

Illumina DRAGEN™ Bio-IT 平台

兼顾性能的准确、超快速二级分析

要点

- **准确的数据**
以高分析灵敏度和高特异性检测小变异
- **超快速的分析**
在约 25 分钟内处理一个 30 倍覆盖度的全基因组，在约 8 分钟内处理一个 100 倍覆盖度的全外显子组
- **经济高效且易于使用的解决方案**
减少硬件投入和云端成本，具有一键式选项或命令行选项
- **稳定可靠的应用程序**
支持本地和 BaseSpace™ Sequence Hub 中的各种应用程序和方法

我们在不断地利用新颖先进的应用程序来释放基因组的力量，在这个过程中，新一代测序（NGS）产生的数据量迅速增长。2018 年，Illumina 系统生成的数据量已超过 100 PB。

客户需要先进的数据分析工具来处理庞大的数据量，这些工具必须能有效地将原始测序数据转换成有意义的结果，同时又不影响准确性和成本。此外，不熟悉这项技术的机构需要易于使用的解决方案，来减少采用该技术的财务障碍和专业知识壁垒，才能更好地利用 NGS 的优势。

Illumina DRAGEN Bio-IT 平台在设计时充分考虑了客户在 NGS 数据分析方面的关键痛点，开发出了一款高度准确且超快速的二级分析解决方案，能满足小型研究实验室和群体规模基因组项目等不同通量的研究需求。

DRAGEN 平台简介

Illumina DRAGEN (Dynamic Read Analysis for GENomics) Bio-IT 平台能对基因组、外显子组和转录组的 NGS 数据进行二级分析。DRAGEN 平台的基本功能解决了基因组分析中的常见难题，例如计算时间长和数据量大。DRAGEN 平台可在不影响精确度的前提下，实现快速、灵活和经济高效，使各种规模和学科的实验室都能利用其基因组数据挖掘更多的信息。

DRAGEN 平台是一款软件 + 硬件组合解决方案，提供多种硬件加速的二级分析流程。DRAGEN 的一系列分析流程采用现场可编程门阵列技术（FPGA），通过硬件加速来实施基因组分析算法，包括 BCL 转换、定位、比对、排序、重复序列标记和单倍型变异检出。

DRAGEN 平台能生成一系列稳定可靠的指标，包括：

- 文库制备质量控制
- 分析质量控制
- 拆分
- 重复 read
- 类似 SAM/PICARD 的原始数据处理工具

DRAGEN 平台的可重编程特性使 Illumina 开发出了一组强大的二级分析流程，并且可以让客户在单个 DRAGEN 服务器上运行多种不同的流程。我们采用了集合多个硬件加速模块的优化了的软件算法来进行全新的设计，不断改进 DRAGEN 流程，发布了多种流程来提供多种分析功能、加强准确性并提高速度。

您可以在本地或通过 Illumina BaseSpace Sequence Hub 在云端使用 DRAGEN 分析流程，所有 DRAGEN 流程都可以进行版本控制。

准确的结果

DRAGEN 平台立足于世界一流的算法，还将通过持续更新紧跟不断演变的行业标准和最佳操作，为基因组应用和 workflows 提供出色的分析灵敏度和特异性。

DRAGEN 平台消除了多种来源的偏差和错误，全面确保准确性。在 2017 年举行的 PrecisionFDA Hidden Treasures 热身挑战赛中，DRAGEN 平台成功识别出所有 50 个隐藏变异，在多种分析方案中脱颖而出，在全基因组变异检出的六项精确度测量中有五项获得了最高分。¹ 每个 DRAGEN 平台的新版本中都设计了改良的算法来提高其准确性（图 1）。

