

illumina®

TruSight Whole Genome Analysis Application

Produktdokumentation

ILLUMINA – EIGENTUMSRECHTLICH GESCHÜTZT

Dokument-Nr. 200049931 v00

April 2024

FÜR IN-VITRO-DIAGNOSTIK.

Versionsverlauf

Dokument	Datum	Beschreibung der Änderung
Dokument-Nr. 200049931 v00	April 2024	Erste Version.

Dieses Dokument und sein Inhalt sind Eigentum von Illumina, Inc. sowie deren Partner-/Tochterunternehmen („Illumina“) und ausschließlich für den bestimmungsgemäßen Gebrauch durch den Kunden in Verbindung mit der Verwendung des hier beschriebenen Produkts/der hier beschriebenen Produkte und für keinen anderen Bestimmungszweck ausgelegt. Dieses Dokument und sein Inhalt dürfen ohne schriftliches Einverständnis von Illumina zu keinem anderen Zweck verwendet oder verteilt bzw. anderweitig übermittelt, offengelegt oder auf irgendeine Weise reproduziert werden. Illumina überträgt mit diesem Dokument keine Lizenzen unter seinem Patent, Markenzeichen, Urheberrecht oder bürgerlichem Recht bzw. ähnlichen Rechten an Drittparteien.

Die Anweisungen in diesem Dokument müssen von qualifiziertem und entsprechend ausgebildetem Personal genau befolgt werden, damit die in diesem Dokument beschriebene Verwendung des Produkts/der Produkte sicher und ordnungsgemäß erfolgt. Vor der Verwendung dieser Produkte muss der Inhalt dieses Dokuments vollständig gelesen und verstanden worden sein.

FALLS NICHT ALLE HIERIN AUFGEFÜHRTEN ANWEISUNGEN VOLLSTÄNDIG GELESEN UND BEFOLGT WERDEN, KÖNNEN PRODUKTSCHÄDEN, VERLETZUNGEN DER BENUTZER UND ANDERER PERSONEN SOWIE ANDERWEITIGER SACHSCHADEN EINTRETEN UND JEDLICHE FÜR DAS PRODUKT/DIE PRODUKTE GELTENDE GEWÄHRLEISTUNG ERLISCHT.

ILLUMINA ÜBERNIMMT KEINERLEI HAFTUNG FÜR SCHÄDEN, DIE AUS DER UNSACHGEMÄSSEN VERWENDUNG DER HIERIN BESCHRIEBENEN PRODUKTE (EINSCHLIESSLICH TEILEN HIERVON ODER DER SOFTWARE) ENTSTEHEN.

© 2024 Illumina, Inc. Alle Rechte vorbehalten.

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

Inhaltsverzeichnis

Versionsverlauf	ii
Überblick	1
Erste Schritte	1
Einstellungen	2
Erstellung eines Laufs	4
Revision ausführen	5
Analyse erneut in die Warteschlange stellen	6
Anzeigen von Lauf und Ergebnissen	7
Zusammenfassung der Ausgabedatei	9
Informationen zum QC-Bericht	9
Informationen zur Varianten-Call-Datei	14
FASTQ-Dateien	22
CRAM-Dateien	23
Technische Unterstützung	25

Überblick

Die TruSight Whole Genome Analysis Application wird verwendet, um Sequenzierumläufe für TruSight Whole Genome zu planen und die Analyse automatisch anzustoßen, nachdem der Lauf abgeschlossen ist. Die Analyse umfasst Demultiplexing, FASTQ-Generierung, Read Mapping, Alignment auf das graphisch aktivierte humane Referenzgenom GrCh38/hg38 und Varianten-Calls unter Verwendung des Illumina DRAGEN-Server for NovaSeq 6000Dx.

In verschiedenen Phasen des Analyse-Workflows führt die Anwendung eine Qualitätskontrolle (QC) gemäß definierter Sequenzierung, FASTQ und Probenbibliotheksmetriken durch und generiert Berichte mit den Ergebnissen. Für Proben, die alle QC-Schritte bestehen, generiert die Anwendung unterstützende Ausgabedateien zur Verwendung in nachfolgenden Keimbahnanwendungen.

Die TruSight Whole Genome Analysis Application führt DRAGEN-Varianten-Caller, u. a. den Small Variant Caller, Copy Number Variant (CNV) Caller und die Repeat Expansion-Bestimmung, mit dem ExpansionHunter aus.

Die Anwendung erstellt auch Annotationen zu niedrigen, mittleren oder hohen Konfidenzstufen für kleine Varianten und fügt diese Annotation der Ausgabedatei hinzu.

Erste Schritte

Stellen Sie sicher, dass die TruSight Whole Genome Analysis Application auf dem NovaSeq 6000Dx-Gerät installiert ist, das für die Sequenzierung im Rahmen des TruSight Whole Genome verwendet wird. Installierte Anwendungen finden Sie auf dem Anwendungsbildschirm auf dem NovaSeq 6000Dx Instrument oder im Illumina Run Manager bei Verwendung eines Browsers auf einem vernetzten Computer. Wenn Sie Hilfe bei der Planung der Installation benötigen, wenden Sie sich an Ihren Illumina-Vertreter vor Ort.

Anforderungen an den Datenspeicher

Siehe NovaSeq 6000Dx Produktdokumentation (Dokument-Nr. 200010105) und DRAGEN Server for NovaSeq 6000Dx Produktdokumentation (Dokument-Nr. 200014171) für allgemeine Informationen zu Datenausgabe und -speicherung.

