

illumina Complete Long Read Prep, Human을 이용한 포괄적인 전장 유전체 시퀀싱

매핑이 어려운 영역에서
뛰어난 성능 제공



소개

차세대 시퀀싱(Next-generation sequencing, NGS)은 과학자에게 유전체(genome)를 해독하여 생명 활동에 대한 이해의 폭을 넓힐 수 있는 기회를 제공합니다. Illumina의 입증된 sequencing by synthesis(SBS) chemistry와 수상 경력이 있는 DRAGEN™ Secondary Analysis가 만나 정확도가 매우 높은 전장 유전체 시퀀싱(whole-genome sequencing, WGS) 데이터를 생성합니다.¹ 그중에서도 극히 일부의 분석이 어려운 유전체 영역에는 긴 리드 길이(read length)를 적용하면 높은 정확도와 해상도로 매핑이 가능하므로 한결 더 좋은 결과를 기대할 수 있습니다. 그러나 다수의 롱 리드(long read) 시퀀싱 솔루션은 워크플로우가 복잡하며, 많은 양의 DNA 사용을 요구하고, 결과의 변동성이 매우 심하다는 단점이 있습니다.²⁻⁵

Illumina Complete Long Reads 제품은 유전체 연구자가 롱 리드 시퀀싱을 간결하고 손쉽게 적용할 수 있도록 해 줍니다. Illumina Complete Long Read Prep, Human은 이러한 새로운 롱 리드 chemistry를 기반으로 한 최초의 제품입니다. 이 고성능 WGS assay는 일반적인 NGS 워크플로우를 사용하여 NovaSeq™ 6000 시스템이나 NovaSeq X 시리즈에서 contiguous 롱 리드 시퀀싱을 생성합니다(그림 1). 하루 안에 완료 가능한 효율적인 라이브러리 준비 프로토콜은 높은 처리량(throughput)을 요구하는 연구에 맞춰 손쉽게 확장이 가능하며, 최소 10 ng의 DNA만을 사용할 수 있고, 전문적인 추출(extraction), 절단(shearing), 크기 선별(size selection) 과정이 필요하지 않습니다.

본 Technical Note는 다양한 DNA 사용량, DNA 품질 및 샘플 타입에서도 생식세포(germline) 변이 분석을 위한 정확도 높고 포괄적인 WGS를 가능케 하는 Illumina Complete Long Read Prep, Human의 뛰어난 성능을 기술하고 있습니다.

방법

Genome in a Bottle(GIAB) 컨소시엄을 통해 유전체 DNA(genomic DNA, gDNA)인 HG002/NA24385 참조 샘플을 얻었습니다. 별도로 혈액 및 타액 샘플을 채취하여 일반적인 키트 또는 고분자량(high molecular weight, HMW) 추출이 가능한 키트를 사용해 제조사의 지침에 따라 gDNA를 추출했습니다.

라이브러리 준비

Illumina Complete Long Read Prep, Human(Illumina, 카탈로그 번호: 20089108)으로 다양한 양의 gDNA를 사용해 "land-mark" 라이브러리를 준비했습니다. 간략하게 말하자면, 긴 단일 분자(single-molecule) DNA 절편을 엔자임(enzyme)을 이용해 고유한 패턴(즉 land-mark)으로 마크하는 과정을 거칩니다. 원래의 단일 분자 절편을 나타내는 긴 contiguous 리드를 생성하기 위해 분석 중에 사용되는 unmarked 라이브러리는 Illumina DNA PCR-Free Prep, Tagmentation(Illumina, 카탈로그 번호: 20041794)을 사용해 준비했습니다.

시퀀싱

Land-mark된 라이브러리와 unmarked 라이브러리는 NovaSeq 6000 시스템에서 개별적인 런(run)을 통해 2 × 150 bp의 리드 길이로 시퀀싱했습니다.

데이터 분석

Land-mark된 리드의 어셈블리(assembly) 및 렌더링(rendering)은 BaseSpace™ Sequence Hub에서 Illumina Complete Long Read WGS 앱을 통해 진행했습니다. 데이터 분석 및 시각화에는 Integrative Genomics Viewer(IGV)를 사용했습니다.

