

DRAGEN™

二级分析

针对新一代测序数据实现准确、
全面、高效的变异检出



通过一套解决方案分析全基因组、
外显子组、甲基化组、转录组和
蛋白组



通过高效的数据分析获得非常准确的
结果，缩短时间并降低数据存储费用



轻松整合因美纳测序仪，实现从
测序到结果的一体化工作流程

简介

利用新一代测序（NGS）技术解码基因组对于推动生物医学研究和精准医疗领域的发展至关重要。为从NGS中获得更多的遗传见解，研究人员需要借助实用的数据分析工具准确且有效地将原始测序数据转化为有意义的结果。此外，为了充分利用NGS的优势，机构需要简便易用的解决方案——不仅要满足各种用户要求，而且方案采用的经济和技术壁垒更低。

Illumina DRAGEN（Dynamic Read Analysis for GENomics，基因组学动态读取分析）二级分析旨在解决与NGS数据分析相关的重大挑战，适用于全基因组测序、外显子组测序、转录组分析和甲基化组研究等一系列广泛的应用。DRAGEN二级分析软件集成了一套应用程序，既可处理NGS数据，又支持三级分析，助力研究人员获取更多见解。高度准确、全面且高效的解决方案基于可用的工具，支持各种规模和学科的实验室利用其基因组数据进行更多研究。

准确的结果

DRAGEN二级分析可生成十分准确的结果。在2020年Precision FDA Truth Challenge V2（PrecisionFDA V2）中，DRAGEN二级分析v3.7脱颖而出，其利用因美纳测序数据分析所有基准区域和难比对区域，展现出高准确性^{1,2}。后续版本继续在准确性方面树立新的标准，并在机器学习（ML）和DRAGEN多基因组技术等领域取得进步*。最新版本的DRAGEN二级分析v4.4提供了更高的小变异检出准确性，在所有基准区域的F1评分（查准率和召回率的综合衡量标准）达到99.90%（图1）。这一准确度的实现得益于新一代DRAGEN多基因组定位技术，该技术基于内部构建的泛基因组参考集开发，整合了128个样本的256种单倍型，能够捕获更广泛的遗传多样性。集成式嵌合检出程序也有助于提高准确性，它可以检测等位基因频率低至3%的嵌合变异。DRAGEN二级分析v4.4采用包含泛基因组参考数据的多基因组定位进行结构变异（SV）检测，提供了更高的SV检出准确度。此外，DRAGEN二级分析v4.4引入了个性化的泛基因组参考，能够更好地代表特定群体并减少祖源偏差。

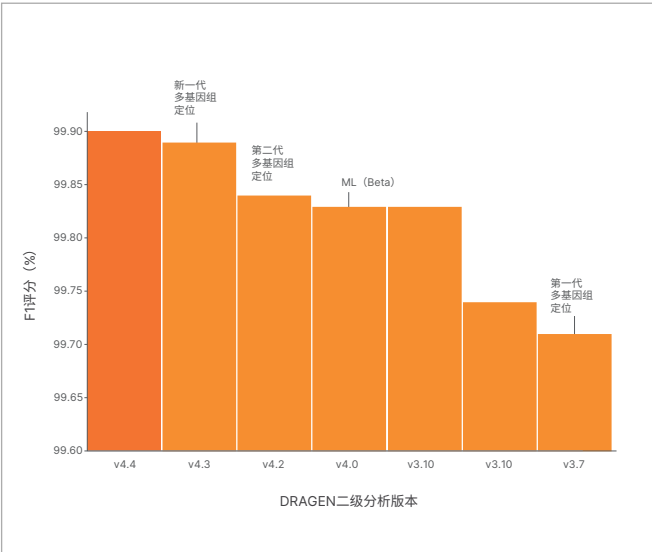


图1：DRAGEN二级分析的准确性
Y轴的F1评分（%）是根据真阳性和真阴性结果占总结果的比例来计算的^{3,4}。

全面的分析

DRAGEN二级分析全面覆盖基因组并支持多种应用程序，可满足开展NGS分析的实验室的多样化需求。DRAGEN流程支持多种实验类型，包括全基因组测序（WGS）、全外显子组测序、富集panel、单细胞RNA-Seq、单细胞ATAC-Seq、bulk RNA-Seq和甲基化分析（表1）。需要30多种开源工具才能部分复制DRAGEN软件的功能广度^{3,4}。此外，贝勒医学院的研究人员发表了一项研究，证明DRAGEN二次分析在速度和准确度方面优于目前最先进的所有变异类型的基因组分析和变异检测工具⁵。

