

# Análisis secundario de DRAGEN™

Llamada de variantes precisa,  
exhaustiva y eficiente para  
datos de secuenciación  
de nueva generación (NGS,  
next-generation sequencing)



## Introducción

Para avanzar en la investigación y la medicina de precisión, es fundamental liberar el potencial del genoma mediante la secuenciación de nueva generación (NGS, Next-Generation Sequencing). Con el fin de obtener el máximo de conocimientos genéticos a través de la NGS, los investigadores precisan herramientas de análisis de datos que permitan una traducción precisa y eficiente de los datos sin procesar de secuenciación en resultados reveladores. Además, para aprovechar los beneficios de la NGS, las organizaciones necesitan soluciones fáciles de usar, que se adapten a distintos usuarios, y con el menor número posible de barreras, tanto técnicas como económicas, para su adopción.

El análisis secundario DRAGEN (Dynamic Read Analysis for GENomics) de Illumina se desarrolló para abordar importantes retos asociados al análisis de datos de NGS para una amplia serie de aplicaciones, incluidos los estudios de genoma completo, exoma, transcriptoma, metiloma y muchos más. El software de análisis secundario DRAGEN es un paquete de aplicaciones que procesa datos de NGS y permite el análisis terciario para obtener información. Las herramientas disponibles conforman una solución muy precisa, exhaustiva y eficiente que facilita a los laboratorios de cualquier tamaño y disciplina sacar más provecho de sus datos genómicos.

## Resultados precisos

El análisis secundario DRAGEN genera resultados excepcionalmente precisos. En el PrecisionFDA Truth Challenge V2 (PrecisionFDA V2) de 2020, el análisis secundario DRAGEN v3.7 obtuvo la mayor precisión en todas las regiones de control y las regiones difíciles de asignar con los datos de secuenciación de Illumina.<sup>1,2</sup> Las versiones posteriores siguen estableciendo nuevos estándares de precisión, con avances en áreas como el aprendizaje automático (ML) y la tecnología multigenómica (grafos) de DRAGEN. La última versión del análisis secundario DRAGEN v4.3 proporciona una precisión sin precedentes en la llamada de variantes pequeñas con una puntuación F1 del 99,89 % (una medida combinada de precisión y retirada) en todas las regiones de control (figura 1). Esto se activa mediante el control multigenómico DRAGEN de tercera generación (gráfico), basado en 128 muestras con 256 haplotipos a partir de datos de control del pangenoma creados internamente, lo que captura una mayor diversidad genética. También contribuye a mejorar la precisión el nuevo llamador de mosaico integrado que se puede habilitar para detectar variantes de mosaico con frecuencias de alelos tan bajas como el 3 %.

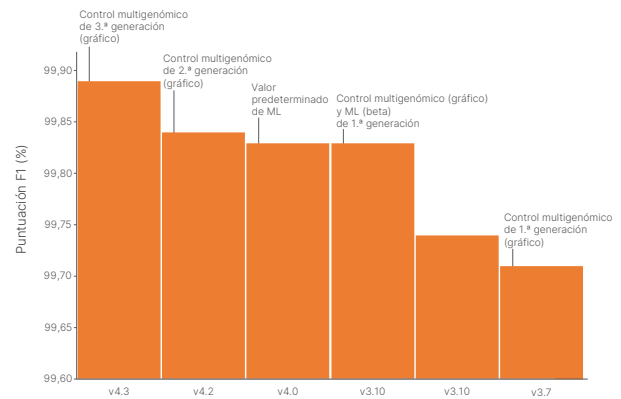


Figura 1: Precisión del análisis secundario DRAGEN: la puntuación F1 (%) en el eje de ordenadas es un cálculo de resultados positivos verdaderos y negativos verdaderos, representado como una proporción de los resultados totales.<sup>3,4</sup>

## Análisis exhaustivo

Con una completa cobertura del genoma y un amplio juego de aplicaciones compatibles, el análisis secundario DRAGEN satisface las diversas necesidades de los laboratorios que realizan análisis NGS. Los procesos de DRAGEN admiten varios tipos de experimentos, entre los que se incluyen la secuenciación del genoma completo (WGS por sus siglas en inglés), secuenciación del exoma completo, paneles de enriquecimiento, RNA-Seq de células individuales, ATAC-Seq de células individuales, RNA-Seq masiva y análisis de metilación (tabla 1). Se necesitarían más de 30 herramientas de código abierto para reproducir parcialmente la amplitud de funciones del software DRAGEN.<sup>3,4</sup>

