

# Illumina Human Comprehensive Panel

단백질 코딩 유전자에서  
분석이 어려운 영역에 대한  
향상된 커버리지를 제공하는  
Illumina Complete Long Read  
Prep with Enrichment, Human



## 소개

전장 유전체 시퀀싱(Whole-genome sequencing, WGS)을 수행할 때 쇼트 리드(short read)만으로는 매핑하기 어려운 유전자 영역이 일부 있을 수 있습니다. 이러한 분석이 어려운 영역으로는 상동성이 높은 영역(homologous region)이나 반복적인 영역(repetitive region), 복잡한 구조적 변이(structural variant, SV), 유사 유전자(pseudogene), 큰 삽입/결실(insertion/deletion, Indel) 등이 있습니다. 이 경우 롱 리드(long read) 시퀀싱으로 표준 쇼트 리드 WGS 데이터를 상호 보완하면 커버리지를 향상시켜 분석이 어려운 영역을 분석하고, 보다 간편하게 변이를 페이징(phasing)하고 하플로타입(haplotype)을 검출할 수 있습니다.

Illumina Complete Long Reads 기술은 일반적인 차세대 시퀀싱(next-generation sequencing, NGS) 워크플로우를 이용해 Illumina 시퀀싱 시스템에서 contiguous 롱 리드 시퀀스를 생성합니다(그림 1). Illumina Complete Long Read Prep with Enrichment 및 Human Comprehensive Panel은 비용 대비 효과적이며 확장이 가능한 솔루션을 제공하므로 연구자가 분석이 어려운 것으로 알려진 유전체 영역을 분석할 수 있도록 해 줍니다. 표적 롱 리드가 정확도가 높은 표준 WGS 데이터와 함께 분석되어 더 많은 영역을 매핑할 수 있게 됩니다.

### 포괄적이고 최적화된 프로브 패널

Illumina Human Comprehensive Panel은 최적화된 하이브리드 캡처(hybrid-capture) 프로브 패널로, 더 긴 리드의 높은 매핑률(mappability)로 검출력이 향상될 수 있는 유전자 영역을 중심으로 설계되었습니다. 이 패널은 6,500개가 넘는 단백질 코딩 유전자(protein-coding gene)에 걸쳐 매핑률이 낮은 영역을 표적화하여(표 1)<sup>1</sup> 표준 쇼트 리드 WGS로는 매핑이 어려운 영역의 해상도를 높여 줄 상호 보완적인 롱 리드를 지원합니다.

표 1: Human Comprehensive Panel의 파라미터<sup>a</sup>

시스템	NovaSeq™ X 시리즈 NovaSeq 6000 시스템
표적 영역 크기	> 95 Mb
샘플당 데이터 아웃풋 <sup>b</sup>	90~120 Gb
프로브 수	약 40,000개
표적 유전자 수	> 6,500개
샘플 종류	유전체 DNA (Genomic DNA, gDNA)
관장 DNA 사용량	50 ng
총 라이브러리 준비 시간	약 2일
수작업 시간	약 6시간
멀티플렉싱(Multiplexing)	최대 64개의 샘플
N50	6.3 kb
페이지 블록(Phase block) N50	15.6 kb
균일성(Uniformity) <sup>c</sup>	95%
패딩된 리드 인리치먼트(PRE) <sup>d</sup>	83%
페이징된 이질접합(heterozygous) SNV의 % <sup>e</sup>	98%

- a. 50 ng의 HG002 gDNA(Coriell, 카탈로그 번호: NA24385)를 사용해 생성된 데이터. 실제 성능은 DNA 사용량과 샘플의 품질에 따라 상이할 수 있음.
- b. 2 × 150 bp 시퀀싱 런 조건에서 약 30x의 최종 Illumina Complete Long Reads 커버리지 확보.
- c. % > 0.2 \* 평균값으로 계산된 커버리지 균일성
- d. 패딩된 리드 인리치먼트(Padded read enrichment, PRE) 값은 100 \* (패딩된 표적 정렬된 리드의 수/총 정렬된 리드의 수)로 계산됨.
- e. SNV = single nucleotide variant(단일 염기서열 변이)