*术语“多基因组（图形）参考”以前用于指定定位方法和参考样本采集。为了更好地描述DRAGEN二次分析版本的更新，现在使用单独的术语。“多基因组定位”指定定位方法，“泛基因组参考”指参考样本采集。

表1: DRAGEN二级分析支持广泛的二级分析应用^a

应用	本地服务器	因美纳测序仪机载			因美纳云端平台	
	DRAGEN服务器	NovaSeq™ X系列	NextSeq™ 1000/2000系列测序仪	MiSeq™ i100系列测序仪	BaseSpace Sequence Hub	Illumina Connected Analytics
BCL Convert	✓	✓	✓	✓	✓	✓
DRAGEN ORA压缩	✓	✓	✓	✓		即将推出
全基因组	生殖系+体细胞	生殖系+体细胞	生殖系+体细胞		生殖系+体细胞	生殖系+体细胞
富集（包括外显子组）	生殖系+体细胞	生殖系+体细胞	生殖系+体细胞		生殖系+体细胞	生殖系+体细胞
DRAGEN Amplicon	✓		仅DNA		✓	✓
RNA	✓	✓	✓		✓	✓
单细胞RNA	✓		✓		✓	✓
NanoString GeoMx NGS			✓		✓	
甲基化	✓	✓			✓	✓
蛋白质定量	✓				✓	✓
宏基因组学	✓ ^b				✓	
COVID、IMAP、IMAP-FLU					✓	
TruSight™ Oncology 500产品系列	✓				✓ ^c	✓
填补	✓				✓	✓
PGx星号等位基因检出程序	✓	✓	✓		✓	✓
Illumina Complete Long Reads					✓	✓
RPIP、RVEK、UPIP、VSP	✓					Beta
小panel全基因组				✓	✓	
血液肿瘤WGS ^d	✓					✓
实体瘤WGS肿瘤-正常对照 ^d	✓					✓

a. 核心DRAGEN软件版本因平台而异，请咨询当地代表获取更多信息。
b. 由Kmer分类方法支持的宏基因组学应用程序，更多工具即将推出。
c. 需要订阅Illumina Connected Analytics。
d. 可通过DRAGEN应用程序管理器获取。
IMAP, Illumina Microbial Amplicon Prep; RPIP, Respiratory Pathogen ID/AMR Panel; RVEK, Respiratory Virus Enrichment Kit; UPIP, Urinary Pathogen ID/AMR Panel; VSP, Urinary Pathogen ID/AMR Panel。

针对生殖系分析，DRAGEN二级分析提供了一套用于基因的变异检出程序（例如ExpansionHunter）和靶向检出程序（例如SMN、GBA、CYP2B6、CYP2D6和HLA）。DRAGEN二级分析v4.3还引入了新的专用检出程序——多区域联合检测（MRJD），它能够覆盖片段重复区域中的复杂基因，例如PMS2、SMN1、SMN2、STRC、NEB、TTN和IKBK。这些工具可以分析广泛的遗传变异，包括单核苷酸位点突变、插入和缺失（indel）、重复扩增和延伸基因组区域中的结构变异。此外，基于泛基因组参考的DRAGEN多基因组定位提高了定位质量，可实现更高的变异检出准确性，并解析由于序列复杂性而难以评估的基因组区域。这不仅提高了潜在医学相关基因的覆盖度，而且能够在难以比对的区域实现单核苷酸变异、小插入缺失、拷贝数变异和结构变异检出。

高效的分析

DRAGEN软件旨在为实验室提供所需的数据分析速度，以优化其NGS数据集的处理效率。DRAGEN二级分析是通过硬件加速的，并使用现场可编程门阵列（FPGA）架构来实现快速周转时间。