Die TruSight Whole Genome Analysis Application gibt Daten in einen Run-Ordner und einen Analyseordner im externen Speicher aus. Die Mindestspeicheranforderungen können anhand der Größe der Datenausgabe in jedem Ordner für einen einzelnen Sequenzierungslauf, wie unten dargestellt, geschätzt werden.

Konfiguration	Laufordner (GB)	Analyseordner (GB)
S2-Fließzelle (6 Proben)	~430	~350
S4-Fließzelle (16 Proben)	~1110	~890

Ungefähre Analysezeit

Die Analyse beginnt automatisch nach Abschluss eines Sequenzierungslaufs und erfolgt nacheinander mit Proben innerhalb eines Laufs. Die Datenausgabedateien sind auf dem externen Speicher verfügbar, sobald die Analyse für alle Proben in einem Lauf abgeschlossen wurde und die Übertragung der Kopie an den externen Speicher abgeschlossen ist. Bei gleichzeitigem Starten eines Sequenzierungslaufs auf Seite A und Seite B wird die Sequenzierung parallel durchgeführt. Die Analyse dieser Sequenzierungsläufe erfolgt nacheinander durch die TruSight Whole Genome Analysis Application nach Abschluss der Sequenzierung. Der Lauf, der Sequenzierung und Übertragung zuerst abschließt, wird zuerst analysiert. Der zweite Sequenzierungslauf wird nach Abschluss der ersten Analyse übertragen und zur Analyse in die Warteschlange gestellt. Informationen zur Bestimmung des Status von aktiven oder fehlgeschlagenen Läufen finden Sie unter [Anzeigen von Lauf und Ergebnissen auf Seite 7](#).

Die ungefähre Zeit bis zum Vorliegen der Analyseergebnisse nach Abschluss der Sequenzierung ist unten für die Situation dargestellt, in der Seite A und Seite B gleichzeitig mit derselben Konfiguration geladen werden.

Konfiguration	Analyselauflauf 1 (Stunden)	Analyselauflauf 2 (Stunden)
S2-Fließzelle (6 Proben)	~12	~24
S4-Fließzelle (16 Proben)	~24	~48

Einstellungen

Wählen Sie die TruSight Whole Genome Analysis Application im Bildschirm Applications (Anwendungen) aus, um die aktuelle Konfiguration anzuzeigen und Berechtigungen zu ändern.

Konfiguration

Im Bildschirm Configuration (Konfiguration) werden diese Anwendungseinstellungen angezeigt:

- **Application Name** (Anwendungsname)
- **Application Version** (Anwendungsversion)
- **DRAGEN Version**
- **RTA Version** (RTA-Version)
- **Release Date** (Versionsdatum)
- **Organization** (Organisation)
- **Device Identifier** (Kennung des Produkts)
- **Production Identifier** (Produktions-ID)
- **Library Prep Kits** (Bibliotheksvorbereitungs-Kits): Zeigt das Bibliotheksvorbereitungs-Kit an. Diese Einstellung kann nicht geändert werden.

- **Index Adapter Kits** (Indexadaptersätze): Zeigt die zur Verwendung verfügbaren Indexadaptersätze an.
- **Index Reads** (Index-Reads)
- **Read Type** (Read-Typ)
- **Index Lengths** (Index-Längen)
- **Read Lengths** (Read-Längen): werden standardmäßig eingestellt, wenn der Indexsatz ausgewählt ist. Diese Einstellung kann nicht geändert werden.

Berechtigungen

Der designierte Administrator hat Berechtigungszugriff und kann die Kontrollkästchen im Bildschirm Permissions (Berechtigungen) verwenden, um den Benutzerzugriff für TruSight Whole Genome Analysis Application zu verwalten.

Weitere Informationen zu Berechtigungen und Benutzerverwaltung finden Sie im Abschnitt Systemkonfiguration von NovaSeq 6000Dx Produktdokumentation (Dokument-Nr. 200010105).

Erstellung eines Laufs

Erstellen Sie neue Läufe im IVD-Modus entweder auf dem Gerät oder durch Zugriff auf Illumina Run Manager (IRM) über einen Browser auf einem Computer im Netzwerk. Um aus der Ferne auf das Gerät zuzugreifen, benutzen Sie die von Ihrem Illumina-Vertreter bereitgestellten Adress- und Benutzerkontoinformationen. Weitere Informationen finden Sie unter NovaSeq 6000Dx Produktdokumentation (Dokument-Nr. 200010105).

Create Run (Lauferstellung) ist die empfohlene Methode für die Laufplanung. Import Sample Sheet (Probenblatt importieren) wird nicht empfohlen. Die in Lauf- und Analyseordnern ausgegebenen Probenblattdateien sind nicht für den Import während der Laufplanung geeignet.

