

DRAGEN™ Bio-IT Plattform

Präzise, umfassende und
effiziente Sekundäranalyse
von Next-Generation
Sequencing-Daten

illumina®

Einleitung

Die Erschließung der im Genom enthaltenen Informationen durch Sequenzierung der nächsten Generation (NGS, Next-Generation Sequencing) ist entscheidend für Fortschritte in Forschung und Medizin. NGS lässt sich bei genetischen Untersuchungen nur optimal einsetzen, wenn die Forscher Datenanalysetools zur Verfügung haben, die durch Sequenzierung gewonnene Rohdaten präzise und effizient in aussagekräftige Ergebnisse überführen. Außerdem können Organisationen die Vorteile von NGS nur nutzen, wenn sie einfach zu verwendende Lösungen erhalten, die für eine Vielzahl von Anwendern geeignet sind und bei denen Preis und technische Hürden keine relevanten Hindernisse für die Einführung darstellen.

Die Illumina DRAGEN (Dynamic Read Analysis for GENomics) Bio-IT Plattform dient zur Bewältigung wichtiger Herausforderungen bei der Analyse von NGS-Daten in einer breiten Palette von Anwendungen, darunter Genom-, Exom-, Transkriptom- und Methylo-m-Studien. Bei der DRAGEN-Plattform handelt es sich um eine Software-Suite für die Sekundäranalyse, die NGS-Daten verarbeitet und die Voraussetzungen für den Erkenntnisgewinn durch Tertiäranalysen schafft. Die verfügbaren Tools bilden eine hochpräzise, umfassende und effiziente Lösung, mit der Labore aller Größen und Fachrichtungen Genomdaten besser auswerten können.

Präzise Ergebnisse

Die DRAGEN Bio-IT Plattform liefert Ergebnisse mit höchster Genauigkeit. 2020 gewann DRAGEN v3.7 bei der PrecisionFDA Truth Challenge V2 (PrecisionFDA V2) mit den präzisesten Daten die Kategorien „Alle Benchmark-Regionen“ und „Schwer zu mappende Regionen“ für Illumina-Sequenzierungsdaten.^{1,2} Innovationen bei Graph Genomes und Illumina Machine Learning (ML) mit DRAGEN 4.0 setzten in Bezug zu allen Sequenzierungstechnologien neue Maßstäbe für die Datengenauigkeit. In der Kategorie „Alle Benchmark-Regionen“ wurde ein F1-Score (kombinierte Messung von Präzision und Recall) von 99,83 % erreicht (Abbildung 1).^{1,2} DRAGEN 4.0 + Graph (ML standardmäßig aktiviert) erzielten außerdem den höchsten F1-Score für das genaueste Calling unter allen Einreichungen bei der PrecisionFDA V2 in der Kategorie „Haupthistokompatibilitätskomplex-Regionen“ (MHC, Major Histocompatibility Complex).

Umfassende Plattform

Die DRAGEN Bio-IT Plattform erfüllt die Anforderungen von Laboren, die ein breites Spektrum von NGS-Anwendungen durchführen, und bietet eine umfassende Coverage für eine Vielzahl von Experimenttypen in einer einzigen Plattform.

