

illumina TruPathTM Genome

Der einfachste Workflow für
die Genomsequenzierung



Der einfachste Probe-zu-Daten-Workflow mit 10 Minuten manuellem Aufwand für die Vorbereitung

TAGCA
CGTAG
ATAGC

Untersuchung schwer zu mappender Regionen des Genoms mit kurzen Reads und Proximity-Daten auf der NovaSeqTM X Series



Gewinnung von Long-Distance-Erkenntnissen mit verbessertem Nachweis struktureller Varianten und ultralanger Phasierung

Bislang unerreichte Zugänglichkeit des Genoms

Die Short-Read-NGS (Next-Generation Sequencing, Sequenzierung der nächsten Generation) ist das in der Forschung am häufigsten genutzte Verfahren zur Untersuchung eines Großteils des Genoms. Dennoch lassen sich bestimmte Regionen und Variantentypen weiterhin nur schwer charakterisieren, darunter extrem repetitive Regionen und solche mit hoher Sequenzhomologie oder struktureller Komplexität. Aus aktuellen Studien geht hervor, dass diese Regionen Varianten enthalten, die eine potenzielle Rolle bei genetischen Erkrankungen des Menschen spielen.^{1,2} Damit sind sie von entscheidender Bedeutung für die Erforschung der Ursachen genetischer Erkrankungen.

Alternative Verfahren wie die Long-Read-Sequenzierung zeichnen sich durch eine höhere Auflösung aus, wodurch sich Erkenntnisse zu zahlreichen dieser schwer zu mappenden Regionen und Varianten gewinnen lassen. Jedoch sind die Workflows bei Long-Read-Methoden häufig komplex und bringen bekannte Probleme mit der Genauigkeit mit sich, die variable Ergebnisse zur Folge haben.

Illumina TruPath Genome mit [Proximity-Read-Mapping](#) revolutioniert den Workflow bei der Genomsequenzierung (WGS, Whole-Genome Sequencing). Dieser innovative Assay erfasst lange DNA-Fragmente. Der manuelle Aufwand für die Vorbereitung des Assays beträgt dabei nur 10 Minuten.

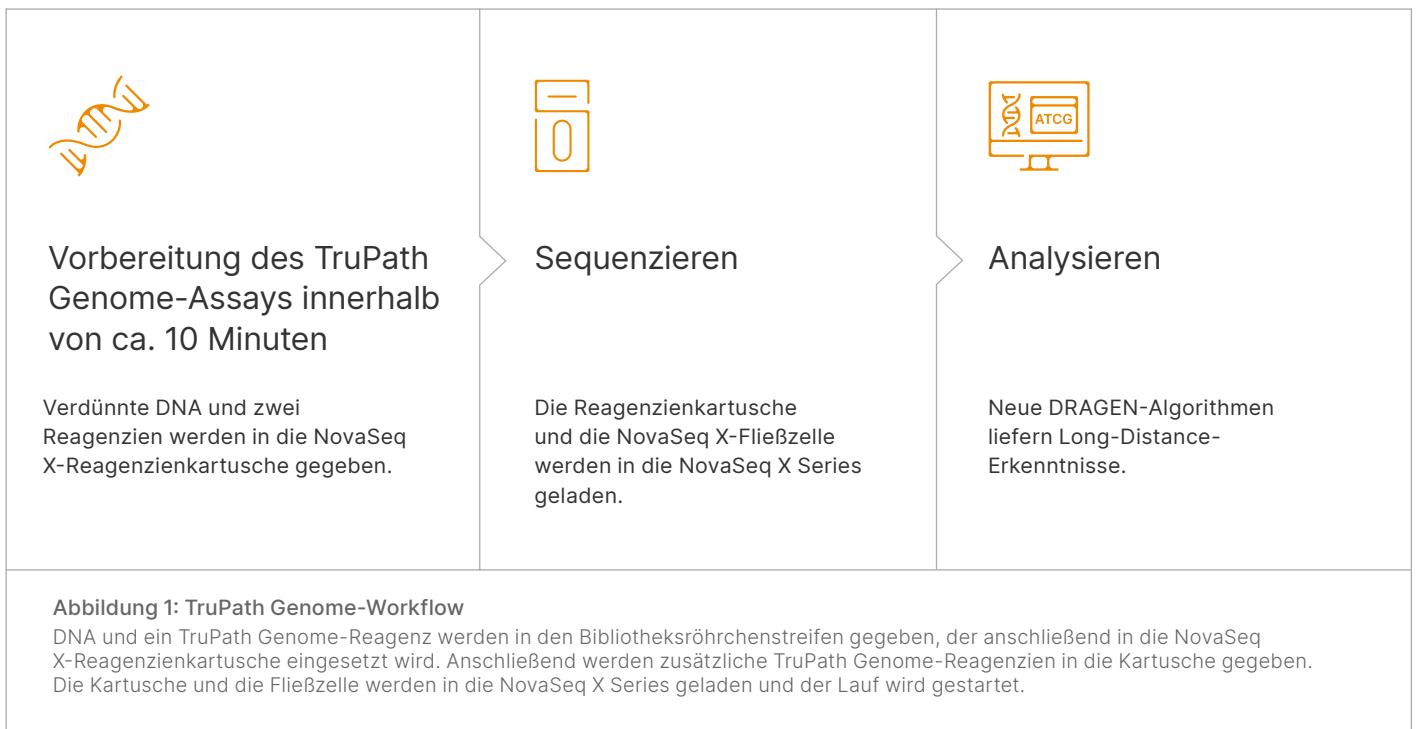
Die Sequenzierung erfolgt auf der vorhandenen NovaSeq X Series mit modernster DRAGEN™-Sekundäranalyse. Dadurch erhält die Forschung ein genaues, umfassendes Genom anhand dessen Erkenntnisse über Spannen hinweg generiert werden können, die größer sind als die bei Long-Read-Methoden üblichen Read-Längen.

Dank TruPath Genome lassen sich die Anwenderfreundlichkeit und Genauigkeit der Short-Read-Sequenzierung jetzt für schwer zu mappende Genomregionen nutzen, der Nachweis struktureller Varianten (SV) optimieren sowie phasierte Reads und Varianten-Calls generieren.

Beispiellose Anwenderfreundlichkeit des Workflows

TruPath Genome vereinfacht den Einsatz von WGS erheblich und verkürzt den manuellen Aufwand auf 10 Minuten ([Abbildung 1](#)). Die bewährte Transposasechemie von Illumina macht herkömmliche Bibliotheksvorbereitungsschritte überflüssig. Fortschrittliche Bioinformatikalgorithmen liefern anhand von Nanowell-Proximity-Informationen und Paired-End-Short-Read-Standarddaten* Erkenntnisse zu Beziehungen zwischen Varianten, die durch mehrere Millionen Basen getrennt sein können.

* Paired-End-Short-Read-Standarddaten sind Daten, die anhand von Genomsequenzierung mit manueller Bibliotheksvorbereitung und standardmäßiger Sequenzierung durch Synthese (SBS, Sequencing by Synthesis) generiert werden.



Proben

Gereinigte doppelsträngige DNA (dsDNA) wird mit im Handel erhältlichen Standard- oder HMW-Kits (High Molecular Weight, hohes Molekulargewicht) extrahiert. Die empfohlene Zugabemenge von 350 ng DNA wird mit TruPath-Tagmentierungspuffer verdünnt und in den Bibliotheksröhrchenstreifen gegeben. Zugabe-DNA kann aus Zelllinien oder Vollblut isoliert werden. Aus FFPE- (formalinfixiert, in Paraffin eingebettet) und cfDNA-Proben (cell-free DNA, zellfreie DNA) isolierte DNA ist nicht geeignet. Für optimale Performance wird empfohlen, DNA mit einem Verfahren zu extrahieren, bei dem große Fragmente erhalten bleiben (40 % mit Länge > 60 kb).³

Hinweise zur Verwendung weiterer Probentypen, zu schwankender Probenqualität und zu optimalen Zugabemengen finden Sie unter [TruPath Genome performance with various sample types, quality, and input technical note \(Technischer Hinweis zur Performance von TruPath Genome mit unterschiedlichen Probentypen, zur Qualität und zur Zugabe\)](#).