Läufe erstellen

1. Wählen Sie im Bildschirm Runs (Läufe) **Create Run** (Lauferstellung) aus.
2. Wählen Sie die Anwendung TruSight Whole Genome Analysis Application und anschließend **Next** (Weiter) aus.
3. Geben Sie im Bildschirm Settings (Einstellungen) einen Laufnamen ein. Der Laufname ist der Name, der den Lauf von der Sequenzierung bis hin zur Analyse identifiziert.
4. [Optional] Geben Sie eine Laufbeschreibung ein, um den Lauf näher zu bestimmen. Das Bibliotheksvorbereitungskit ist standardmäßig auf TruSight Whole Genome eingestellt und kann nicht geändert werden.
5. Wählen Sie den gewünschten TruSight Whole Genome-Indexsatz aus dem Dropdown-Menü **Index Adapter Kit** (Indexadaptersätze) aus.
Die Readlänge wird standardmäßig eingestellt und kann nicht geändert werden. (Für Read 1 und 2 werden 151 Zyklen verwendet; für Index 1 und 2 werden 10 Zyklen verwendet).
6. Geben Sie eine Bibliotheksröhrchen-ID ein (empfohlenes Format: DX1234567-LIB) und wählen Sie dann **Next** (Weiter) aus.
Wenn in diesem Schritt keine Bibliotheksröhrchen-ID angegeben ist, muss der geplante Lauf vor dem Laden von Sequenzierungs-Verbrauchsmaterialien ausgewählt werden. Wird in diesem Schritt die falsche Bibliotheksröhrchen-ID eingegeben, muss der geplante Lauf vor dem Laden von Verbrauchsmaterialien korrigiert werden. Das Protokoll zur Korrektur des Laufs, wenn das Verbrauchsmaterial geladen werden kann, finden Sie im Abschnitt [Revision ausführen auf Seite 5](#).
7. Auf den Bildschirmen Sample Data (Probanden) und Sample Settings (Probeneinstellungen) werden Probeninformationen eingegeben. Probanden können manuell oder durch Importieren einer Probendatei eingegeben werden. Die Proben-ID muss für jede Probe eindeutig sein und darf nur alphanumerische Zeichen, Unterstriche und Bindestriche enthalten. Verwenden Sie keine Leerzeichen. Die Well-Position bezieht sich auf das Well im Format A01 bis H04 der Indexplatte. Die Indexsequenzinformationen werden automatisch ausgefüllt, wenn die Well-Position der Indexplatte

eingetragen wird. Das Geschlecht muss als männlich, weiblich oder unbekannt eingegeben werden. Library Plate ID (Bibliotheksplatten-ID) und Library Well ID, (Bibliotheks-Well-ID) (z. B. Format A01) sind Pflichtfelder.

- Um Probanddaten manuell einzugeben, fügen Sie Zeilen hinzu (insgesamt 6 für S2- oder 16 für S4-Fließzellen) und geben Sie die erforderlichen Informationen in die Felder Sample ID (Proben-ID) und Well Position (Well-Position) ein. Informationen können auch aus Excel kopiert und eingefügt werden. Wählen Sie **Next** (Weiter) aus. Geben Sie im Bildschirm Sample Settings (Probeneinstellungen) die Bibliotheksplatten-ID, die Bibliotheks-Well-ID und das Geschlecht ein. Wählen Sie **Next** (Weiter) aus.
 - Um eine Probandendatei zu importieren, wählen Sie **Import Samples** (Proben importieren) und laden Sie die Probandendatei hoch. Die Informationen werden automatisch in Zeilen eingetragen. Auf diesem Bildschirm steht eine Vorlage (*.csv) zum Download zur Verfügung. Wählen Sie **Next** (Weiter) aus. Auf dem Bildschirm Probeneinstellungen werden Informationen automatisch aus der importierten Probandendatei in Zeilen eingetragen. Wählen Sie **Next** (Weiter) aus.
8. Geben Sie im Bildschirm Analysis Settings (Analyseinstellungen) den Batch Name (Chargennamen) ein, der während der Chargen- und Laufplanung aufgezeichnet wurde.
 9. [Optional] Wählen Sie den Fließzellentyp S2 oder S4 aus.
 10. Bestätigen oder deaktivieren Sie das Kontrollkästchen, um ORA-komprimierte FASTQs zu generieren, und wählen Sie **Next** (Weiter) aus.

HINWEIS Der TruSight Whole Genome Analysis Application generiert standardmäßig ORA-komprimierte FASTQs. Wenn Sie diese Einstellung ändern, wird die Größe der endgültigen Datenausgabe erhöht.

11. Überprüfen Sie im Bildschirm Run Review (Lauf überprüfen) die eingegebenen Informationen. Wenn keine Änderungen erforderlich sind, wählen Sie **Save** (Speichern). Wenn Änderungen erforderlich sind, wählen Sie ggfs. **Back** (Zurück), um zum entsprechenden Bildschirm zurückzukehren.



VORSICHT

TruSight Whole Genome wurde bei Verwendung der NovaSeq 6000Dx-S2-Fließzelle für 6 Proben und bei Verwendung der NovaSeq 6000Dx-S4-Fließzelle für 16 Proben validiert. Stellen Sie sicher, dass die richtige Anzahl von Proben für die ausgewählte Fließzellenkonfiguration eingegeben wurde.

Revision ausführen

Wenn nach der Erstellung des Laufs und vor dem Laden von Verbrauchsmaterialien für die Sequenzierung Änderungen erforderlich sind, überarbeiten Sie die Läufe im IVD-Modus entweder auf dem Gerät oder indem Sie auf Illumina Run Manager (IRM) über einen Browser auf einem vernetzten Computer zugreifen.

1. Wählen Sie **Runs** (Läufe).
2. Wählen Sie den Laufnamen in der Registerkarte Planned Runs (Geplante Läufe) aus.
3. Wählen Sie **Edit** (Bearbeiten).
4. Aktualisieren Sie die Lauf- oder Probeninformationen nach Bedarf. Geben Sie beispielsweise die Bibliotheksröhrchen-ID ein oder korrigieren Sie sie, damit sie mit der übereinstimmt, die beim Abschluss des Workflows verwendet wurde.
5. Wählen Sie **Next** (Weiter), um den Lauf zu revidieren.
6. Wählen Sie **Save** (Speichern) aus.
7. Wählen Sie **Exit** (Beenden) aus.