高效的DRAGEN分析算法创造了两项基因组数据分析世界纪录^{6,7}。在实际应用中，本地DRAGEN二级分析可以使用所有检出程序*在35分钟之内处理相当于40×覆盖度的全基因组NGS数据，而常用的开源方法检出的变异类型有限，且所需的时间超过8小时⁸。

为了更轻松地存储、管理和共享大型NGS数据文件，DRAGEN Original Read Archive（ORA）技术可将传统fastq.gz格式的FASTQ文件无损压缩为原来的1/5。DRAGEN ORA的无损压缩保留了FASTQ文件的详细内容，并且速度非常快，压缩50-70 GB的FASTQ[†]文件仅需8分钟左右，支持各种常见的研究物种。DRAGEN二级分析具有一组通用的流程，这些流程还可以在流程的不同阶段接收输入数据文件并创建输出文件（图2）。

* 根据DRAGEN Server v4上基于HG001-HG007标准的因美纳内部数据，不包括DRAGEN二级分析v4.3中提供的MRJD和可变数目串联重复序列等新的专用检出程序。
† 在DRAGEN Server v3上。

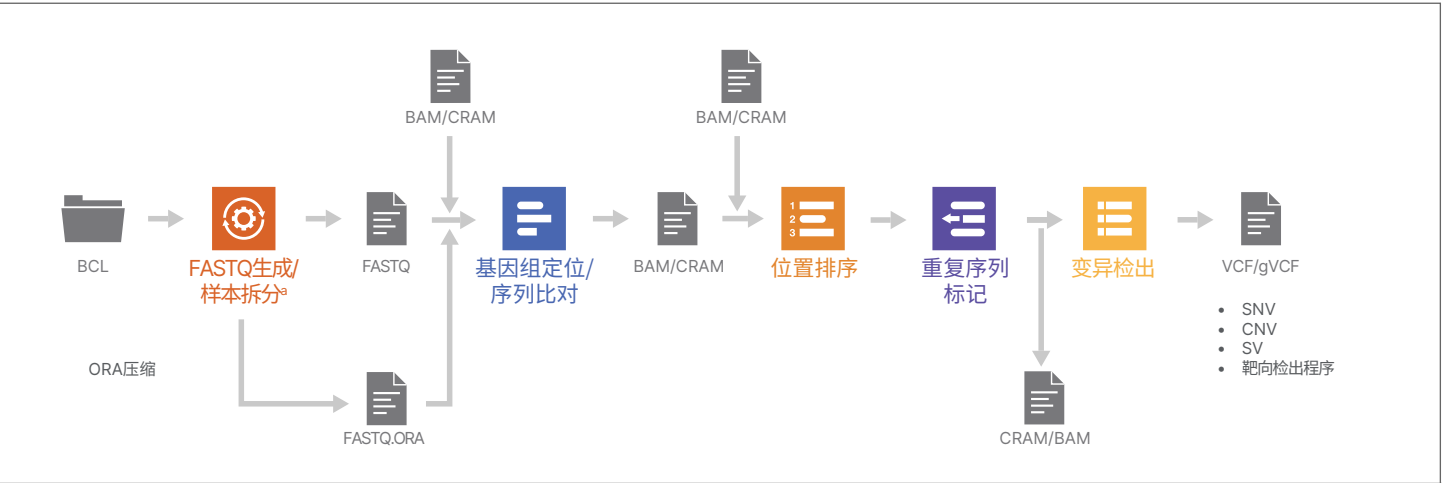


图2：DRAGEN二级分析流程的灵活性
每种DRAGEN流程都有一组独特的步骤，以支持准确高效的分析。例如，DRAGEN全基因组生殖系示例流程可以灵活地接受多种输入文件格式并生成一系列输出文件，为用户提供个性化的体验并生成他们所需的文件格式。
a. BCL Convert也可作为独立工具使用。

FPGA和硬件加速

高度可配置的FPGA允许通过超高效的加速硬件来实施基因组分析算法，例如碱基检出（BCL）文件转换、定位、比对、分选、重复标记和单倍型变异检出。FPGA的灵活特性使因美纳开发出了一组广泛的DRAGEN应用流程，通过频繁的更新和功能添加，提供出色的准确性、全面性和效率。

