

# 采用 NovaSeq™ 6000 测序系统进行人类全基因组测序

凭借简单直观的设计，NovaSeq 6000 系统为人类全基因组测序提供了极高的单日通量和出色的数据质量

## 要点

### • 可扩展通量

使用简单的一体化、自动化工作流程，在双流动槽模式下生成高达 6 Tb 和 200 亿条 read

### • 极富灵活性

该系统可在 1 天里对一组 3 个样本进行测序，也可支持在 2 天左右的时间里对多达 48 个基因组进行测序

### • 出色的数据质量

高度准确的 Illumina 边合成边测序 (SBS) 化学技术可提供非常优秀的的数据质量

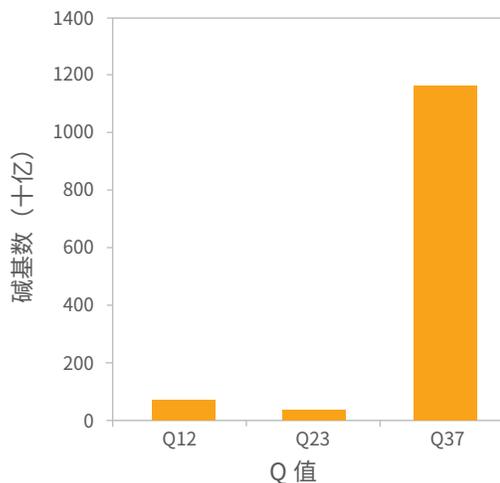


图 1 : NovaSeq 质量分值分布 - NovaSeq 6000 系统上人类基因组以 2×151 bp 运行的质量分值。这个例子表明超过 91% 的测序碱基高于 Q30。

## 简介

NovaSeq 6000 系统用突破性的创新开启了测序的新时代，其提供的通量、速度和灵活性可让用户非常快速的、经济地完成项目。NovaSeq 6000 系统结合之前 Illumina 平台的最优特性，采用了更多创新技术，可在大约 2 天内提供最高达 6 Tb 和 200 亿条 read 的可调产出。NovaSeq 6000 系统利用最广泛采用的新一代测序 (NGS) 技术——成熟的 Illumina 边合成边测序 (SBS) 化学技术\*。SBS 化学技术提供了业界出色的数据准确性、极高产的无错 read，以及最高比例的分值大于 Q30 的碱基检出<sup>1</sup>。

本应用说明介绍并比较了 NovaSeq 6000 仪器上单次运行产生的人类全基因组测序 (WGS) 数据与之前在 HiSeq® 2500、HiSeq 4000 和 HiSeq X® 系统上产生的人类 WGS 数据。结果表明，在现有的平台上执行测序，该仪器的数据质量已达到或超过之前的数据质量。

## 方法

文库是用 100 ng Coriell NA12878 基因组 DNA，使用 TruSeq® DNA PCR-Free Library Preparation Kit (Illumina，货号 FC-121-3001) 制备的。PCR 定量和稀释后，在 NovaSeq 6000 系统上使用 NovaSeq 6000 S2 Reagent (Illumina，货号 200012860) 进

行簇生成及测序。所有文库均以 2×150 bp 的配置使用双重标签运行，只有 HiSeq 2500 以 2×125 bp 的配置使用双重标签运行 (表 1)。每个 NovaSeq S2 流动槽装载六个样本。包括 Read 1、标签序列 1 read (i7)、标签序列 2 read (i5)、双端运行和 Read 2 的簇生成和测序步骤在无需用户干预的情况下自动进行。

NovaSeq 6000 系统产生的数据文件与人类参考基因组 GRCh38 进行比对<sup>3</sup>。在 BaseSpace® Sequence Hub 中使用 Illumina Whole Genome Sequencing Analysis App v5 对原始数据和降低覆盖度的数据进行分析得到 BAM 文件和变异检出文件<sup>4</sup>。表 2 示出了变异检出、构建深度及其他二级分析指标。BaseSpace Sequence Hub 的公共数据部分提供了此运行产生的原始数据和变异检出文件。

## 结果

NovaSeq 6000 系统产生了高质量的全基因组数据。NovaSeq 6000 测序运行在单个 NovaSeq S2 流动槽上产生了 1268 Gb 数据，其中 91% 的碱基数据分值高于 Q30 (图 1 和表 1)。NovaSeq S2 流动槽的产量更高，每个流动槽可以容纳最高达八个人类全基因组样本或两个 WGS 三重样本，具体取决于所需的覆盖深度。