Kehren Sie zur Sequenzierung im IVD-Modus zurück, um das Laden von Verbrauchsmaterialien zu wiederholen. Der Lauf sollte nun automatisch hervorgehoben sein.

Wenn Sie die Bibliotheksröhrchen-ID beim Laden von Verbrauchsmaterialien aktualisieren, kehren Sie zur Run-Auswahl in der Steuerungssoftware zurück und wählen Sie **Refresh** (Aktualisieren) für die zugehörige Spalte A oder B. Der Lauf sollte jetzt automatisch hervorgehoben sein. Wenn nicht, wählen Sie **Back** (Zurück), um erneut Verbrauchsmaterialien zu laden.

Analyse erneut in die Warteschlange stellen

Im Abschnitt Fehlerbehebung im TruSight Whole Genome Package Insert (Dok.-Nr. 200050132) finden Sie Informationen darüber, welche Art von Warteschlangenanalyse am besten geeignet ist.

Analyse ohne Änderungen wieder in die Warteschlange stellen

1. Wählen Sie den Namen des Completed Run (abgeschlossener Lauf) aus, um Run Details (Details zum Lauf) anzuzeigen.
2. Wählen Sie **Requeue Analysis** (Analyse erneut in die Warteschlange stellen).
3. Wählen Sie **Requeue Analysis with no changes** (Analyse ohne Änderungen wieder in die Warteschlange stellen).
4. Geben Sie im Feld Reanalysis Reason (Grund für die erneute Analyse) Details ein.
5. Wählen Sie **Requeue Analysis** (Analyse erneut in die Warteschlange stellen).
6. Verlassen Sie die Seite und navigieren Sie zur Seite Active Runs (Aktive Läufe), um zu bestätigen, dass die Warteschlange in Bearbeitung ist.

Analyse mit Änderungen wieder in die Warteschlange stellen

1. Wählen Sie den Namen des Completed Run (abgeschlossener Lauf) aus, um Run Details (Details zum Lauf) anzuzeigen.
2. Wählen Sie **Requeue Analysis** (Analyse erneut in die Warteschlange stellen).

3. Wählen Sie **Edit run settings** (Laufeinstellungen bearbeiten) und **Requeue Analysis** (Analyse erneut in die Warteschlange stellen).
4. Geben Sie im Feld Reanalysis Reason (Grund für die erneute Analyse) Details ein.
5. Wählen Sie **Requeue Analysis** (Analyse erneut in die Warteschlange stellen).
6. Bestätigen oder aktualisieren Sie die Lauf-Einstellungen und wählen Sie dann **Next** (Weiter).
7. Korrigieren Sie die Beispielinformationen nach Bedarf, indem Sie die Felder manuell aktualisieren, oder wählen Sie **Download Template** (Vorlage herunterladen), um eine Datei `sampledata.csv` mit aktuellen Informationen zu erstellen. Korrigieren Sie Informationen und löschen Sie vorhandene Zeilen auf der Registerkarte Sample Data (Probendaten), bevor Sie Import Samples (Proben importieren) verwenden, um die korrigierten Probendaten einzupflegen.
8. Überprüfen Sie auf dem Bildschirm Review (Überprüfen) die Informationen zum Lauf und wählen Sie **Save** (Speichern) aus.
9. Wählen Sie **Exit** (Beenden) und navigieren Sie zur Seite Active Runs (Aktive Läufe), um zu bestätigen, dass die Warteschlange ausgeführt wird.
Der ursprüngliche Laufdatenordner muss an dem externen Speicherort vorhanden sein, der in den Laufdetails angegeben ist, damit die erneute Analyse erfolgreich abgeschlossen werden kann. Wenn die erneute Analyse fehlschlägt, stellen Sie sicher, dass der Lauf nicht verschoben oder gelöscht wurde.

Anzeigen von Lauf und Ergebnissen

1. Wählen Sie auf dem Illumina Run Manager Hauptbildschirm im IVD-Modus **Runs** (Läufe) aus.
2. Wählen Sie auf der Registerkarte Completed Runs (Abgeschlossene Läufe) den Laufnamen aus. Auf dieser Registerkarte werden auch Läufe angezeigt, die aufgrund einer fehlgeschlagenen Sequenzierung, Datenübertragung oder Analyse abgeschlossen wurden. Aktive Läufe und ihr Status werden auf der Registerkarte Active Runs (Aktive Läufe) angezeigt. Weitere Informationen finden Sie unter NovaSeq 6000Dx Produktdokumentation (Dokument-Nr. 200010105).
3. Wählen Sie auf der Registerkarte Completed Runs (Abgeschlossene Läufe) den Laufnamen aus, um die Laufdetails und Ergebnisse für den Pfad zum Analyseausgabeordner anzuzeigen. Überprüfen Sie bei fehlgeschlagenen Läufen den Status für jeden Schritt und lesen Sie dann den Abschnitt Fehlerbehebung im TruSight Whole Genome Package Insert (Dok.-Nr. 200050132).
4. Navigieren Sie zum Analyseordner auf Ihrem lokalen Laufwerk und öffnen Sie den konsolidierten Bericht, um das Ergebnis PASS/FAIL (bestanden/fehlgeschlagen) für jeden QC-Schritt wie folgt zu überprüfen:
 - Informationen zur Sequenzierungslauf-QC finden Sie unter Summary Sequencing QC Result (Zusammenfassung des QC-Ergebnisses zur Sequenzierung)
 - Informationen zur FASTQ-QC für jede Probe im Lauf finden Sie unter Summary FASTQ QC Result (QC-Ergebnis für FASTQ gesamt)

- Informationen zur Bibliotheks-QC für jede Probe im Lauf finden Sie unter Summary Sample Library QC Result (Zusammenfassung des QC-Ergebnisses der Probenbibliothek)
Wenn das Ergebnis FAIL (fehlgeschlagen) lautet, notieren Sie den QC-Schritt und lesen Sie im Abschnitt Fehlerbehebung im TruSight Whole Genome Package Insert (Dok.-Nr. 200050132) nach.

