

# 采用 Illumina DNA PCR-Free Prep, Tagmentation 进行高效的全基因组测序

快速灵活的文库制备，实现卓越的变异检出性能。

## 简介

全基因组测序 (WGS) 是检测人类基因组 32 亿对碱基最全面、无偏向性的方法<sup>1,2</sup>。测序成本的迅速下降和 WGS 快速生成大量数据的能力，使全基因组测序成为许多人类基因组学研究应用的强大工具。在为 WGS 选择文库制备解决方案时，要重点考虑工作流程的性能、成本和可扩展性。Illumina DNA PCR-Free Prep, Tagmentation 是一种先进的解决方案，采用磁珠固化转座酶 (On-Bead Tagmentation) 作为一体化、可扩展且快速的 WGS 工作流程的一部分 (图 1)。

PCR-Free 文库制备被公认为是 WGS 的金标准，与基于 PCR 的方法相比，其在检出许多不同类型的变异时，具有更均一的覆盖度和更出色的性能<sup>3</sup>。这是因为 PCR 偏向性会导致基因组中测序困难的区域覆盖不均，例如碱基组成极不均匀的启动子区域或重复区域 (图 2)。然而，过往的 PCR-Free 文库制备实验方案繁琐且耗时。Illumina DNA PCR-Free Prep, Tagmentation 快速且灵活，同时能够提供 PCR-Free 实验的高质量变异检出性能。

## Illumina DNA PCR-Free Prep 工作流程

Illumina DNA PCR-Free Prep 采用磁珠固化转座酶实现简单、快速的文库制备工作流程<sup>4</sup>。酶切法片段化是一种酶促反应，它将接头连接 (即“加标记”) 和 DNA 片段化整合到了一个快速步骤中，而无需与机械打断相关的昂贵设备和耗材。当 DNA 起始量大于 300 ng 时，磁珠固化转座酶中与磁珠结合的转座酶达到饱和状态，可实现自均一化，从

而各个文库可以等体积混合<sup>5</sup>。然后使用 Qubit 对混合池进行一次定量 (约 10 分钟)，并进行稀释以备测序，减少了与 qPCR 均一化相关的时间和成本。

该工作流程十分快速，约 90 分钟即可完成 16 个样本的文库制备和混合，并且具有可扩展性，提供 24 个 /96 个样本的试剂盒配置，其体积适合进行自动化。由于很少进行板转移，并且不需要手动干预、机械剪切或 qPCR 定量，因此其有望成为完全无人值守的自动化解决方案。此外，在使用饱和工作流程时，Illumina DNA PCR-Free Prep, Tagmentation 能够始终为 NovaSeq™ S2 或 S4 流动槽中的两次标准运行，或 NovaSeq S1 流动槽中的一次标准运行和一次 DFL 运行提供足够的产量。这种重新进行测序运行的选项，为实验室的测试流程提供了更大的灵活性和稳健性。本应用白皮书展示了 Illumina DNA PCR-Free Prep, Tagmentation 的出色性能，其可作为 WGS 应用整体工作流程的一部分。

## 方法

### 文库制备

使用 300 ng 人类参考 DNA (Coriell, 货号 NA12878) 和 Illumina DNA PCR-Free Prep, Tagmentation 工作流程制备文库。生成的文库按体积混合，并上样至 NovaSeq 6000 测序系统进行测序。



图 1: Illumina DNA PCR-Free Prep 工作流程——Illumina DNA PCR-Free Prep, Tagmentation 可用作全面且精益化的人类 WGS 工作流程的一部分，包括文库制备、测序和数据分析。

## 测序

根据所显示的混合样本数（表 1），将制备好的文库混合物上样至 NovaSeq 6000 测序系统的 S4 流动槽中，99% 以上的样本（163/164）覆盖度  $\geq 40\times$ ，其平均运行产量通常高于 NovaSeq 6000 性能规范。通过优化标签组合以及降低目标区域的覆盖度，可增加混合的样本数。