Sequenzierung

Der Bibliotheksröhrchenstreifen mit verdünnter DNA und zwei TruPath Genome-Reagenzien wird in die NovaSeq X-Reagenzienkartusche gegeben und gemeinsam mit einer NovaSeq X C2- oder C8-Fließzelle zur Sequenzierung von zwei bzw. acht Humangenomenen in die NovaSeq X Series geladen.

Illumina-Transposome werden als Primer für die Nanowells auf der Fließzellenoberfläche verwendet.[†] Die DNA wird in die Fließzelle eingebracht, wo sie von den Transposomen über mehrere Wells hinweg erfasst wird. Die gefüllten Nanowells bilden dabei ein sternbildähnliches Muster ([Abbildung 2A](#)). Transposome spalten die erfasste DNA und heften die Fragmente in einem als Tagmentierung bezeichneten Prozess[‡] ([Abbildung 2B](#)) an der Oberfläche der Fließzelle an. In einem Waschschrift werden ungebundene DNA und Transposome entfernt.

Aus der tagmentierten DNA in den Nanowells werden Cluster[§] für die Sequenzierung gebildet. Aus DNA desselben ursprünglichen langen DNA-Fragments werden Cluster in benachbarten Nanowells gebildet. Die Sequenzierung erfolgt auf der NovaSeq X Series mit der bewährten XLEAP-Chemie zur Sequenzierung durch Synthese (XLEAP-SBS™, XLEAP-Sequencing by Synthesis) und einem Lauf mit 2 × 150 bp.

Analyse

Im Anschluss an die Sequenzierung aligniert die DRAGEN™ Germline-Pipeline mithilfe eines neuartigen Algorithmus, der Sequenzierungsdaten mit Nanowell-Proximity-Informationen kombiniert, die DNA auf ein Referenzgenom. Proximity-Informationen ermöglichen die korrekte Zuordnung von Reads, die zuvor nicht auf das Referenzgenom gemappt werden konnten oder bei denen das Alignment mit geringer Konfidenz erfolgt ist.

[†] Transposome sind DNA-Transposasekomplexe, die als Dimer vorliegen.

[‡] Tagmentierung bezeichnet den Prozess, bei dem DNA-Fragmente geschnitten und mit einem Transposom eine Adaptersequenz hinzugefügt wird (Tagging).

[§] Bei einem Cluster handelt es sich um einen amplifizierten DNA-Punkt auf einer Fließzelle, der sequenziert wird.

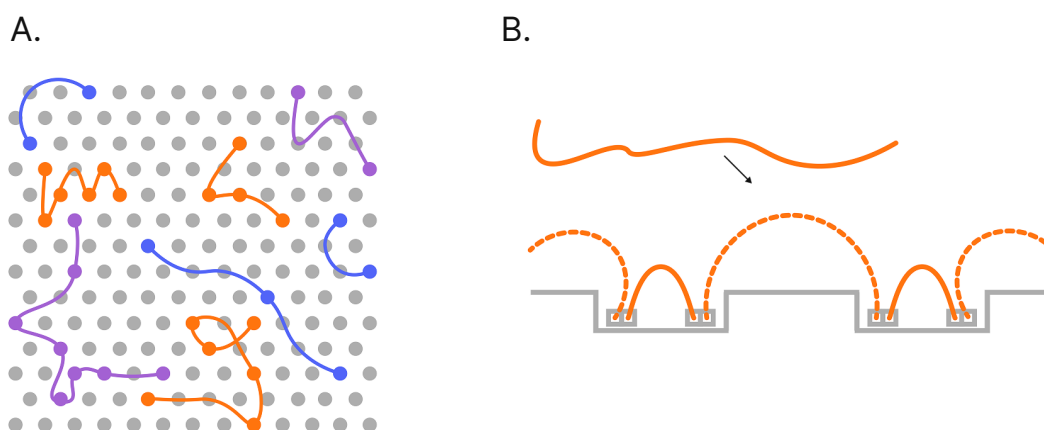


Abbildung 2: Funktionsweise des Proximity-Read-Mappings

(A) DNA fließt über mehrere benachbarte Nanowells, wodurch Proximity-Informationen die Generierung von Long-Distance-Erkenntnissen ermöglichen. (B) DNA wird von Transposomen in den Nanowells (graue Kästchen) erfasst, wo sie Tagmentierung und Clusterbildung durchläuft.

Auf diese Weise können herkömmliche Short Reads probabilistisch einander zugeordnet werden, was einen umfassenden Nachweis kleiner Varianten (Einzelnukleotid-Varianten (SNVs, Single-Nucleotide Variants)), Insertionen/Deletionen (Indels), großer Varianten (> 50 bp) und bekannter homologer Regionen sowie die Phasierung von Calls kleiner Varianten ermöglicht, die durch Millionen von Basenpaaren getrennt sein können.

Untersuchung eines umfassenden Genoms mit hochpräzisen kurzen Reads

Die standardmäßige Sequenzierung durch Synthese (SBS, Sequencing by Synthesis) zeichnet sich zwar durch hohe Genauigkeit bei kleinen Varianten (SNVs, Indels und Kopienzahlvarianten (CNVs, Copy Number Variations)) aus, bei bestimmten Regionen und Variantentypen wie strukturellen Varianten ist dies jedoch nicht der Fall. TruPath Genome bringt neue Kapazitäten für die Short-Read-Sequenzierung mit, die ein umfassenderes Genom liefern. Ab sofort lassen sich bislang herausfordernde Genomregionen mithilfe kurzer Reads und dank zusätzlicher Variantentypen mit höherer Genauigkeit untersuchen.

Verbesserte Coverage in schwer zu mappenden Regionen

Komplexe, stark polymorphe oder duplizierte Bereiche des Genoms lassen sich aufgrund geringer Mappability, die auf geringe Coverage oder unzureichende Mappingqualität (MapQ) zurückzuführen ist, unter Umständen nur schwer assemblieren bzw. auf ein Referenzgenom alignieren. Bislang zeigten sich bei der Coverage dieser „dunklen“ Regionen⁴ niedrige MapQ-Scores, Lücken in den Genomsequenzdaten und eine insgesamt begrenzte Auflösung des Varianten-Callings. TruPath Genome verwendet Proximity-Informationen aus benachbarten Clustern zur Zuweisung der korrekten Position im Genom. Dies ermöglicht das zuverlässige Mapping dieser nicht eindeutig gemappten Reads ([Abbildung 3](#) und [Abbildung 4](#)).

Höhere Genauigkeit beim Varianten-Calling

TruPath Genome wurde zur Demonstration der Performance beim Calling kleiner Varianten, dem SV-Nachweis und der phasierten Sequenzierung einem Benchmarktest mit dem HG002-Genom und dem Referenzdatensatz T2T-Q100 v1.1 v0.019 des Telomere-to-Telomere-Konsortiums (T2T-Konsortiums) unterzogen.