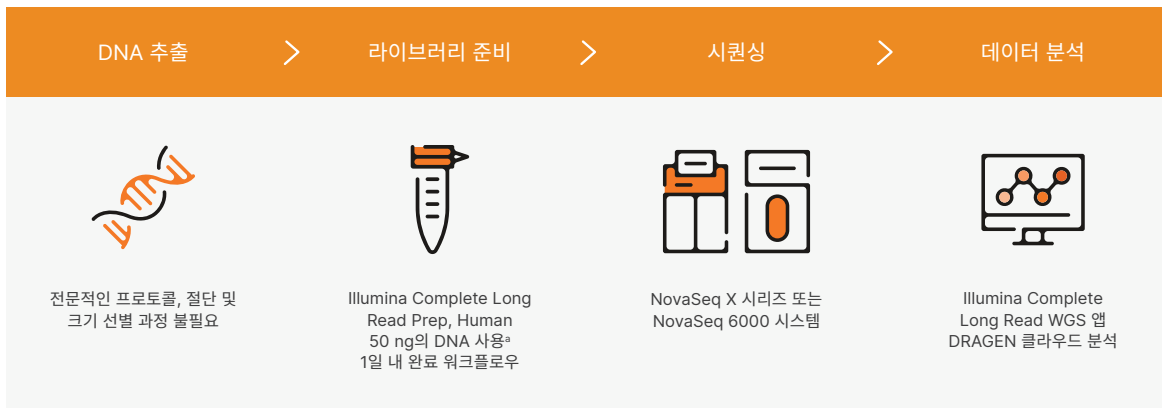


그림 1: Illumina Complete Long Read Prep, Human 워크플로우 — DNA 추출, 라이브러리 준비, 시퀀싱, 데이터 분석의 네 단계로 구성된 간소한 워크플로우를 제공하는 Illumina Complete Long Read Prep, Human

a. 최소 10 ng으로 사용 가능하나, 50 ng의 DNA 사용이 권장됨.

문서 번호: M-KR-00239 KOR

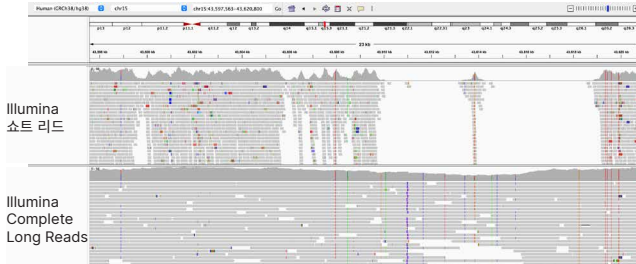
2 | 원본 문서: M-GL-01421 v2.0

연구 전용입니다. 진단 절차에는 사용할 수 없습니다.

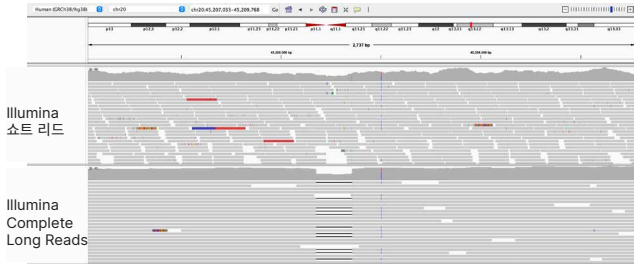
분석이 어려운 영역에서 고품질의 데이터 제공

ILLUMINA COMPLETE LONG READ PREP, HUMAN은 쇼트 리드(short read)가 매핑될 수 없는 영역에서도 명확한 커버리지를 제공합니다. 이를 토대로 *STRC*(그림 2A), *SEMG1*(그림 2B), *SULTIA1*(그림 2C) 등 쇼트 리드만으로는 매핑 시 문제가 있는 것으로 알려진 질병 관련 유전자 및 유사 유전자(pseudogene)의 contiguous 시퀀싱이 가능해집니다.

A. STRC 유전자



B. SEMG1 유전자



C. SULTIA1 유전자



그림 2: 분석이 어려운 영역에서 다양한 변이의 정확한 검출을 지원하는 포괄적인 커버리지 — ILLUMINA COMPLETE LONG READ PREP, HUMAN으로 (A) *STRC*, (B) *SEMG1*, (C) *SULTIA1* 등 분석이 어려운 유전체 영역에서 질병 관련 유전자의 정확한 시퀀싱이 가능함.