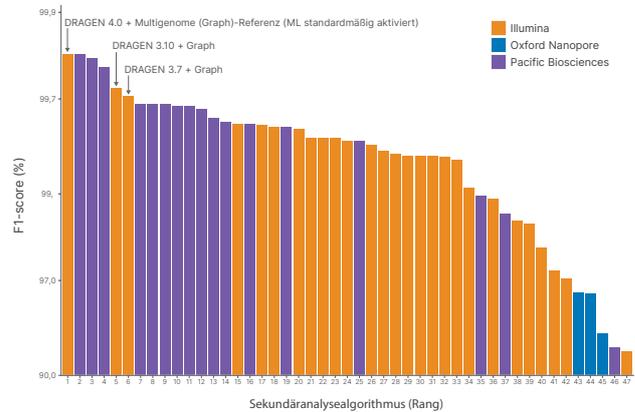


Abbildung 1: Genauigkeit von DRAGEN 4.0 + Graph (ML standardmäßig aktiviert) im Vergleich zu den PrecisionFDA Truth Challenge v2-Einreichungen im Datensatz „Alle Benchmark Regionen“: DRAGEN 4.0 + Graph (ML standardmäßig aktiviert) zeigt eine außergewöhnliche Genauigkeit, gleichauf mit Google DeepVariant auf Pacific Biosciences-Sequenzierungsdaten. DRAGEN 3.10 + Graph ist DRAGEN 3.7 + Graph aufgrund von Verbesserungen bei der Handhabung von Graphen und Referenz/alternativem Contig überlegen. Die Y-Achse, F1-Score (%), gibt die berechneten richtig positiven und richtig negativen Ergebnisse als Anteil an den Gesamtergebnissen an.^{3,4}

DRAGEN-Pipelines für Sekundäranalysen eignen sich für verschiedene Versuchsarten, einschließlich Genomsequenzierung (WGS, Whole-Genome Sequencing), Anreicherungspanels, Einzelzell-RNA-Seq, Einzelzell-ATAC-Seq, Einzelzell-Multiomik, Massen-RNA-Seq und Methylierungsanalyse (Tabelle 1). Es wären über 30 Open-Source-Tools erforderlich, um den Funktionsumfang von DRAGEN-Software auch nur annähernd zu erreichen.^{3,4}

Mit der mitgelieferten Suite von Varianten-Callern – Repeat-Expansion, strukturelle Variante (SV), Kopienzahlvariante (CNV, Copy Number Variation), ExpansionHunter sowie gezielte Caller wie *SMN*, *GBA*, *CYP2B6*, *CYP2D6*, und *HLA* – zeichnet sich die DRAGEN-Software durch eine breite genomische Coverage aus. Darüber hinaus dehnt DRAGEN Graph mit paralleler Rechenarchitektur die Illumina-Reads effektiv auf Regionen mit geringer Komplexität aus, wodurch Bereiche des Genoms verwertbar gemacht werden, die aufgrund von Repeat-Sequenzen schwer zu untersuchen sind. Dies verbessert die Coverage potenziell medizinisch relevanter Gene und ermöglicht das Calling von Einzelnukleotid-, Kopienzahl- und Strukturvarianten in schwer zu mappenden Regionen.

Effiziente Analyse

Die DRAGEN Bio-IT Plattform wurde mit dem Ziel entwickelt, Laboren die optimale Nutzung von NGS-Datensätzen zu ermöglichen. Die Plattform bietet die hierfür notwendige schnelle Datenanalyse sowie die erforderlichen Dateioptionen. Die DRAGEN-Software zeichnet sich dank Hardwarebeschleunigung und FPGA-Architektur (Field-Programmable Gate Array, feldprogrammierbarer Gate-Array) durch kurze Durchlaufzeiten aus. Dank der effizienten DRAGEN-Analysealgorithmen wurden zwei Weltrekorde bei der Analyse genomischer Daten aufgestellt.^{5, 6} In praktischen Anwendungen kann die lokale DRAGEN Bio-IT Plattform NGS-Daten im Umfang eines Genoms bei 34-facher Coverage in ca. 30 Minuten analysieren. Ein herkömmliches CPU-basiertes System benötigt dagegen über 15 Stunden.⁷

Zur Bewältigung der Speicherplatzanforderungen in Zusammenhang mit großen NGS-Datendateien ermöglicht die DRAGEN Original Read Archive (ORA)-Technologie die verlustfreie 5-fache Komprimierung von FASTQ-Dateien. Dank der beeindruckend schnellen verlustfreien Komprimierung von DRAGEN ORA bleiben alle Informationen in FASTQ-Dateien erhalten. Für die Komprimierung von FASTQ-Dateien mit 50 bis 70 GB werden ca. 8 Minuten benötigt. Die vielseitigen Pipelines von DRAGEN können auch Eingabedateien verarbeiten und Ausgabedateien in verschiedenen Stadien der Pipelines erstellen ([Abbildung 2](#)).