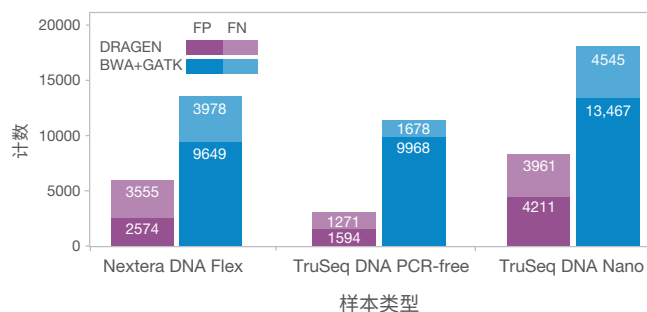


图 1：高度准确的单核苷酸位点变异检测 - 我们将 DRAGEN 平台与其他热门的变异检出平台进行了比较，使用三种 Illumina 文库制备试剂盒对参考基因组 DNA 进行了测序，再分别通过 DRAGEN 平台或 BWA+GATK 进行分析。将变异检出结果与参考基因组真值集合进行了对比，确定误检（假阳性，FP）或漏检（假阴性，FN）的变异。

超快速的分析

DRAGEN 平台凭借 FPGA 架构成功地加快了周转时间。经比较，传统的基于 CPU 的系统通过执行软件代码行来实现算法功能，而 FPGA 能够以逻辑电路的方式实现上述算法，从而基本达到结果的即时输出。

DRAGEN 平台可以在约 25 分钟之内在本地处理 30 倍覆盖度的人类基因组 NGS 数据，在约 8 分钟之内在本地处理 100 倍覆盖度的人类外显子组 NGS 数据，而传统的基于 CPU 的系统需要的时间超过 10 小时。DRAGEN 平台还为基因组数据分析创造了两项世界速度记录（图 2）。^{2,3}



图 2：基因组分析跻身吉尼斯记录的速度 - 两个独立机构使用 DRAGEN 平台创下了基因组分析的速度记录。^{1,2}

经济高效的解决方案

DRAGEN 平台可以减少服务器集群的硬件投资和云计算资源的利用。本地 DRAGEN 平台可代替多达 30 个传统计算实例，从而降低了硬件和维护成本，以及电力消耗和冷却等其他费用。

BaseSpace Sequence Hub 上的 DRAGEN 提供了同样高质量的 DRAGEN 流程，而且还具备 BaseSpace 的灵活性和安全性，平均价格为 5 美元 / 基因组和 3 美元 / 外显子组。*

DRAGEN 流程可以将 read 的比对结果本地压缩为 CRAM 文件格式，与标准 BAM 文件相比，能减少了 50% 的数据占用空间。

* 近似成本。随所用样本的起始量而变化

稳定可靠的应用程序

DRAGEN 平台具有一组稳定可靠的二级分析流程（表 1），支持基因组、外显子组和 RNA 分析。能在流程的不同阶段接收输入文件或创建输出文件（图 3），还能在单个本地服务器或 BaseSpace Sequence Hub 上运行多种不同流程。

如需查看完整的 DRAGEN 流程列表，请访问 www.illumina.com/DRAGEN。

表 1：DRAGEN 超快速分析流程支持多种应用

Illumina DRAGEN Germline Pipeline^a
DRAGEN Germline Pipeline 提供端到端（BCL → VCF）NGS 分析（包括用于提高精确度的高级误差模型校准）以及通过 Illumina Expansion Hunter 进行的短序列重复片段检测和基因分型。
Illumina DRAGEN Somatic Pipeline^a
DRAGEN Somatic Pipeline 包括仅肿瘤模式和肿瘤 - 对照模式，用于检测肿瘤样本中的体细胞变异。
Illumina DRAGEN RNA Pipeline^a
DRAGEN RNA Pipeline 可进行快速比对、定量、可变剪切和融合检测。
Illumina DRAGEN Joint Genotyping/Population Pipeline^a
DRAGEN Joint Genotyping/Population Pipeline 可在不影响精确度的前提下，从多个样本中联合检出变异并快速扩展到数千个样本。
Illumina DRAGEN CNV Pipeline^b
DRAGEN CNV Pipeline 可进行生殖系和体系外显子组和基因组层面的拷贝数变异（CNV）分析。在获得最终检出结果之前，可使用不同水平的过滤条件来减少假阳性。
Illumina DRAGEN Methylation Pipeline^a
DRAGEN Methylation Pipeline 提供多种操作模式，包括仅参考比对模式和注释辅助模式。
Illumina DRAGEN Map + Align Pipeline^c
DRAGEN Map + Align Pipeline 能够超快速地定位和比对外显子组和基因组的 DNA 和 RNA。
Illumina DRAGEN Reference builder^a
DRAGEN Reference Builder 可使用 FASTA 文件，并且可构建供 DRAGEN 应用程序使用的专有参考基因组索引。