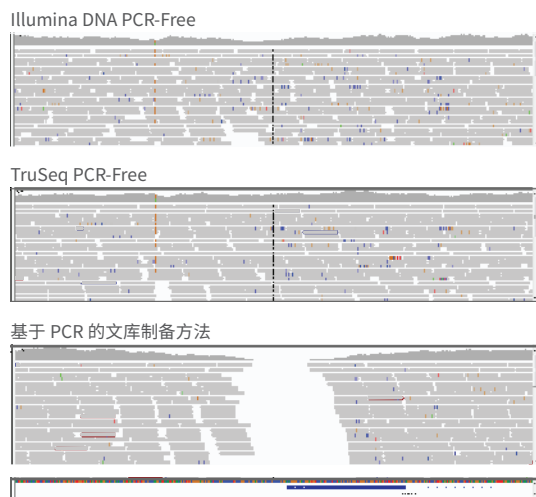


图 2: 高 GC 含量区域的 read 覆盖度比较——与 TruSeq™ DNA PCR-Free 和基于 PCR 的文库制备方案相比，Illumina DNA PCR-Free Prep 在人类 *RNPEPL1* 基因高 GC 启动子区域的 read 覆盖度更高。read 图谱可通过 Integrative Genomics Viewer (IGV, 集成式基因组学浏览器) App 查看，该软件可通过 BaseSpace Sequence Hub 获取。

表 1: 覆盖度指标

指标	S4 流动槽	S2 流动槽	S1 流动槽
样本文库数量	16	6	3
运行次数	8	4	4
每个样本的平均产量 <sup>a</sup>	200.7 Gbp	199.1 Gbp	182.2 Gbp
每个样本的平均 read 数	669.2M	663.6M	607.3M
每次运行的平均产量 <sup>a</sup>	3643 Gbp	1333 Gbp	606 Gbp
平均常染色体覆盖度	45.19	46.26	46.76
唯一双标签对 (UDP, 板 A)	UDP0037-44 + UDP0065-72 UDP0081-96	UDP0081-86 UDP0089-94	UDP0089-91 UDP0081-83
标签变异系数 (CV), 中位数 % (范围)	13.9% (8.4-26)	5.0% (3.3-7.0)	3.7% (1.3-15.3)

a. 平均运行和样本产量均高于 NovaSeq 6000 性能规范。

## 数据分析

测序完成后，数据从仪器直接传输到云端生态系统，使用 BaseSpace™ Sequence Hub 提供的 DRAGEN™ 应用程序进行一键式分析。DRAGEN Germline app v3.5.7 用于与 hg38 参考基因组组装进行比对。使用 Variant Calling Assessment Tool v4.0.1 BaseSpace 应用程序，根据

Platinum Genomes v2017.1 真值集评估单核苷酸变异 (SNV) 和插入 / 缺失 (indel) 检出。将每个样本的变异检出与使用 witty.er tool v0.3.2 和长 read 技术生成的该细胞系的内部真值集进行比较，以评估结构变异 (SV) 检出 (50 bp–10 Kbp 大小范围内的插入缺失)。

## 结果

为了评估 Illumina DNA PCR-Free Prep, Tagmentation 用于 WGS 的性能，将其结果与使用 PCR 文库制备试剂盒生成的文库的数据进行比较。

### 高质量变异检出

与基于 PCR 的实验方案相比，本研究中评估的几乎每个指标性能都得到了改善 (表 2)。这些改善包括对 SNV 和插入缺失的查准率和查全率，以及对可以实现基因型检出的基因组位置比例的评估 (检出率)。

### 均匀覆盖基因组

本研究中对 CNV 性能的评估是有限的。在 NA12878 细胞系中观察到的变异检出很少 (约 60 个 CNV)，并且缺乏评估变异检出的完整、可靠的真值集。在这种有限的评估中，Illumina DNA PCR-Free Prep 和基于 PCR 的文库制备方案之间没有观察到明显的性能差异。然而，与基于 PCR 的实验方案相比，Illumina DNA PCR-Free Prep 的覆盖度明显更加均匀，这通常是预测 CNV 检出性能良好的一个指标 (图 3)。

