

Flujo de trabajo de enriquecimiento para la detección del coronavirus mediante los sistemas NGS de Illumina

Detección y caracterización
extremadamente sensibles de virus
respiratorios comunes, incluidas
las cepas de coronavirus



Vigilancia vírica con NGS

Las infecciones víricas son un problema de salud global importante, con nuevas variantes de enfermedades infecciosas que no cesan de aparecer. El brote del nuevo coronavirus en 2019 (SARS-CoV-2) que comenzó en Wuhan, China, y se extendió rápidamente a varios países es un ejemplo importante de los posibles riesgos de las pandemias virales globales. Los coronavirus (CoV) son una gran familia de virus que pueden infectar a los seres humanos y causar enfermedades respiratorias que van desde el resfriado común hasta enfermedades más graves, como el síndrome respiratorio de Oriente Medio (MERS-CoV) y el síndrome respiratorio agudo grave (SARS-CoV). La Organización Mundial de la Salud (OMS) designó la enfermedad asociada al SARS-CoV-2 (COVID-19) como una emergencia de salud pública de interés internacional, lo que destaca la necesidad de una detección vírica rápida y precisa.¹

La secuenciación de nueva generación (NGS, Next-generation sequencing) proporciona una forma eficaz y novedosa de detectar muestras y virus sin el conocimiento previo del agente infeccioso.² El enriquecimiento de objetivos es un método de secuenciación que captura regiones genómicas de interés mediante la hibridación de ácidos nucleicos en sondas biotiniladas específicas según el objetivo. El enriquecimiento de objetivos de captura híbrida permite una detección muy sensible, sin necesidad de la alta profundidad de lectura que se exige para la secuenciación metagenómica indiscriminada. Además, el flujo de trabajo de NGS de enriquecimiento de objetivos permite datos de secuencias casi completos de los objetivos y abre aplicaciones como el análisis de variantes para la evolución vírica o la vigilancia vírica.³ En comparación con otros métodos de secuenciación de objetivos, como la secuenciación de amplicones, el enriquecimiento a través de la captura híbrida permite perfiles de sondas mucho más grandes con perfiles más completos de las regiones objetivo. Además, las sondas

de oligonucleótidos usadas para los protocolos de captura híbrida siguen siendo eficaces dentro de las regiones altamente mutágenas, lo que permite dirigirse a virus que evolucionan rápidamente, como los virus de ARN.

Flujo de trabajo del enriquecimiento vírico

Esta nota de aplicación destaca un flujo de trabajo optimizado para detectar y analizar el coronavirus con el kit de preparación de librerías Illumina DNA Prep with Enrichment, los perfiles de selección de objetivos víricos, la secuenciación probada de Illumina y el análisis de datos simplificado (figura 1). Este flujo de trabajo está concebido para enriquecer los objetivos de ADN y ARN víricos a partir de la extracción total de ácido nucleico. El protocolo comienza con la extracción de muestras de ácido nucleico y la posterior transcripción inversa del ARN extraído en un ADN complementario (ADNc) bicatenario. La preparación de librerías se realiza con Illumina DNA Prep with Enrichment con IDT for Illumina DNA/RNA UD Indexes. En este protocolo, el ADN y el ADN complementario (ADNc) se someten a tagmentación, limpieza y amplificación previa al enriquecimiento. Después de la amplificación, se pueden agrupar hasta 12 muestras para una reacción de enriquecimiento utilizando un perfil de oligonucleótidos que se dirigen a secuencias víricas. A la hibridación de la sonda le sigue la captura de esta, la amplificación de enriquecimiento, la cuantificación y la secuenciación.

