

DRAGENTM

二级分析

利用新一代测序数据实现准确、
全面、高效的变异检出

illumina[®]

简介

利用新一代测序（NGS）技术解码基因组对于推动研究和医学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，适用于基因组、外显子组、转录组和甲基化组研究等一系列应用。DRAGEN 平台集成了多种二级分析软件，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 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绘制区域，展现出高准确性^{1,2}。Graph 基因组、因美纳机器学习（ML）与 DRAGEN 4.0 软件不断创新，与其他测序技术相比，展现了出色的数据准确性，在所有基准区域中的 F1 评分（精度和查全率的组合测量）为 99.83%（图 1）^{1,2}。针对主要组织相容性复合物（MHC）区域，与所有 PrecisionFDA V2 提交数据相比，DRAGEN 4.0 + Graph（默认启用 ML）在检出准确性方面也获得了最高的 F1 评分。

全面的分析

对于需要执行各种 NGS 应用的实验室而言，DRAGEN 二级分析可满足其需求，利用单一平台即可全面覆盖各类实验。DRAGEN 流程支持多种实验类型，包括全基因组测序（WGS）、

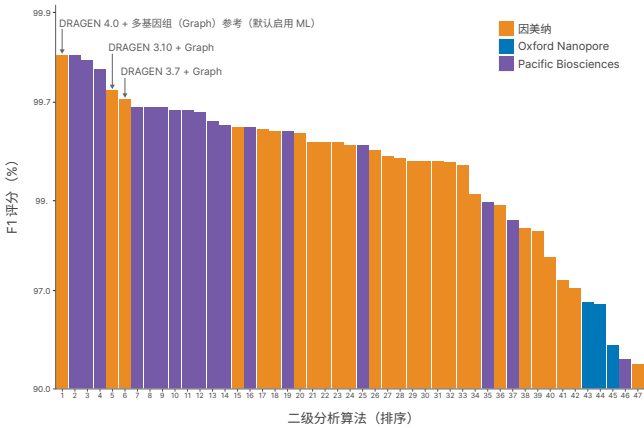


图 1: 与所有基准区域数据集中的 PrecisionFDA Truth Challenge v2 提交数据相比，DRAGEN 4.0 + Graph（默认启用 ML）的准确性——DRAGEN 4.0 + Graph（默认启用 ML）展现了出色的准确性，与 Pacific Biosciences 测序数据的 Google DeepVariant 并列。相对于 DRAGEN 3.7 + Graph，DRAGEN 3.10 + Graph 有所改进，这是由于图形和参考基因组重叠群处理方面得到优化。Y 轴的 F1 评分 (%) 是根据真阳性和真阴性结果占总结果的比例来计算的^{3,4}。

富集 panel、单细胞 RNA-Seq、单细胞 ATAC-Seq、单细胞多组学、细胞群 RNA-Seq 和甲基化分析（表 1）。该流程需要 30 多个开源工具才能部分复制 DRAGEN 软件中的丰富功能^{3,4}。

DRAGEN 软件利用随附的变异检出程序组合——重复扩增、结构变异（SV）、拷贝数变异（CNV）、ExpansionHunter 和靶向检出程序（如 SMN、GBA、CYP2B6、CYP2D6 和 HLA），可提供广泛的基因组覆盖范围。此外，DRAGEN 多基因组（Graph）参考有效延长了因美纳 read，并探索低复杂度区域，从而解析由于重复序列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，而且能够在难以绘制的区域实现单核苷酸、拷贝数和结构变异检出。

表 1: DRAGEN 二级分析支持广泛的二级分析应用

应用	DRAGEN 本地服务器	机载 NovaSeq X 系列平台	机载 NextSeq 1000、NextSeq 2000 系统	BaseSpace Sequence Hub	Illumina Connected Analytics	
					预配置	定制
BCL 转换	✓	✓	✓	✓	✓	✓
DRAGEN ORA 压缩	✓	✓	✓			✓
DRAGEN FASTQ + MultiQC	✓	✓	✓	✓	✓	✓
全基因组	生殖系 + 体细胞	仅包含生殖系	仅包含生殖系	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	生殖系 + 体细胞
DNA 扩增子	✓		✓	✓	✓	✓
RNA	✓	✓	✓	✓	✓	✓
单细胞 RNA	✓		✓	✓	✓	✓
差异表达		✓	✓	✓		
NanoString GeoMx NGS				✓		
RNA 扩增子	✓			✓	即将推出	即将推出
甲基化	✓			✓	✓	✓
宏基因组学				✓		
RNA 病原体检测				✓		
COVID 谱系	✓			✓	即将推出	
TruSight Oncology 500	✓				✓	
ScATAC-Seq	✓			✓	✓	✓
填补	✓			✓	✓	✓
PGx 星号等位基因检出程序	✓			✓	✓	✓
Illumina Complete Long Reads				✓		