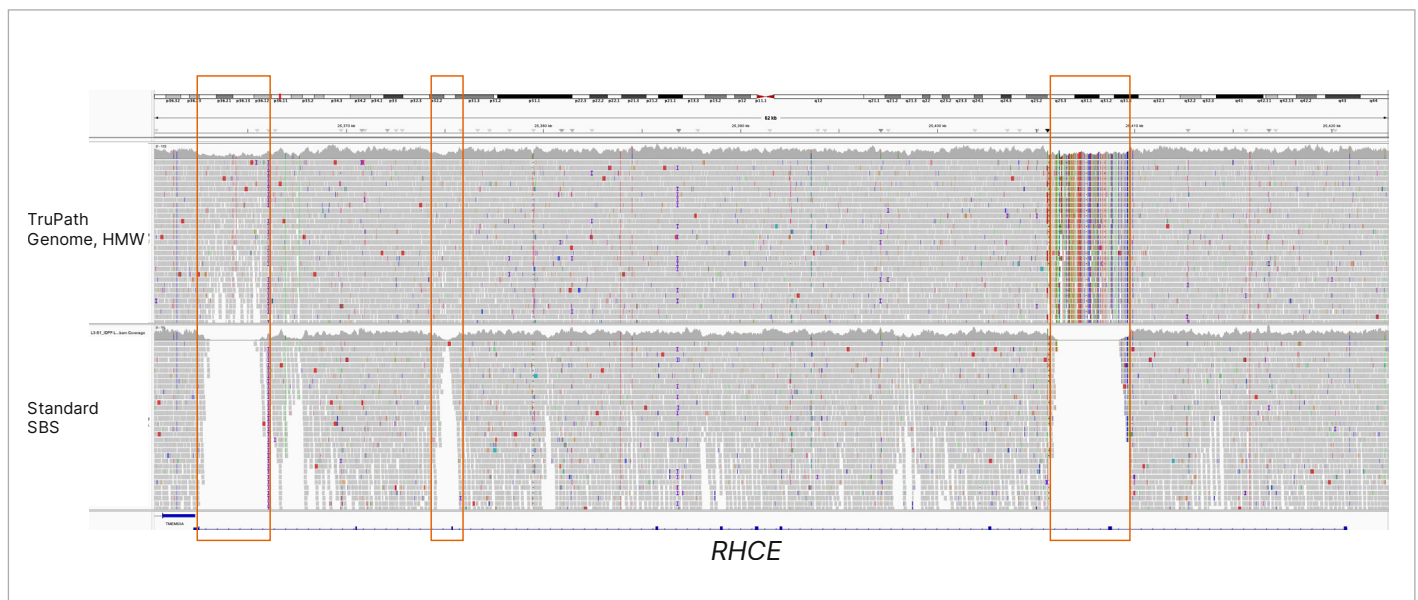


Abbildung 3: Effektive Untersuchung schwer zu mappenden, medizinisch relevanter Gene mit TruPath Genome

Wiederherstellung der Coverage im Gen *RHCE*, das sich im Rh-Blutgruppenlocus neben *RHD* befindet und eine extrem hohe Sequenzhomologie mit häufiger Genkonversion und hybriden Rearrangements aufweist, wodurch das Short-Read-Mapping mitunter mehrdeutig wird. Die genaue *RHCE*-Genotypisierung ist klinisch bedeutsam für die Bestimmung von Rh-Antigenen (C/c, E/e), da Fehlpaarungen das Risiko für hämolytische Transfusionsreaktionen und hämolytische Erkrankungen des Fötus/Neugeborenen (HDFN, Hemolytic Disease of the Fetus/Newborn) erhöhen können. Die Standard-SBS wurde mit Illumina DNA PCR-Free Prep durchgeführt.

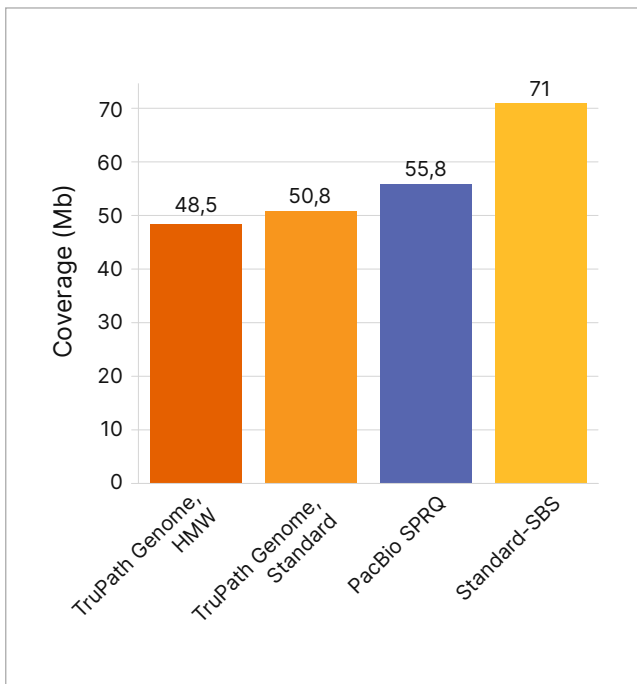


Abbildung 4: TruPath Genome verbessert die Coverage „dunkler Regionen“ des Genoms

TruPath Genome verbessert die Coverage „dunkler Regionen“, die in Ebbert et al.¹ als „dark-by-MAPQ“-Regionen definiert sind, bei denen 90 % der Reads, die die Coverage der Region darstellen, eine Mapping-Qualität (MAPQ) von weniger als 10 aufweisen. TruPath Genome reduzierte die Größe der „dunklen Regionen“ des Genoms dank des verbesserten Mappings von Reads in schwierigen Genomregionen. Die Standard-SBS wurde mit Illumina DNA PCR-Free Prep durchgeführt.

Die dank Proximity-Read-Mapping höhere Mapping-Auflösung erhöht die Performance beim Calling kleiner Varianten, insbesondere in bislang nur schwer zu mappenden (beispielsweise homologen oder repetitiven) Regionen des Genoms. TruPath Genome zeigt im Vergleich zur standardmäßigen Short-Read-Sequenzierung und dem entsprechenden Varianten-Calling eine höhere Genauigkeit, da deutlich weniger falsch negative (FN) und falsch positive (FP) Varianten-Calls auftreten ([Abbildung 5](#)).

Bei TruPath Genome zeigt sich im Vergleich zu herkömmlichen Short-Read-Verfahren eine erheblich höhere Performance beim Calling von SVs (> 50 bp). Die SV-Recall-Rate stieg von 86 % mit standardmäßiger SBS auf 94 % mit Proximity-Read-Mapping ([Abbildung 6](#)).

Ultralange Phasierung

Phasierte Sequenzierung liefert Haplotypinformationen, die eine Unterscheidung zwischen Allelen auf mütterlichen und väterlichen Chromosomen ermöglichen.⁸ Diese Informationen unterstützen in der Forschung die Untersuchung von Varianten mit Compound-Heterozygotie, die wichtig für die Interpretation autosomal-rezessiver Erkrankungen sind.

Proximity-Read-Mapping eignet sich besonders gut für die Phasierung des Humangenoms. Die auf der Fließzelle erfasste DNA bildet phasierte Blöcke mit einer Größe von Hunderten Kilobasen bis hin zu mehreren Megabasen ([Abbildung 7](#) und [Abbildung 8](#)), was zur Generierung besserer Erkenntnisse zu Haplotypen und Compound-Heterozygotie beiträgt.

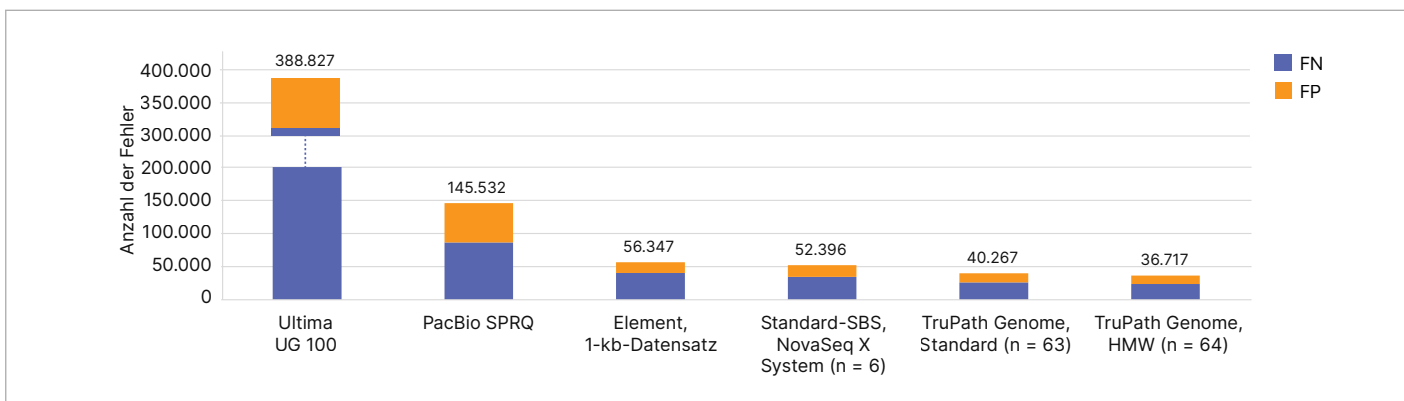


Abbildung 5: Höhere Genauigkeit des Callings kleiner Varianten mit TruPath Genome