Métodos

Preparación de muestras

Para demostrar el rendimiento del perfil de enriquecimiento de objetivos vírico, se analizó una cepa de la muestra de coronavirus desactivado, la cepa OC43 de CoV

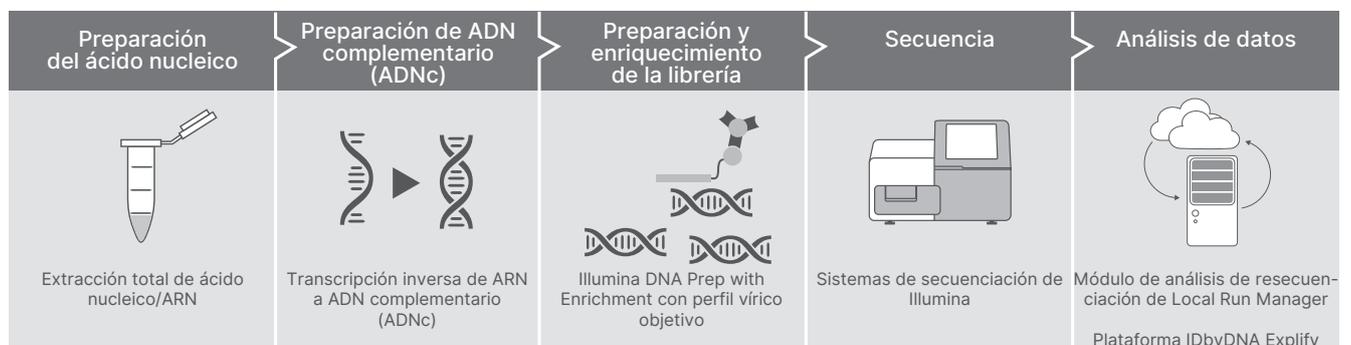


Figura 1: Flujo de trabajo de enriquecimiento para la detección del coronavirus: el flujo de trabajo de NGS optimizado para la detección del coronavirus integra la preparación de muestras, la preparación de librerías, el enriquecimiento de objetivos, la secuenciación y el análisis de datos.

de Microbiologics (QC Sets and Panels: Helix Elite, n.º de catálogo 8217) (tabla 1). La muestra de cultivo vírico de coronavirus se extrajo con QIAGEN QIAmp Viral Mini Kit (QIAGEN, n.º de catálogo 52904) en un entorno de laboratorio BSL2. Se invirtieron 150 ng de ARN extraído transcritos en ADN complementario (ADNc) mediante dos flujos de trabajo diferentes: uno derivado de los reactivos de ARN TruSeq de Illumina y el otro mediante el uso de Thermo Fisher Scientific Maxima H Minus Double-Stranded cDNA Synthesis Kit (Thermo Fisher Scientific, n.º de catálogo K2561). Para imitar los resultados de secuenciación de las muestras de pacientes, la muestra vírica también se añadió al ARN de fondo de referencia humana universal (UHR) al 95 % (Agilent Technologies, n.º de catálogo 74000) y se transcribió inversamente en ADN complementario [ADNc] utilizando los métodos descritos anteriormente (tabla 1).

Además, se extrajo un grupo de virus respiratorios que contenía cuatro virus de ARN y un virus de ADN (tabla 2) mediante QIAGEN AllPrep PowerViral DNA/RNA Kit (QIAGEN, n.º de catálogo 28000-50) en un entorno de laboratorio BSL2. El volumen máximo de ácido nucleico total se invirtió y se transcribió en ADN complementario (ADNc) (sin tratamiento con desoxirribonucleasa) mediante un flujo de trabajo derivado de los reactivos de ARN TruSeq de Illumina. Para imitar los resultados de secuenciación de las muestras de pacientes con varios patógenos presentes, la muestra de grupo vírico también se añadió al ARN de fondo de referencia humana universal (UHR) al 95 % (Agilent Technologies, n.º de catálogo 74000) y se transcribió inversamente en ADN complementario [ADNc] (tabla 1).