- a. 可在本地或在 BaseSpace Sequence Hub 中使用。
- b. 仅在本地使用。
- c. 仅在本地使用。如需在 BaseSpace Sequence Hub 中使用该功能，可通过 Germline Pipeline 进行定位和比对，导出的 BAM 文件即可用于变异检出。

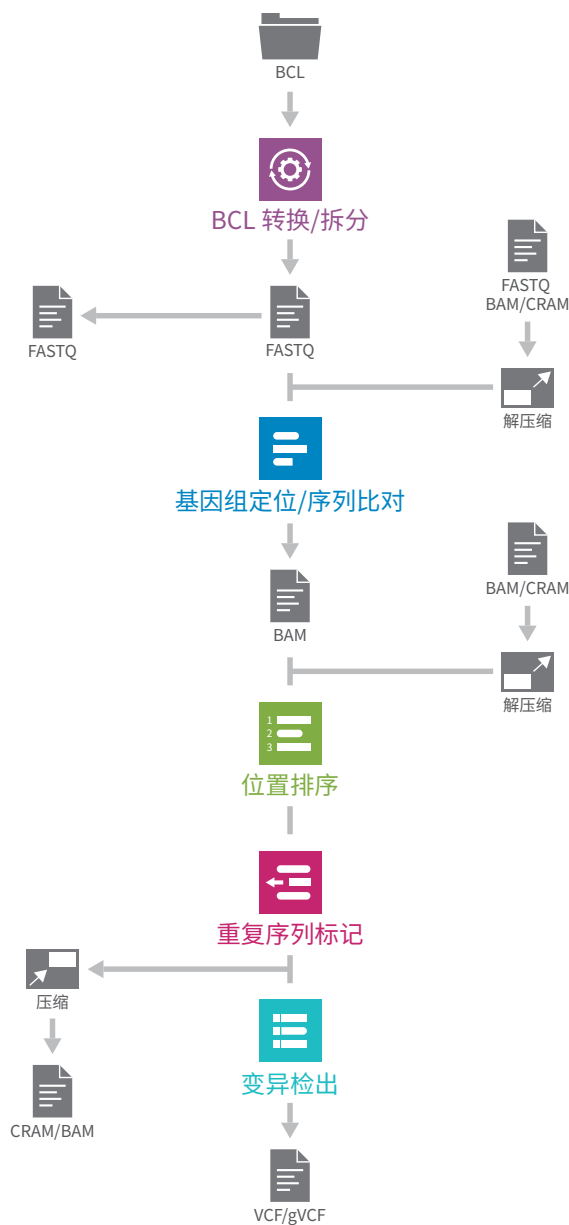


图 3：DRAGEN 流程的灵活性 - 根据 DRAGEN 流程功能的不同，每种流程都有一组独特的步骤。正如图上 DRAGEN Germline Pipeline 所示，DRAGEN 可以灵活地接受多种输入文件格式并生成一系列输出文件，为用户提供个性化的体验并生成他们所需的文件格式。

定制参考基因组

客户可以使用 DRAGEN Reference Builder（也被称为哈希表）来生成非人类或非标准的参考基因组索引。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。DRAGEN Reference Builder 需要 FASTA 文件。大多数 DRAGEN 流程都包含对 hg19、hg38（包含或不包含 HLA）、GRCh37 和 Hs37d5 的内置支持。

易于实施

本地 DRAGEN 和 BaseSpace Sequence Hub 上的云端 DRAGEN 的解决方案适用于具备不同生物信息学专业知识的实验室——支持命令行执行和一键式执行（图 4）。DRAGEN 平台不需其他配置，可以直接使用。