Para el análisis germinal, el análisis secundario DRAGEN incluye un paquete de llamadores de variantes como ExpansionHunter y llamadores selectivos como *SMN*, *GBA*, *CYP2B6*, *CYP2D6* y *HLA*. DRAGEN v4.3 también presenta MRJD, un nuevo llamador especializado que permite la cobertura de genes difíciles en regiones de duplicación segmentaria, como *PMS2*, *SMN1*, *SMN2*, *STRC*, *NEB*, *TTN* y *IKBK*. Estas herramientas permiten el análisis de una amplia gama de variaciones genéticas, incluidas variaciones de nucleótido único, inserciones y deleciones (indels), expansiones repetidas y variaciones estructurales en regiones genómicas ampliadas. Además, el control multigenómico de DRAGEN (grafos) mejora la calidad de la asignación, lo que conduce a una mayor precisión en la llamada de variantes y a la resolución de áreas del genoma que son difíciles de evaluar debido a las complejidades de la secuencia. Esto mejora la cobertura de genes con potencial médicamente relevante y permite la llamada de variantes de nucleótido único, de indel pequeño, de variación de número de copias y estructurales en regiones difíciles de asignar.

Tabla 1: El análisis secundario DRAGEN es compatible con multitud de aplicaciones de análisis secundario<sup>a</sup>

| Aplicación                                  | Servidor en las instalaciones | Sistemas de secuenciación de Illumina integrados |  | Plataformas en la nube de Illumina |                              |
|---|-------------------------------|--|--|------------------------------------|------------------------------|
|   | Servidor DRAGEN               | NovaSeq X Series                                 | NextSeq 1000 System, NextSeq 2000 System | BaseSpace Sequence Hub             | Illumina Connected Analytics |
| BCL convert (Conversión de BCL)             | ✓                             | ✓  | ✓  | ✓                                  | Solo personalizado           |
| DRAGEN Ora Compression                      | ✓                             | ✓  | ✓  |                                    | Solo personalizado           |
| DRAGEN FASTQ + MultiQC                      | ✓                             | ✓  | ✓  | ✓                                  | ✓                            |
| Genoma completo                             | Germinal + somático           | Germinal + somático                              | Germinal + somático                      | Germinal + somático                | Germinal + somático          |
| Enriquecimiento (incluido el exoma)         | Germinal + somático           | Germinal + somático                              | Germinal + somático                      | Germinal + somático                | Germinal + somático          |
| DRAGEN Amplicon                             | ✓                             |  | Solo ADN                                 | ✓                                  | ✓                            |
| ARN   | ✓                             | ✓  | ✓  | ✓                                  | ✓                            |
| ARN de células únicas                       | ✓                             |  | ✓  | ✓                                  | ✓                            |
| NanoString GeoMx NGS                        |                               |  | ✓  | ✓                                  |                              |
| Metilación                                  | ✓                             | ✓  |  | ✓                                  | ✓                            |
| Metagenómica                                | ✓ <sup>b</sup>                |  |  | ✓                                  |                              |
| Detección de patógenos de ARN               |                               |  |  | ✓                                  |                              |
| COVID                                       | COVIDSeq. Estirpe de COVID    |  | COVIDSeq (solo en la nube)               | COVIDSeq. Estirpe de COVID         |                              |
| Gama de soluciones TruSight Oncology 500    | ✓                             |  |  | ✓ <sup>c</sup>                     | ✓                            |
| scATAC-Seq                                  | ✓                             |  |  | ✓                                  | ✓                            |
| Atribución                                  | ✓                             |  |  | ✓                                  | ✓                            |
| PGx Star Allele Caller                      | ✓                             | ✓  | ✓  | ✓                                  | ✓                            |
| Illumina Complete Long Reads                |                               |  |  | ✓                                  | ✓                            |
| Análisis secundario DRAGEN para RPIP y UPIP | ✓                             |  |  | ✓                                  | Beta                         |

a. La versión principal del software DRAGEN varía de una plataforma a otra; para más información, consulte a un representante local.

b. Aplicaciones de metagenómica habilitadas por clasificador K-mer, más herramientas disponibles próximamente.

c. Se requiere una suscripción a Illumina Connected Analytics.