Tabla 1: Composición de muestras víricas para el análisis

Muestra	Composición ^a	Genoma de referencia
CoVOC43	150 ng de ARN de CoV OC43	Coronavirus humano OC43 AY391777.1
CoVOC43_95UHR	7,5 ng de ARN de CoV OC43 y 142,5 ng de ARN de UHR	Coronavirus humano OC43 AY391777.1
ViralPool	Volumen máximo de ácido nucleico total del grupo vírico (8,5 µl)	Consulte la tabla 2
ViralPool_95UHR	5 % de ácido nucleico total del grupo vírico y 95 % de ARN de UHR por volumen	Consulte la tabla 2

a. El ARN mínimo/entrada de ácido nucleico total recomendados para la transcripción inversa es de 10 ng. Para obtener los mejores resultados, se debe realizar la transcripción inversa en muestras de ácido nucleico recién extraídas.

Tabla 2: Composición del grupo vírico

Virus	Tipo de ácido nucleico	Genoma de referencia
Virus de la gripe A (H1N1)	ARN	Gripe A/ Michigan/45/2015
Virus de la gripe B	ARN	Gripe B/ Colorado/06/2017
Virus paragripal humano 3	ARN	NC_001796.2
RSV B9320	ARN	AY353550.1
Adenovirus 7	ADN	AY594255.1

Preparación de librerías

Las librerías preparadas para secuenciación se prepararon con Illumina DNA Prep with Enrichment (Illumina, n.º de catálogo 20025524) e IDT for Illumina DNA UD Indexes (Illumina, n.º de catálogo 20027213)* mediante ADN complementario (ADNc) de la muestra de CoV (CoVOC43), la muestra de grupo vírico (ViralPool) y ambas muestras de adición (CoVOC43_95UHR y ViralPool_95UHR). La entrada de ADN total recomendada para la tagmentación es de 10-1000 ng.

Después de la amplificación, las muestras se dividieron en grupos independientes para el enriquecimiento en función del tipo de muestra (CoVOC43, CoVOC43_95UHR, ViralPool y ViralPool_95UHR). Las reacciones de enriquecimiento se realizaron con el perfil de oligonucleótidos de virus respiratorios (Illumina, n.º de catálogo 20042472), que cuenta con aproximadamente 7800 sondas diseñadas para detectar virus respiratorios, cepas de gripe recientes y SARS-CoV-2, así como sondas humanas para actuar como controles positivos (tabla 3). Tras el enriquecimiento (figura 2), las librerías preparadas se cuantificaron, agruparon y cargaron en el sistema MiSeq™ para la secuenciación.

* El n.º de catálogo 20027213 ha quedado obsoleto, se ha sustituido por Illumina DNA/RNA UD Indexes Set A, Tagmentation (96 indexes, 96 samples), n.º de catálogo 20091654.

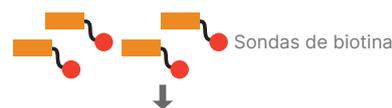
Tabla 3: Virus objetivo del perfil de oligonucleótidos de virus respiratorios

Coronavirus humano 229E	Virus paragripal humano 1
Coronavirus humano NL63	Virus paragripal humano 2
Coronavirus humano OC43	Virus paragripal humano 3
Coronavirus humano HKU1	Virus paragripal humano 4a
SARS-CoV-2	Metapneumovirus humano (CAN97-83)
Adenovirus humano B1	Virus respiratorio sincicial (tipo A)
Adenovirus humano C2	Virus respiratorio sincicial humano 9320 (tipo B)
Adenovirus humano E4	Virus de la gripe A (A/Puerto Rico/8/1934[H1N1])
Bocavirus humano 1 (cepa vírica aislada st2 del bocaparvovirus primate 1)	Virus de la gripe A (A/Korea/426/1968[H2N2])
Cepa vírica aislada PK-5510 del bocavirus humano 2c PK	Virus de la gripe A (A/New York/392/2004[H3N2])
Bocavirus humano 3	Virus de la gripe A (A/ganso/Guangdong/1/1996[H5N1])
Cepa HBoV4-NI-385 del bocavirus humano 4 NI	Virus de la gripe A (A/Zhejiang/DTID-ZJU01/2013[H7N9])
Poliomavirus KI Estocolmo 60	Virus de la gripe A (A/Hong Kong/1073/99[H9N2])
Poliomavirus WU	Virus de la gripe A (A/Texas/50/2012[H3N2])
Parecovirus humano tipo 1	Virus de la gripe A PicoBank/HPeV1/a (A/Michigan/45/2015[H1N1])
Parecovirus humano tipo 6	Virus de la gripe B (B/Lee/1940)
Rinovirus humano A89	Virus de la gripe B (B/Wisconsin/01/2010)
Rinovirus humano C (cepa 024)	Virus de la gripe B (B/Brisbane/60/2008)
Rinovirus humano B14	Virus de la gripe (B/Colorado/06/2017)
Cepa C104 de enterovirus humano: AK11	Virus de la gripe B (B/Washington/02/2019)
Cepa vírica aislada NICA08-4327 de enterovirus humano C109	Genes de control humanos

