. Célula. Infectar. Microbiol. **11** , 609160 (2021).

[Google Académico](#)

41. [↵](#)

1. H. Kaessmann ,
2. N. Vinckenbosch ,
3. M. Long

, Duplicación de genes basada en ARN: conocimientos mecanicistas y evolutivos . Nat. Rev. Genet. **10** , 19 - 31 (2009).

[CrossRefPubMedGoogle Académico](#)

42. [↵](#)

1. S. Lanciano ,
2. G. Cristofari

, Medición e interpretación de la expresión de elementos transponibles . Nat. Rev. Genet. **21** , 721 - 736 (2020).

[Google Académico](#)

43. [↵](#)

1. TA Morrish y col.

., Reparación del ADN mediada por retrotransposición LINE-1 independiente de endonucleasas . Nat. Genet. **31** , 159 - 165 (2002).

[CrossRefPubMedGoogle Académico](#)

44. [↵](#)

1. JC Venter y col.

., La secuencia del genoma humano . Science **291** , 1304 - 1351 (2001).

[Resumen / Texto completo GRATISGoogle Académico](#)

45. [↵](#)

1. T. Sultana y col.

., El panorama de los retrotransposones L1 en el genoma humano está determinado por los sesgos de secuencia de preinserción y la selección posterior a la inserción . Mol. Celda **74** , 555 - 570.e7 (2019).

[CrossRefGoogle Académico](#)

46. [↵](#)

1. DA Flasch y col.

., La retrotransposición de novo L1 en todo el genoma conecta la actividad de la endonucleasa con la replicación . Celda **177** , 837 - 851.e28 (2019).

[Google Académico](#)

47. [↵](#)

1. DL Stern

, Mapeo basado en etiquetado (TagMap) de sitios de inserción genómica de ADN móvil . bioRxiv [Preprint] (2017). <https://doi.org/10.1101/037762> (consultado el 16 de febrero de 2021).

[Google Académico](#)

48. [↵](#)

1. S. Picelli y col.

., Tn5 transposasa y procedimientos de etiquetado para proyectos de secuenciación a gran escala . Genome Res. **24** , 2,033 mil - 2,040 (2014).

[Resumen / Texto completo GRATIS](#)[Google Académico](#)

49. [↵](#)

1. L. Zhang y col.

., ARN del SARS-CoV-2 transcrito de forma inversa e integrado en el genoma humano . bioRxiv [Preprint] (2020). <https://doi.org/10.1101/2020.12.12.422516> (consultado el 16 de marzo de 2021).

[Google Académico](#)

50. [↵](#)

1. D. Blanco-Melo y col.

., La respuesta desequilibrada del host al SARS-CoV-2 impulsa el desarrollo de COVID-19 . Celda **181** , 1036 - 1045.e9 (2020).

[CrossRefPubMed](#)[Google Académico](#)

51. [↵](#)

1. J. Huang y col.

., La infección por SARS-CoV-2 de células alveolares de pulmón humano tipo 2 derivadas de células madre pluripotentes provoca una rápida respuesta inflamatoria epitelial-intrínseca . Cell Stem Cell **27** , 962 - 973.e7 (2020).

[Google Académico](#)

52. [↵](#)

1. JA Perez-Bermejo y col .

., La infección por SARS-CoV-2 de células cardíacas derivadas de iPSC humanas refleja características citopáticas en corazones de pacientes con COVID-19 . Sci. Transl. Medicina. , doi: 10.1126 / scitranslmed.abf7872 (2021).

[Texto completo GRATISGoogle Académico](#)

53. [↵](#)

1. F. Jacob y col.

., Las células nerviosas derivadas de células madre pluripotentes humanas y los organoides cerebrales revelan que el neurotropismo del SARS-CoV-2 predomina en el epitelio del plexo coroideo . Cell Stem Cell **27** , 937 - 950.e9 (2020).

[Google Académico](#)

54. [↵](#)

1. GG Giobbe y col.

., Infección y replicación del SARS-CoV-2 en organoides gástricos fetales y pediátricos humanos . bioRxiv [Preprint] (2020). <https://doi.org/10.1101/2020.06.24.167049> (Consultado el 28 de octubre de 2020).