FPGA und Hardwarebeschleunigung

Die hochgradig konfigurierbare FPGA-Programmierung ermöglicht hocheffiziente, hardwarebeschleunigte Implementierungen genomischer Analysealgorithmen, z. B. Konvertierung von Base-Call-Dateien (BCL), Mapping, Alignment, Sortierung, Markierung von Duplikaten und Calling von Haplotyp-Varianten. Dank der Flexibilität von FPGAs kann Illumina eine umfangreiche Suite von DRAGEN-Anwendungspipelines entwickeln, die fortlaufend aktualisiert und ergänzt wird. Damit sind optimale Genauigkeit, Vollständigkeit und Effizienz gewährleistet.

Anwendungsspezifische Referenzen

Mit dem DRAGEN Reference Builder können Human-, Nichthuman- oder Nichtstandardreferenzen (auch als Hash-Tabellen bezeichnet) generiert werden. Die erstellten Referenzen können als Eingabe für alle DRAGEN-Anwendungen verwendet werden, die Kundenreferenzdateien unterstützen. Die DRAGEN Reference Builder-Anwendung auf BaseSpace™ Sequence Hub erfordert eine FASTA-Datei. Die meisten DRAGEN-Pipelines verfügen über eine integrierte Unterstützung für hg19, hg38 (mit und ohne HLA*), GRCh37 und hs37d5. Mit dem DRAGEN Graph Toolkit können die Benutzer die Möglichkeiten der Graphreferenzen auch auf verschiedenste menschliche Graphreferenzen ausweiten.

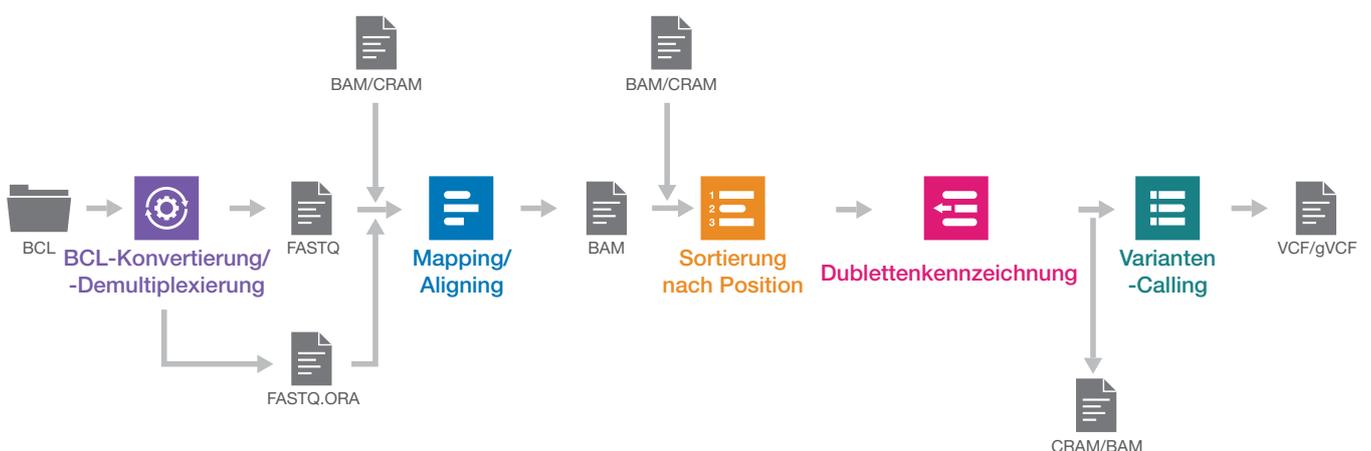


Abbildung 2: Flexibilität der DRAGEN-Pipelines: Alle DRAGEN-Pipelines enthalten eine Reihe spezifischer Schritte, die eine genaue und effiziente Analyse ermöglichen. Die DRAGEN-Pipeline ist flexibel und ermöglicht so die Verarbeitung unterschiedlicher Eingabedateien sowie die Generierung zahlreicher Ausgabebetypen. Damit haben Anwender die Möglichkeit, den Prozess und das Format der Ausgabedateien individuell anzupassen.

