

Illumina Human Comprehensive Panel

Höhere Coverage schwieriger
Regionen in proteincodierenden
Genen mit Illumina Complete
Long Read Prep with
Enrichment, Human



Einleitung

Bei der Sequenzierung des Humangenoms (WGS, Whole-Genome Sequencing) lässt sich häufig ein kleiner Teil der Genregionen mit Short-Reads allein nur schwer mappen. Schwierige Regionen sind hochgradig homologe oder repetitive Regionen, komplexe strukturelle Varianten, Pseudogene und große Insertionen/Deletionen (Indels). In diesen Fällen kann die Long-Read-Sequenzierung die mit der herkömmlichen Short-Read-WGS generierten Daten ergänzen, was für eine bessere Coverage zur Analyse schwieriger Regionen sorgt und die Phasierung von Varianten sowie das Haplotypen-Calling vereinfacht.

Die Illumina Complete Long Reads-Technologie generiert anhand eines herkömmlichen NGS-Workflows (Next-Generation Sequencing, Sequenzierung der nächsten Generation) auf Illumina-Sequenziersystemen zusammenhängende Long-Read-Sequenzen (Abbildung 1). Illumina Complete Long Read Prep with Enrichment und das Human Comprehensive Panel bieten Forschern ein kostengünstiges, skalierbares Verfahren für die Analyse bekannter schwieriger Genregionen. Gezielte Long-Reads werden gemeinsam mit herkömmlichen hochgenauen analysiert, was das Mapping von mehr Regionen ermöglicht.

Umfassendes, optimiertes Sondenpanel

Beim Illumina Human Comprehensive Panel handelt es sich um ein optimiertes Hybrid-Capture-Sondenpanel für Genregionen, bei denen die höhere Mapping-Fähigkeit längerer Reads vorteilhaft ist. Das Panel zielt auf Regionen mit geringer Mapping-Fähigkeit in über 6.500 proteincodierenden Genen ab (Tabelle 1)¹ und ermöglicht damit komplementäre Long-Reads, die die Auflösung von Bereichen verbessern, die sich mit herkömmlichen Short-Read-WGS-Methoden nur schwer mappen lassen.

Tabelle 1: Human Comprehensive Panel-Parameter^a

System	NovaSeq X Series NovaSeq 6000 System
Größe der Zielregion	> 95 Mb
Sequenzierungsausgabe je Probe ^b	90–120 Gb
Anzahl der Sonden	ca. 40.000
Anzahl der Zielgene	> 6.500
Probentyp	Genomische DNA
Empfohlene DNA-Zugabe	50 ng
Gesamtdauer der Bibliotheksvorbereitung	ca. 2 Tage
Manueller Aufwand	ca. 6 h
Multiplexing	Bis zu 64 Proben
N50	6,3 kb
Phasenblock N50	15,6 kb
Einheitlichkeit	95 %
Padded-Read-Anreicherung (PRE, Padded Read Enrichment) ^c	83 %
Anteil phasierter heterozygoter SNV in % ^d	98 %

- Daten, die mit 50 ng genomischer HG002-DNA (Coriell, Katalog-Nr. NA24385) generiert wurden. Die Leistung kann je nach DNA-Zugabe und Probenqualität variieren.
- Erfordert einen Sequenzierungslauf mit 2×150 bp. Dieser sorgt für die ca. 30-fache finale Coverage von Illumina Complete Long Reads.
- Coverage-Einheitlichkeit berechnet als $\% > 0,2 \times$ Mittelwert.
- PRE berechnet als $100 \times (\text{alignierte Padded-Target-Reads} / \text{insgesamt alignierte Reads})$.
- SNV, Single Nucleotide Variants (Einzelnukleotidvarianten)



Abbildung 1: Teil eines integrierten Workflows: Zugriff auf kostengünstige, zielgerichtete Long-Read-WGS-Daten mithilfe eines skalierbaren, optimierten Bibliotheksvorbereitungsprotokolls mit Anreicherung, bewährter Illumina-Sequenzierungschemie und DRAGEN-Sekundäranalyse. Erfordert herkömmliche Short-Read-WGS-Daten mit ≥ 30 -facher Coverage aus derselben Probe für die Analyse. Es können FASTQ-Dateien einer Probe aus einem vorherigen Lauf verwendet werden.