Die Performance beim Calling kleiner Varianten für unterschiedliche NGS-Systeme und Assays wurde mit dem Genome in a Bottle-Referenzdatensatz NIST T2T-Q100 HG002 SV v1.1 bestimmt. Die Daten für Ultima UG 100, PacBio mit SPRQ-Chemie und Element AVITI mit einem Datensatz von 1 kb wurden veröffentlichten Materialien entnommen.⁵⁻⁷ Die Daten zur Standard-SBS wurden mit Illumina DNA PCR-Free Prep-Bibliotheken generiert, die auf einem NovaSeq X System mit v1.4-Software sowie der 10B-Fließzelle sequenziert und mit DRAGEN Germline v4.5.2 (sechs technische Replikate) analysiert wurden. Die mit „TruPath Genome, Standard“ bezeichneten Daten wurden anhand von mit Standardverfahren extrahierter DNA generiert und auf einem NovaSeq X System mit v1.4-Software sowie der C8-Fließzelle sequenziert und mit DRAGEN Germline v4.5.2 (63 technische Replikate) analysiert. Die mit „TruPath Genome, HMW“ bezeichneten Daten wurden anhand von mit HMW-Verfahren extrahierter DNA generiert und auf einem NovaSeq X System mit v1.4-Software sowie der C8-Fließzelle sequenziert und mit DRAGEN Germline v4.5.2 (64 technische Replikate) analysiert. FN: falsch negative Varianten-Calls; FP: falsch positive Varianten-Calls.

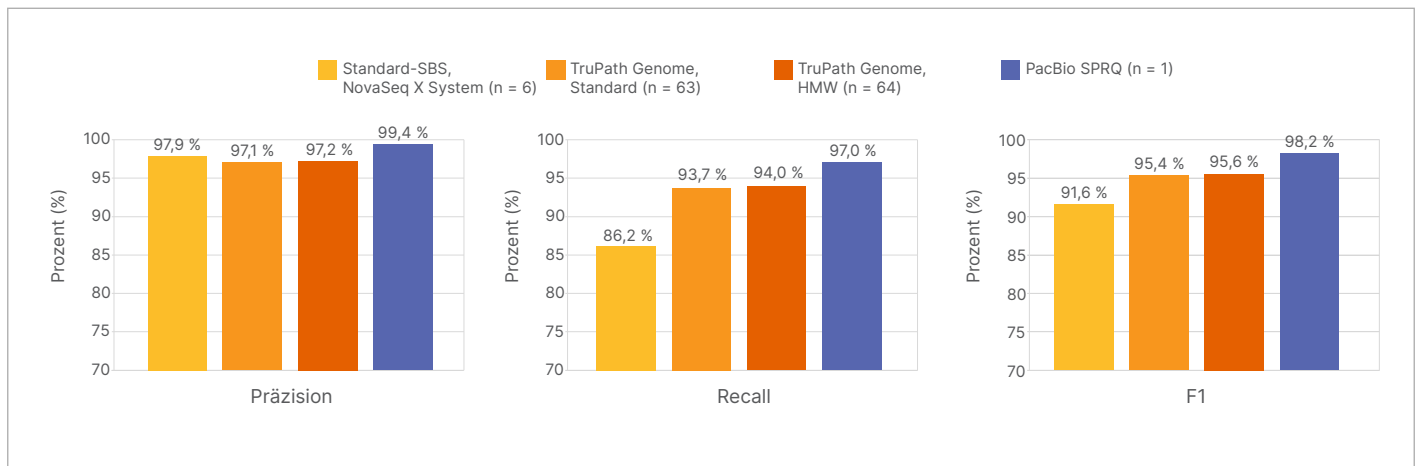


Abbildung 6: Verbessertes Calling struktureller Varianten mit TruPath Genome

Die Performance von TruPath Genome für strukturelle Varianten (SVs) im Vergleich zu Standard-SBS mit DRAGEN Germline v4.5.2. Bei der Analyse wurde der Genom in a Bottle-Referenzdatensatz NIST T2T-Q100 HG002 SV v1.1 mit der BED-Datei mit bekannten SVs durchgeführt. Das Benchmarking wurde gemäß der [Genome in a Bottle-Anleitung für das Benchmarking in Bezug auf strukturelle Varianten mit den Befehlen „bench“ und „refine“](#) von Truvari v4.2.2 durchgeführt. Die Daten für Standard-SBS wurden mit Illumina DNA PCR-Free Prep-Bibliotheken generiert, die auf einem NovaSeq X System mit v1.4-Software sowie der 10B-Fließzelle (sechs technische Replikate) sequenziert wurden. Die mit „TruPath Genome, Standard“ bezeichneten Daten wurden anhand von mit Standardverfahren extrahierter DNA generiert und auf einem NovaSeq X System mit v1.4-Software sowie der C8-Fließzelle (63 technische Replikate) sequenziert. Die mit „TruPath Genome, HMW“ bezeichneten Daten wurden anhand von mit HMW-Verfahren extrahierter DNA generiert und auf einem NovaSeq X System mit v1.4-Software sowie der C8-Fließzelle (64 technische Replikate) sequenziert. Die Daten zu PacBio mit SPRQ-Chemie wurden veröffentlichten Materialien entnommen.⁶ Die Präzision bezeichnet den Anteil richtig positiver Ergebnisse. Recall bezeichnet den prozentualen Anteil richtig negativer Ergebnisse im Vergleich zur Gesamtzahl der Varianten in einer Probe. Der F1-Score (%) gibt die berechneten richtig positiven und richtig negativen Ergebnisse als Anteil an den Gesamtergebnissen an. Höhere Scores weisen auf eine gegenüber Referenzdaten verbesserte Genauigkeit hin.

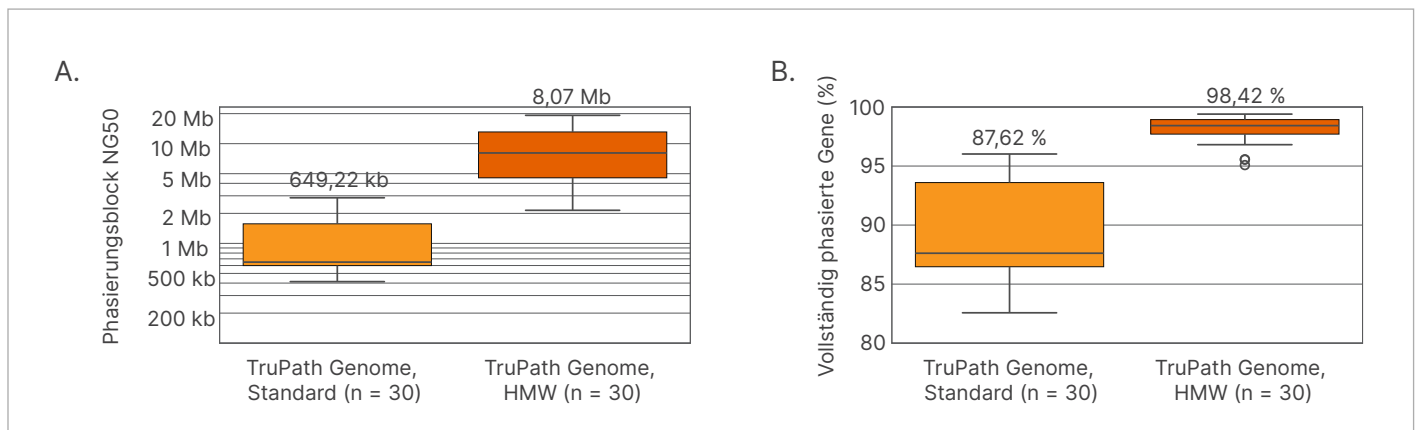


Abbildung 7: Ultralange Phasierung mit TruPath Genome auf 30 einander zugeordneten Zelllinien mit HMW- und Standard-Extraktion