[Google Académico](#)

55. [↵](#)

1. SE Gill y col.

., Perfil transcripcional de leucocitos en pacientes con COVID19 críticamente enfermos: implicaciones para la respuesta al interferón y la coagulación . Med de cuidados intensivos. Exp. **8** , 75 (2020).

[Google Académico](#)

56. [↵](#)

1. D. Kim y col.

., La arquitectura del transcriptoma del SARS-CoV-2 . Celda **181** , 914 - 921.e10 (2020).

[CrossRefPubMedGoogle Académico](#)

57. [↵](#)

1. B. Yan y col.

., Los eventos quiméricos del huésped-virus en las células infectadas con SARS-CoV2 son poco frecuentes y artefactos . bioRxiv [Preprint] (2021). <https://doi.org/10.1101/2021.02.17.431704> (Consultado el 20 de febrero de 2021).

[Google Académico](#)

58. [↵](#)

1. Y. Han y col.

., Identificación de agentes terapéuticos candidatos a COVID-19 utilizando organoides pulmonares derivados de hPSC . bioRxiv [Preprint] (2020). <https://doi.org/10.1101/2020.05.05.079095> (Consultado el 10 de marzo de 2021).

[Google Académico](#)

59. [↵](#)

1. S. Alexandersen ,
2. A. Chamings ,
3. TR Bhatta

., Los ARN genómicos y subgenómicos del SARS-CoV-2 en muestras de diagnóstico no son un indicador de replicación activa . Nat. Comun. **11** , 6059 (2020).

[CrossRefPubMedGoogle Académico](#)

60. [↵](#)

1. N. Desai y col.

., Heterogeneidad temporal y espacial de la respuesta del huésped a la infección pulmonar por SARS-CoV-2 . Nat. Comun. **11** , 6319 (2020).

[Google Académico](#)

61. [↵](#)

1. M. Liao y col.

., Paisaje unicelular de células inmunes broncoalveolares en pacientes con COVID-19 . Nat. Medicina. **26** , 842 - 844 (2020).

[PubMedGoogle Académico](#)

62. [↵](#)

1. MC Dalakas

, Síndrome de Guillain-Barré: la primera enfermedad neurológica autoinmune desencadenada por COVID-19 documentada: más por venir con miositis a la vista . Neurol. Neuroimmunol. Neuroinflamm. **7** , e781 (2020).

[Resumen / Texto completo GRATISGoogle Académico](#)

63. [↵](#)

1. S. Pfeuffer y col.

., Autoinmunidad que complica la infección por SARS-CoV-2 en la deficiencia selectiva de IgA . Neurol. Neuroimmunol. Neuroinflamm. **7** , e881 (2020).

[Texto completo GRATISGoogle Académico](#)

64. [↵](#)

1. A. Baum y col.

., Los anticuerpos REGN-COV2 previenen y tratan la infección por SARS-CoV-2 en macacos rhesus y hámsteres . Ciencia **370** , 1110 - 1115 (2020).

[Resumen / Texto completo GRATISGoogle Académico](#)

65. [↵](#)

1. BJ Wagstaff ,
2. M. Barnerssoi ,
3. Soy Roy-Engel

, Conservación evolutiva de la modularidad funcional de elementos LINE-1 de primates y murinos . PLoS One **6** , e19672 (2011).

[CrossRefPubMedGoogle Académico](#)

66. [↵](#)

1. EA Farkash ,
2. GD Kao ,
3. SR Horman ,
4. ET Prak

, La radiación gamma aumenta la retrotransposición de L1 dependiente de endonucleasas en un ensayo de células cultivadas . Ácidos nucleicos Res. **34** , 1196 - 1204 (2006).

[CrossRefPubMedGoogle Académico](#)

67. [↵](#)

1. OMS

, Recurso de la Organización Mundial de la Salud (OMS) sobre ensayos moleculares desarrollados

internamente. https://www.who.int/docs/default-source/coronaviruse/whoinhouseassays.pdf?sfvrsn=de3a76aa_2 . Consultado el 6 de junio de 2020 .