Bei der Entwicklung dieses Panels hat Illumina alle der über 20.000 proteincodierenden Gene berücksichtigt, einschließlich Introns sowie untranslatierter Regionen (UTRs), und zielte auf 70 % aller schwer zu mappenden Basenpaare in diesen Regionen ab. Gene, die allein durch Short-Reads umfassend abgedeckt sind, wurden nicht berücksichtigt. Zur Maximierung der Effizienz und Leistung von Design und Sequenzierung wurde dieses Panel im Nasslabor getestet und optimiert.

Hochgradig skalierbare Sequenzierung und Analyse

Der Illumina Complete Long Read Prep with Enrichment-Workflow ist hochgradig skalierbar, einfach zu automatisieren und eignet sich so für die umfassende WGS bei einer größeren Anzahl von Proben (Tabelle 2). Auf dem NovaSeq[™] X Plus System können Anwender bis zu 15.000 hochgenaue Genome pro Jahr generieren.* Für kleinere Panels und geringere Anforderungen hinsichtlich der Batchherstellung bieten sich Verbrauchsmaterialien für geringeren Durchsatz wie die NovaSeq X Series 1.5B-Fließzelle an. Mit Illumina Complete Long Read Prep with Enrichment und dem Human Comprehensive Panel lassen sich vorhandene WGS-Datensätze als Reflex-Tool für eine weitreichendere Variantenerkennung ergänzen.

* Potenzieller Durchsatz bei Verwendung von Illumina Complete Long Read Prep with Enrichment und Human Comprehensive Panel mit dem NovaSeq X Plus System und einem Lauf mit zwei 25B-Fließzellen.

Die Datenanalyse für Illumina Complete Long Read Prep with Enrichment, Human ist mit einer BaseSpace[™] Sequence Hub-App oder über Illumina Connected Analytics möglich. Die Daten werden in die Cloud gestreamt und gemeinsam mit herkömmlichen Short-Read-WGS-Daten (≥ 30 -fach) aus derselben Probe (vorab oder parallel generiert) analysiert. Die DRAGEN[™]-Pipeline führt Ergebnisse zu einem einzigen Satz von Ausgabedateien zusammen, einschließlich gezielten DRAGEN-Callern.³

Verbesserte Coverage bei geringen Kosten

Illumina Complete Long Read Prep with Enrichment und das Human Comprehensive Panel verbessern im Vergleich zur herkömmlichen Short-Read-WGS die Genauigkeit des Varianten-Callings sowie die Coverage der Zielregionen (Abbildung 2, Abbildung 3, Tabelle 3). Gezielte Long-Reads liefern mit dem Human Comprehensive Panel umfassende Genauigkeit mit einem F1-Score (SNVs + Indels) von 99,87 %² über das gesamte Genom hinweg. Dies stellt eine erhebliche Verbesserung gegenüber der ausschließlichen Verwendung von Short-Reads dar. Hieraus ergibt sich eine Leistung, die mit der von Illumina Complete Long Read Prep, Human und anderen Long-Read-Genomverfahren bei proteincodierenden Genen vergleichbar ist, jedoch zu geringeren Kosten und mit höherem Durchsatz (Tabelle 3).