Der Phasierungsblock wird über die Chromosomen 1–22 mit DRAGEN-Phasierung in DRAGEN Germline v4.5.2 gemessen. Die TruPath Genome-Daten wurden mit 30 verschiedenen Coriell-Zelllinienproben mit HMW-Extraktionsmethoden (High Molecular Weight, hohes Molekulargewicht) und Standard-Extraktionsmethoden gewonnen. Der Phasierungsblock NG50 entspricht der Länge des Phasierungsblocks, sobald 50 % der Zielregion (Genom usw.) phasiert wurden. Nicht zur Phasierung von 50 % einer bestimmten Zielregion geeignete Technologie liefert einen NG50 von null Basenpaaren. Der Prozentsatz vollständig phasierter Gene ist der prozentuale Anteil der Genregionen aus einer vorgegebenen Genliste (Gencode v44 genes.gtf), die vollständig in einem einzelnen Phasierungsblock enthalten sind.

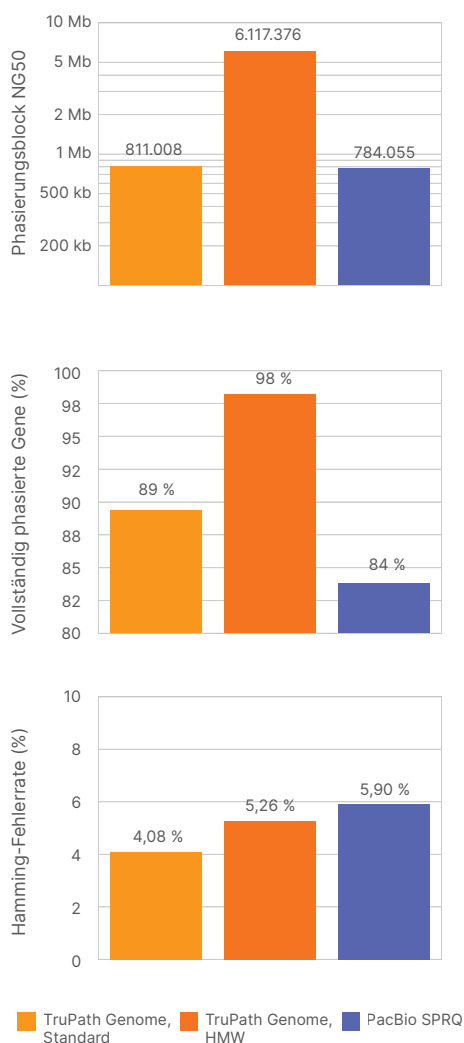


Abbildung 8: Mit TruPath Genome erzielte phasierte Sequenzierung, die einem Benchmarking mit HG002-Referenzdaten unterzogen wurde

Der Phasierungsblock wird über die Chromosomen 1–22 mit DRAGEN-Phasierung in DRAGEN Germline v4.5.2 gemessen. Die HiFi-Daten (PacBio SPRQ) Phased VCF wurden https://downloads.pacbcloud.com/public/revio/2024Q4/WGS/GIAB_trio/HG002_rep1/analysis/v3.0.2/ entnommen. Die TruPath Genome-Daten wurden mit Sequenzierungsreplikaten aus HG002-Zelllinienproben mit HMW- und Standard-Extraktion gewonnen. Der Phasierungsblock NG50 entspricht der Länge des Phasierungsblocks, sobald 50 % der Zielregion (Genom usw.) phasiert wurden. Der Prozentsatz vollständig phasierter Gene ist der prozentuale Anteil der Genregionen aus einer vorgegebenen Genliste (Gencode v44 genes.gtf), die vollständig in einem einzelnen Phasierungsblock enthalten sind. Die Phasing-Hamming-Fehlerrate wurde anhand von T2T Q100-Referenzdaten⁹ ermittelt, die von https://ftp-trace.ncbi.nlm.nih.gov/ReferenceSamples/giab/data/AshkenazimTrio/analysis/NIST_HG002_DraftBenchmark_defrabbV0.020-20250117/GRCh38_HG2-T2TQ100-V1.1_smvar.vcf.gz heruntergeladen wurden.

TruPath Genome phasiert ca. 91 % aller Gene (Median) mit standardmäßiger DNA-Extraktion und ca. 97 % aller Gene mit HMW-DNA-Extraktion. Darüber hinaus phasiert TruPath Genome ca. 98 %[†] aller heterozygoten SNVs sowohl bei der standardmäßigen als auch bei der HMW-DNA-Extraktion.

TruPath Genome-Daten mit DRAGEN-Phasierung für HMW-DNA reduzieren Hamming-Fehler** (ca. 4 %) im Vergleich zu öffentlich verfügbaren Tools wie HapCut2 (ca. 6 %).¹⁰

Varietennachweis in Regionen mit hoher Homologie

Die genaue Variantenerkennung in paralogen Regionen mit standardmäßigen kurzen Reads ist aufgrund der hohen Sequenzhomologie schwierig und hat mehrdeutiges Read-Mapping sowie Fehler beim Varietennachweis zur Folge. Die Untersuchung erfordert häufig kostspielige Reflexassays wie MLPA (Multiplex Ligation-Dependent Probe Amplifikation, multiplexe ligationsabhängige Sondenamplifikation) oder LR-PCR (Long-Range Polymerase Chain Reaction, Polymerase-Kettenreaktion mit hoher Reichweite). TruPath Genome vereint Fließzellen-Proximity-Informationen mit einem neuartigen MRJD-Algorithmus (Multiregion Joint Detection, regionsübergreifender Joint-Nachweis) zum zuverlässigen, haplotypspezifischen Calling von *De-novo*-Varianten in paralogen Genen ([Abbildung 9](#) und [Abbildung 10](#)).

TruPath Genome nimmt elf Gene ([Tabelle 1](#)) ins Visier, die bekanntermaßen über das gesamte Genom verteilte Tandem- oder Nicht-Tandem-Kopien aufweisen, ermöglicht so das hochpräzise Mapping mit MRJD und liefert haplotypspezifische Rekonstruktionen sowie das Calling kleiner Varianten ([Abbildung 10](#)).

Tabelle 1: Gene, die TruPath Genome in die MRJD-Analyse einbezieht

| | | |
|----------------|--------------------------------|---------------|
| <i>CFHR</i> | <i>NCF1</i> | <i>SMN1/2</i> |
| <i>CYP2D6</i> | <i>PMS2</i> | <i>STRC</i> |
| <i>CYP11B1</i> | <i>RCCX (CYP21A2 und TNXB)</i> | <i>USP8</i> |

[†] Beim Prozentsatz phasierter heterozygoter Varianten handelt es sich um den prozentualen Anteil phasierter heterozygoter SNVs, d. h. den Quotienten aus der Anzahl phasierter SNVs und der Anzahl heterozygoter SNVs.

** Hamming-Fehler treten auf, wenn Varianten dem falschen mütterlichen oder väterlichen Strang zugeordnet werden.

Verbesserte Auflösung bei kurzen Tandemwiederholungen (STRs, Short Tandem Repeats)

STRs sind repetitive Regionen des Genoms, die Längen über den Normalbereich hinaus erreichen und Mutationen verursachen können, die mit zahlreichen genetischen Erkrankungen assoziiert sind, darunter das Fragile-X-Syndrom, amyotrophe Lateralsklerose und Chorea Huntington.¹¹ Eine genaue Bestimmung der Größe und Read-Wiederherstellung dieser Regionen war bislang schwierig, da die relevantesten expandierten Allele häufig länger sind als die Reads standardmäßiger Daten der Short-Read-Sequenzierung. TruPath Genome ermöglicht durch eine Kombination aus Proximity-Informationen und hochwertigen Short-Read-Daten eine verbesserte Wiederherstellung der Reads, die für eine genaue Bestimmung der Größe von STRs sowie höhere Auflösung erforderlich sind, indem eine Genotypisierung der beiden Kopien mit phasierten Informationen erfolgt (Abbildung 11).