一键式选项：BaseSpace Sequence Hub 上的 DRAGEN 可让具备不同信息学专业知识的实验室都能以低廉的成本在内部进行二级分析。

命令行选项：本地 DRAGEN 提供了命令行界面，能通过基于简单易学的 Linux 的命令行界面（CLI）进行单命令启动或用于高级命令行。

BaseSpace Sequence Hub	本地	
一键式	单命令启动	高级命令行
简单的图形用户界面（GUI） 管理服务 HIPAA 合规 * 工作组功能 简单的数据共享	基于简单易学的 Linux 的 命令行界面（CLI） 简单的命令行执行	编写连续作业脚本 为不同应用生成不同的 配置文件

图 4：DRAGEN 的实施选项 - 在 BaseSpace Sequence Hub 中，用户只需选择应用程序、输入信息并启动运行。本地 DRAGEN 采用的是命令行界面。新用户可选择易于学习和操作的命令行界面。更有经验的用户可以使用高级命令行界面进行自定义设置。* HIPAA 兼容性仅适用于美国的企业级 BaseSpace 客户。

可扩展性

DRAGEN 平台能让实验室在保持低成本和短周转时间的同时实现规模化运行。DRAGEN 可以通过多种方式来扩展研究能力：

- 1) **满足 NovaSeq™ 6000 系统的需求：**一台 DRAGEN 服务器可以处理两台 NovaSeq 6000 仪器采用双 S4 流动槽满负荷运行产生的数据。
- 2) **可扩展容量：**在高运算需求且样本量增加的情况下，实验室可以使用 BaseSpace Sequence Hub 上的 DRAGEN 来扩展计算能力。并行的 DRAGEN 流程可以让您将分析转移到 BaseSpace Sequence Hub。
- 3) **扩大运行规模：**单个 DRAGEN 平台就可运行所有的 DRAGEN 流程和支持的样本类型。DRAGEN 的速度、准确性和成本效益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：**从全外显子组测序（WES）扩大到全基因组测序（WGS）涉及到生成数据量的大幅度增长。DRAGEN 让客户轻松地 从外显子组扩展到基因组，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。

可在本地或通过 BaseSpace Sequence Hub 使用

稳定可靠的 DRAGEN 流程既可以在本地使用，也可以通过 BaseSpace Sequence Hub 在云端使用，让实验室能使用最符合其需求的解决方案。

本地 DRAGEN

对于希望在本地进行分析的机构，本地 DRAGEN 提供了稳定可靠的二级分析解决方案，可整合到现有的储存解决方案中（图 5）。

本地 DRAGEN 非常适合用于：

- **将数据保存在本地：**对于希望将数据储存在本地的机构，本地 DRAGEN 是一个理想的解决方案。
- **有限的网络连接：**在网络连接受限或紧张的地区，本地 DRAGEN 可以离线运行。
- **利用现有的基础设施：**本地 DRAGEN 能够让实验室使用现有的储存设备。

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储存设备后，DRAGEN 会再将数据从储存设备传输到 DRAGEN 服务器并执行所选的工作流程，然后将生成的分析输出文件写入本地储存解决方案。DRAGEN 服务器采用基于 Linux 的命令行界面（CLI），可以设置为单命令启动或高级命令行。

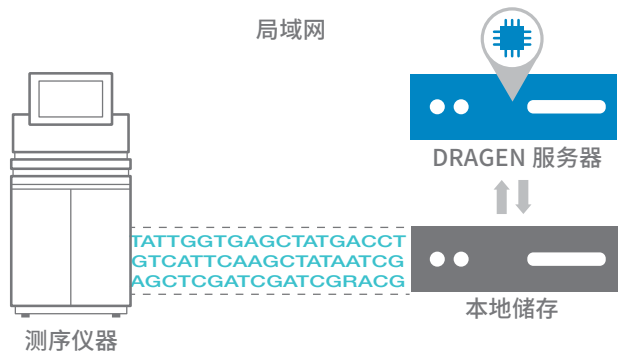


图 5：本地 DRAGEN 解决方案 - 数据由测序仪器直接传输到本地储存解决方案，再传输到 DRAGEN 服务器以进行拆分和二级分析。DRAGEN 服务器会将分析结果返回本地储存解决方案。