또한 ILLUMINA COMPLETE LONG READ PREP, HUMAN은 일반적인 유전자 크기보다 훨씬 더 큰 페이즈 블록(phase block)을 제공합니다. 따라서 연구자는 높은 레벨의 시퀀스 상동성(homology), 유사 유전자, 변동성 및 수많은 대립유전자(allele)가 밀집되어 존재하는 다형적(polymorphic) 영역인 인간 백혈구 항원(human leukocyte antigen, HLA) 유전자좌위(locus) 전체에서 균일한 커버리지를 확인할 수 있습니다. ILLUMINA COMPLETE LONG READ PREP, HUMAN은 이처럼 정확한 시퀀싱이 어려운 HLA 영역에서의 균일한 커버리지와 유전체 페이징(phasing)을 지원함으로써 모계 염색체와 부계 염색체 간의 대립유전자를 명확하게 구분하고 이를 시각화할 수 있습니다(그림 3).

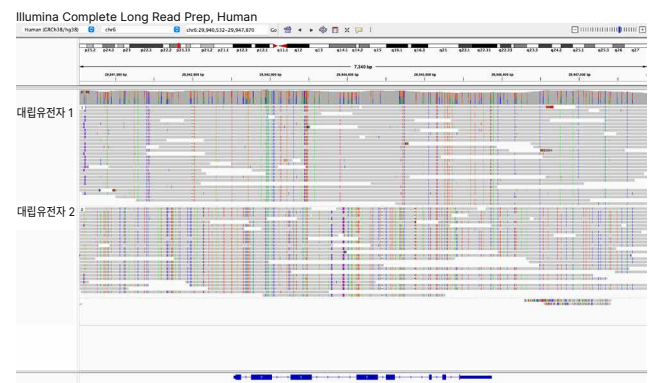


그림 3: 명확한 하플로타입(haplotype) 페이징을 실현하는 균일한 커버리지 — ILLUMINA COMPLETE LONG READ PREP, HUMAN은 HLA 영역에서 포괄적인 커버리지를 제공하며 HLA 영역 내 명확하고 정확한 대립유전자 페이징을 지원함.

더 종합적인 변이 검출 정확도 평가를 위해 ILLUMINA LONG READ PREP, HUMAN을 precisionFDA Truth Challenge V2에 제출된 ILLUMINA의 쇼트 리드 및 다른 롱 리드 플랫폼과 비교했습니다. precisionFDA, GIAB 컨소시엄 그리고 미국 국립 표준 기술 연구소(National Institute of Standards and Technology, NIST)의 주최로 진행된 이 챌린지는 매핑이 어려운(difficult-to-map) 영역에 초점을 두고 일반적인 레퍼런스 기준에 대한 작은 변이 검출 파이프라인의 성능을 확인하기 위해 개최되었습니다. 이 데이터 세트를 활용했을 때 ILLUMINA COMPLETE LONG READ PREP, HUMAN은 99.90%의 F1 점수를 기록하여 다른 롱 리드 플랫폼보다 우수한 정확도를 보였습니다(그림 4).

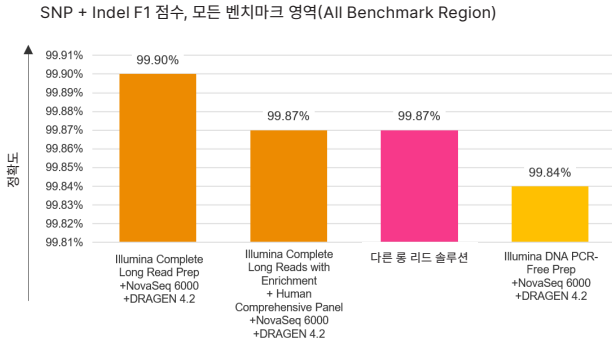


그림 4: 정확도의 새로운 기준 — WGS에 대한 정밀도(precision) 및 재현율(recall)을 반영하는 F1 점수(%)를 기준으로 볼 때, Illumina Complete Long Read Prep, Human은 훨씬 더 우수한 변이 검출 정확도를 보임.¹

다양한 양의 DNA 사용 시 확인된 우수한 성능

Illumina Complete Long Read Prep, Human의 성능을 다양한 양의 DNA를 사용해 평가했습니다. 5 ng, 10 ng, 25 ng, 50 ng, 100 ng, 600 ng 및 1,200 ng의 DNA로 라이브러리를 준비해 3회 반복 실험을 진행했습니다. 시퀀싱 결과, 최소 5 ng의 DNA를 사용해 준비한 라이브러리에서도 N50, 페이징 블록 N50 등 고품질의 매트릭스(metrics)를 확인할 수 있었습니다(그림 5). Illumina는 50 ng의 DNA 사용을 권장하나, 이보다 적은 양으로도 사용 가능합니다.