Erkenntnisse zu komplexen SVs über ultralange Spannen hinweg

Mit TruPath Genome lassen sich dank neuer Funktionen, die über herkömmliche Performance-Benchmarks für das Varianten-Calling hinausgehen, große strukturelle Rearrangements analysieren. TruPath Genome erfasst Informationen über Reads aus proximalen Clustern^{††} zwischen jedem Paar von Genomregionen und ermöglicht damit die Erstellung hochauflösender grafischer Genomstrukturkarten, die als „Colocation-Plots“ bezeichnet werden (Abbildung 12).

†† Proximale Cluster befinden sich auf der Fließzelle in physischer Nähe zueinander.

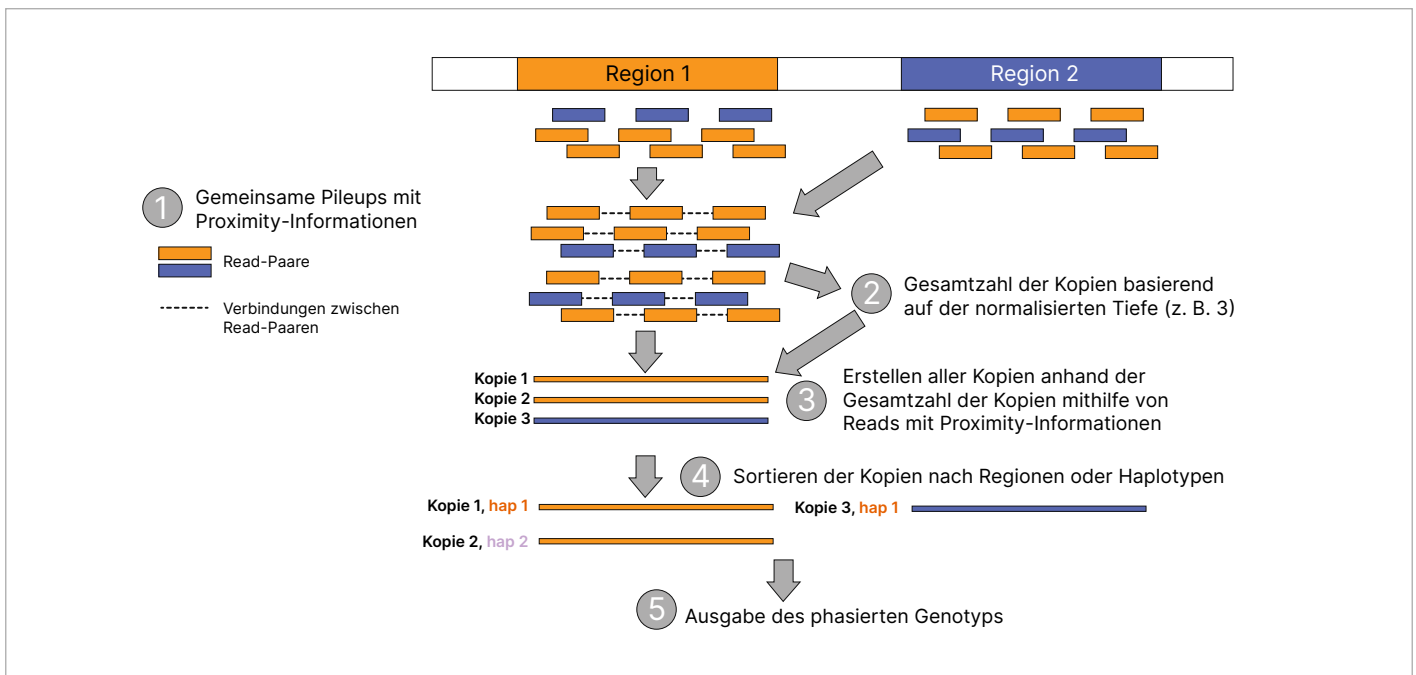
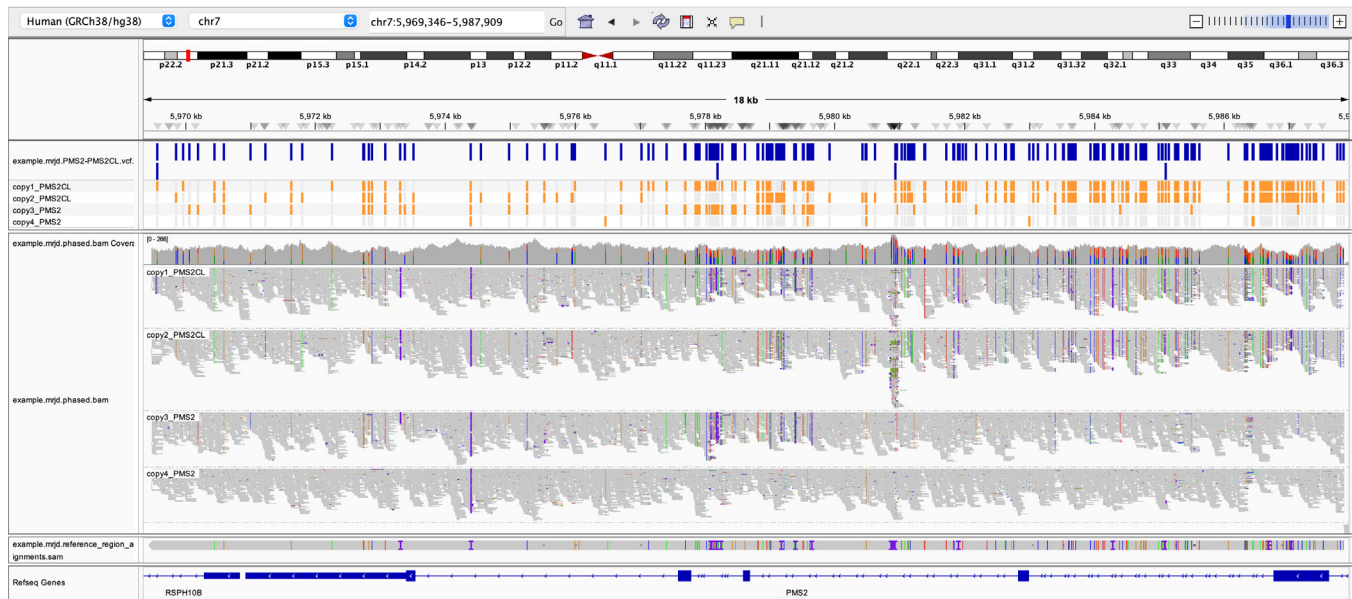


Abbildung 9: Schematische Darstellung des MRJD-Verfahrens zum Nachweis von Varianten in Regionen mit segmentaler Duplizierung

1. MRJD extrahiert für einen gegebenen Satz segmental duplizierter Regionen alle Reads, die auf eine der Kopien in diesem Regionssatz für die gemeinsame Analyse gemappt sind. Diese Reads werden anhand von Proximity-Informationen in Beziehung zueinander gesetzt und es werden Read-Sätze ermittelt, die aus demselben ursprünglichen Molekül stammen. 2. Anschließend wird die Anzahl aller in Schritt 1 extrahierten Reads normalisiert, GC-korrigiert und zur Ermittlung der Gesamtzahl der Kopien der segmental duplizierten Sequenz in der Zugabeprobe verwendet. 3. Ein Ansatz zur gemeinsamen Genotypisierung, der auf der Anzahl der Kopien der Region basiert, die in Schritt 2 ermittelt und auf den vollständigen Satz proximityverknüpfter Reads angewendet wird, dient zur Ermittlung und Phasierung sämtlicher Varianten in jeder Kopie der segmental duplizierten Sequenz in der Zugabeprobe. 4. Jede vollständig phasierte Kopie wird einer bestimmten Referenzgenomposition zugewiesen (wenn sich Regionen an distalen Referenzpositionen befinden) oder zu den anderen Kopien in Haplotypen phasiert (nur möglich, wenn Regionen mit segmentaler Duplikation in der Referenz im Tandem/dicht nebeneinander liegen). 5. Vollständig phasierte Varianten für alle Kopien, die Haplotypzuweisung jeder Kopie, die Zuweisung von Kopien zu jeder Referenzgenomposition und die jeder Kopie zugewiesenen Reads werden in die MRJD-spezifischen VCF-, JSON- und BAM-Dateien ausgegeben.