Zusammenfassung der Ausgabedatei

TruSight Whole Genome Analysis Application speichert die folgenden Hauptausgabedateien. Die Speicherorte der Hauptausgabedateien finden Sie in den Dateiinformationsabschnitten unten.

Läufe und Proben, die die Gültigkeitskriterien nicht erfüllen, erzeugen keine CRAM-, ROH-BED- oder *geome.vcf-Dateien.

Ausgabedatei	Beschreibung
Konsolidierter Bericht (*.csv)	Enthält Qualitätsmetriken, die für die Laufzeitgültigkeit (einschließlich Gesamtergebnis und Q30), Probenvaliditätsmetriken (einschließlich FASTQ-Ergebnis), Bibliotheks-QC-Metriken und Nur zur Information (FIO)-Metriken für alle Proben im Lauf verwendet werden.
Probenbericht (*.csv)	Enthält die Sequenzierungsergebnisse für QC, FASTQ und für die QC der Probenbibliothek. Der Bericht enthält auch Ploidie-Konkordanz- und FIO-Metriken für die einzelne Probe sowie den zugehörigen Sequenzierungslauf.
Small Variant und mSNV VCF (*.annotated.hard-filtered-gvcf.gz)	Enthält Varianten-Call-Informationen für kleine Varianten (SNVs, Indels) und mitochondriale SNVs.
CNV VCF (*.cnv.vcf.gz)	Enthält Informationen zu Varianten-Calls für Zunahmen und Verluste von Kopienzahlen.
Repeats VCF (*.repeats.vcf.gz)	Enthält Informationen zu Varianten-Calls für STR-Expansionen und SMN1.
ROH BED (*.roh.bed)	Enthält Informationen für Homozygotie-Regionen.
FASTQ (*.fastq.gz oder *.fastq.ora)	Temporäre Dateien mit qualitativ benoteten Base-Calls. FASTQ-Dateien sind die primären Eingabedateien für den Alignment-Schritt. Wenn die ORA-Komprimierung ausgewählt ist, wird das im Dateinamen wiedergegeben.
Alignment CRAM (*.cram)	Enthält alignierte Reads für eine bestimmte Probe.

Informationen zum QC-Bericht

Der konsolidierte Bericht <<RunID>>_Consolidated_Report.csv befindet sich im Verzeichnis TruSightWholeGenomeAnalysis_x.x.x_run-complete und enthält Informationen über Qualitätsmetriken, die verwendet werden, um Proben in verschiedenen Analysephasen bestehen oder

fehlschlagen zu lassen. Einzelne Probenberichte <<Sample_ID>>_Sample_Report.csv befinden sich in den Ordnern <Sample_ID> im Verzeichnis TruSightWholeGenomeAnalysis_x.x.x_run-complete.

Die Kopfzeilen der Berichte enthalten die folgenden Informationen über den Lauf: die App-Version, den Chargennamen, die Röhrchen-ID des Bibliothekspools, den Namen des Sequenzierungslaufs, die Sequenzierungslauf-ID und den Fließzellentyp. Die folgenden Tabellen beschreiben die im konsolidierten Bericht enthaltenen Informationen. Der einzelne Probenbericht enthält die gleichen Informationen mit Ausnahme der Demultiplex-Metriken.

Tabelle 1 Metriken für die Sequenzierungsqualitätssicherung

Metriken	Vorgabe	Beschreibung
Nicht indiziertes Gesamtergebnis (GB)	N. z.	Keine Spezifikation, da Läufe mit niedrigerem Ergebnis zum Bestehen von Probenbibliotheken führen können. Gehen Sie von ≥ 3000 Gbp für eine S4-Fließzelle und von ≥ 1000 GB für eine S2-Fließzelle aus.
Gesamt % \geq Q30	≥ 85	Messung der Basenqualität auf Lafebene. Die Mindestanforderung ist spezifiziert, da zu niedrige %Q30-Läufe die Q30-Basen der Probenbibliotheks-QC nicht bestehen.
QC-Ergebnis der Sequenzierung gesamt	PASS (bestanden) oder FAIL (fehlgeschlagen)	Informationen zum Fehlschlagen der Sequenzierungs-QC finden Sie im Abschnitt Fehlerbehebung im TruSight Whole Genome Package Insert (Dok.-Nr. 200050132).

Tabelle 2 Demultiplexing-Metriken

Metriken	Vorgabe	Beschreibung
Prozent identifizierter Reads	N. z.	Der Anteil gesamt an bestandenen Filter-Reads im Lauf, die Proben während des Demultiplex-Verfahrens zugewiesen wurden.
Prozent CV	N. z.	Bietet ein Maß für die Gleichmäßigkeit von Reads, die für jedes Indexpaar auf dem Lauf demultiplexiert wurden. Gehen Sie von < 25 % für Läufe ohne FASTQ-QC-Ergebnisfehler aus.