다양한 품질의 DNA 지원

손상된 DNA를 모방하기 위해 Covaris ultrasonicator를 사용해 양질의 gDNA를 1초 또는 3초 동안 절단했습니다. 이후 Fragment Analyzer를 사용해 DNA 품질을 평가했습니다(그림 6). 라이브러리는 절단된 gDNA와 Illumina Complete Long Read Prep, Human을 사용해 준비했습니다. 시퀀싱 결과, 1초간 절단한 DNA로 고품질의 데이터가 생성되었으며 정확한 변이 검출이 가능했습니다(그림 7). 또한 동결 및 해동 횟수가 많은 gDNA에서도 고품질의 데이터가 생성되었습니다(그림 8).

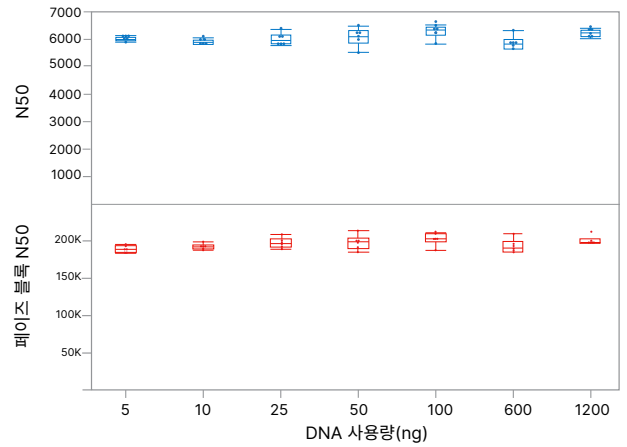


그림 5: 다양한 양의 DNA 사용 시 확인된 고성능 — Illumina Complete Long Read Prep, Human은 5~1,200 ng(3회 반복 실험)의 DNA로부터 제작된 양질의 라이브러리에서 비슷한 품질의 N50 및 페이징 블록 N50 데이터를 생성함. N50이란 전체 어셈블리 길이의 50%일 때 가장 짧은 콘티그(contig)의 시퀀스 길이로 정의됨. N50은 페이징 블록을 측정하는 척도로도 사용 가능.

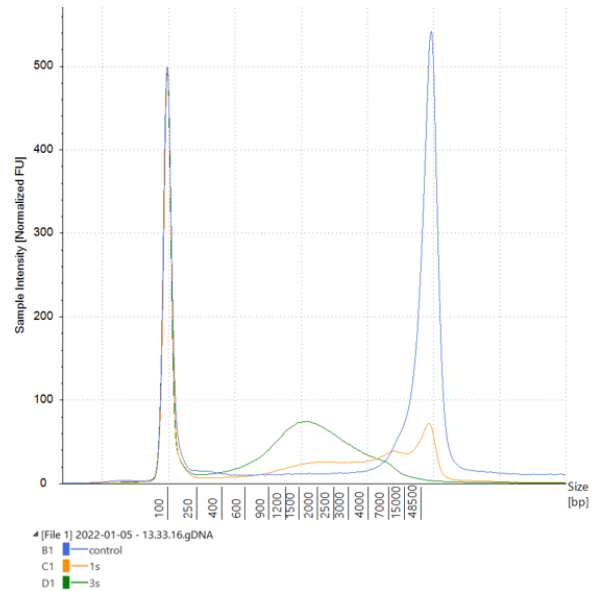


그림 6: 절단된 DNA의 품질 — Fragment Analyzer의 DNA 분석 결과, 절단 시간이 길수록 품질의 저하가 관찰됨. Covaris로 DNA 샘플을 1초 또는 3초 동안 절단한 후 Fragment Analyzer로 품질을 측정함.