定制参考基因组

DRAGEN二级分析使用户能够生成人类、非人类或非标准定制参考基因组。创建的参考基因组索引可以用作所有DRAGEN应用程序的输入，这些应用程序都能支持定制参考文件。大多数DRAGEN流程都包含对基因组hg19、hg38（包含或不包含HLA）、GRCh37、CHM13v2和hs37d5的内置支持。DRAGEN软件使用户能够扩展图形标准多基因定位功能，以应用于各种不同的群体以及特定的群体。

可扩展性

DRAGEN二级分析能让实验室保持低成本和短周转时间，同时根据需要进行规模化运行。DRAGEN软件可以通过多种方式来扩展研究能力：

- 1. 使用NovaSeq™ X系列测序仪实现同步运行——机载DRAGEN可通过单次运行在每个流动槽中同时执行多个应用程序（可同时执行4个应用，最多1个BCL Convert和3个您选择的其他流程）。
- 2. 可扩展的处理能力——在工作量增加、样本量大的情况下，实验室可以通过Illumina Connected Analytics上的DRAGEN二级分析或BaseSpace™ Sequence Hub上的DRAGEN应用程序获得额外的云端处理能力（图3）。
- 3. 扩大运行规模——一个DRAGEN实例可以运行各种DRAGEN流程和支持的样本类型。DRAGEN软件的全面性和高效性使用户可以扩大运行规模，而不会影响周转时间或结果的质量。
- 4. 向基因组过渡——DRAGEN预设流程轻松实现从靶向panel到外显子组再到基因组的过渡。