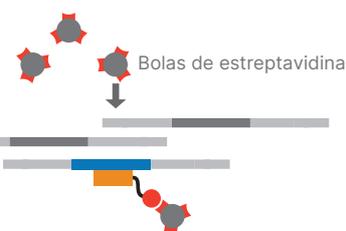
A. Librerías de Pool Illumina DNA Prep with Enrichment



B. Hibridación de sondas biotinilada para regiones objetivo



C. Captura con bolas de estreptavidina



D. Elución de las bolas

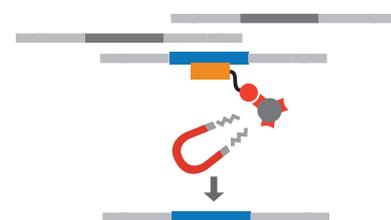


Figura 2: Química de enriquecimiento: las librerías preparadas se someten a enriquecimiento con un perfil de objetivos víricos, seguido de amplificación, cuantificación y secuenciación.

Secuenciación

Las librerías preparadas se pueden secuenciar en cualquier instrumento de Illumina. Los sistemas iSeq™ 100, MiniSeq™ y MiSeq de sobremesa son muy adecuados debido a los bajos requisitos de lectura de estas muestras. En este estudio, las librerías preparadas a partir de las muestras virales se desnaturalizaron y diluyeron hasta una concentración de carga final de 10 pM, de acuerdo con la Guía de librerías de desnaturalización y dilución del sistema MiSeq (n.º de documento 15039740 v10) y se secuenciaron en dicho sistema a una longitud de lectura de 2 × 151 pb con reactivos MiSeq v3 (Illumina, n.º de catálogo MS-102-3001). Las partes alícuotas de todas las librerías

también se secuenciaron antes de la hibridación de la sonda para determinar el cambio múltiple de la reacción de enriquecimiento.

El valor del virus, la calidad del ARN y el número de lecturas por muestra afectan al número de lecturas específicas del virus obtenidas y a la cobertura del genoma vírico. Como pauta general, la recomendación de lectura para este flujo de trabajo es de 500 000 lecturas por muestra, pero estos números pueden ser variables y esta cifra es solo un punto de partida recomendado.

Análisis de datos

Para el análisis integrado en el instrumento, los experimentos de secuenciación se configuraron en Local Run Manager mediante el módulo de análisis de resecuenciación. Este módulo permite introducir toda la información del experimento y la selección del genoma de referencia para la alineación de secuencias posterior. Los usuarios pueden cargar genomas de referencia directamente en el instrumento, lo que permite una fácil personalización. El análisis se inicia automáticamente una vez finalizada la secuenciación para que los usuarios puedan interpretar los resultados lo más rápido posible. El módulo de resecuenciación proporciona datos de alineación, cobertura y variantes pequeñas, así como archivos FASTQ, BAM y VCF para su uso en otras canalizaciones de análisis de datos, si se desea (figura 3).