Tabelle 2: Empfohlener Probendurchsatz zur Erzielung der finalen 30-fachen Coverage bei Human Comprehensive Panel und Illumina Complete Long Read Prep with Enrichment, Human^{a, b, c, d}

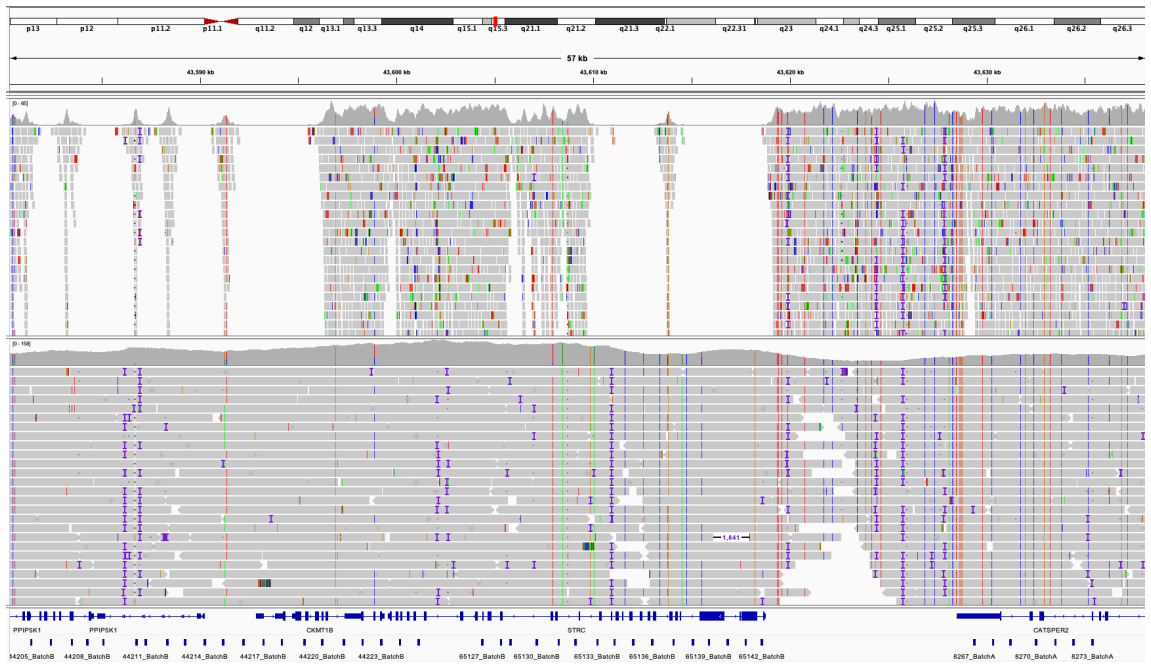
	NovaSeq 6000				NovaSeq X		
	SP	S1	S2	S4	1.5B	10B	25B
Reagenzien-Kits für 300 Zyklen							
Proben pro Fließzelle	2	4	10	24	4	24	64
Ausgabe pro Fließzelle	ca. 250 Gb	ca. 500 Gb	ca. 1,25 Tb	ca. 3 Tb	ca. 500 Gb	ca. 3 Tb	ca. 8 Tb
Laufzeit	ca. 25 h	ca. 25 h	ca. 36 h	ca. 44 h	ca. 21 h	ca. 25 h	ca. 48 h

- Die Größe der Zielregion des Human Comprehensive Panel liegt bei > 95 Mb und erfordert eine Sequenzierungsausgabe von 90–120 Gb je Probe.
- Erfordert einen Sequenzierungslauf mit 2×150 bp. Dieser sorgt für die ca. 30-fache finale Coverage von Illumina Complete Long Reads.
- Erfordert herkömmliche Short-Read-Humangenomdaten mit 30-facher Coverage aus derselben Probe für die Analyse. Illumina DNA PCR-Free Prep wird empfohlen. WGS-Kits von Drittanbietern sind ebenfalls geeignet. Die Bibliothek ohne Kennzeichnung muss nicht parallel vorbereitet oder sequenziert werden. Die Verwendung von FASTQ-Dateien einer Probe aus einem vorherigen Lauf ist möglich.
- Bei der Sequenzierung von Illumina Complete Long Read-Bibliotheken auf NovaSeq-Geräten kann der gemeldete Q30-Score eines Laufs unter der NovaSeq-Spezifikation liegen. Dies deutet nicht auf Leistungsprobleme in Zusammenhang mit dem Sequenzierungslauf oder der Bibliothek hin.