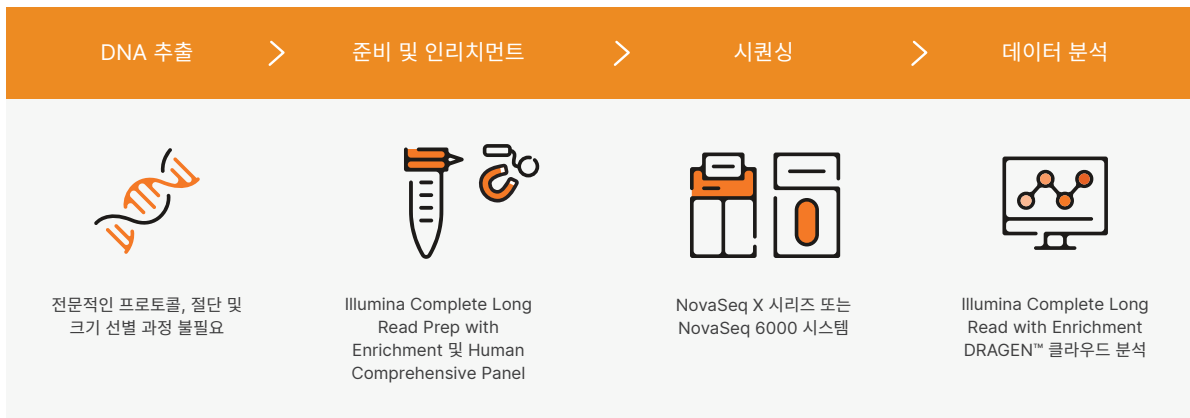


그림 1: 통합 워크플로우의 한 구성 요소 — 연구 규모에 따라 조정이 가능한 최적화된 라이브러리 준비 및 인리치먼트 프로토콜, 입증된 Illumina의 시퀀싱 chemistry 및 DRAGEN Secondary Analysis를 활용해 비용 대비 효과적으로 표적 롱 리드 WGS 데이터를 생성하는 워크플로우. 분석 시 동일한 샘플로 얻은 ≥ 30x 표준 쇼트 리드 WGS 데이터 필요. 이전 런 샘플로 생성한 FASTQ 파일 사용 가능.

ILLUMINA는 이 패널의 개발 단계에서 인트론(intron)과 비번역 영역(untranslated region, UTR)을 비롯한 20,000개가 넘는 단백질 코딩 유전자를 모두 고려해 이러한 영역에서 매핑이 어려운 염기쌍(base pair, bp) 중 70%를 표적화했습니다. 이때 쇼트 리드만으로도 포괄적으로 커버되는 유전자는 제외했습니다. ILLUMINA Human Comprehensive Panel은 연구자가 패널 디자인, 시퀀싱 효율성 및 성능을 최대한 활용할 수 있도록 실제 젯 랩(wet lab)을 통한 검증 및 최적화 절차를 거쳤습니다.

### 확장성이 뛰어난 시퀀싱 및 분석 역량

ILLUMINA Complete Long Read Prep with Enrichment 워크플로우는 확장성이 뛰어나고 자동화가 쉬워 더 많은 샘플을 처리해야 하는 포괄적인 WGS 연구를 지원할 수 있습니다(표 2). NovaSeq X Plus 시스템은 연간 최대 15,000개의 정확도 높은 유전체를 생성할 수 있습니다.\* 요구되는 배치(batch)별 샘플 수를 줄이려는 경우에는 NovaSeq X 1.5B 플로우 셀과 같이 처리량이 낮은 소모품을 사용해 더 적은 수의 샘플을 시퀀싱하면 됩니다. ILLUMINA Complete Long Read Prep with Enrichment 및 Human Comprehensive Panel은 보다 광범위한 변이 검출이 필요할 때 기존의 WGS 데이터 세트를 상호 보완하는 추가 도구로 활용할 수 있습니다.

