

Sekundäranalyse mit DRAGEN[™]

Präzises, umfassendes
und effizientes Varianten-
Calling für Next-Generation-
Sequencing-Daten



Einleitung

Die Erschließung der im Genom enthaltenen Informationen durch Sequenzierung der nächsten Generation (NGS, Next-Generation Sequencing) ist entscheidend für Fortschritte in der biomedizinischen Forschung und der Präzisionsmedizin. 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 Sekundäranalyse mit Illumina DRAGEN (Dynamic Read Analysis for GENomics) dient zur Bewältigung wichtiger Herausforderungen bei der Analyse von NGS-Daten in einer breiten Palette von Anwendungen, darunter Genom-, Exom-, Transkriptom-, Methylom-Studien u. v. m. Bei der DRAGEN-Software-Suite für die Sekundäranalyse handelt es sich um eine Reihe zusammengehöriger Anwendungen, die NGS-Daten verarbeiten und die Voraussetzungen für den Erkenntnisgewinn durch Tertiäranalysen schaffen. Die verfügbaren Tools bilden eine hochpräzise, umfassende und effiziente Lösung, mit der Labore aller Größen und Fachrichtungen Genomdaten besser auswerten können.

Präzise Ergebnisse

Die Sekundäranalyse mit DRAGEN liefert herausragend präzise Ergebnisse. 2020 gewann die Sekundäranalyse mit DRAGEN v3.7 bei der PrecisionFDA Truth Challenge V2 (PrecisionFDA V2) mit den präzisesten Daten bei allen Benchmark-Regionen und schwer zu mappende Regionen mit Illumina-Sequenzierungsdaten.^{1,2} Nachfolgende Versionen setzen weiterhin neue Maßstäbe bei der Genauigkeit und zeichnen sich durch Fortschritte in Bereichen wie den Technologien maschinelles Lernen (ML) und DRAGEN Multigenome (graph) aus. Die neueste Version der Sekundäranalyse mit DRAGEN (v4.3) zeichnet sich durch bislang unerreichte Genauigkeit beim Calling kleiner Varianten aus. Der F1-Score liegt bei 99,89 % (ein kombiniertes Maß für Präzision und Recall) in allen Benchmark-Regionen (Abbildung 1). Möglich macht dies die DRAGEN Multigenome (graph)-Referenz der dritten Generation, die auf 128 Proben mit 256 Haplotypen aus intern zusammengestellten genübergreifenden Referenzdaten basiert und damit eine größere genetische Vielfalt erfasst. Mit dem neuen integrierten Mosaik-Caller lassen sich Mosaikvarianten mit Allelfrequenzen von nur 3 % bestimmen.

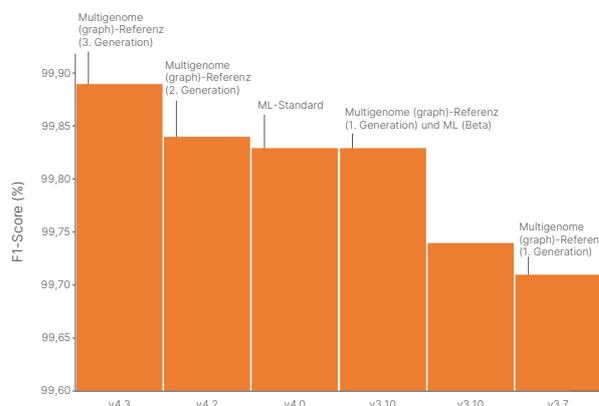


Abbildung 1: Genauigkeit der Sekundäranalyse mit DRAGEN: Die Y-Achse, F1-Score (%), gibt die berechneten richtig positiven und richtig negativen Ergebnisse als Anteil an den Gesamtergebnissen an.^{3, 4}

Umfassende Analyse

Mit umfassender Coverage des Genoms und einer breiten Palette unterstützter Anwendungen erfüllt die Sekundäranalyse mit DRAGEN die vielfältigen Anforderungen von Laboren, die NGS-Analysen durchführen. DRAGEN-Pipelines eignen sich für verschiedene Versuchsarten, einschließlich Genomsequenzierung (WGS, Whole-Genome Sequencing), Exomsequenzierung, Anreicherungspanels, Einzelzell-RNA-Seq, Einzelzell-ATAC-Seq, 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}