本地 DRAGEN 提供了几种不同的许可证通量级别，从 100,000 Gb/ 年到 2,000,000 Gb/ 年不等（表 2）。定期发布更新版本和新功能，可通过 DRAGEN 客户门户网站获取。我们提供本地 DRAGEN 付费安装服务。

BaseSpace Sequence Hub 上的 DRAGEN

DRAGEN 可以通过 Illumina BaseSpace Sequence Hub 在云端使用，为各种规模、各个学科的实验室提供快捷、准确、经济高效的一键式二级分析。BaseSpace 上的 DRAGEN 采用了亚马逊网络服务（AWS）EC2 F1 实例，可为基因组、外显子组、转录组等提供更快的二级分析。

BaseSpace Sequence Hub 上的 DRAGEN 具有以下特性：

- **易于使用：**用户可以直接将数据从测序仪器传输到 BaseSpace Sequence Hub，然后按下按键即可启动 DRAGEN 流程
- **低成本：**没有硬件投入，样本分析的成本约为 5 美元 / 基因组和 3 美元 / 外显子组。
- **云端的安全性与合规性：**BaseSpace Sequence Hub 的理念是安全优先
- **共享数据：**可在云端与合作者安全地共享项目
- **灵活性：**应用程序可以用于小型研究，也可以根据实验室需求扩展分析规模。

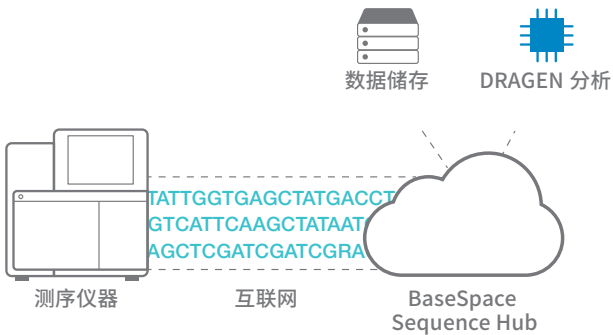


图 6：云端 DRAGEN - 测序数据可以实时传输到 BaseSpace Sequence Hub，在此储存并使用选定的 DRAGEN 流程进行分析。

BaseSpace Sequence Hub 上的云端 DRAGEN 将 DRAGEN 准确快速的分析与 BaseSpace Sequence Hub 的安全生态系统和多种功能结合在了一起。

紧凑的仪器整合使加密数据可以直接从仪器传输至 BaseSpace Sequence Hub，用于分析、储存、共享和其他形式的数据管理（图 6）。BaseSpace Sequence Hub 通过无线互联网连接仪器，可以在仪器上机设置期间或软件安装后通过机载设置菜单轻松启用（图 7）。

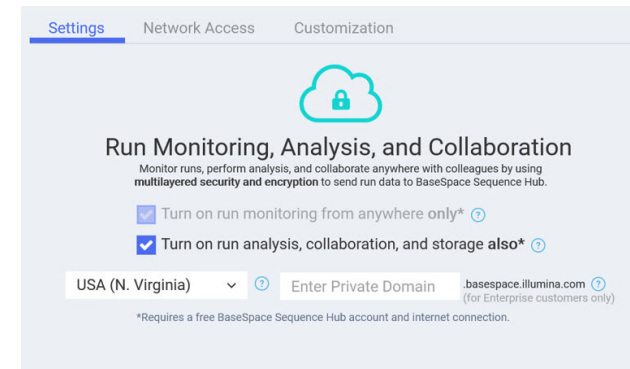


图 7：在 BaseSpace Sequence Hub 上轻松设置 DRAGEN - 在上机设置期间或软件安装后，可以通过设置菜单在仪器上轻松设置从仪器传输至 BaseSpace Sequence Hub 的加密数据。

云端的安全性与合规性

BaseSpace Sequence Hub 一直秉承安全优先的理念，它经过了独立的审查，通过了 HIPAA 合规、ISO 27001（信息安全管理）和 ISO 13485（医疗器械质量管理体系）* 的认证。它的构建旨在保护数据隐私和符合 GDPR 的规范。BaseSpace Sequence Hub 包括端到端加密、审查和精细的访问控制。BaseSpace Sequence Hub 允许用户更新至新版本、还原至旧版本，对于环境受控的实验室，亦可选择维持版本恒定不变。