ILLUMINA Complete Long Read Prep with Enrichment, Human 데이터는 BaseSpace™ Sequence Hub 앱 또는 ILLUMINA Connected Analytics를 통해 분석할 수 있습니다. 데이터는 클라우드로 업로드된 후 동일한 샘플로 얻은 ≥ 30X 표준 쇼트 리드 WGS 데이터(이전 또는 동시에 생성)와 함께 분석됩니다. DRAGEN 파이프라인은 이러한 결과를 DRAGEN targeted caller를 포함해 하나의 결과 파일 세트로 병합합니다.<sup>3</sup>

### 비용 대비 효과적이며 향상된 커버리지

ILLUMINA Complete Long Read Prep with Enrichment 및 Human Comprehensive Panel은 표준 쇼트 리드 WGS와 비교했을 때 변이 검출 정확도를 높이고 표적 영역 전체에 걸쳐 커버리지를 향상시킵니다(그림 2, 그림 3, 표 3). Human Comprehensive Panel 사용 시 표적 롱 리드는 F1 점수(SNV + Indel)가 99.87%인 포괄적인 전장 유전체 정확도를 달성하여,<sup>2</sup> 쇼트 리드만을 사용했을 때보다 훨씬 높은 점수를 기록했습니다. 또 ILLUMINA Complete Long Read Prep, Human 및 단백질 코딩 유전자의 다른 롱 리드 전장 유전체와 비슷한 수준의 성능을 제공하면서 비용은 낮추고 처리량은 높여줍니다(표 3).

\* ILLUMINA Complete Long Read Prep with Enrichment 및 Human Comprehensive Panel을 NovaSeq X Plus 시스템 및 25B 플로우 셀과 함께 사용해 듀얼 플로우 셀 런(dual flow cell run)을 수행한 경우 달성 가능한 처리량

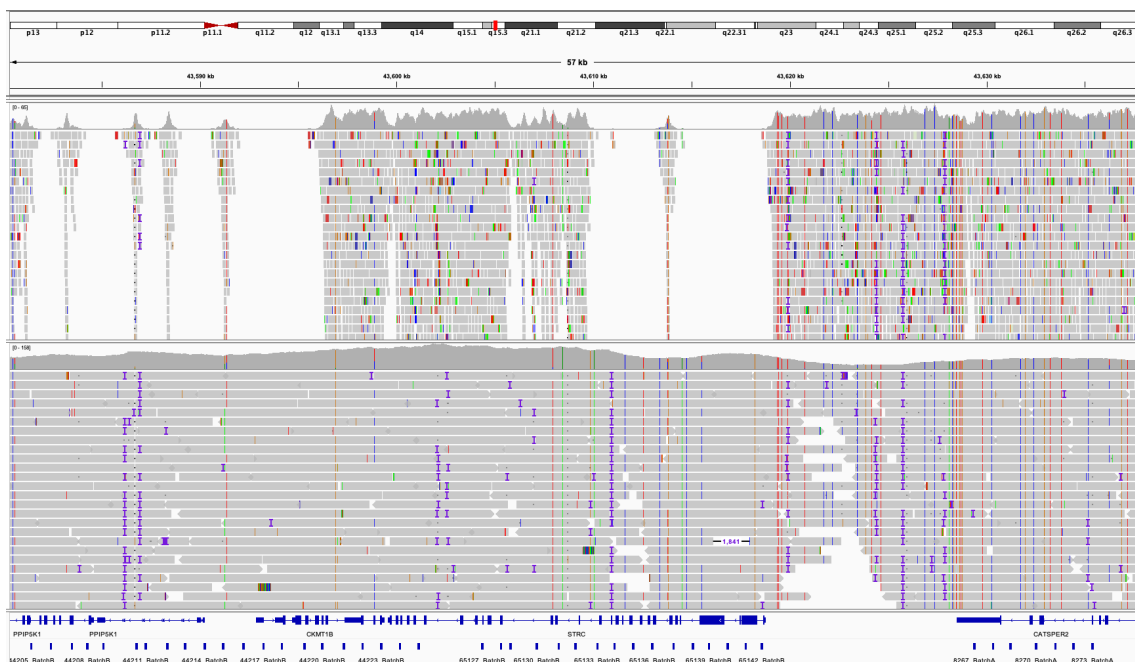
표 2: 30x의 최종 Human Comprehensive Panel 및 ILLUMINA Complete Long Read Prep with Enrichment, Human 커버리지에 권장되는 샘플 처리량<sup>a,b,c,d</sup>