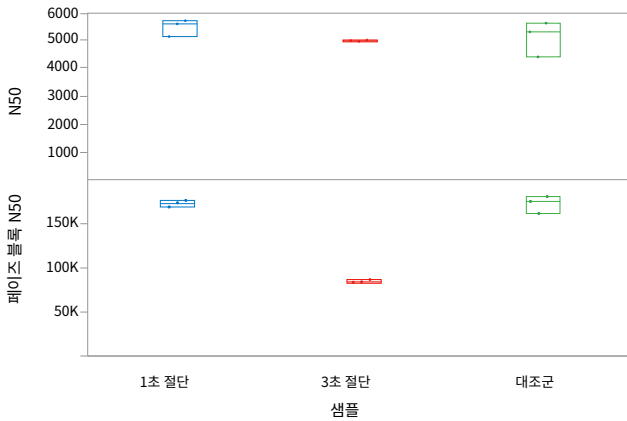


그림 7: 절단된 DNA의 시퀀싱 매트릭스 — 1초 및 3초 동안 절단한 DNA를 사용해 Illumina Complete Long Read Prep, Human으로 준비한 라이브러리는 N50 및 페이지 블록 N50을 기준으로 측정했을 때 고품질의 롱 리드 시퀀싱 데이터를 생성함.

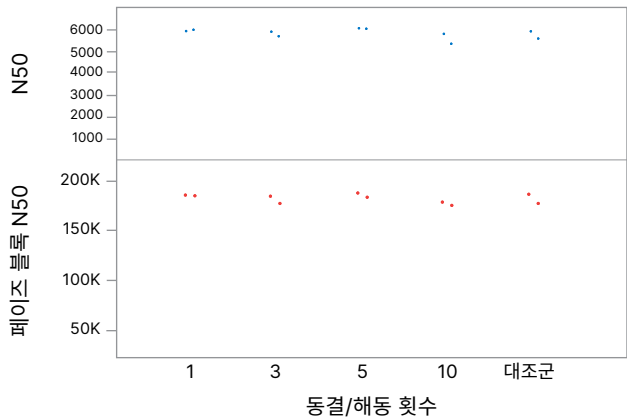


그림 8: 동결 및 해동한 DNA의 시퀀싱 매트릭스 — 1회, 3회, 5회 및 10회의 동결 및 해동 과정을 거친 DNA를 사용해 Illumina Complete Long Read Prep, Human으로 준비한 라이브러리는 N50 및 페이지 블록 N50을 기준으로 측정했을 때 고품질의 롱 리드 시퀀싱 데이터를 생성함.

다양한 샘플 타입을 지원하는 유연성

Illumina Complete Long Read Prep, Human은 혈액과 타액을 포함한 다양한 샘플 타입을 지원합니다. 일반적인 키트 또는 HMW 키트를 사용해 추출한 DNA의 품질을 Fragment Analyzer를 사용해 평가했습니다(그림 9). 시퀀싱 결과, 혈액과 타액에서 추출한 DNA로 고품질의 데이터가 생성되었습니다(그림 10). 흥미롭게도 일반적인 키트로 추출한 DNA와 HMW 키트로 추출한 DNA의 성능에는 별다른 차이가 확인되지 않았습니다(그림 10).

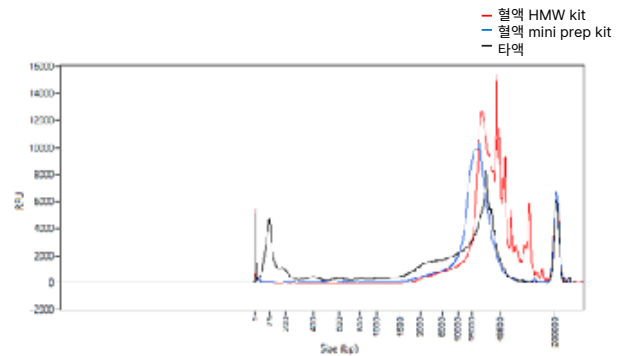


그림 9: 혈액 및 타액에서 추출한 DNA의 품질 — 혈액(HMW, mini) 또는 타액에서 추출한 DNA를 Fragment Analyzer로 분석한 결과

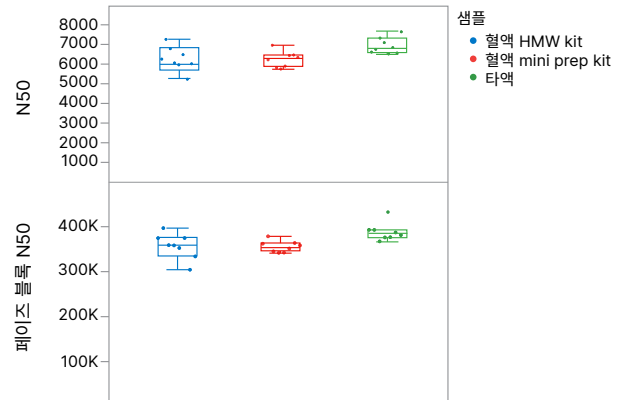


그림 10: 다양한 샘플 타입 사용 시 확인된 고성능 — N50 및 페이지 블록 N50을 기준으로 측정했을 때 혈액(HMW, mini) 또는 타액을 사용해 Illumina Complete Long Read Prep, Human으로 준비한 라이브러리는 고품질의 데이터를 생성함.