高效分析

DRAGEN 软件经过专门设计，可为实验室提供所需的数据分析速度和文件选项，以便从 NGS 数据集中获得最大收益。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记录^{5,6}。在实际应用中，本地 DRAGEN 二级分析可以在 25 分钟之内以 30× 的覆盖度在本地处理整个人类基因组的 NGS 数据，而基于 CPU 的传统系统需要的时间超过 15 小时⁷。

为了满足大型 NGS 数据文件的存储需求，DRAGEN Original Read Archive (ORA) 技术可将 FASTQ 文件无损压缩至原来的 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。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)。

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 Reference Builder 使用户能够生成人类、非人类或非标准参考基因组，所有这些都称为哈希表。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客户的参考文件。BaseSpace™ Sequence Hub 上的 DRAGEN Reference Builder 应用程序需要使用 FASTA 文件。大多数 DRAGEN 流程都包含对 hg19、hg38（包含或不含 HLA*）、GRCh37 和 hs37d5 的内置支持。DRAGEN Graph 工具包使用户能够将图形参考基因组功能扩展到更多样化的人类图形参考基因组。

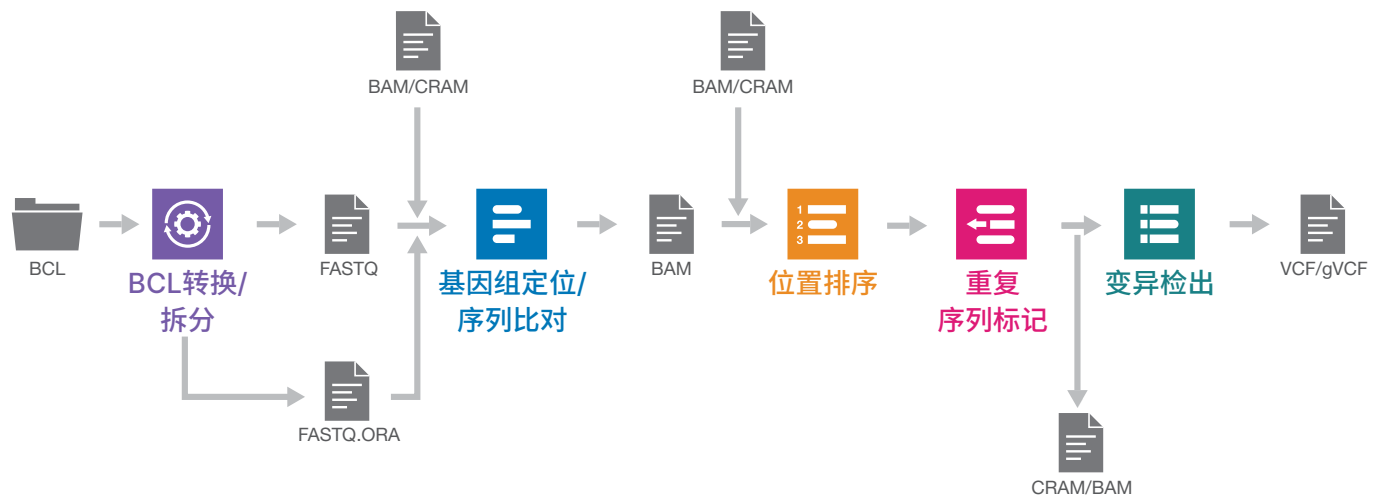


图 2: DRAGEN 流程的灵活性——每种 DRAGEN 流程都有一组独特的步骤, 以支持准确高效的分析。DRAGEN 流程可以灵活地接受多种输入文件格式并生成一系列输出文件, 为用户提供个性化的体验并生成他们所需的文件格式。

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达 4 个应用程序。
- 2. **可扩展容量**——在工作量增加且样本量大的时候，实验室可以通过可用的并行 DRAGEN 软件访问选项扩展容量（图 3）。
- 3. **扩大运行规模**——一个 DRAGEN 实例就可运行所有 DRAGEN 流程和支持的样本类型。DRAGEN 软件的准确性、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. **从外显子组到基因组**——从全外显子测序（WES）扩大到 WGS 涉及到生成数据量的大幅度增长。DRAGEN 软件能让客户将外显子组分析扩展到基因组分析，而且无需在额外的硬件基础设施或云端解决方案上进行大量投资。

5. **超大型数据集**——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，以高精度从队列抽样中检出小型和大型变异。DRAGEN 软件能够对数千到数百万个基因组变异检出格式（gVCF）文件进行汇总和基因分型，整合新批次无需重新处理已有批次。DRAGEN Joint Genotyping Pipeline 可在不影响精度的前提下，在多个基因组中联合检出变异，并通过快速分析扩展到大型队列⁸。例如，针对千人基因组计划数据的 DRAGEN 二次分析能够对不同样本进行准确的大规模变异检出，并确定覆盖数据不均一或偏离假设的区域。