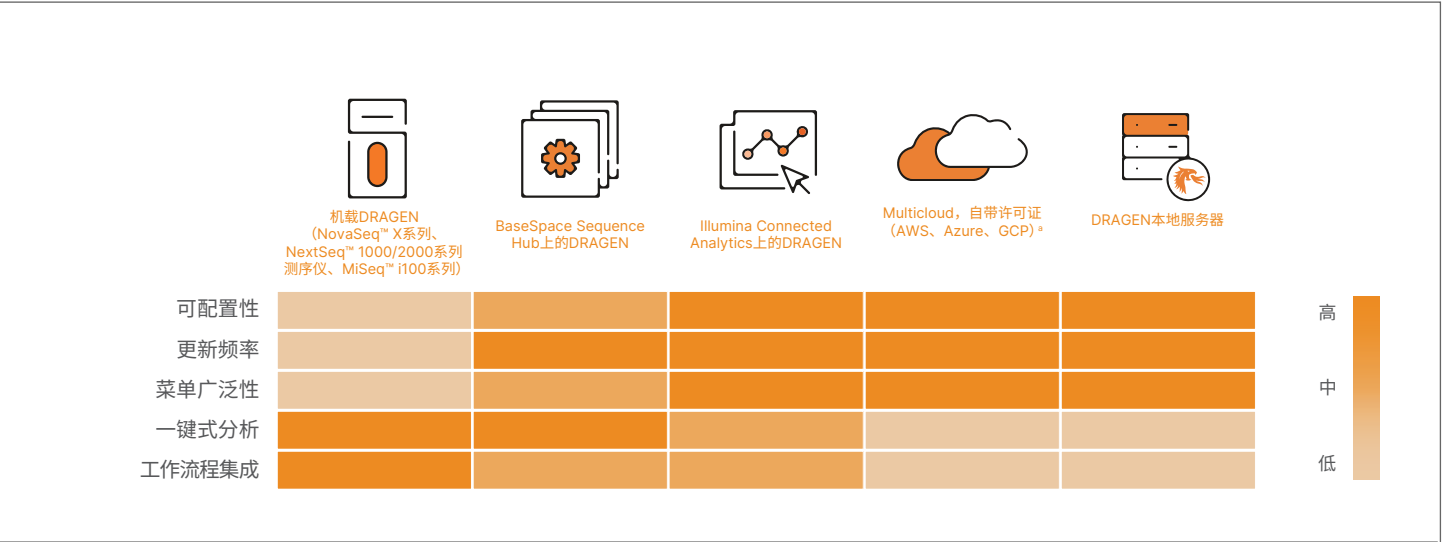


图3：DRAGEN流程访问选项，旨在满足每个实验室的NGS分析需求

a. 请联系您的因美纳代表，了解有关Amazon Web Service（AWS）、Azure或谷歌云平台（GCP，早期试用）访问因美纳产品的信息。

5. **大型群体基因组学计划**——DRAGEN二级分析为大规模队列分析提供了简化的工作流程，其中包含多个流程，通过结合使用这些流程，以高精度检出遗传变异。DRAGEN gVCF基因分型程序能够汇总数千到数百万个基因组变异检出格式（gVCF）文件，整合新批次无需重新处理已有批次。ORA压缩技术节省了存储成本。
6. **深度测序应用**——DRAGEN二级分析支持高效的深度测序数据分析，可以处理平均覆盖度超过300×的基因组数据和平均覆盖度超过1000×的外显子组数据。深度测序功能对于肿瘤学研究和罕见遗传病研究等应用非常有价值。

多平台可访问性

DRAGEN流程组合可通过可用的本地、仪器或云端解决方案获取，便于实验室选择最适合其需求的解决方案（图3）。

DRAGEN本地服务器

DRAGEN本地服务器需要使用本地储存解决方案来收集和储存NGS数据。DRAGEN Server v4的规格列于表2中。原始测序数据通过本地网络连接从测序仪器传输到本地存储，用户可通过DRAGEN服务器访问这些数据，以执行选定的工作流程。分析完成后，软件将生成的输出文件写入本地存储位置。DRAGEN本地服务器：

- 支持通过命令行界面灵活配置DRAGEN功能
- 可代替多达30个传统计算实例
- 在约35分钟内处理40×覆盖度下的整个人类基因组的NGS数据
- 支持Illumina Connected Insights-Local，为肿瘤学临床研究提供变异解读和报告

表2：DRAGEN Server v4规格

组件	规格/描述
CPU	双核Intel Xeon Gold 6226R 2.9 GHz，16C/32T
系统内存	512 GB DDR4
Scratch驱动器	2 × 7.68 TB NVMe
OS驱动器	2× 480 GB SSD（RAID 1）
硬件加速	1× FPGA卡
尺寸规格	2U
尺寸	高8.8 cm（3.5 in），宽43.8 cm（17.2 in），深76.4 cm（29.9 in）
电源	1968 W双路热插拔冗余电源

NovaSeq™ X系列的机载DRAGEN

NovaSeq™ X系列包括机载DRAGEN二级分析，提供准确、自动化、简化的分析，旨在支持NovaSeq™ X系列生成的海量数据。机载DRAGEN软件套装可提供常规NGS应用的二级分析和ORA压缩（表1）。机载DRAGEN：

- 并行运行多个二级分析流程
- 包括BCL Convert、生殖系、体细胞、富集、RNA和甲基化流程
- 实现高达5×的无损数据压缩，节省存储成本

NextSeq™ 1000/2000系列测序仪的机载DRAGEN

NextSeq™ 1000/2000系列测序仪包括机载DRAGEN软件，可实现准确的二级分析。该软件可通过用户友好的图形界面访问，支持专家和非专家用户执行所需的分析并快速生成结果。机载DRAGEN软件提供了一组精选的流程，涵盖一系列常见的NGS应用（表1），并包括屡获殊荣的ML和多基因组定位分析，可实现高质量变异检出。机载DRAGEN：

- 在所有桌面式测序仪中，通过机载DRAGEN二级分析实现更高的准确度
- 提供多种DRAGEN信息学流程可供选择
- 两小时内就能生成结果
- 使用直观的流程算法可减少对外部信息学专家的依赖

MiSeq™ i100系列上的机载DRAGEN

MiSeq™ i100系列可提供直观、非常快速的分析，包括DRAGEN BCL Convert、DRAGEN文库QC、DRAGEN小panel WGS和DRAGEN Microbial Enrichment Plus流程。机载DRAGEN：

- 可在两小时内快速生成全面的二级分析结果[‡]
- 包括高效的工作流程，通过单一用户接触点实现VCF输出和/或报告，并且无中间文件传输
- 为非专家用户提供用户友好、直观的界面