Tabelle 1: Die DRAGEN Bio-IT Plattform eignet sich für eine breite Palette von Sekundäranalyseanwendungen.

Anwendung	DRAGEN auf einem lokalen Server	NextSeq 1000, NextSeq 2000 System	BaseSpace Sequence Hub	Illumina Connected Analytics		Multi-Cloud mit eigener Lizenz
				Vorkonfiguriert	Selbst erstellt	
BCL-Konvertierung	✓	✓	✓		✓	Nicht empfohlen auf FPGA
DRAGEN ORA-Komprimierung	✓	✓			✓	✓
DRAGEN FASTQ + MultiQC	✓	✓	✓	✓	✓	✓
Genom	Keimbahn + somatisch	Nur Keimbahn	Keimbahn + somatisch	Keimbahn + somatisch	Keimbahn + somatisch	Keimbahn + somatisch
Anreicherung (einschließlich Exom)	Keimbahn + somatisch	Keimbahn + somatisch	Keimbahn + somatisch	Keimbahn + somatisch	Keimbahn + somatisch	Keimbahn + somatisch
DNA-Amplikon	✓	✓	✓	✓	✓	✓
RNA	✓	✓	✓	✓	✓	✓
Einzelzell-RNA	✓	✓	✓	✓	✓	✓
Differenzialexpression		✓	✓			
NanoString GeoMx NGS		✓	✓			
RNA-Amplikon	✓		✓	In Kürze verfügbar	✓	✓
Methylierung	✓		✓	✓	✓	✓
Metagenomik			✓			
Nachweis von RNA-Erregern			✓			
COVID-Lineage	✓		✓	In Kürze verfügbar		
TruSight Oncology 500 ctDNA	✓			✓		
ScATAC-Seq	✓		✓	✓	✓	✓
Imputation	✓		✓	✓	✓	✓
PGx Star Allele Caller	✓		✓	✓	✓	✓

Skalierbarkeit

Die DRAGEN Bio-IT Plattform ermöglicht den bedarfsabhängigen Ausbau des Einsatzes bei niedrigen Kosten und schneller Verarbeitung. Die DRAGEN-Software ermöglicht die Erweiterung der Forschungskapazitäten auf mehreren Ebenen:

- Schritt halten mit dem NovaSeq™ 6000 System:** Ein einzelner DRAGEN-Server kann einen mit einer S4-Fließzelle auf dem NovaSeq 6000 System durchgeführten Lauf in weniger als zwei Stunden demultiplexieren.
- Burst-Kapazität:** In Phasen mit erhöhtem Arbeitsaufkommen mit vielen Proben können Labore dank der Optionen zum parallelen Zugriff auf die DRAGEN-Plattform zusätzliche Kapazitäten nutzen (Abbildung 3).
- Erweiterung des Einsatzes:** Eine einzelne DRAGEN-Plattform kann für alle DRAGEN-Pipelines und alle unterstützten Probenarten verwendet werden. Die Genauigkeit, der Umfang und die Effizienz der DRAGEN-Tools ermöglichen Anwendern den Ausbau des Einsatzes ohne Abstriche bei der Verarbeitungsdauer und der Qualität der Ergebnisse.
- Von Exomen bis zu Genomen:** Ein Wechsel von der Exomsequenzierung (WES, Whole-Exome Sequencing) zur Genomsequenzierung (WGS, Whole-Genome Sequencing) bedeutet, dass wesentlich mehr Daten generiert werden. Die DRAGEN-Software ermöglicht Kunden bei der Analyse den einfachen Umstieg von Exomen auf Genome ohne teure Anschaffung neuer Hardware oder cloudbasierter Lösungen.