## Análisis eficiente

El software DRAGEN está diseñado para proporcionar a los laboratorios la velocidad de análisis de datos que necesitan para optimizar la eficiencia del procesamiento de sus conjuntos de datos de NGS. El análisis secundario DRAGEN está optimizado con aceleración por hardware y utiliza una arquitectura de array de puertas programables por campo (FPGA, Field-Programmable Gate Array) para lograr tiempos de procesamiento rápidos. La eficiencia de los algoritmos de análisis DRAGEN dio lugar a dos registros de velocidad mundial para el análisis de datos genómicos.<sup>5,6</sup> En aplicaciones prácticas, el análisis secundario DRAGEN en las instalaciones puede procesar datos de NGS para un genoma completo equivalente con una cobertura de 40x en aproximadamente 35 minutos con todos los llamadores\* frente a las más de 8 horas que tardan los métodos de código abierto más utilizados en llamar a un número limitado de tipos de variantes.<sup>7</sup>

\* Basado en datos internos de Illumina basados en los estándares HG001-HG007 en el servidor DRAGEN v4, sin nuevos llamadores especializados como MRJD y VNTR disponibles en DRAGEN v4.3.

Para facilitar el almacenamiento, la gestión y el intercambio de archivos de datos de NGS grandes, la tecnología DRAGEN Original Read Archive (ORA) proporciona una compresión sin pérdidas de hasta 5 veces de los archivos FASTQ en el formato fastq.gz tradicional. La compresión sin pérdidas de DRAGEN ORA mantiene los detalles de los archivos FASTQ y es considerablemente más rápida; se requieren aproximadamente 8 minutos para comprimir archivos FASTQ de 50 a 70 GB, lo que cubre una amplia gama de especies de estudio habitual. El análisis secundario DRAGEN cuenta con un juego versátil de procesos que también pueden aceptar archivos de datos de entrada y generar archivos de rendimiento en diferentes etapas de los procesos (figura 2).

### FPGA y optimización con aceleración por hardware

La FPGA de gran capacidad de configuración permite instalaciones ultraeficientes optimizadas con aceleración por hardware de algoritmos de análisis genómico, como la conversión de archivos de llamada de bases (BCL, Base Call), asignaciones, alineaciones, clasificaciones, marcado de duplicados y llamada de variantes de haplotipo. La naturaleza flexible de las FPGA permite a Illumina desarrollar un amplio paquete de procesos de DRAGEN para diversas aplicaciones, con actualizaciones y adiciones frecuentes para ofrecer la mayor precisión, exhaustividad y eficiencia posibles.



Figura 2: Flexibilidad de los procesos de análisis secundario DRAGEN: cada proceso de DRAGEN contiene un juego específico de etapas para respaldar un análisis preciso y eficiente. El proceso del ejemplo germinal de genoma completo de DRAGEN proporciona flexibilidad para aceptar diversos archivos de entrada y producir una amplia variedad de tipos de rendimiento, lo que permite a los usuarios personalizar su experiencia y generar el formato de archivo deseado.

a. La conversión de BCL también está disponible como herramienta independiente.

## Referencias personalizadas

El análisis secundario DRAGEN permite a los usuarios generar una referencia humana, no humana o no estándar personalizada. Las referencias creadas se pueden utilizar como entrada en todas las aplicaciones de DRAGEN compatibles con los archivos de referencia personalizados. La mayoría de procesos de DRAGEN incluyen compatibilidad incorporada para hg19, hg38 (con o sin HLA), GRCh37, CHM13v2 y hs37d5. El software DRAGEN permite a los usuarios ampliar las capacidades de control multigenómico estándar de gráficos tanto para poblaciones diversas como específicas.