[Google Académico](#)

68. [↵](#)

1. H. Li

, Minimap2: alineación por pares para secuencias de nucleótidos . Bioinformática **34** , 3094 - 3100 (2018).

[CrossRefPubMedGoogle Académico](#)

69. [↵](#)

1. WJ Kent

, BLAT: la herramienta de alineación similar a BLAST . Genome Res. **12** , 656 - 664 (2002).

[Resumen / Texto completo GRATISGoogle Académico](#)

70. [↵](#)

1. A. Dobin y col.

., STAR: Alineador universal ultrarrápido de RNA-seq . Bioinformática **29** , 15 - 21 (2013).

[CrossRefPubMedGoogle Académico](#)

71. [↵](#)

1. H. Li y col .; Subgrupo de procesamiento de datos del proyecto 1000 Genome

, El formato de alineación / mapa de secuencia y SAMtools . Bioinformática **25** , 2078 - 2079 (2009).

[CrossRefPubMedGoogle Académico](#)

72. [↵](#)

1. F. Ramírez y col.

., deepTools2: un servidor web de próxima generación para análisis de datos de secuenciación profunda . Ácidos nucleicos Res. **44** , W160 - W165 (2016).

[CrossRefPubMedGoogle Académico](#)

73. [↵](#)

1. AR Quinlan ,

2. Sala de mensajería instantánea

, BEDTools: un conjunto flexible de utilidades para comparar características genómicas . Bioinformática **26** , 841 - 842 (2010).

[CrossRefPubMedGoogle Académico](#)

74. ↵

1. GX Zheng y col.

., Perfiles transcripcionales digitales masivamente paralelos de células individuales . Nat. Comun. **8** , 14049 (2017).

[CrossRefPubMedGoogle Académico](#)

75. ↵

1. T. Stuart y col.

., Integración completa de datos de una sola celda . Cell **177** , 1888 - 1902.e21 (2019).

[CrossRefPubMedGoogle Académico](#)

Nosotros recomendamos

1. [El SARS-CoV-2 induce respuestas inmunitarias innatas mediadas por ARN bicatenario en células derivadas del epitelio respiratorio y cardiomiocitos](#)

Yize Li y otros, Proc Natl Acad Sci USA, 2021

2. [Base estructural para el retroceso por el complejo de replicación-transcripción del SARS-CoV-2](#)

Brandon Malone y otros, Proc Natl Acad Sci EE. UU.

3. [Secuenciación de ARN mediante etiquetado directo de híbridos de ARN / ADN](#)

Lin Di et al., Proc Natl Acad Sci EE. UU., 2020

4. [El análisis funcional y genético de los ortólogos del receptor viral ACE2 revela una amplia gama de huéspedes potenciales de SARS-CoV-2](#)

Yinghui Liu y otros, Proc Natl Acad Sci EE. UU.

5. [Parada traslacional y evasión de la respuesta inmune innata por la proteína NSP14 del SARS-CoV-2](#)

Jack Chun-Chieh Hsu y otros, Proc Natl Acad Sci EE. UU.

1. [N 6 -metiladenosina regula la abundancia de ARN de SARS-CoV-2](#)

Ting Zhang y otros, Cell Discovery, 2021

2. [Informe de Nature Papers sobre los factores del huésped necesarios para la infección por SARS-CoV-2, arquitectura de cromatina genómica de levadura](#)

Ciara Curtin, GenomeWeb

3. [La secuenciación de ARN unicelular revela la dinámica de la infección por SARS-CoV-2 en los pulmones de los monos verdes africanos](#)

Emily Speranza y otros, Sci Transl Med, 2021

4. [Panorama 3D de las interacciones del virus de la hepatitis B con las cromatinas humanas](#)

Bo Yang et al., Cell Discovery, 2020

5. [AstraZeneca obtiene la aprobación de la FDA para el adyuvante Tagrisso en cáncer de pulmón mutado en EGFR](#)

Noticias de oncología de precisión, 2020