A. PMS2 gene



B. SMN gene

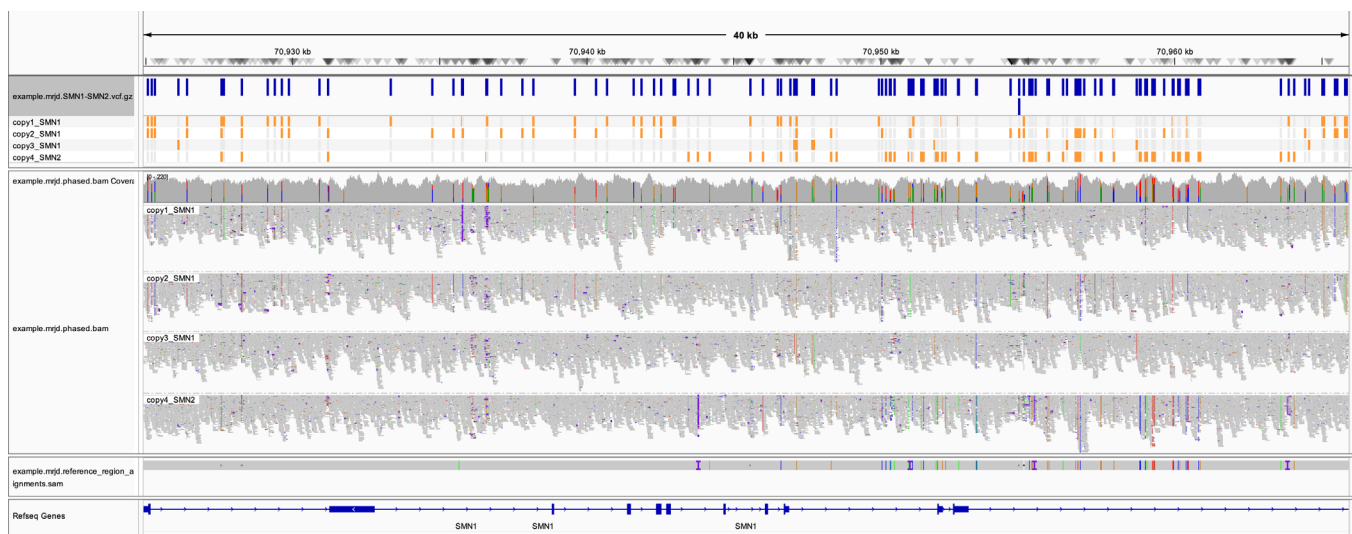


Abbildung 10: Beispiel für die Ausgabe von Varianten-Calls und Read-Zuweisung in paralogen Genen mit TruPath Genome und MJRD

(A) Phasierte Varianten-Calls und zugehörige bestätigende Reads für alle Kopien des Gen-Pseudogen-Paars *PMS2/PMS2CL* an der Position von *PMS2* im Referenzgenom. Zwei der vier Kopien sind als *PMS2CL*-Kopien (Pseudogen) gekennzeichnet und zwei als *PMS2*-Kopien (aktives Gen). Diese Kennzeichnung basiert auf Proximity-Verknüpfungen mit den *PMS2/PMS2CL*-Flanken. Copy3 (*PMS2*) enthält ein mutmaßliches Genkonversionsereignis, das die Exons 13 und 14 überlappt, was das Vorhandensein von Varianten zeigt, die der *PMS2CL*-Referenzsequenz entsprechen. Das Alignment der *PMS2CL*- und der *PMS2*-Referenzsequenz ist im dritten Bereich (von oben) dargestellt. (B) Phasierte Varianten-Calls und zugehörige bestätigende Reads für alle Kopien des Gen-Pseudogen-Paars *SMN1/SMN2* an der Position von *SMN1* im Referenzgenom. Drei der vier Kopien sind als *SMN1* (aktives Gen) und zwei als *SMN2* (inaktives Gen) gekennzeichnet. Diese Kennzeichnung erfolgt anhand der Base an der Position SMN1 c.840 in jeder der phasierten Kopien. (*SMN1*-Kopien haben an dieser Position die Base C und *SMN2*-Kopien die Base T).

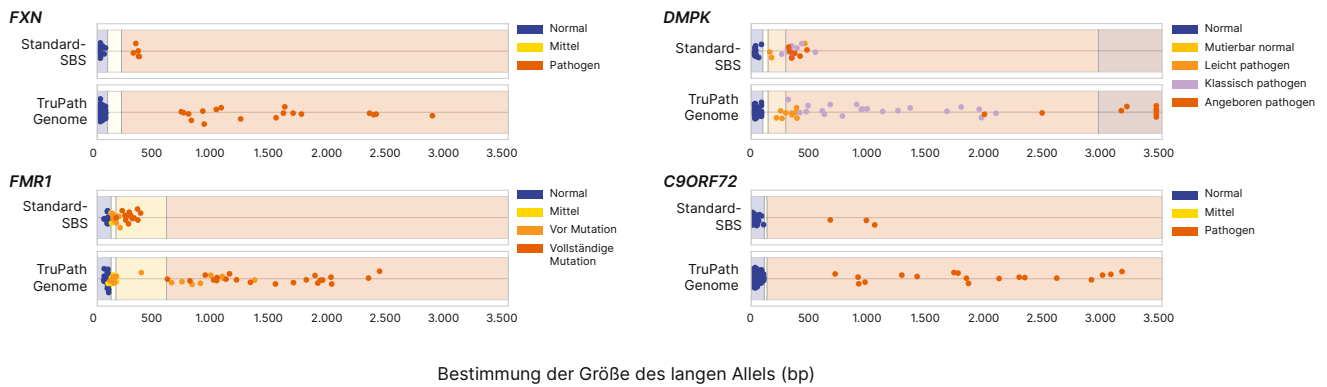


Abbildung 11: Genauere Bestimmung von STR-Expansionen mit TruPath Genome

Tatsächliche Klassifizierungen von STRs werden durch die Farbe der einzelnen Punkte dargestellt. Farbbänder stellen die unterschiedlichen Klassifizierungsbereiche dar, die der Definition der unterschiedlichen Expansionsklassen an den einzelnen Loci dienen. Die X-Achse gibt die mit DRAGEN STR ermittelte STR-Größe an, wobei entweder TruPath Genome oder Standard-SBS (Illumina DNA PCR-Free Prep) als Zugabe verwendet wird. TruPath Genome-Klassifizierungen entsprechen wesentlich deutlicher den tatsächlichen Klassifizierungen und umfassen ein breiteres Spektrum an STR-Längen. Gesamtzahl der Replikate für Standard-SBS: 78 Bioproben (78 eindeutige Einzelpersonen). Gesamtzahl der Replikate für TruPath Genome: 135 Bioproben (42 eindeutige Einzelpersonen).