Zur Keimbahnanalyse umfasst die Sekundäranalyse mit DRAGEN eine Reihe von Varianten-Callern wie ExpansionHunter sowie gezielte Caller wie *SMN*, *GBA*, *CYP2B6*, *CYP2D6* und *HLA*. Mit DRAGEN v4.3 wird zudem MRJD eingeführt, ein neuer spezialisierter Caller, der die Coverage schwieriger Gene in segmentalen Duplikationsregionen wie *PMS2*, *SMN1*, *SMN2*, *STRC*, *NEB*, *TTN* und *IKBK* gewährleistet. Diese Tools ermöglichen die Analyse einer breiten Palette genetischer Variationen, einschließlich Einzelnukleotidvarianten, Insertionen und Deletionen (Indels), Repeat-Expansionen und struktureller Varianten in größeren genomischen Regionen. Darüber hinaus erhöht die DRAGEN Multigenome (graph)-Referenz die Qualität des Mappings, was die Genauigkeit des Varianten-Callings erhöht und die Untersuchung von Bereichen des Genoms ermöglicht, die aufgrund der komplexen Sequenz nur schwer zugänglich sind. Dies erhöht die Coverage potenziell medizinisch relevanter Gene und ermöglicht das Calling von Einzelnukleotidvarianten und kleinen Indels sowie von Kopienzahl- und Strukturvarianten in schwer zu mappenden Regionen.

Tabelle 1: Die Sekundäranalyse mit DRAGEN eignet sich für eine breite Palette von Sekundäranalyseanwendungen.^a

Anwendung	Auf einem lokalen Server	Auf Sequenziersystemen von Illumina		Illumina-Cloudplattformen	
	DRAGEN-Server	NovaSeq X Series	NextSeq 1000, NextSeq 2000 System	BaseSpace Sequence Hub	Illumina Connected Analytics
BCL-Konvertierung	✓	✓	✓	✓	Nur anwendungsspezifisch
DRAGEN ORA Compression	✓	✓	✓		Nur anwendungsspezifisch
DRAGEN FASTQ + MultiQC	✓	✓	✓	✓	✓
Genom	Keimbahn + somatisch	Keimbahn + somatisch	Keimbahn + somatisch	Keimbahn + somatisch	Keimbahn + somatisch
Anreicherung (einschließlich Exom)	Keimbahn + somatisch	Keimbahn + somatisch	Keimbahn + somatisch	Keimbahn + somatisch	Keimbahn + somatisch
DRAGEN Amplicon	✓		Nur DNA	✓	✓
RNA	✓	✓	✓	✓	✓
Einzelzell-RNA	✓		✓	✓	✓
NanoString GeoMx NGS			✓	✓	
Methylierung	✓	✓		✓	✓
Metagenomik	✓ ^b			✓	
Nachweis von RNA-Erregern				✓	
COVID	COVIDSeq. COVID-Lineage		COVIDSeq (nur Cloud)	COVIDSeq. COVID-Lineage	
TruSight Oncology 500-Portfolio	✓			✓ ^c	✓
scATAC-seq	✓			✓	✓
Imputation	✓			✓	✓
PGx Star Allele Caller	✓	✓	✓	✓	✓
Illumina Complete Long Reads				✓	✓
DRAGEN-Sekundäranalyse für RPIP und UPIP	✓			✓	Beta

a. Grundlegende DRAGEN-Softwareversion abhängig von der jeweiligen Plattform, weitere Informationen erhalten Sie vom zuständigen Vertriebsmitarbeiter.

b. Metagenomikanwendungen ermöglicht durch Kmer-Klassifizierer, weitere Tools demnächst verfügbar.

c. Illumina Connected Analytics-Abonnement erforderlich.

Effiziente Analyse

Die DRAGEN-Software ermöglicht Laboren die Datenanalysegeschwindigkeit, die zur Optimierung der Verarbeitung von NGS-Datensätzen erforderlich ist. Die Sekundäranalyse mit DRAGEN 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 Sekundäranalyse mit DRAGEN NGS-Daten im Umfang eines Genoms bei 40-facher Coverage in ca. 35 Minuten mit allen Callern verarbeiten.⁷ Bei einem herkömmlichen Open-Source-Verfahren benötigt dagegen das Calling einer begrenzter Anzahl von Variantentypen mehr als 8 Stunden.⁷

Die DRAGEN Original Read Archive (ORA)-Technologie ermöglicht eine bis zu 5-fache verlustfreie Komprimierung von FASTQ-Dateien im herkömmlichen „Fastq.gz“-Format.

* Auf Basis interner Illumina-Daten auf Grundlage der Standards HG001-HG007 auf einem DRAGEN-Server der Version 4, ohne die neuen spezialisierten Caller wie MRJD und VNTR, die in DRAGEN v4.3 zur Verfügung stehen.