- Besonders große Datenbestände:** Die DRAGEN-Plattform bietet einen vereinfachten Workflow für umfangreiche Kohortenanalysen. Der Workflow beinhaltet mehrere Pipelines, die gemeinsam eingesetzt zur hochgenauen Bestimmung kleiner und großer Varianten in einer Kohortenstichprobe dienen. Die DRAGEN-Software ermöglicht die Aggregation und Genotypisierung von Tausenden bis hin zu Millionen gVCF-Dateien (genomic Variant Call Format) und aggregiert neue Batches, ohne vorhandene Batches erneut zu verarbeiten. Die DRAGEN Joint Genotyping Pipeline dient zum Varianten-Calling in mehreren Genomen und eignet sich dank schneller Analyse und höchster Genauigkeit auch für große Kohorten.⁸ Der Einsatz der DRAGEN-Plattform zur Auswertung der Daten des [1000 Genomes Project](#) ermöglichte beispielsweise in großem Maßstab das genaue Varianten-Calling bei unterschiedlichen Proben sowie die Bestimmung von Regionen, die uneinheitliche Coverage-Daten aufweisen oder von Annahmen abweichen.

Zugriff über unterschiedliche Plattformen

Die Suite mit DRAGEN-Pipelines kann als lokale, geräteintegrierte oder als Cloudlösung genutzt werden, sodass Labore die für sie geeignetste Lösung wählen können.

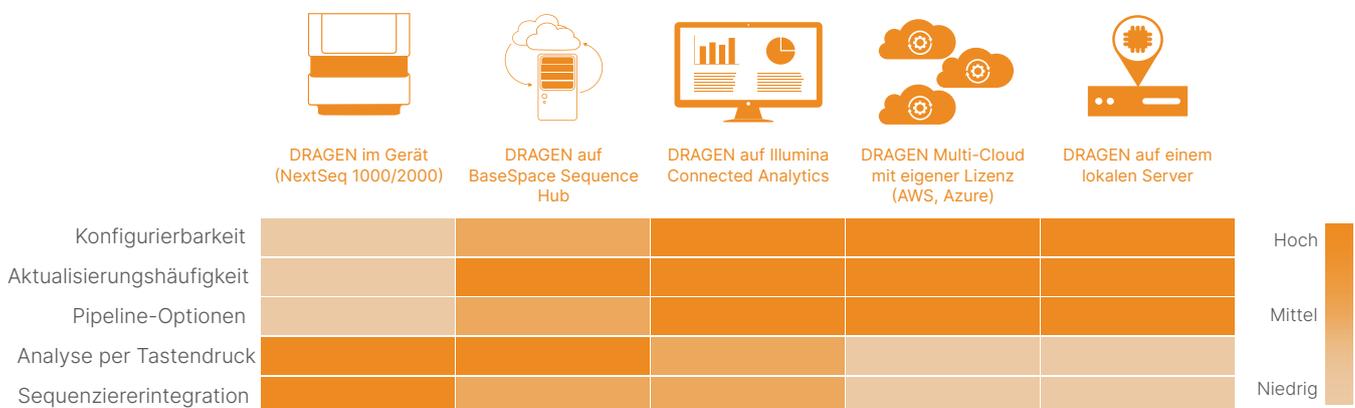


Abbildung 3: Optionen für den Zugriff auf DRAGEN-Pipelines mit Funktionen passend zu den Anforderungen an die NGS-Analyse aller Labore.