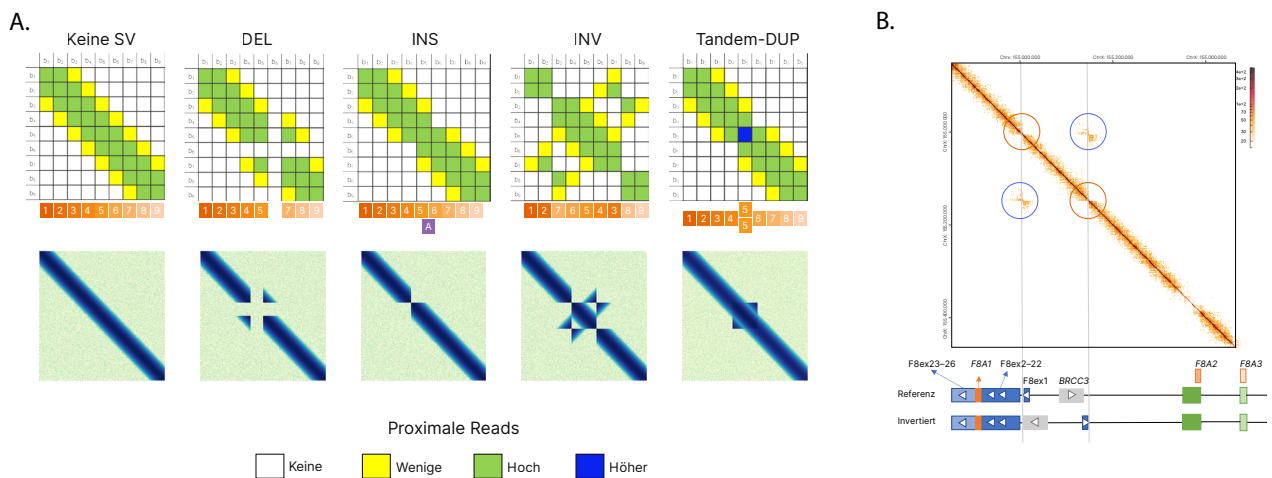


Abbildung 12: Visualisierung struktureller Variantenergebnisse mit TruPath Genome-Colocation-Plots

(A) Die Anzahl der proximalen Reads (Ort auf der Fließzelle) für alle Genom-Bin-Paare wird zur Generierung einer zweidimensionalen Colocation-Karte bestimmt. Die X- und Y-Achse stellen Referenzgenom-Bins dar. Eine hohe Anzahl an Verbindungen zwischen zwei Bins deutet darauf hin, dass sich diese Bins im Genom in der Zugabeprobe in räumlicher Nähe zueinander befinden. In einer Region ohne strukturelle Varianten befinden sich Genom-Bins, die sich im Referenzgenom und auch in der Zugabeprobe in räumlicher Nähe zueinander befinden und daher als eine hohe Anzahl von Verbindungen nahe der Diagonale des Colocation-Plots erscheinen. Ist eine strukturelle Variante vorhanden, befinden sich Genom-Bins, die sich im Referenzgenom in räumlicher Nähe zueinander befinden, in der Zugabeprobe unter Umständen nicht mehr in räumlicher Nähe zueinander. Gleichermäßen können Genom-Bins, die in der Referenz weit entfernt voneinander sind, in der Zugabeprobe proximal sein. Derartige Änderungen der Verbindung zeigen sich auf den Colocation-Karten als nicht diagonale Signale unterschiedlicher Formen sowie als fehlende diagonale Verbindung. (B) Ein Beispiel für eine Colocation-Karte für eine Intron-1-Geninversion am Genlocus *F8*, die mit schwerer Hämophilie A assoziiert ist. Beachtenswert sind das erwartete sanduhrförmige Signal in der Region abseits der Diagonale (blaue Kreise) und das Fehlen eines diagonalen Signals (orange-farbene Kreise) in den Ereignisgrenzen. Unter dem Plot befindet sich eine schematische Darstellung des Locus im Referenzgenom und der invertierten Genregion.

Der einfachste Workflow für die Genomsequenzierung

Illumina TruPath Genome revolutioniert mit dem bahnbrechenden Proximity-Read-Mapping die Humangenomsequenzierung. Beim innovativen Workflow entfällt die herkömmliche Bibliotheksvorbereitung, was für größere Einfachheit sorgt, während durch die Proximity-Informationen die Long-Distance-DNA-Stranginformationen als hochpräzise, standardmäßige Short-Read-Sequenzierungsdaten erhalten bleiben. Das Ergebnis ist ein Assay, der die Vorteile der Datenqualität kurzer Reads, den hochpräzisen Nachweis kleiner Varianten mit verbesserter Mapping-Coverage des Humangenoms und ein leistungsstarkes Tool zur Erforschung genetischer Erkrankungen vereint.

Weitere Informationen →

[Illumina TruPath Genome](#)

[Proximity-Read-Mapping](#)

[NovaSeq X Series](#)

| Bestellinformationen | |
|--|-------------|
| Produkt | Katalog-Nr. |
| Illumina TruPath Genome (NovaSeq X C8 flow cell) | 20157405 |
| Illumina TruPath Genome (NovaSeq X C2 flow cell) | 20157406 |

Quellen

1. Ebbert MTW, Jensen TD, Jansen-West K, et al. [Systematic analysis of dark and camouflaged genes reveals disease-relevant genes hiding in plain sight](#). *Genome Biol.* 2019;20(1):97. Veröffentlicht am 20. Mai 2019. doi:10.1186/s13059-019-1707-2
2. Ryan NM, Corvin A. [Investigating the dark-side of the genome: a barrier to human disease variant discovery?](#) *Biol Res.* 2023;56(1):42. Veröffentlicht am 20. Juli 2023. doi:10.1186/s40659-023-00455-0
3. Illumina. TruPath Genome performance with samples of varying type and quality. [illumina.com/content/dam/illumina-marketing/documents/products/technotes/trupath-genome-sample-info-tech-note-m-gl-03932](https://www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/trupath-genome-sample-info-tech-note-m-gl-03932). Veröffentlicht im Februar 2026. Aufgerufen am XX. Februar 2026.
4. Wadsworth ME, Page ML, Aguzzoli Heberle B, Miller JB, Steely C, Ebbert MTW. [Sequencing the gaps: dark genomic regions persist in CHM13 despite long-read advances](#). Vorabdruck. *bioRxiv.* 2025;2025.05.23.655776. Veröffentlicht am 28. Mai 2025. doi:10.1101/2025.05.23.655776
5. Ultima Genomics. Genome-in-a-Bottle (GIAB) samples HG001-HG005. <https://cdn.sanity.io/files/l7780ks7/production-2024/0a1b6a62a6da3e3fcafb81cad4c8ff2ffe85dd41.pdf>. Veröffentlicht im Februar 2025. Aufgerufen am 29. Januar 2026.
6. PacBio. Index of /public/revio/2024Q4/WGS/GIAB_trio/HG002_rep1/analysis/v3.0.2. https://downloads.paccloud.com/public/revio/2024Q4/WGS/GIAB_trio/HG002_rep1/analysis/v3.0.2/. Aufgerufen am 29. Januar 2026.
7. Asri M, Chang PC, Mier JC, et al. [Pangenome-aware DeepVariant](#). Vorabdruck. *bioRxiv.* 2025;2025.06.05.657102. Veröffentlicht am 6. Juni 2025. doi:10.1101/2025.06.05.657102
8. Browning SR, Browning BL. [Haplotype phasing: existing methods and new developments](#). *Nat Rev Genet.* 2011; 12(10):703-714. doi: 10.1038/nrg3054
9. Hansen NF, Dwarshuis N, Ji HJ, et al. [A complete diploid human genome benchmark for personalized genomics](#). Vorabdruck. *bioRxiv.* 2025;2025.09.21.677443. Veröffentlicht am 21. September 2025. doi:10.1101/2025.09.21.677443
10. Archivdaten. Illumina, Inc., 2026.
11. Steely CJ, Watkins WS, Baird L, Jorde LB. [The mutational dynamics of short tandem repeats in large, multigenerational families](#). *Genome Biol.* 2022;23(1):253. Veröffentlicht am 12. Dezember 2022. doi:10.1186/s13059-022-02818-4



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

© 2026 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-03931 DEU v1.0