Tabelle 3 FASTQ QC-Metriken

Metriken	Vorgabe	Beschreibung
Ergebnis pro Probe (bps)	$\geq 90.000.000.000$	Das Minimum ist so eingestellt, dass es der ~ 26 -fachen durchschnittlichen Autosomen-Coverage entspricht, um Proben zu triagieren, die die Bibliotheks-QC nicht bestehen, damit die Analysedauer reduziert wird.

Metriken	Vorgabe	Beschreibung
QC-Ergebnis für FASTQ gesamt	PASS (bestanden) oder FAIL (fehlgeschlagen)	Informationen zum Fehlschlagen der FASTQ QC finden Sie im Abschnitt Fehlerbehebung im TruSight Whole Genome Package Insert (Dok.-Nr. 200050132).

Tabelle 4 QC-Metriken für die Probenbibliothek

Metriken	Vorgabe	Beschreibung
Durchschnittliche Autosomen-Coverage	≥ 35	Durchschnittliche Coverage der Autosomen. Es gibt eine spezifizierte Mindestanforderung, um die analytische Leistung sicherzustellen.
Autosomen % mit einer Coverage über 20X	$\geq 93,94$	Maß für die Gleichförmigkeit der Coverage, mit dem Probleme erkannt werden, die nicht unbedingt mit dem GC-Bias zusammenhängen. Es gibt eine spezifizierte Mindestanforderung, um die analytische Leistung sicherzustellen.
Normalisierte Coverage bei 60 % bis 79 % der GC-Klassen	$0,82 \leq x \leq 1,13$	Maß für die Gleichförmigkeit der Coverage, das GC-Bias erkennt, insbesondere einen Verlust der Coverage in Bereichen des Genoms mit höherer prozentualer GC- und niedrigerer prozentualer AT-Basenzusammensetzung. Es gibt eine spezifizierte Mindest- und Maximalanforderung, um die analytische Leistung sicherzustellen.
Normalisierte Coverage bei 20 % bis 39 % der GC-Klassen	$0,97 \leq x \leq 1,06$	Maß für die Gleichförmigkeit der Coverage, das GC-Bias erkennt, insbesondere einen Verlust der Coverage in Bereichen des Genoms mit niedrigerer prozentualer GC- und höherer prozentualer AT-Basenzusammensetzung. Es gibt eine spezifizierte Mindest- und Maximalanforderung, um die analytische Leistung sicherzustellen.
Durchschnittliche Mitochondrien-Coverage	≥ 500	Coverage des mitochondrialen Chromosoms. Es gibt eine spezifizierte Mindestanforderung, um die mitochondriale SNV-Nachweisgrenze sicherzustellen.
Prozent der Q30-Basen	≥ 85	Messwert zur Basenqualität. Es gibt eine spezifizierte Mindestanforderung, um die analytische Leistung sicherzustellen.

Metriken	Vorgabe	Beschreibung
Geschätzte Probenkontamination	≤ 0,005	Erkennt kontaminierende Reads von anderen Proben. Die maximale Anforderung ist spezifiziert, um die mitochondriale SNV-Nachweisgrenze (der Variantentyp mit der höchsten Empfindlichkeit für eine Kontamination) sicherzustellen.
Zusammenfassung des QK-Ergebnisses der Probenbibliothek	PASS (bestanden) oder FAIL (fehlgeschlagen)	Informationen zum Fehlschlagen der Probenbibliotheks-QC finden Sie im Abschnitt Fehlerbehebung im TruSight Whole Genome Package Insert (Dok.-Nr. 200050132).

Tabelle 5 Ploidie-QC-Metriken

Metriken	Vorgabe	Beschreibung
Bereitgestellte Geschlechtschromosomen-Ploidy	N. z.	Geschlecht, das der Bediener während der Lauferstellung angegeben hat (female [weiblich], male [männlich], unknown [unbekannt]).
Ploidy estimation	N. z.	Geschlechts-Ploidie geschätzt durch DRAGEN.
Zusammenfassendes Ploidy-Ergebnis	CONCORDANT (konkordant), DISCORDANT (diskordant) oder ND (nicht bestimmt)	CONCORDANT zeigt die Übereinstimmung zwischen bereitgestellter und geschätzter Geschlechts-Ploidie an. ND gibt das Geschlecht als unbekannt oder als andere Schätzung als XX oder XY an. Bei Ergebnissen mit DISCORDANT wurde entweder ein falsches Geschlecht während der Lauferstellung eingegeben oder es hat möglicherweise ein Probentausch stattgefunden. Siehe Abschnitt Fehlerbehebung in TruSight Whole Genome Package Insert (Dok.-Nr. 200050132).

Tabelle 6 Nur zu Informationszwecken – Metriken

Metriken	Beschreibung
Mittlere Insert-Länge	Das Target ist 450 bp, aber man sollte von einer Variation durch Sequenzierungslauf und Operator ausgehen. Ein Bereich von etwa 360 bis 550 bp ist akzeptabel. Ein konstanter Betrieb außerhalb dieses Bereichs kann zu einer höheren Inzidenz von fehlschlagenden Proben führen.