	NovaSeq 6000				NovaSeq X		
	SP	S1	S2	S4	1.5B	10B	25B
300사이클 시약 키트	SP	S1	S2	S4	1.5B	10B	25B
플로우 셀당 처리 샘플 수	2개	4개	10개	24개	4개	24개	64개
플로우 셀당 데이터 아웃풋	약 250 Gb	약 500 Gb	약 1.25 Tb	약 3 Tb	약 500 Gb	약 3 Tb	약 8 Tb
런 타임	약 25시간	약 25시간	약 36시간	약 44시간	약 21시간	약 25시간	약 48시간

a. Human Comprehensive Panel의 표적 영역 크기는 > 95 Mb이며 샘플당 90~120 Gb의 시퀀싱 데이터 아웃풋이 요구됨.  
 b. 2 × 150 bp 시퀀싱 런 조건에서 약 30x의 최종 ILLUMINA Complete Long Reads 커버리지 확보.  
 c. 분석 시 동일한 샘플로 얻은 30x 표준 쇼트 리드 인간 전장 유전체 데이터 필요. ILLUMINA DNA PCR-Free Prep 사용 권장. 타사 WGS 키트도 호환 가능. Unmarked 라이브러리의 준비 또는 시퀀싱은 동시에 진행될 필요 없음. 이전 런 샘플로 생성한 FASTQ 파일 사용 가능.  
 d. ILLUMINA Complete Long Read 라이브러리를 NovaSeq 플랫폼에서 시퀀싱할 경우 런의 Q30 점수가 NovaSeq 사양보다 낮게 보고될 수 있으나, 이는 시퀀싱 런이나 라이브러리의 성능 문제를 의미하지 않음.

A.