Dank der beeindruckend schnellen verlustfreien Komprimierung von DRAGEN ORA bleiben alle Informationen in FASTQ-Dateien erhalten. Für die Komprimierung von FASTQ-Dateien mit 50–70 GB für eine Vielzahl häufig untersuchter Spezies werden ca. 8 Minuten benötigt. Bei der Sekundäranalyse mit DRAGEN können die vielseitigen Pipelines 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 regelmäßig aktualisiert und ergänzt wird. Damit sind optimale Genauigkeit, Vollständigkeit und Effizienz gewährleistet.

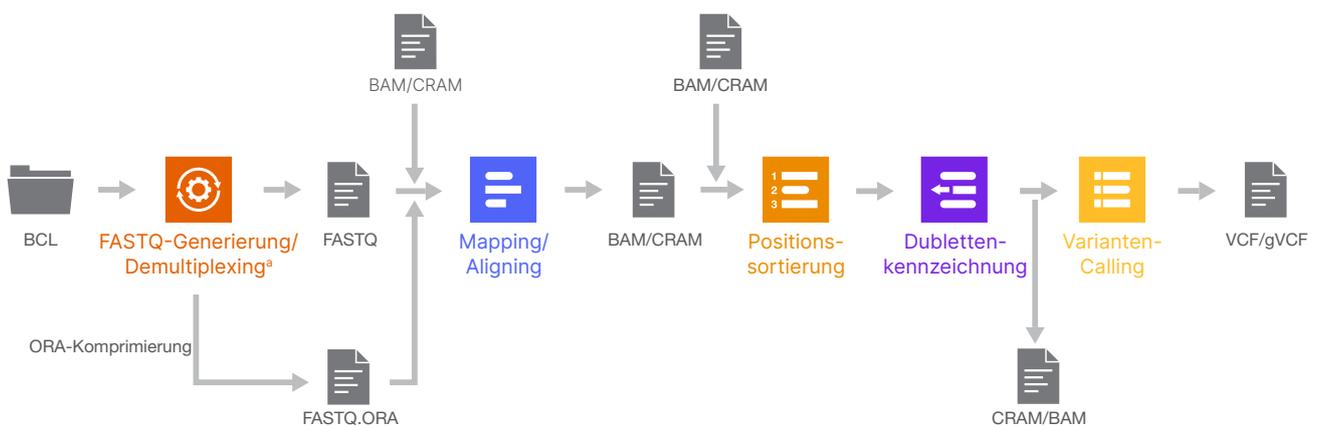


Abbildung 2: Flexibilität der DRAGEN-Pipelines für die Sekundäranalyse: Alle DRAGEN-Pipelines enthalten eine Reihe spezifischer Schritte, die eine genaue und effiziente Analyse ermöglichen. Die DRAGEN-Beispielpipeline für die Genom-Keimbahnanalyse ist flexibel und ermöglicht so die Verarbeitung unterschiedlicher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

a. Die BCL-Konvertierung ist auch als eigenständiges Tool verfügbar.

Anwendungsspezifische Referenzen

Die Sekundäranalyse mit DRAGEN ermöglicht Anwendern die Erstellung einer anwendungsspezifischen Referenz (human, nicht human oder nicht standardmäßig). Die erstellten Referenzen können als Eingabe für alle DRAGEN-Anwendungen verwendet werden, die anwendungsspezifische Referenzdateien unterstützen. Die meisten DRAGEN-Pipelines verfügen über eine integrierte Unterstützung für hg19, hg38 (mit und ohne HLA), GRCh37, CHM13v2 und hs37d5. Mithilfe der DRAGEN-Software profitieren Anwender von erweiterten Einsatzmöglichkeiten der Multigenome (graph)-Standardreferenz sowohl für vielfältige als auch für spezifische Populationen.