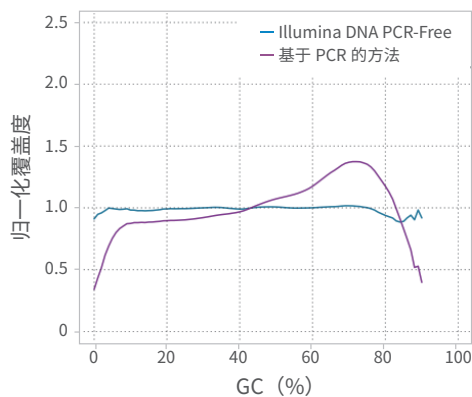


图 3: Illumina DNA PCR-Free 覆盖度均匀性——Illumina DNA PCR-Free 在人类基因组的各种 GC 含量下均可实现更均匀的覆盖。使用 DRAGEN Germline app v3.5.7 计算归一化覆盖度数据。

### 高度准确的 STR 检测

我们知道，基于 PCR 的方案无法全面检测富含 GC 的 STR 扩增，这类基因组片段的 read 几乎不会出现在数据集中。在一项研究中评估了 Illumina DNA PCR-Free Prep 在检测 STR 扩增方面的性能，该研究对含有 9 种不同致病 STR 扩增的 16 个不同细胞系进行了一式两份或四份的制备。通过 8 次 NovaSeq S4 运行对文库进行测序。Illumina

表 2: Illumina DNA PCR-Free Prep 变异检出性能

	Illumina DNA PCR-Free (平均值, n=20)	基于 PCR 的文库制备 (平均值, n=4)	差值 (PCR-Free 减去 PCR)
常染色体基因检出率	97.54%	97.32%	0.21%
常染色体中无法检出的碱基 (bp)	67,564,679	73,463,941	-5,899,261
常染色体编码外显子检出率	99.06%	98.914%	0.16%
常染色体编码外显子中无法检出的碱基 (bp)	571,201	666,716	-95,514
SNV 查全率	99.37%	99.31%	0.06%
假阴性 SNV	22,753	24,835	-2081
SNV 查准率	99.82%	99.82%	0.00%
假阳性 SNV	6468	6469	-1
插入缺失查全率	98.74%	97.96%	0.77%
假阴性插入缺失	7242	11,671	-4429
插入缺失查准率	97.99%	97.75%	0.24%
假阳性插入缺失	11,623	12,932	-1308
缺失 (50 bp-10 Kbp) 查全率	40.40%	39.29%	1.1%
缺失 (50 bp-10 Kbp) 查准率	85.82%	86.67%	-0.8%
插入 (50 bp-10 Kbp) 查全率	35.13%	31.67%	3.46%
插入 (50 bp-10 Kbp) 查准率	94.35%	94.42%	-0.07%