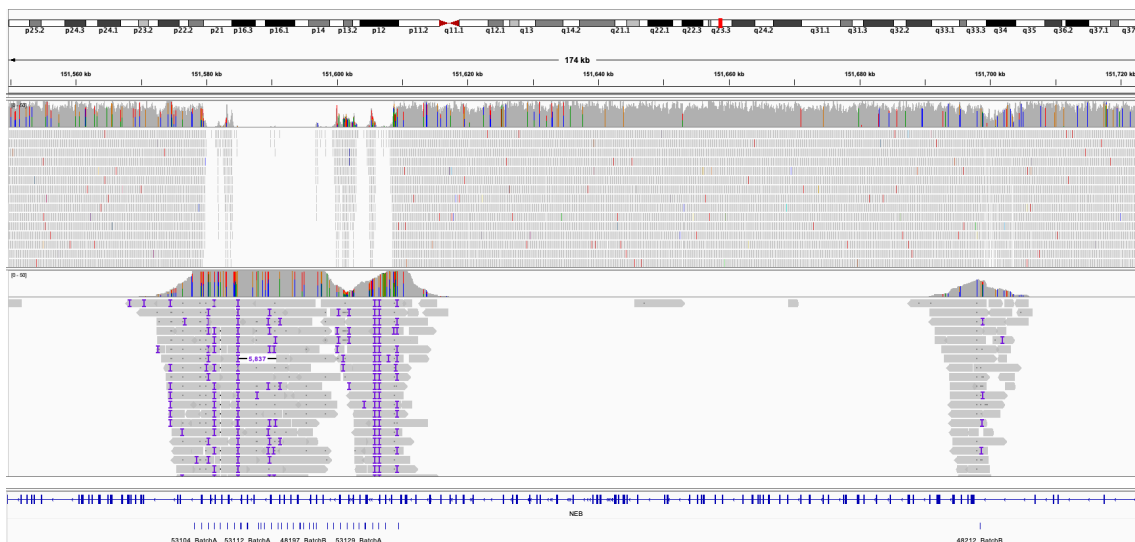
표준 쇼트 리드 WGS



Illumina Complete Long Read Prep with Enrichment

B.

표준 쇼트 리드 WGS



Illumina Complete Long Read Prep with Enrichment

그림 2: Human Comprehensive Panel과 표적 롱 리드로 매핑이 어려운 영역 분석 — Illumina Complete Long Read Prep with Enrichment, Human 및 Human Comprehensive Panel은 분석이 어려운 유전자 영역의 커버리지를 향상시켜 표준 쇼트 리드 인간 WGS를 상호 보완함. 표준 쇼트 리드 WGS 및 Illumina Complete Long Reads with Enrichment를 사용해 얻은 (A) *STRC* 및 (B) *NEB*의 Integrative Genomics Viewer(IGV) 플롯.

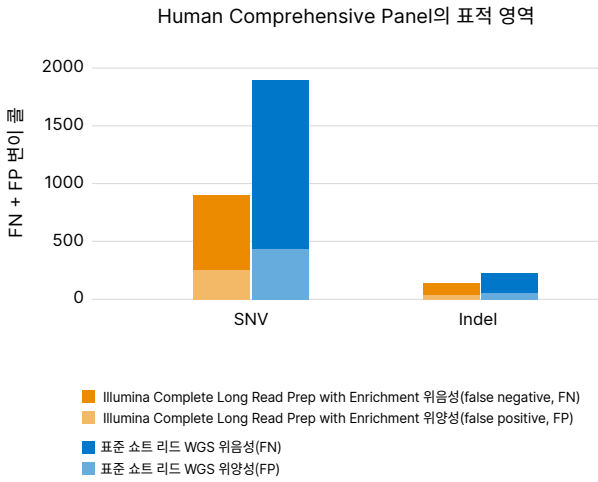


그림 3: 변이 검출 정확도를 향상시키는 표적 롱 리드 — 표준 쇼트 리드 WGS(파란색)와 Illumina Complete Long Read Prep with Enrichment 및 Human Comprehensive Panel(주황색)을 사용해 SNV 및 Indel 검출 시 Human Comprehensive Panel이 표적으로 하는 유전자 영역에서의 위음성(FN) + 위양성(FP) 결과

## 요약

Illumina Human Comprehensive Panel은 쇼트 리드만으로는 매핑하기 어려운 일부 유전자 영역에서 정확도가 높은 롱 리드 커버리지를 제공합니다. Illumina Complete Long Read Prep with Enrichment 및 Human Comprehensive Panel은 Illumina의 WGS 기술을 상호 보완하고 완전한 워크플로우 솔루션을 통해 롱 리드로 가장 유용한 정보를 얻을 수 있는 영역에 초점을 맞춘 비용 대비 효과적이고 최적화된 전장 유전체 assay입니다.