Skalierbarkeit

Die Sekundäranalyse mit DRAGEN ermöglicht die bedarfsgerechte Skalierung der Nutzung bei niedrigen Kosten und schneller Verarbeitung. Die DRAGEN-Software ermöglicht die Erweiterung der Forschungskapazitäten auf mehreren Ebenen:

- Schritthalten mit der NovaSeq™ X Series sowie dem NextSeq™ 1000 System und dem NextSeq 2000 System:** DRAGEN auf dem Gerät kann in einem Lauf mehrere Anwendungen gleichzeitig (vier mit maximal einer BCL-Konvertierung und drei weiteren Pipelines Ihrer Wahl) pro Fließzelle ausführen.
- Burst-Kapazität:** Bei erhöhter Auslastung mit hohen Probenvolumina können Labore zusätzliche Cloud-Kapazitäten für die Sekundäranalyse mit DRAGEN auf
- Skalierung der Nutzung:** Eine einzelne DRAGEN-Instanz kann für zahlreiche unterschiedliche DRAGEN-Pipelines und unterstützte Probenarten verwendet werden. Der Umfang und die Effizienz der DRAGEN-Software ermöglichen Anwendern die Skalierung der Nutzung ohne Abstriche bei der Verarbeitungsdauer oder der Qualität der Ergebnisse.
- Übergang zu Genomen:** Vorkonfigurierte DRAGEN-Pipelines ermöglichen den einfachen Übergang von gezielten Panels zur Untersuchung von Exomen und Genomen.
- Genomische Untersuchungen großer Populationen:** Die Sekundäranalyse mit DRAGEN bietet einen vereinfachten Workflow für umfangreiche Kohortenanalysen. Der Workflow beinhaltet mehrere Pipelines, die gemeinsam eingesetzt zur hochgenauen Bestimmung genetischer Varianten dienen. DRAGEN gVCF Genotyper ermöglicht die Aggregation von Tausenden bis hin zu Millionen gVCF-Dateien (genomic Variant Call Format) und bezieht neue Batches ein, ohne vorhandene Batches erneut zu verarbeiten. ORA-Komprimierung spart Speicherkosten.
- Deep-Sequencing-Anwendungen:** Die Sekundäranalyse mit DRAGEN ermöglicht die Analyse im Rahmen der High-Depth-Sequenzierung mit hoher Effizienz. Hierbei wird eine durchschnittliche Coverage von über 300-fach für Genome und 1.000-fach für Exome erreicht. Die Deep-Sequencing-Funktionen sind besonders wertvoll für Anwendungen im Bereich der Onkologieforschung sowie bei Studien zu seltenen genetischen Erkrankungen.

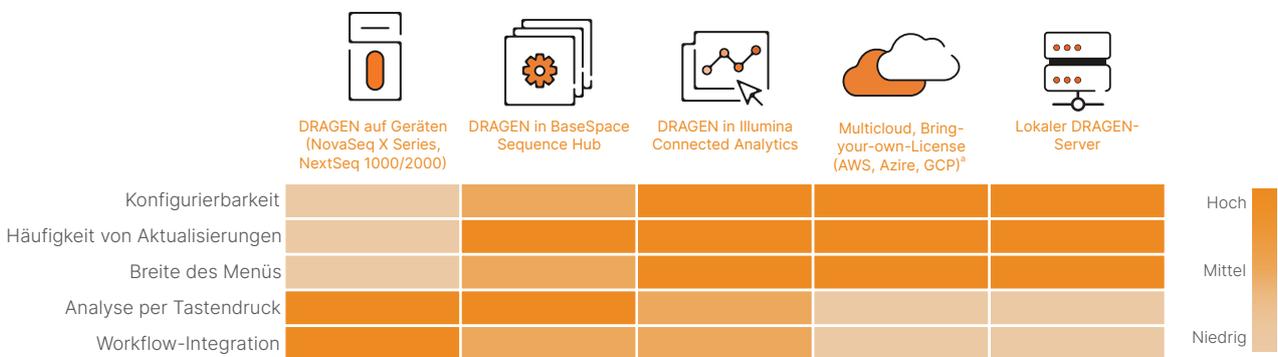


Abbildung 3: Optionen für den Zugriff auf DRAGEN-Pipelines mit Funktionen passend zu den Anforderungen an die NGS-Analyse aller Labore.

a. Wenden Sie sich an Ihren Ansprechpartner bei Illumina, wenn Sie mehr über den Zugriff auf Amazon Web Service (AWS), Azure oder Google Cloud Platform (GCP, Early Access) erfahren möchten.

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](#)).