BaseSpace Sequence Hub

BaseSpace Sequence Hub上提供的基于云的DRAGEN组合将准确、高效的分析与安全的生态系统和多功能可扩展性相结合。BaseSpace Sequence Hub上的DRAGEN软件可为各种规模和学科的实验室提供一键式二级分析。BaseSpace Sequence Hub是因美纳仪器的直接扩展。仪器生成的加密数据可以直接导入BaseSpace Sequence Hub，您可以利用一组精选的应用程序来轻松地管理和分析数据。BaseSpace Sequence Hub由Amazon Web Services（AWS）提供支持：

- 为DRAGEN分析提供简便易用的一键式解决方案
- 使用直观的图形用户界面，便于专家和非专家用户高效操作
- 可访问强大的计算资源，无需为额外的基础架构支出资金

Illumina Connected Analytics

Illumina Connected Analytics是一个基于云的全面生物信息学平台，它使研究人员能够在安全、可扩展且灵活的环境中管理、分析并解读大量多组学数据。访问Illumina Connected Analytics上的DRAGEN二级分析组合，可将其作为预设流程或单独的工具整合到定制流程中。

总结

DRAGEN二级分析是一组功能强大的软件工具，可实现准确、全面且高效的NGS数据分析。多种DRAGEN软件部署选项能帮助实验室选择最适合其项目类型和规模的解决方案。此外，用户可以组合各种部署选项以更好地满足其性能和 workflows 需求。随着NGS技术的不断发展，及时更新DRAGEN二级分析可确保当前流程的理想性能，同时随着应用日益广泛，将不断扩展新流程。

[‡] 根据样本建议运行时。

了解更多 →

[DRAGEN二级分析](#)

[DRAGEN二级分析支持页面](#)

[联系我们](#)

参考文献

1. US Food and Drug Administration. Truth Challenge V2: Calling Variants from Short and Long Reads in Difficult-to-Map Regions. precision.fda.gov/challenges/10/results. Accessed March 31, 2025.
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. Published January 12, 2022 Accessed March 31, 2025.
3. Illumina. DRAGEN wins at PrecisionFDA Truth Challenge V2 showcase accuracy gains from alt-aware mapping and graph reference genomes. illumina.com/science/genomicsresearch/articles/dragen-wins-precisionfda-challengeaccuracy-gains.html. Published November 9, 2020. Accessed March 31, 2025.
4. Internal data on file. Illumina, Inc., 2025.
5. Behera S, Catreux S, Rossi M, et al. [Comprehensive genome analysis and variant detection at scale using DRAGEN](#). *Nat Biotechnol*. doi:10.1038/s41587-024-02382-1
6. 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-edicoset-world-record-for-secondary-analysis-speed. Published October 23, 2017. Accessed March 31, 2025.
7. Sisson P. Rady Children's Institute sets Guinness world record. The San Diego Union-Tribune. sandiegouniontribune.com/2018/02/12/rady-childrens-institute-sets-guinness-world-record/. Published February 12, 2018. Accessed March 31, 2025.
8. Betschart RO, Thiéry A, Aguilera-Garcia 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. doi:10.1038/s41598-022-26181-3
9. Illumina. Accurate and Efficient Calling of Small and Large Variants from PopGen data sets Using the DRAGEN Bio-IT Platform. illumina.com/science/genomics-research/articles/popgen-variant-calling-with-dragen.html. Published May 24, 2021. Accessed March 31, 2025.

Illumina中国

上海办公室 • 电话 (021) 6032-1066 • 传真 (021) 6090-6279

北京办公室 • 电话 (010) 8441-6900 • 传真 (010) 8455-4855

技术支持热线 400-066-5835 • chinasupport@illumina.com

市场销售热线 400-066-5875 • china_info@illumina.com • www.illumina.com.cn

© 2025 Illumina, Inc. 保留所有权利。所有商标均为因美纳公司或其各自所有者的财产。
关于具体的商标信息, 请访问www.illumina.com.cn/company/legal.html。

M-GL-00680 v12.0



illumina[®]