有关 BaseSpace Sequence Hub 数据安全的更多详情，请参阅[安全性与合规性白皮书](#)。

* 可通过企业级 BaseSpace Sequence Hub 帐户获取

表 2：本地 DRAGEN 的订购信息

产品名称	说明		货号
DRAGEN Server（DRAGEN 服务器）	采用 FPGA 芯片来加速 NGS 二级分析		20027360
DRAGEN Server Advance Exchange Support Plan（DRAGEN 服务器高级维保换机支持计划）	包含针对 DRAGEN 服务器的高级维保换机服务；远程技术支持（8 × 5）		20032797
DRAGEN Server Installation（DRAGEN 服务器安装）			20031995
产品名称	吞吐量	估计的样本量（30×WGS）	货号
DRAGEN Level 1 License（DRAGEN 1 级许可证）	100,000 Gb	1000 个样本	20027361
DRAGEN Level 2 License（DRAGEN 2 级许可证）	250,000 Gb	2500 个样本	20027361
DRAGEN Level 3 License（DRAGEN 3 级许可证）	500,000 Gb	5000 个样本	20027361
DRAGEN Level 4 License（DRAGEN 4 级许可证）	1,000,000 Gb	10,000 个样本	20027361
DRAGEN Level 5 License（DRAGEN 5 级许可证）	2,000,000 Gb	20,000 个样本	20027361
每种许可证的有效期均为 1 年。			

云端 DRAGEN 流程

BaseSpace Sequence Hub 上有所有的 DRAGEN 流程，并且会定期发布新版本（图 8）。

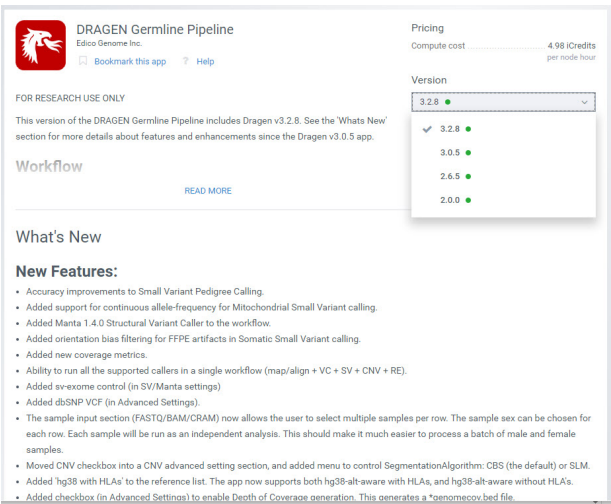


图 8：BaseSpace Sequence Hub 上的 DRAGEN 流程会定期更新，各个版本均已锁定持续可用。

了解更多

如需了解更多有关 Illumina DRAGEN™ Bio-IT 平台的信息，请发送电子邮件至 Informatics_GC@illumina.com 或者拨打电话 4000665835。

当前的用户指南和安装指南等支持文档，均可通过 [Illumina 支持网站](#) 获取。

参考文献

1. Precision FDA Hidden Treasures Warm Up. precision.fda.gov/challenges/1/view/results. Accessed September 14, 2018.
2. Bio IT World. Children's Hospital Of Philadelphia, Edico Set World Record For Secondary Analysis Speed. October 23, 2017. www.bio-itworld.com/2017/10/23/childrens-hospital-of-philadelphia-edico-set-world-record-for-secondary-analysis-speed.aspx. Accessed September 19, 2018.
3. The San Diego Union Tribune. Rady Children's Institute sets Guinness world record. February 12, 2018. www.sandiegouniontribune.com/news/health/sd-no-rady-record-20180209-story.html. Accessed September 19, 2018.

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 · techsupport@illumina.com · www.illumina.com.cn

仅供研究使用。不得用于诊断。

© 2019 Illumina, Inc. 保留所有权利。所有商标均为 Illumina 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。Pub. No. 970-2018-002-B QB 6572



Illumina Academy



@Illumina

illumina®