Figura 3: Flujo de trabajo de análisis de enriquecimiento: los experimentos de secuenciación se configuran en Local Run Manager y el análisis de datos se realiza automáticamente una vez finalizado el experimento. Si lo desea, puede disponer de análisis adicionales con aplicaciones de terceros.

IDbyDNA Explify Platform es una solución fácil de usar para el análisis de datos en profundidad que cuenta con un sólido control de calidad de los datos (QC), una interpretación estandarizada de los resultados, bases de datos cuidadosamente seleccionadas y la generación de informes personalizados. El análisis de datos se basa en los k-mer y los pasos de alineación, incluida la detección de virus a nivel de proteínas, lo que aumenta la capacidad

de identificar virus nuevos y altamente divergentes. Se puede acceder a la plataforma en BaseSpace™ Sequence Hub.

Resultados

Después de la preparación y secuenciación de librería, se utilizó el módulo de análisis de resecuenciación Local Run Manager (v2.5.56.11) para alinear cada muestra con el genoma de referencia OC43 del coronavirus, como se describe en la [Guía de flujo de trabajo del módulo de análisis de resecuenciación Local Run Manager](#).

Transcripción inversa y detección de CoV con el módulo de resecuenciación Local Run Manager

En este estudio, se utilizaron dos kits de transcripción inversa diferentes para generar ADN complementario (ADNc) para la preparación de librerías: un kit disponible comercialmente de Thermo Fisher Scientific y un flujo de trabajo derivado de los reactivos de ARN TruSeq de Illumina (consulte la sección Métodos para obtener más detalles). En el caso de las muestras de CoVOC43, los resultados mostraron que el porcentaje de alineación con el genoma de referencia era superior al 99 % para el método de transcripción inversa de Illumina y superior al 97 % para el método de Thermo Fisher Scientific (figura 4). El porcentaje de alineación de las muestras de adición fue superior al 64 % y superior al 19 % para el perfil de oligonucleótidos de virus respiratorios, independientemente del método de transcripción inversa (figura 4).

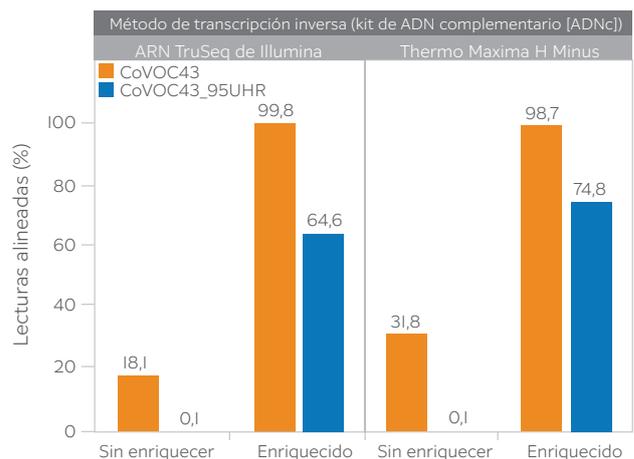


Figura 4: Rendimiento equivalente con diferentes kits de ADN complementario (ADNc): el porcentaje de lecturas alineadas para las muestras víricas transcritas de forma inversa con los dos kits de ADN complementario (ADNc) diferentes (consulte la sección Métodos) se traza antes (sin enriquecer) y después (enriquecidas) del enriquecimiento con el perfil de oligonucleótidos de virus respiratorios.

Como medida del éxito de la reacción de captura e hibridación del objetivo, se compararon los porcentajes de lecturas alineadas de las muestras antes y después del enriquecimiento para calcular el cambio múltiplo de enriquecimiento. Los resultados mostraron aumentos significativos en el enriquecimiento múltiplo para las muestras de CoV analizadas independientemente del método de transcripción inversa. Esto fue especialmente cierto para las muestras de CoVOC43_95UHR, que mostraron un enriquecimiento superior a 640 veces para el perfil de oligonucleótidos de virus respiratorios (tabla 4).