A.

Herkömmliche Short-Read-WGS

Illumina Complete Long Read Prep with Enrichment



B.

Herkömmliche Short-Read-WGS

Illumina Complete Long Read Prep with Enrichment

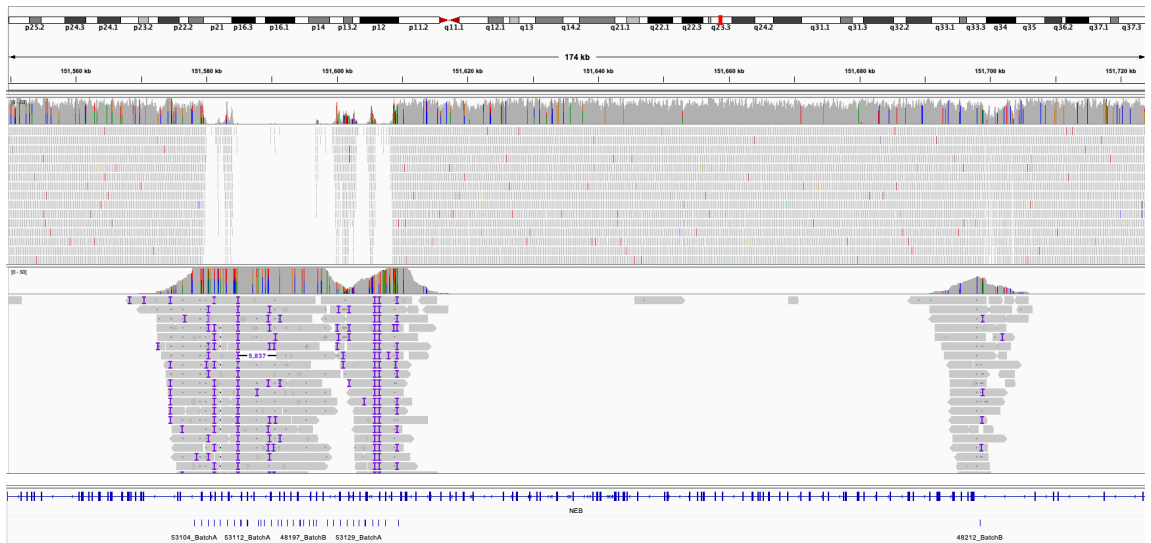


Abbildung 2: Mit dem Human Comprehensive Panel lassen sich schwer zu mappende Regionen mit gezielten Long-Reads analysieren: Illumina Complete Long Read Prep with Enrichment, Human und das Human Comprehensive Panel helfen bei der Verbesserung der Coverage anspruchsvoller Genregionen und ergänzen die herkömmliche Short-Read-Sequenzierung des Humangenoms. Integrative Genomics Viewer(IGV)-Plots von (A) *STRC* und (B) *NEB* unter Verwendung von herkömmlicher Short-Read-Sequenzierung und Illumina Complete Long Reads with Enrichment.