## Flexibilidad

El análisis secundario DRAGEN permite flexibilizar las operaciones del laboratorio según las necesidades, al tiempo que mantienen bajos los costes y los tiempos de procesamiento. El software DRAGEN puede facilitar la expansión de las funciones de investigación de varias formas:

- 1. Mantenerse al día con los sistemas NovaSeq™ X Series y NextSeq™ 1000 y NextSeq 2000:** DRAGEN integrado puede realizar múltiples aplicaciones simultáneas (cuatro aplicaciones simultáneas con un máximo de una conversión de BCL y tres canales más de su elección) por celda de flujo en un experimento único.
- 2. Transmisión por ráfagas:** durante los tiempos de aumento de las cargas de trabajo con altos volúmenes de muestras, los laboratorios pueden aprovechar la capacidad adicional en la nube con el análisis

secundario DRAGEN en Illumina Connected Analytics o en las aplicaciones DRAGEN en BaseSpace Sequence Hub (figura 3).

- 3. Ampliación de las operaciones:** se puede utilizar una única instancia de DRAGEN para ejecutar todos los procesos de DRAGEN y los tipos de muestras compatibles. La exhaustividad y la eficiencia del software DRAGEN permiten a los usuarios ampliar las operaciones, sin comprometer los tiempos de procesamiento ni la calidad de los resultados.
- 4. Transición a genomas:** los procesos preconstruidos de DRAGEN permiten una transición sencilla de paneles selectivos a exomas y a genomas.
- 5. Iniciativas de genómica para grandes poblaciones:** el análisis secundario DRAGEN ofrece un flujo de trabajo simplificado para análisis de cohortes a gran escala, con múltiples procesos que se utilizan conjuntamente para la llamada de variantes genéticas con una alta precisión. DRAGEN gVCF Genotyper permite la agregación de miles a millones de archivos de formato de llamada de variantes genómicas (gVCF, Genomic Variant Call Format) e incorpora nuevos lotes sin volver a procesar los existentes. La compresión ORA ahorra costes de almacenamiento.
- 6. Aplicaciones de secuenciación profunda:** el análisis secundario de DRAGEN admite el análisis de secuenciación de alta profundidad con alta eficiencia para una cobertura media de más de 300× para genomas y 1000× para exomas. Las capacidades de secuenciación profunda resultan valiosas para aplicaciones tales como la investigación oncológica y los estudios de enfermedades genéticas raras.

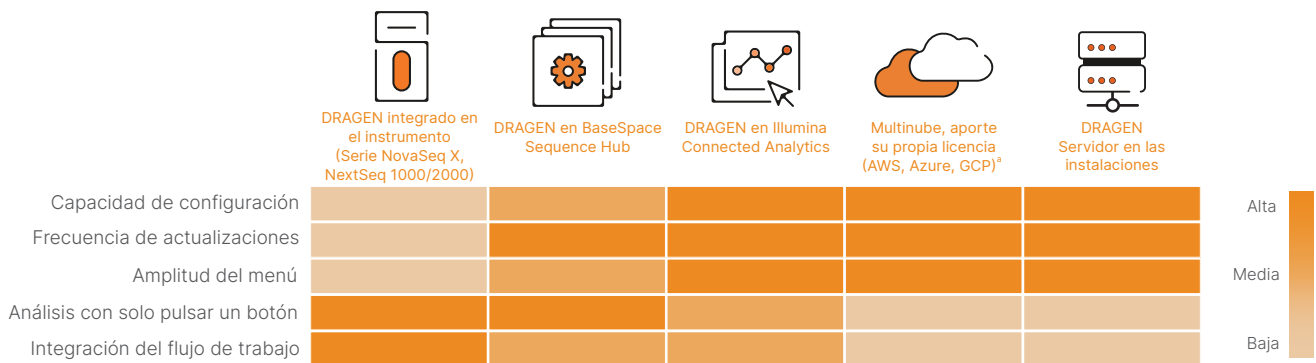


Figura 3: Opciones de acceso a los procesos de DRAGEN con funciones diseñadas para adaptarse a las necesidades de análisis de NGS de cada laboratorio.

a. Póngase en contacto con su representante de Illumina para obtener información sobre el acceso a Amazon Web Service (AWS), Azure o Google Cloud Platform (GCP, acceso temprano).

## Accesibilidad a través de diversas plataformas

El paquete de procesos de DRAGEN es accesible a través de soluciones en las instalaciones, integradas en el instrumento o basadas en la nube, para que cada laboratorio elija la solución que mejor se adapte a sus necesidades (figura 3).