La cobertura media proporciona una descripción general de la profundidad de cobertura de cada base en el genoma de referencia, identificando cualquier región que no se haya secuenciado. El enriquecimiento con sondas de objetivo vírico aumentó en gran medida la cobertura media de la secuencia vírica para todas las muestras, independientemente del método de transcripción inversa (tabla 5).

Tabla 4: Criterios de medición de enriquecimiento del perfil de oligonucleótidos de virus respiratorios

Muestra	ARN TruSeq de Illumina		Thermo Maxima H Minus	
	CoVOC43	CoVOC43_95UHR	CoVOC43	CoVOC43_95UHR
Enriquecimiento múltiplo	5,5x	646x	3,1x	748x

Tabla 5: Cobertura media con el perfil de oligonucleótidos de virus respiratorios^a

Kit de ADN complementario (ADNc)	CoVOC43		CoVOC43_95UHR	
	Sin enriquecer	Enriquecido	Sin enriquecer	Enriquecido
ARN TruSeq de Illumina	427,1	3283,5	2,6	2437,3
Thermo Maxima H Minus	987,0	3861,4	4,2	3035,0

a. Los criterios de medición de cobertura media se han normalizado a 1 millón de lecturas por muestra.

Análisis con la plataforma IDbyDNA Explify

La plataforma Explify identificó el coronavirus añadido (figura 5) y los cinco virus del grupo vírico (figura 6). La plataforma Explify proporcionó secuencias genómicas consensuadas víricas, gráficos de cobertura y la capacidad demostrada de detectar coinfecciones con otros virus, bacterias, hongos o parásitos (figura 5 y figura 6).

La plataforma Explify también demostró que los datos de enriquecimiento de CoVOC43_95UHR del perfil de oligonucleótidos de virus respiratorios mostraron una cobertura equivalente a la de los datos de secuenciación metagenómica indiscriminada⁴ (figura 7),

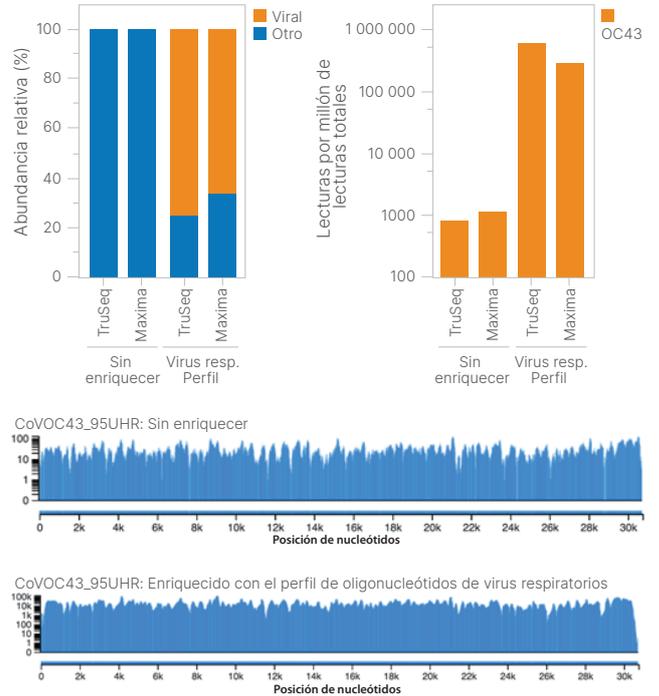


Figura 5: Identificación de OC43: se identificó OC43, independientemente del kit de ADN complementario (ADNc), con una cobertura genómica elevada. Los datos se redujeron a 1 millón de lecturas por muestra.



Figura 6: Caracterización de grupos víricos: se identificaron todas las cepas víricas de ARN y ADN en el grupo vírico. El grupo vírico no se verificó con técnicas ortogonales, por lo que no hay garantía de representación equivalente de cada virus. Los datos se redujeron a 1 millón de lecturas por muestra.