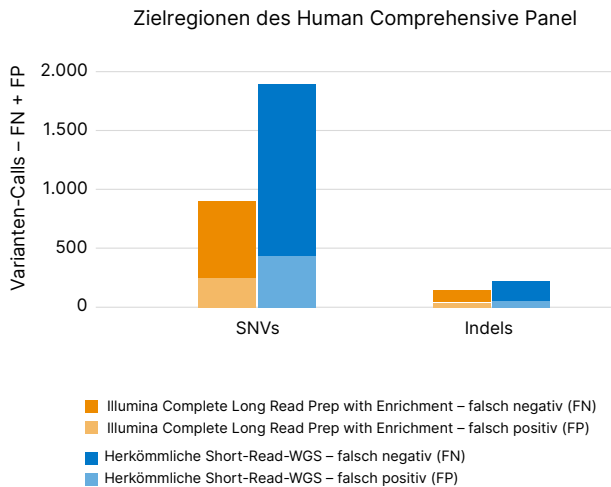


Abbildung 3: Zielgerichtete Long-Reads zur Verbesserung der Genauigkeit des Varianten-Callings in schwierigen Regionen: Falsch negative (FN) und falsch positive (FP) Varianten-Calls für SNVs und Indels in den Zielregionen des Human Comprehensive Panel mit Illumina Complete Long Read Prep with Enrichment und dem Human Comprehensive Panel (orange) im Vergleich zur herkömmlichen Short-Read-WGS (blau).

Zusammenfassung

Das Illumina Human Comprehensive Panel ermöglicht eine hochgenaue Long-Read-Coverage des geringen Anteils von Genregionen, die sich mit Short-Reads allein nur schwer mappen lassen. Illumina Complete Long Read Prep with Enrichment und das Human Comprehensive Panel bieten einen optimierten, kostengünstigen Genom-Assay, der in einer Lösung mit umfassendem Workflow die Illumina-WGS ergänzt und sich auf Long-Reads stützt, wo diese den größten Vorteil bieten.

Weitere Informationen

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

[Human Comprehensive Panel](#)

[Technologie für die Long-Read-Sequenzierung](#)

Tabelle 3: Leistung des Human Comprehensive Panel

Genauigkeit für die Zielregionen des Human Comprehensive Panel ^a		
	Illumina Complete Long Read Prep with Enrichment und Human Comprehensive Panel ^b	Herkömmliche Short-Read-WGS ^c
F1-Score (SNVs)	99,54 %	99,04 %
F1-Score (Indels)	99,47 %	99,15 %
F1-Score (SVs)	80,43 %	60,18 %

Genauigkeit für das gesamte Genom, alle Benchmark-Regionen ^d F1-Score (SNVs + Indels)			
	Illumina Complete Long Read Prep with Enrichment und Human Comprehensive Panel ^b	Illumina Complete Long Read Prep, Human ^e	Herkömmliche Short-Read-WGS ^c
	99,87 %	99,90 %	99,84 %

a. Nur Autosome. SNV, Single Nucleotide Variants (Einzelnukleotidvarianten); SV, strukturelle Varianten.
b. Illumina Complete Long Read Prep with Enrichment, Human Comprehensive Panel + Illumina DNA PCR-Free Prep + NovaSeq 6000 System + DRAGEN v4.2.
c. Illumina DNA PCR-Free Prep + NovaSeq 6000 System + DRAGEN v4.2.
d. Im Vergleich zu allen teilnehmenden Lösungen anhand von PrecisionFDA Truth Challenge v2-Benchmark-Daten², internen Daten für DRAGEN v4.
e. Illumina Complete Long Read Prep, Human + Illumina DNA PCR-Free Prep + NovaSeq 6000 System + DRAGEN v4.2.

Quellen

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. Veröffentlicht 2021. Aufgerufen am 30. August 2023.
2. PrecisionFDA. Truth Challenge V2: Calling Variants from Short and Long Reads in Difficult-to-Map Regions. precision.fda.gov/challenges/10. Aufgerufen am 2. Oktober 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. Veröffentlicht 2023. Aufgerufen am 22. September 2023.

Bestellinformationen

Produkt	Katalog-Nr.
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



1 800 8094566 (USA, gebührenfrei) | +1 858 2024566 (Tel. außerhalb der USA)
techsupport@illumina.com | www.illumina.com

© 2024 Illumina, Inc. Alle Rechte vorbehalten. Alle Marken sind Eigentum von Illumina, Inc. bzw. der jeweiligen Inhaber. Spezifische Informationen zu Marken finden Sie unter www.illumina.com/company/legal.html.

M-GL-02191 DEU v1.0