요약

롱 리드 시퀀싱 chemistry는 쇼트 리드 NGS로는 분석하기 어려운 유전체 영역에 대한 추가적인 정보를 제공합니다. Illumina Complete Long Reads 제품은 유전체 연구자가 롱 리드 시퀀싱을 간결하고 손쉽게 적용할 수 있도록 해 줍니다. Illumina Complete Long Read Prep, Human은 이러한 새로운 chemistry를 기반으로 최초로 개발한 제품으로, 간소화되고 익숙한 워크플로우를 제공하며 Illumina의 입증된 SBS chemistry 및 DRAGEN 분석 기능과 함께 시너지 효과를 발휘합니다. 이 혁신적인 chemistry를 통해 한 대의 기기에서 쇼트 리드 및 롱 리드 시퀀싱을 수행할 수 있으므로 유전체 랩에서는 원활하게 롱 리드 NGS를 도입할 수 있습니다. Illumina Complete Long Read Prep, Human은 여러 가지 체액 샘플에서 추출한 품질 및 양이 다양한 DNA를 사용했을 때 전반적으로 우수한 성능을 보입니다. Illumina Complete Long Read Prep, Human은 뛰어난 확장성과 정확성을 갖춘 인간 WGS 솔루션입니다.

상세 정보

[Illumina Complete Long Read Prep, Human](#)

참고 문헌

1. Mehio R, Ruehle M, Catreux S, et al. DRAGEN Wins at Precision- FDA Truth Challenge V2 Showcase Accuracy Gains from Alt-aware Mapping and Graph Reference Genomes. [illumina.com/science/genomics-research/dragen-wins-precisionfda-challenge-showcase-accuracy-gains.html](https://www.illumina.com/science/genomics-research/dragen-wins-precisionfda-challenge-showcase-accuracy-gains.html). Accessed January 12, 2023.
2. Pacific Biosciences. Preparing DNA for PacBio HiFi sequencing—Extraction and quality control. [pacb.com/wp-content/uploads/Technical-Note-Preparing-DNA-for-PacBio-HiFi-Sequencing-Extraction-and-Quality-Control.pdf](https://www.pacb.com/wp-content/uploads/Technical-Note-Preparing-DNA-for-PacBio-HiFi-Sequencing-Extraction-and-Quality-Control.pdf). Published 2022. Accessed January 12, 2023.
3. Pacific Biosciences. Preparing whole genome and metagenome libraries using SMRTbell prep kit 3.0. [pacb.com/wp-content/uploads/Procedure-checklist-Preparing-whole-genome-and-metagenome-libraries-using-SMRTbell-prep-kit-3.0.pdf](https://www.pacb.com/wp-content/uploads/Procedure-checklist-Preparing-whole-genome-and-metagenome-libraries-using-SMRTbell-prep-kit-3.0.pdf). Published 2022. Accessed January 12, 2023.
4. Oxford Nanopore Technologies. Ligation Sequencing Kit. [store.nanoporetech.com/us/ligation-sequencing-kit110.html](https://www.store.nanoporetech.com/us/ligation-sequencing-kit110.html). Accessed January 12, 2023.
5. Pacific Biosciences. Low Yield Troubleshooting Guide. [pacb.com/wp-content/uploads/Guide-Low-Yield-Troubleshooting.pdf](https://www.pacb.com/wp-content/uploads/Guide-Low-Yield-Troubleshooting.pdf). Published 2018. Accessed January 12, 2023.



무료 전화(한국) 080-234-5300
techsupport@illumina.com | www.illumina.com

© 2024 Illumina, Inc. All rights reserved.
모든 상표는 Illumina, Inc. 또는 각 소유주의 자산입니다.
특정 상표 정보는 www.illumina.com/company/legal.html을 참조하십시오.
문서 번호: M-KR-00239 KOR