变异检出的差异用颜色标出, Illumina DNA PCR-Free 的性能改善以绿色突出显示。

表 3: Illumina DNA PCR-Free Prep STR 扩增准确性

细胞系	具有 STR 扩增的基因位点	STR 基序	疾病关联	扩增阈值 (重复单元)	检测到扩增的重复实验次数
ND12161	<i>C9ORF72</i>	CCCCGG	肌萎缩性脊髓侧索硬化症 1 型 (ALS1)	≥ 31	2/2
ND11917	<i>C9ORF72</i>	CCCCGG		≥ 31	2/2
NA23709	<i>AR</i>	CAG	X-连锁脊髓延髓肌萎缩症 (SMAX1)	≥ 36	4/4
NA13716	<i>ATN1</i>	CAG	脊髓小脑性共济失调 1 型	≥ 36	2/2
NA13717	<i>ATN1</i>	CAG		≥ 36	2/2
NA06151	<i>ATXN3</i>	CAG	马查多-约瑟夫病 (MJD)	≥ 45	4/4
NA04567	<i>DMPK</i>	CAG	肌强直性营养不良 (DM1)	≥ 35	2/2
NA04648	<i>DMPK</i>	CAG		≥ 35	2/2
NA04025	<i>FMR1</i>	CGG	脆性 X 综合征 (FXS)	≥ 55	2/2
NA07862	<i>FMR1</i>	CGG		≥ 55	2/2
NA03816	<i>FXN</i>	GAA	弗里德赖希共济失调 1 型 (FRDA)	≥ 66	2/2
NA15850	<i>FXN</i>	GAA		≥ 66	2/2
NA13507	<i>HTT</i>	CAG	亨廷顿病 (HD)	≥ 36	2/2
NA13509	<i>HTT</i>	CAG		≥ 36	2/2
NA13536	<i>ATXN1</i>	CAG	脊髓小脑性共济失调 1 型 (SCA1 ATAXIN1, ATX1)	≥ 36	2/2
NA13537	<i>ATXN1</i>	CAG		≥ 36	2/2

DNA PCR-Free Prep 的 STR 扩增检测灵敏度为 100% (表 3)。该研究包括在 *C9ORF72* 基因位点扩增的两种细胞系 (ND12161 和 ND11917) 和在 *FMR1* 基因位点扩增的两种细胞系 (NA04025 和 NA07862)。基于 PCR 的方法无法检测这些较长的 GC 扩增;然而,使用 Illumina DNA PCR-Free Prep 可以实现准确检测 (表 3)。该研究包

括来自 19 个不同个体的细胞系和血液样本的其他 72 个 Illumina DNA PCR-Free 文库, 之前没有证据表明, 这些个体在 DRAGEN Germline Pipeline 靶向的基因位点中存在致病性 STR 扩增。在这 72 个样本中未检测到假阳性扩增 (特异性 100%) (数据未显示)。

## 总结

Illumina DNA PCR-Free Prep, Tagmentation 是一种先进的文库制备解决方案，采用磁珠固化转座酶作为一体化、可扩展且快速的 WGS 工作流程的一部分。文库制备与该实验方案的全自动化选项兼容。对于受监管或进行大规模操作的实验室，可以集成 BaseSpace Clarity LIMS，帮助追踪样本和管理工作流程，从而优化实验室操作，提高效率。Illumina DNA PCR-Free Prep 文库可生成高质量的测序数据，实现均匀覆盖，并在变异检出方面提供出色性能。

## 了解更多

如需了解有关 Illumina DNA PCR-Free Prep, Tagmentation 的更多信息，请访问 [www.illumina.com/products/by-type/sequencing-kits/library-prep-kits/dna-pcr-free-prep.html](http://www.illumina.com/products/by-type/sequencing-kits/library-prep-kits/dna-pcr-free-prep.html)

## 参考文献

1. Wang K, Yuen ST, Xu J, et al. Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nat Genet.* 2014;46(6):573-582.
2. Zahir FR, Mwenifumbo JC, Chun HE, et al. Comprehensive whole genome sequence analyses yields novel genetic and structural insights for Intellectual Disability. *BMC Genomics.* 2017;18(1):403.
3. Fungtammasan A, Ananda G, Hile SE, et al. [Accurate typing of short tandem repeats from genome-wide sequencing data and its applications.](#) *Genome Res.* 2015;25(5):736-749.
4. Illumina (2018). [Nextera DNA Flex Library Preparation Kit.](#) Accessed April 10, 2020.
5. Bruinsma S, Burgess J, Schlingman D, Czyz A, Morrell N, et al. [Bead-linked transposomes enable a normalization-free workflow for NGS library preparation.](#) *BMC Genomics.* 2018;19(1):722.

## 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)

© 2021 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](http://www.illumina.com/company/legal.html)。770-2020-006-A QB10058

仅供研究使用。不得用于诊断。



因美纳 因美纳讲堂

**illumina**<sup>®</sup>