Metriken	Beschreibung
Prozentsatz der gemappten Reads	Prozentsatz der Reads, die mit dem Referenzgenom gemappt wurden Kann verringert sein, als Reaktion auf eine nicht-humane gDNA-Kontamination, schlechte Probenqualität oder eine zu kleine Insertionslänge, die zu Mapping-Problemen führt.
Prozentsatz der Reads mit zusätzlichen Alignments	Prozentsatz der Reads mit Mapping, die sich über verschiedene Loci im Referenzgenom verteilen.
Prozentsatz doppelt markierter Reads	Gehen Sie < 20 % aus. Kann höher liegen, wenn das Ergebnis aus der Bibliotheksvorbereitung niedrig ist oder das Pooling niedriger als das erforderliche Volumen ist, oder als Reaktion auf sequenzierungsbezogene Probleme.
Prozentsatz der Soft-Clip-Basen, Read 1	Nützlich bei der Diagnose der Ursache für das Fehlschlagen einer durchschnittlichen Autosomen-Coverage.
Prozentsatz der Soft-Clip-Basen, Read 2	Nützlich bei der Diagnose der Ursache für das Fehlschlagen einer durchschnittlichen Autosomen-Coverage.
Prozentsatz der Basen mit Trimming, Read 1	Nützlich bei der Diagnose der Ursache für das Fehlschlagen einer durchschnittlichen Autosomen-Coverage.
Prozentsatz der Basen mit Trimming, Read 2	Nützlich bei der Diagnose der Ursache für das Fehlschlagen einer durchschnittlichen Autosomen-Coverage.

Informationen zur Varianten-Call-Datei

VCF-Dateien

Dateien im Varianten-Call-Format (*.vcf) enthalten Informationen über Varianten, die an spezifischen Positionen in einem Referenzgenom vorhanden sind und befinden sich im Verzeichnis `<Sample_ID>/Analysis`.

In der Kopfzeile der VCF-Datei sind die VCF-Dateiformatversion und die Varianten-Caller-Version enthalten und die Annotations angegeben, die im Rest der Datei verwendet werden. In der letzten Zeile in der Kopfzeile sind die Spaltenüberschriften für die Datenzeilen enthalten. In allen Datenzeilen der VCF-Datei sind Informationen zu einer einzelnen Referenzposition enthalten.

Alle VCF-Dateien enthalten eine Kopfzeile mit Beschreibungen der Ausgabespalten und Varianten-Call-Daten in Spalten, die als CHROM, POS, ID, REF, ALT, QUAL, FILTER, INFO, FORMAT, PROBE gekennzeichnet sind. Definitionen von Spaltenwerten können je nach Varianten-Caller unterschiedlich sein.

Small Variant und mSNV VCF

Die Ausgabe wird unter der Datei `<Sample_ID>.annotated.hard-filtered.gvcf.gz` im Verzeichnis `<Sample_ID>/Analysis` gespeichert.

Eine genomische VCF-Datei (gVCF) enthält Informationen zu den Varianten und Positionen, die als homozygot zum Referenzgenom bestimmt wurden. Für homozygote Regionen enthält die gVCF-Datei Statistiken zur readbasierten Evidenz für das Nichtvorhandensein von Varianten oder alternativen Allelen. Die gVCF-Datei enthält ein künstliches `<NON_REF>`-Allel. Reads, die die Referenz oder etwaige Varianten nicht unterstützen, werden dem Allel `<NON_REF>` zugewiesen. DRAGEN verwendet diese Reads, um zu bestimmen, ob die Position als homozygote Referenz aufgerufen werden kann, anstatt sie unaufgerufen zu lassen. Der resultierende Score stellt das nach Phred skalierte Konfidenzintervall in einem homozygoten Referenz-Call dar. Im Keimbahnmodus gilt der Score `FORMAT/GQ`.

DRAGEN bietet eine Post-VCF-Filterung nach Varianten anhand von Annotationen in den VCF-Datensätzen. Die harte Filterung von Varianten wird nachfolgend beschrieben. Aufgrund der Beschaffenheit der DRAGEN-Algorithmen, die die Hypothese korrelierter Fehler aus dem Kern des Varianten-Caller berücksichtigen, wurde jedoch die Unterscheidung zwischen echten Varianten und Geräuschen in der Pipeline deutlich verbessert, was die Abhängigkeit von der Post-VCF-Filterung erheblich verringert.

Der TruSight Whole Genome Analysis Application liefert Annotations zum Zuverlässigkeitswert (Konfidenz-Score) und -niveau (Konfidenzstufe) für kleine Varianten, die zur weiteren Leistungsverbesserung herangezogen werden können. Die Annotation zur Konfidenzstufe ist kein Qualitätsfilter und spiegelt sich daher nicht direkt im Qualitätsstatus der Varianten-Calls wider. Daher ist es möglich, dass Varianten-Calls aufgenommen werden, obwohl sie in der Annotation einen niedrigen KI-Score erhielten.

Tabelle 7 Kopfzeilen in VCF-Dateien

Kopfzeile	Beschreibung
CHROM	Das Chromosom des Referenzgenoms. Chromosomen erscheinen in der gleichen Reihenfolge wie die Referenz-FASTA-Datei.
POS	Die Single-Base-Position der Variante im Referenzchromosom. Für Einzelnukleotidvarianten (SNV) ist diese Position die Referenz-Base mit der Variante. Für Indels ist diese Position die Referenzbasis unmittelbar vor der Variante.
ID	Immer .
REF	Der Referenzgenotyp. Beispielsweise wird die Deletion eines einzelnen T als Referenz TT und alternatives T dargestellt. Die Einzelnukleotidvarianten A bis T werden als Referenz A und alternatives T dargestellt.
ALT	Die Allele, die sich vom Referenz-Read unterscheiden. Beispielsweise wird eine Insertion eines einzelnen T als Referenz A und alternatives AT dargestellt. Die Einzelnukleotidvarianten A bis T werden als Referenz A und alternatives T dargestellt.
QUAL	Ein vom Varianten-Caller zugewiesener Phred-skaliertes Qualitäts-Score. Höhere Scores weisen auf eine höhere Zuverlässigkeit der Variante und eine geringere Fehlerwahrscheinlichkeit hin. Bei einem Qualitäts-Score von Q beträgt die geschätzte Fehlerwahrscheinlichkeit $10^{-(Q/10)}$. Beispielsweise hat die Reihe der Q30-Calls eine Fehlerrate von 0,1 %. Viele Varianten-Caller weisen Qualitäts-Scores auf Basis ihrer statistischen Modelle zu, die im Verhältnis zur beobachteten Fehlerrate hoch sind.