### Servidor DRAGEN en las instalaciones

El servidor DRAGEN en las instalaciones depende de una solución de almacenamiento local para recopilar y almacenar datos de NGS. Una vez que los datos de secuenciación sin procesar se hayan transferido del instrumento de secuenciación al almacenamiento local mediante una conexión de red local, se transfieren al servidor DRAGEN para ejecutar el flujo de trabajo seleccionado. Después del análisis, el software guarda los archivos de rendimiento generados en la ubicación de almacenamiento local. El servidor DRAGEN en las instalaciones:

- Admite la configuración flexible de las funciones de DRAGEN mediante una interfaz de línea de comandos
- Sustituye hasta 30 instancias computacionales tradicionales
- Procesa datos de NGS para un genoma humano completo a una cobertura de 40× en aproximadamente 35 minutos

### DRAGEN integrado en el instrumento en NovaSeq X Series

NovaSeq X Series incluye un análisis secundario DRAGEN integrado, que ofrece un análisis preciso, automatizado y optimizado, diseñado para admitir el extraordinario volumen de datos generados por NovaSeq X Series. El paquete de software DRAGEN integrado en el instrumento proporciona análisis secundario y compresión de ORA con aplicaciones de NGS (tabla 1) que abarcan la conversión de BCL, la línea germinal, somático, enriquecimiento, ARN y metilación. DRAGEN integrado en el instrumento:

- Ejecuta varios procesos de análisis secundario en paralelo
- Realiza hasta cuatro aplicaciones simultáneas por celda de flujo en un experimento único
- Permite ahorrar hasta 5× en compresión de datos sin pérdidas y en costes de almacenamiento
- Proporciona una reducción del coste de los análisis que, en cinco años, puede superar el precio de compra de NovaSeq X System

### DRAGEN integrado en el instrumento en NextSeq 1000 System y NextSeq 2000 System

NextSeq 1000 System y NextSeq 2000 System incluyen el software DRAGEN integrado en el instrumento para un análisis secundario rápido y preciso. Se accede al software mediante una interfaz gráfica fácil de usar que permite a usuarios expertos y no expertos realizar los análisis necesarios y producir resultados con suma rapidez. El software DRAGEN integrado en el instrumento ofrece un juego selecto de procesos para cubrir una gama de aplicaciones de NGS comunes (tabla 1) e incluye un galardonado análisis de control de ML y multigenómico (gráfico) para la llamada de variantes de alta calidad. DRAGEN integrado en el instrumento:

- Ofrece la mayor precisión de cualquier sistema de secuenciación de sobremesa con el análisis secundario DRAGEN integrado en el instrumento
- Proporciona acceso a determinados procesos informáticos de DRAGEN
- Permite a los usuarios generar resultados en tan solo dos horas
- Utiliza algoritmos de procesos intuitivos para reducir la dependencia de expertos informáticos externos

### BaseSpace Sequence Hub

El paquete DRAGEN basado en la nube disponible en BaseSpace Sequence Hub combina el análisis preciso y eficiente con un ecosistema seguro y una gran flexibilidad. El software DRAGEN en BaseSpace Sequence Hub permite realizar análisis secundarios con solo pulsar un botón en laboratorios de cualquier tamaño y disciplina. BaseSpace Sequence Hub es una extensión directa de sus instrumentos de Illumina. Los datos cifrados se transfieren del instrumento a BaseSpace Sequence Hub, para que pueda gestionar y analizar fácilmente sus datos con un juego seleccionado de aplicaciones. BaseSpace Sequence Hub, respaldado por Amazon Web Services (AWS):

- Ofrece una solución fácil de usar con solo pulsar un botón para análisis DRAGEN.
- Utiliza una interfaz de usuario gráfica intuitiva para un funcionamiento eficiente por parte de usuarios expertos e inexpertos.
- Proporciona acceso a potentes recursos informáticos sin necesidad de invertir en infraestructuras adicionales.

## ILLUMINA CONNECTED ANALYTICS

ILLUMINA CONNECTED ANALYTICS es una exhaustiva plataforma bioinformática basada en la nube que permite a los investigadores gestionar, analizar e interpretar grandes volúmenes de datos multiómicos en un entorno seguro, flexible y adaptable. Acceda al paquete de análisis secundario DRAGEN en ILLUMINA CONNECTED ANALYTICS, disponible como procesos preempaquetados o herramientas individuales para incorporarlos a procesos personalizados.