## 상세 정보

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

[Human Comprehensive Panel](#)

[롱 리드 시퀀싱 기술](#)

표 3: Human Comprehensive Panel의 성능

Human Comprehensive Panel의 표적 영역에 대한 정확도 <sup>a</sup>		
	Illumina Complete Long Read Prep with Enrichment 및 Human Comprehensive Panel <sup>b</sup>	표준 쇼트 리드 WGS <sup>c</sup>
F1 점수(SNV)	99.54%	99.04%
F1 점수(Indel)	99.47%	99.15%
F1 점수(SV)	80.43%	60.18%

전장 유전체, 모든 벤치마크 영역(all benchmark region)에 대한 정확도 <sup>d</sup> F1 점수(SNV + Indel)		
Illumina Complete Long Read Prep with Enrichment 및 Human Comprehensive Panel <sup>b</sup>	Illumina Complete Long Read Prep, Human <sup>e</sup>	표준 쇼트 리드 WGS <sup>c</sup>
99.87%	99.90%	99.84%

a. 상염색체(Autosome)에 국한됨. SNV = single nucleotide variant(단일 염기서열 변이), SV = structural variant(구조적 변이)  
 b. Illumina Complete Long Read Prep with Enrichment, Human Comprehensive Panel + Illumina DNA PCR-Free Prep + NovaSeq 6000 시스템 + DRAGEN v4.2  
 c. Illumina DNA PCR-Free Prep + NovaSeq 6000 시스템 + DRAGEN v4.2  
 d. precisionFDA Truth Challenge v2의 벤치마크 데이터를 사용해 모든 참가 솔루션을 측정된 결과<sup>2</sup>(출처: internal data on file for DRAGEN v4)  
 e. Illumina Complete Long Read Prep, Human + Illumina DNA PCR-Free Prep + NovaSeq 6000 시스템 + DRAGEN v4.2

## 참고 문헌

1. Bekritsky MA, Bekritsky MA, Colombo C, Eberle MA. Identifying genomic regions with high quality single nucleotide variant calling. [illumina.com/science/genomics-research/articles/identifying-genomic-regions-with-high-quality-single-nucleotide-.html](https://illumina.com/science/genomics-research/articles/identifying-genomic-regions-with-high-quality-single-nucleotide-.html). Published 2021. Accessed August 30, 2023.
2. PrecisionFDA. Truth Challenge V2: Calling Variants from Short and Long Reads in Difficult-to-Map Regions. [precision.fda.gov/challenges/10](https://precision.fda.gov/challenges/10). Accessed October 2, 2023.
3. Roessler K. Illumina Complete Long Reads software analysis workflow for human WGS. [illumina.com/science/genomics-research/articles/complete-long-read-software-analysis.html](https://illumina.com/science/genomics-research/articles/complete-long-read-software-analysis.html). Published 2023. Accessed September 22, 2023.

## 제품 목록

제품명	카탈로그 번호
Illumina Complete Long Read Prep with Enrichment, Human (24 samples)	20113832
Illumina Complete Long Read Prep with Enrichment, Human (96 samples)	20113833
Illumina Complete Long Read Prep with Enrichment, Human Comprehensive Panel (24 samples)	20113834
Illumina Complete Long Read Prep with Enrichment, Human Comprehensive Panel (96 samples)	20113835
Illumina Human Comprehensive Panel (24 samples)	20113836
Illumina Human Comprehensive Panel (96 samples)	20113837
Illumina Unique Dual Indexes, LT (48 indexes, 48 samples)	20098166
Illumina DNA/RNA UD Indexes Set A, Tagmentation (96 indexes, 96 samples)	20091654
Illumina DNA/RNA UD Indexes Set B, Tagmentation (96 indexes, 96 samples)	20091656
Illumina DNA/RNA UD Indexes Set C, Tagmentation (96 indexes, 96 samples)	20091658
Illumina DNA/RNA UD Indexes Set D, Tagmentation (96 indexes, 96 samples)	20091660
Illumina Analytics - 1 iCredit	20042038



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.html](http://www.illumina.com/company/legal.html)을 참조하십시오.

M-GL-02191 v1.0 KOR