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）。

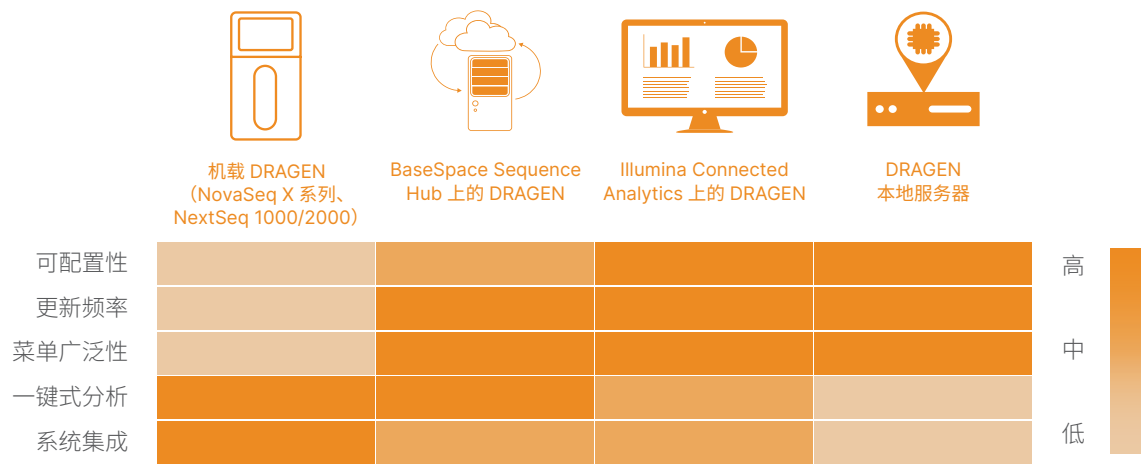


图 3： DRAGEN 流程访问选项，旨在满足每个实验室的 NGS 分析需求。

DRAGEN 本地服务器

本地 DRAGEN 需要使用本地储存解决方案来收集和储存 NGS 数据。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，再将其传输到 DRAGEN 服务器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：

- 支持不同水平的命令行界面
- 可代替多达 30 个传统计算实例
- 在约 25 分钟内以 30× 的覆盖度处理整个人类基因组的 NGS 数据

NovaSeq™ X 系列机载 DRAGEN

NovaSeq X 系列包括更强大的 DRAGEN 软件，提供准确、自动化且全面的二级分析。机载 DRAGEN 软件组合通过 NGS 应用程序（表 1）提供变异检出和 ORA 压缩功能，包含 BCL Convert、Germline、RNA 和 Enrichment 等应用程序。机载 DRAGEN：

- 并行运行多个二级分析流程
- 在每次运行中，每个流动槽最多可同时执行 4 个应用程序
- 通过支持的应用程序实现高达 5× 的无损数据压缩以及分析
- 节省分析成本，五年内可超过 NovaSeq X 系统的购买价格

NextSeq™ 1000 和 NextSeq 2000 系统 DRAGEN 机载分析

NextSeq 1000 和 NextSeq 2000 系统包括机载 DRAGEN 软件，可实现快速、准确的二级分析。机载 DRAGEN 软件组合提供了一组精选的流程，涵盖一系列常见的 NGS 应用（表 1），其用户友好的界面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：

- 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

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 由亚马逊网络服务（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 上的 DRAGEN 二级分析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。Illumina Connected Analytics：

- 可访问完整的 DRAGEN 软件，该软件可作为预设流程或用于定制流程的单个工具
- 支持高度自动化的工作流程和定制解决方案，以便优化高通量研究
- 提供高度安全的环境，保证数据存储、单点登录访问、审查日志和访问控制功能，支持健康保险便携性与责任法案（HIPAA）规定和欧盟通用数据保护法规（GDPR）原则

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。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最佳性能，同时随着应用日益广泛，将不断扩展新流程。

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. Food and Drug Administration. Truth Challenge V2: Calling Variants from Short and Long Reads in Difficult-to-Map Regions. precision.fda.gov/challenges/10. Accessed March 14, 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. Accessed March 14, 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. Accessed March 14, 2022.
4. Internal data on file. 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. Accessed March 14, 2022.
6. San Diego Union Tribune. Rady Children's Institute sets Guinness world record. <https://www.sandiegouniontribune.com/95899028-132.html>. Published February 12, 2018. Accessed March 14, 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. Accessed March 14, 2022.

illumina®

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) 8455-4866 • 传真 (010) 8455-4855

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

© 2023 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/company/legal.html。

M-GL-00680 v5.0



illumina®