## RESUMEN

El análisis secundario DRAGEN es un potente paquete de herramientas de software que proporciona un análisis preciso, exhaustivo y eficiente de los datos de NGS. Las múltiples opciones de instalación del software DRAGEN permiten a los laboratorios seleccionar la solución que mejor se adapte al tipo y a la escala de sus proyectos. Además, los usuarios pueden combinar varias opciones de instalación para adaptarse mejor a sus necesidades de rendimiento y flujo de trabajo. A tenor de la continua evolución de la tecnología de NGS, las actualizaciones puntuales del análisis secundario DRAGEN garantizan el mejor rendimiento posible de los procesos actuales, sin cesar en la incorporación de nuevos procesos a medida que se disponen de nuevas aplicaciones.

## INFORMACIÓN ADICIONAL

[Análisis secundario DRAGEN](#)

[Página de asistencia del análisis secundario DRAGEN](#)

[Póngase en contacto con nosotros](#)



1 800 809 4566 (llamada gratuita, EE. UU.) | tel.: +1 858 202 4566  
techsupport@illumina.com | www.illumina.com

© 2024 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](http://www.illumina.com/company/legal.html).  
M-GL-00680 ESP v10.0

## Bibliografía

1. The Food and Drug Administration. Truth Challenge V2: Calling Variants from Short and Long Reads in Difficult-to-Map Regions. <https://precision.fda.gov/challenges/10/results>. Fecha de consulta: 14 de marzo de 2022.
2. Catreux S, Jain V, Murray L, et al. DRAGEN sets new standard for data accuracy in PrecisionFDA benchmark data. Optimizing variant calling performance with Illumina machine learning and DRAGEN graph. Sitio web de Illumina. <https://www.illumina.com/science/genomics-research/articles/dragen-shines-again-precisionfda-truth-challenge-v2.html>. Fecha de publicación: 12 de enero de 2022. Fecha de consulta: 11 de julio de 2024.
3. Mehio R, Ruehle M, Catreux S, et al. DRAGEN wins at PrecisionFDA Truth Challenge V2 showcase accuracy gains from alt-aware mapping and graph reference genomes. Sitio web de Illumina. [illumina.com/science/genomics-research/articles/dragen-wins-precisionfda-challenge-accuracy-gains.html](https://www.illumina.com/science/genomics-research/articles/dragen-wins-precisionfda-challenge-accuracy-gains.html). Fecha de publicación: 9 de noviembre de 2020. Fecha de consulta: 11 de julio de 2024.
4. Datos internos archivados. Illumina, Inc. 2022.
5. BioIT World. Children's Hospital Of Philadelphia, Edico Set World Record For Secondary Analysis Speed. [bio-itworld.com/news/2017/10/23/children-s-hospital-of-philadelphia-edico-set-world-record-for-secondary-analysis-speed](https://www.bio-itworld.com/news/2017/10/23/children-s-hospital-of-philadelphia-edico-set-world-record-for-secondary-analysis-speed). Fecha de publicación: 23 de octubre de 2017. Fecha de consulta: 11 de julio de 2024.
6. San Diego Union Tribune. Rady Children's Institute sets Guinness world record. <https://www.sandiegouniontribune.com/2018/02/12/rady-childrens-institute-sets-guinness-world-record/>. Fecha de publicación: 12 de febrero de 2018. Fecha de consulta: 11 de julio de 2024.
7. Betschart RO, Thiéry A, Aguilera-García D, et al. Comparison of calling pipelines for whole genome sequencing: an empirical study demonstrating the importance of mapping and alignment. *Sci Rep.* 2022;12(1):21502. Fecha de publicación: 13 de diciembre de 2022. doi:10.1038/s41598-022-26181-3
8. Gross A, Maciucă S, Cox A, et al. Accurate and Efficient Calling of Small and Large Variants from PopGen data sets Using the DRAGEN Bio-IT Platform. Sitio web de Illumina. [www.illumina.com/science/genomics-research/articles/popgen-variant-calling-with-dragen.html](https://www.illumina.com/science/genomics-research/articles/popgen-variant-calling-with-dragen.html). Fecha de publicación: 24 de mayo de 2021. Fecha de consulta: 14 de marzo de 2022.