DRAGEN auf einem lokalen Server

DRAGEN auf einem lokalen Server nutzt eine lokale Speicherlösung zur Zusammenführung und Speicherung von NGS-Daten. Nachdem die Rohdaten der Sequenzierung vom Sequenzierungsgerät über ein lokales Netzwerk in den lokalen Speicher übertragen wurden, werden die Daten zur Durchführung des gewählten Workflows auf den DRAGEN-Server übertragen. Nach der Analyse schreibt die Software die generierten Ausgabedateien zurück auf den lokalen Speicherplatz. DRAGEN auf einem lokalen Server:

- Kann über unterschiedliche Befehlszeilenniveaus bedient werden
- Ersetzt bis zu 30 herkömmliche Recheninstanzen
- Verarbeitet NGS-Daten für ein vollständiges Humangenom bei 34-facher Coverage in ca. 30 min

Integrierte DRAGEN Bio-IT Plattform in den Systemen NextSeq™ 1000 und NextSeq 2000

Die Systeme NextSeq 1000 und NextSeq 2000 sind mit einer integrierten DRAGEN-Software für die schnelle und präzise Sekundäranalyse ausgestattet. Die integrierte DRAGEN-Software-Suite umfasst eine Reihe ausgewählter Pipelines für gängige NGS-Anwendungen (Tabelle 1). Über die benutzerfreundliche Oberfläche können sowohl Experten als auch andere Bediener benötigte Analysen schnell durchführen und Ergebnisse generieren. Die integrierte DRAGEN Bio-IT Plattform:

- Ermöglicht die Nutzung ausgewählter DRAGEN-Informatik-Pipelines
- Ermöglicht Anwendern, in nur zwei Stunden Ergebnisse zu generieren
- Nutzt intuitive Pipeline-Algorithmen und macht Anwender so unabhängiger von externen Informatikexperten

BaseSpace Sequence Hub

Die cloudbasierte DRAGEN-Suite auf BaseSpace Sequence Hub vereint präzise, effiziente Analysen mit einem sicheren Ökosystem und flexibler Skalierbarkeit. Die DRAGEN-Software auf BaseSpace Sequence Hub ermöglicht Laboren jeder Größe und Fachrichtung Sekundäranalysen per Tastendruck. BaseSpace Sequence Hub bildet eine direkte Erweiterung Ihrer Illumina-Geräte. Die Daten werden verschlüsselt vom Gerät in BaseSpace Sequence Hub übertragen und lassen sich so einfach verwalten und in ausgewählten Anwendungen analysieren. BaseSpace Sequence Hub auf Basis von Amazon Web Services (AWS):

- Bietet eine anwenderfreundliche Lösung für die DRAGEN-Analyse per Tastendruck
- Stellt eine intuitive grafische Benutzeroberfläche für die effiziente Bedienung durch Experten und andere Anwender bereit
- Bietet Zugang zu leistungsstarken Rechenressourcen ohne umfangreiche Kosten für zusätzliche Infrastruktur

Illumina Connected Analytics

Bei der DRAGEN Bio-IT Plattform auf Illumina Connected Analytics handelt es sich um eine umfassende Cloudplattform für Datenmanagement und -analyse, mit der Forscher große Mengen an Multiomikdaten in einer sicheren, skalierbaren und flexiblen Umgebung verwalten, analysieren und auswerten können. Illumina Connected Analytics:

- Bietet Zugang zur kompletten DRAGEN Bio-IT-Suite mit vorkonfigurierten Pipelines oder zu einzelnen Tools zur Verwendung anwendungsspezifischer Pipelines
- Unterstützt hochgradig automatisierte Workflows und anwendungsspezifische Lösungen für optimierte Studien mit hohem Durchsatz
- Stellt eine hochsichere Umgebung mit garantierter Datenresidenz, Single-Sign-On-Zugang, Audit-Protokollen und Zugriffskontrolle bereit, die die Einhaltung des Health Insurance Portability and Accountability Act (HIPAA) sowie der Datenschutz-Grundverordnung der Europäischen Union (DSGVO) ermöglicht

DRAGEN Multi-Cloud mit eigener Lizenz (BYOL, Bring Your Own License)