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:

- Unterstützt die flexible Konfiguration von DRAGEN-Funktionen über eine Befehlszeilenschnittstelle
- Ersetzt bis zu 30 herkömmliche Recheninstanzen
- Verarbeitet NGS-Daten für ein vollständiges Humangenom bei 40-facher Coverage in ca. 35 min

DRAGEN auf der NovaSeq X Series

Bei der NovaSeq X Series ist die Sekundäranalyse mit DRAGEN im Gerät integriert. Diese bietet eine genaue, automatisierte und optimierte Analyse, die speziell zur Bewältigung der großen mit der NovaSeq X Series generierten Datenmengen entwickelt wurde. Die integrierte DRAGEN-Software-Suite dient bei NGS-Anwendungen zur Sekundäranalyse und ORA-Komprimierung ([Tabelle 1](#)) und deckt dabei BCL-Konvertierung, Keimbahn- und somatische Analysen, RNA und Methylierung ab. DRAGEN auf Geräten:

- Führt mehrere Sekundäranalyse-Pipelines parallel durch
- Führt in einem Einzellauf bis zu vier Anwendungen pro Fließzelle gleichzeitig aus
- Erzielt bis zu 5-fache verlustfreie Datenkomprimierung sowie Speicherkosteneinsparungen
- Sorgt für Einsparungen bei der Analyse, deren Höhe über fünf Jahre die Anschaffungskosten des NovaSeq X System übersteigen kann

DRAGEN auf dem NextSeq 1000 System und NextSeq 2000 System

Das NextSeq 1000 System und das NextSeq 2000 System sind mit einer integrierten DRAGEN-Software für die schnelle und präzise Sekundäranalyse ausgestattet. Der Zugriff auf die Software erfolgt über eine anwenderfreundliche grafische Benutzeroberfläche, über die Experten und Laien die gewünschten Analysen durchführen und schnell Ergebnisse erzielen können. Die integrierte DRAGEN-Software bietet eine Reihe ausgewählter Pipelines für zahlreiche gängige NGS-Anwendungen ([Tabelle 1](#)) und umfasst preisgekrönte ML- und Multigenome (graph)-Referenzanalysen für das Varianten-Calling mit hoher Qualität. DRAGEN auf Geräten:

- Bietet die höchste Genauigkeit aller Tischsequenziersysteme mit integrierter DRAGEN-Sekundäranalyse
- 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 Illumina Connected Analytics handelt es sich um eine umfassende Bioinformatik-Cloudplattform, mit der Forscher große Mengen an Multiomikdaten in einer sicheren, skalierbaren und flexiblen Umgebung verwalten, analysieren und auswerten können. Sie erhalten Zugriff auf die DRAGEN-Sekundäranalyseanwendungen auf Illumina Connected Analytics, die als vordefinierte Pipelines oder einzelne Tools zur Integration in anwendungsspezifische Pipelines verfügbar sind.

Zusammenfassung

Bei der Sekundäranalyse mit DRAGEN handelt es sich um eine leistungsstarke Suite mit Software-Tools, die eine genaue, umfassende und effiziente Analyse von NGS-Daten ermöglichen. Dank der unterschiedlichen Bereitstellungsoptionen der DRAGEN-Software können Labore die Lösung wählen, die am besten zu Art und Umfang ihrer Projekte passt. Darüber hinaus können Anwender zur Erfüllung ihrer Anforderungen an Performance und Workflow unterschiedliche Bereitstellungsoptionen kombinieren. Die NGS-Technologie entwickelt sich stetig weiter. Fortlaufende Updates der Sekundäranalyse mit DRAGEN gewährleisten, dass Anwender stets von aktuellen Pipelines mit optimaler Performance profitieren. Zusätzlich werden kontinuierlich neue Pipelines hinzugefügt, sobald Anwendungen verfügbar sind.

Weitere Informationen

[Sekundäranalyse mit DRAGEN](#)

[Supportseite zur Sekundäranalyse mit DRAGEN](#)

[Kontakt](#)



1 800 8094566 (USA, gebührenfrei) |
 +1 858 2024566 (Tel. außerhalb der USA)
 techsupport@illumina.com | www.illumina.com

© 2024 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 v10.0

Quellen

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>. Aufgerufen am 14. März 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. Illumina-Website. <https://www.illumina.com/science/genomics-research/articles/dragen-shines-again-precisionfda-truth-challenge-v2.html>. Veröffentlicht am 12. Januar 2022. Aufgerufen am 11. Juli 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. Illumina-Website. [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). Veröffentlicht am 9. November 2020. Aufgerufen am 11. Juli 2024.
4. Interne archivierte Daten. 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). Veröffentlicht am 23. Oktober 2017. Aufgerufen am 11. Juli 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/>. Veröffentlicht am 12. Februar 2018. Aufgerufen am 11. Juli 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. Veröffentlicht am 13. Dezember 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. Illumina-Website. www.illumina.com/science/genomics-research/articles/popgen-variant-calling-with-dragen.html. Veröffentlicht am 24. Mai 2021. Aufgerufen am 14. März 2022.