\* Illumina 公司 2015 年存档数据。

表 1 :单流动槽运行的测序运行指标比较

指标 <sup>a</sup>	HiSeq 2500 v4	HiSeq 4000	HiSeq X	NovaSeq 6000
运行配置 (read 数 × 读长)	2×125	2×150	2×150	2×150
通过过滤的 read (百万)	2095	2937	3862	4037
≥ Q30 的碱基 <sup>b</sup>	87.7%	89.7%	91.9%	91.3%
产量 (Gb)	524	883	1190	1268

a. 单个流动槽单次运行产生的所有数据

b. 两种 read 的平均对于质量分值为 30 (Q30) 的碱基检出, 预测每 1000 个碱基检出中有一个不正确

表 2 :覆盖度和变异分析指标

指标 <sup>a,b</sup>	HiSeq 2500	HiSeq 4000	HiSeq X	NovaSeq 6000
构建深度 <sup>c</sup>	30×	30×	30×	30×
总 SNP	3,628,296	3,684,413	3,706,366	3,677,388
Het:Hom 比例	1.54	1.57	1.57	1.55
Ti:Tv 比例	2.07	2.06	2.06	2.06
dbSNP 数据库中匹配的位置	95.04%	94.64%	94.53%	94.87%
覆盖度 ≥ 10× 的百分比	97.81%	98.33%	98.44%	98.40%
SNP 查准率	99.85	99.79	99.80	99.87
SNP 查全率	96.28	96.97	97.23	97.08
插入缺失查准率	96.93	96.06	97.34	97.42
插入缺失查全率	87.09	90.40	94.90	95.28

a. 缩写: **SNP** = 单核苷酸多态性, **Indel** = 插入缺失突变, **Het:Hom 比例** = 杂合子 / 纯合子比例, **Ti:Tv 比例** = 转换 / 颠换比例, **查准率** (准确度) = 按 [ 真阳性检出数 / (真阳性检出数 + 假阳性检出数) ] 的比值计算, **查全率** (灵敏度) = 按 [ 真阳性检出数 / (真阳性检出数 + 假阴性检出数) ] 的比值计算

b. 指标是指所有样本的平均数值: HiSeq 2500 为 4 个样本, HiSeq 4000 为 2 个样本, HiSeq X 为 8 个样本, NovaSeq 为 6 个样本

c. 构建深度降低至 30×

此外, 与 HiSeq 2500 或 HiSeq 4000 平台相比, 这些运行的成本更低、速度更快、覆盖度更高\*\*。

除了使用 NovaSeq S2 试剂获得的高数据产量, 比对上的 read 百分比数据也显示出在不同平台上的相似性能。在 S2 流动槽上测序的所有样本中, 有百分之九十五的 read (Read 1 和 Read 2 平均) 比对到了参考基因组上。对于发现变异, 所有平台的数据集降低至 30× 覆盖度, 并对所有样本的构建指标进行了分析。NovaSeq S2 流动槽的 30× 构建指标证明性能与 HiSeq X 流动槽相当 (表 1), 并且超过了 HiSeq 2500 和 HiSeq 4000 系统上的变异分析指标 (表 2)。BaseSpace Sequence Hub 提供了此数据集的比对和 read 统计资料。

总体而言, 对 NovaSeq 6000 系统产生的 WGS 数据进行分析, 并与现有的 HiSeq 系统产生的数据对比, 证明了数据质量非常高, 具有始终如一且高度一致的性能。

## 结论

需要大量数据的应用, 如人类 WGS, 如今能够以更加经济高效的方式轻松完成。简单的即插即用操作、方便的仪器内簇生成, 以及数据存储与分析工具的自动无缝整合, 这些都简化了整个实验流程。NovaSeq 6000 系统的出现, 让所有研究人员得以扩展 NGS 的可能性。无论是运行单个系统, 还是运行一个庞大的组合, NovaSeq 系统为各种类型的样本和应用开辟了新途径。凭借精简的操作、出色的样本可扩展性以及支持各种应用的出色灵活性, NovaSeq 6000 系统已成为迄今为止最灵活、强大的高通量 Illumina 测序系统之一。

\*\* Illumina 公司 2017 年存档数据。

## 了解更多

如需了解更多有关 NovaSeq 6000 的信息，请访问 [NovaSeq 系统页面](#)。

如需获取本应用说明中使用的原始数据和变异检出文件，请访问 [BaseSpace Sequence Hub](#) 中的公共数据部分。

## 参考文献

1. Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, et al. [Characterizing and measuring bias in sequence data](#). *Gen Biol*. 2013;14(5):R51.
2. Liu L, Li Y, Li S, et al. [Comparison of next-generation sequencing systems](#). *J Biomed Biotechnol*. 2012: 251364.
3. National Center for Biotechnology Information (NCBI) [Human Genome Build GRCh38](#). Accessed July 2017.
4. Illumina BaseSpace Sequence Hub App. [Whole Genome Sequencing Analysis App v5](#). Accessed July 2017.

## 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](mailto:chinasupport@illumina.com) · [www.illumina.com.cn](http://www.illumina.com.cn)

仅供研究使用。不得用于诊断。

© 2017 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。请访问 [www.illumina.com/company/legal.html](http://www.illumina.com/company/legal.html) 获取具体商标的信息。Pub. No. 770-2017-015-A



因美纳

因美纳讲堂

**illumina**<sup>®</sup>