Labore, die bereits Clouddienste nutzen, können die DRAGEN Bio-IT Plattform als virtuelles Image erwerben und so DRAGEN-Analysen auf AWS oder Microsoft Azure ausführen. DRAGEN Multi-Cloud mit eigener Lizenz:

- Verwendet dieselbe Software und dieselben Befehlszeilenooptionen wie DRAGEN auf einem lokalen Server (ausgenommen TruSight™ Oncology 500-Pipelines)
- Eignet sich für fortgeschrittene Anwender, die DRAGEN-Pipelines in vorhandene Cloudumgebungen und Workflows integrieren möchten

Zusammenfassung

Bei der DRAGEN Bio-IT Plattform handelt es sich um eine leistungsstarke Suite mit Software-Tools, die eine genaue, umfassende und effiziente Sekundäranalyse von NGS-Daten ermöglichen. Die DRAGEN-Plattform bietet unterschiedliche Optionen, mit denen Labore die Lösung wählen können, die am besten zu Art und Umfang ihrer Projekte passt. Die NGS-Technologie entwickelt sich stetig weiter. Fortlaufende Updates der DRAGEN Bio-IT Plattform gewährleisten, dass Anwender stets von aktuellen Pipelines mit optimaler Performance profitieren. Zusätzlich werden kontinuierlich neue Pipelines hinzugefügt, sobald Anwendungen verfügbar sind.

Weitere Informationen

ILLUMINA DRAGEN Bio-IT Plattform: [illumina.com/dragen](https://www.illumina.com/dragen)

Supportseite zur DRAGEN-Plattform: support.illumina.com/sequencing/sequencing_software/dragen-bio-it-platform

Kontakt: [illumina.com/company/contact-us.html#/united-states/customer-care](https://www.illumina.com/company/contact-us.html#/united-states/customer-care)

Quellen

1. Food and Drug Administration. Truth Challenge V2: Calling Variants from Short and Long Reads in Difficult-to-Map Regions. precision.fda.gov/challenges/10. Aufgerufen am 14. März 2022.
2. Illumina. DRAGEN Sets New Standard for Data Accuracy in PrecisionFDA Benchmark Data. Optimizing Variant Calling Performance with Illumina Machine Learning and DRAGEN Graph. [illumina.com/science/genomics-research/articles/dragen-shines-again-precisionfda-truth-challenge-v2.html](https://www.illumina.com/science/genomics-research/articles/dragen-shines-again-precisionfda-truth-challenge-v2.html). Aufgerufen am 14. März 2022.
3. Illumina. DRAGEN Wins at PrecisionFDA Truth Challenge V2 Showcase Accuracy Gains from Alt-aware Mapping and Graph Reference Genomes. [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). Aufgerufen am 14. März 2022.
4. Interne Archivdaten. 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). Aufgerufen am 14. März 2022.
6. San Diego Union Tribune. Rady Children's Institute sets Guinness world record. <https://www.sandiegouniontribune.com/95899028-132.html>. Veröffentlicht am 12. Februar 2018. Aufgerufen am 14. März 2022.
7. Miller NA, Farrow EG, Gibson M, et al. A 26-hour system of highly sensitive whole genome sequencing for emergency management of genetic diseases. *Genome Med.* 2015;7:100. doi: 10.1186/s13073-015-0221-8.
8. Illumina. Accurate and Efficient Calling of Small and Large Variants from PopGen Datasets Using the DRAGEN Bio-IT Platform. www.illumina.com/science/genomics-research/articles/popgen-variant-calling-with-dragen.html. Aufgerufen am 14. März 2022.

illumina®

+1.800.809.4566 (USA, gebührenfrei) | +1.858.202.4566 (Tel. außerhalb der USA)
techsupport@illumina.com | www.illumina.com

© 2022 Illumina, Inc. Alle Rechte vorbehalten. Alle Marken sind Eigentum von Illumina, Inc. bzw. der jeweiligen Inhaber. Spezifische Informationen zu Marken finden Sie unter www.illumina.com/company/legal.html.
M-GL-00680 DEU v4.0