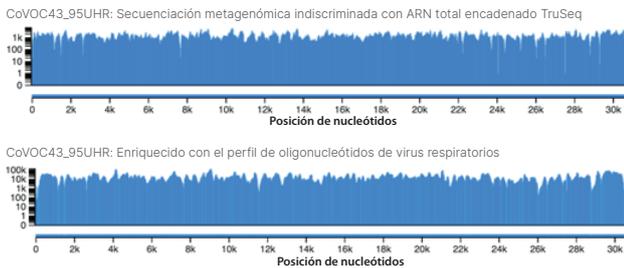
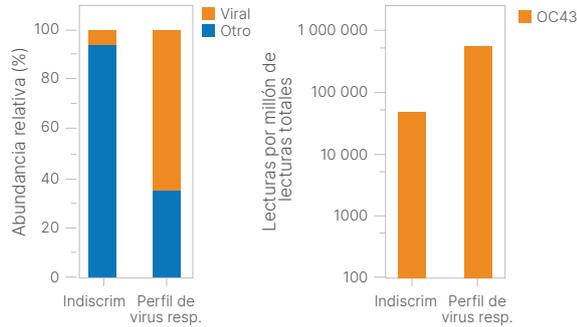


Figura 7: Rendimiento equivalente con el método indiscriminado: el enriquecimiento mostró un rendimiento y una cobertura equivalentes a la secuenciación metagenómica indiscriminada.

lo que demuestra que el enriquecimiento con grupos de sondas de objetivo víricas no introduce sesgos de cobertura en comparación con el método indiscriminado.

Resumen

La identificación y caracterización de virus emergentes es fundamental para mejorar la salud pública. En estas situaciones, NGS es un método potente para la detección de amplio alcance para identificar virus conocidos y emergentes. El uso de Illumina DNA Prep with Enrichment con perfiles dirigidos a patógenos víricos permite a los investigadores obtener datos genómicos que pueden confirmar la presencia de CoV y avanzar en los análisis, como el genotipado y el análisis de variantes. El diseño

independiente permite la identificación generalizada de virus patógenos en todos los tipos de muestras de interés y el uso de índices dobles únicos reduce el riesgo de cualquier cruce de indexado de muestras de multiplexado. Este flujo de trabajo fácil de seguir permite la detección y caracterización de brotes de patógenos, como el nuevo SARS-CoV-2.

Información adicional

Métodos de secuenciación de virus

Bibliografía

1. Organización Mundial de la Salud. COVID 19 Public Health Emergency of International Concern (PHEIC) Global research and innovation forum: towards a research roadmap. Fecha de publicación: 12 de febrero de 2020. Fecha de consulta: 9 de septiembre de 2023.
2. Bulcha B. Review on viral metagenomics and its future perspective in zoonotic and arboviral disease surveillance. *J Biol Agric Healthcare*. 2017;7(21):35–41.
3. Gaudin M, Desnues C. Hybrid Capture-Based Next Generation Sequencing and Its Application to Human Infectious Diseases. *Front Microbiol*. 2018;9:2924. Fecha de publicación: 27 de noviembre de 2018. doi:10.3389/fmicb.2018.02924
4. Illumina. Flujo de trabajo exhaustivo para la detección del coronavirus mediante los sistemas de sobremesa de Illumina. Año de publicación: 2020. Fecha de acceso: 29 de agosto de 2023.



1.800.809.4566 (llamada gratuita, EE. UU.) | Tel.: +1 858 202 4566
 techsupport@illumina.com | www.illumina.com

© 2023 Illumina, Inc. Todos los derechos reservados. Todas las marcas comerciales pertenecen a Illumina, Inc. o a sus respectivos propietarios. Si desea consultar información específica sobre las marcas comerciales, consulte www.illumina.com/company/legal.html.
 M-GL-02146 ESP v1.0