Tabelle 8 Annotationen in VCF-Dateien

Kopfzeile	Beschreibung
FILTER	<ul style="list-style-type: none"> • PASS: Alle Filter bestanden. • DRAGENSnpHardQUAL: Wenn wahr einstellen auf:QUAL < 10,41 • DRAGENIndelHardQUAL: Wenn wahr einstellen auf:QUAL < 7,83 • LowDepth: Wenn wahr, einstellen:DP <= 1 • LowGQ: Wenn wahr, einstellen:GQ = 0 • PloidyConflict: Genotyp-Call vom Varianten-Caller stimmt nicht mit Chromosomen-Ploidie überein. • base_quality: Position gefiltert, da die mittlere Basisqualität von Alt-Reads an diesem Locus den Schwellenwert nicht erreicht. • filtered_reads: Position gefiltert, da ein zu großer Teil der Reads herausgefiltert wurde. • fragment_length: Position gefiltert, da der absolute Unterschied zwischen der mittleren Fragmentlänge von Alt-Reads und der mittleren Fragmentlänge von Ref-Reads an diesem Locus den Schwellenwert überschreitet. • low_af: Allelfrequenz erreicht Schwellenwert nicht. • low_depth: Position gefiltert, da die Read-Tiefe zu gering ist. • low_frac_info_reads: Position gefiltert, da der Anteil informativer Reads unter dem Schwellenwert liegt. • low_normal_depth: Position gefiltert, da die Read-Tiefe normaler Proben zu gering ist. • long_indel: Position gefiltert, da die Indel-Länge zu lang ist. • mapping_quality: Position gefiltert, da die mediane Zuordnungsqualität von Alt-Reads an diesem Locus den Schwellenwert nicht erreicht. • multiallelic: Position gefiltert, da mehr als zwei alt-Allele die Tumor-LOD passieren. • non_homref_normal: Position gefiltert, da der normale Probengenotyp keine homozygote Referenz ist. • no_reliable_supporting_read: Position gefiltert, da kein zuverlässiger unterstützender somatischer Read vorhanden ist. • panel_of_normals: In mindestens einer Probe der Normalpanel-VCF festgestellt. • read_position: Position gefiltert, da der Median der Reichweiten zwischen Start/Ende des Reads und diesem Locus unter dem Schwellenwert liegt. • RMxNRepeatRegion: Position gefiltert, da das Varianten-Allel oder ein Teil davon eine Wiederholung der Referenz ist. • strand_artifact: Position gefiltert aufgrund starker Strang-Verzerrung. • str_contraction: Position gefiltert aufgrund eines vermuteten PCR-Fehlers, bei dem das alt-Allel eine Wiederholungseinheit weniger als die Referenz ist. • too_few_supporting_reads: Position gefiltert, da zu wenige unterstützende Reads in der Tumorprobe vorhanden sind. • weak_evidence: Score der somatischen Variante erreicht Schwellenwert nicht.

Kopfzeile	Beschreibung
INFO	<ul style="list-style-type: none"> • DB: dbSNP-Mitgliedschaft. • FS: Phred-skaliertes p-Wert anhand des exakten Fisher-Tests zur Erkennung von Strang-Verzerrungen. • QD: Variantenzuverlässigkeit/-qualität nach Tiefe. • R2_5P_bias: Score basierend auf Mate-Verzerrung und Reichweite von 5 Prime Ends. • SOR: Symmetrisches Chancenverhältnis der 2x2-Kontingenztafel zur Erkennung von Strang-Verzerrungen. • DP: Ungefähre Read-Tiefe (informativ und nicht informativ); einige Reads wurden möglicherweise basierend auf Mapq usw. gefiltert. • END: Endposition des Intervalls. • FractionInformativeReads: Der Anteil informativer Reads an den Gesamt-Reads. • MQ: RMS-Zuordnungsqualität. • MQRankSum: Z-Score aus dem Wilcoxon-Rangsummentest von Alt- im vs. Qualitätswerte des Ref-Read-Mapping. • ReadPosRankSum: Z-Score aus dem Wilcoxon-Rangsummentest von Alt- vs. Ref-Read-Positionsverzerrungen. • SOMATIC: Mindestens eine Variante an dieser Position ist somatisch. • IENS: Indel-Längen für jede ALT-Variante. • SCORE: Zuverlässigkeitswert für jeden am Prüfczentrum vorhandenen Variantentyp als (Variantentyp):(Konfidenz-Score). • TIER: Zuverlässigkeitsniveau für jeden am Prüfczentrum vorhandenen Variantentyp als (Variantentyp):(Konfidenzstufe).

Kopfzeile	Beschreibung
FORMAT	<p>In der Spalte FORMAT werden durch Doppelpunkte getrennte Felder angegeben, z. B. GT:GQ.</p> <ul style="list-style-type: none"> • AD: Alleltiefen (wobei nur informative Reads von den Gesamt-Reads gezählt werden) für die ref- und alt-Allele in der angegebenen Reihenfolge. • AF: Allelfraktionen für Alt-Allele in der angegebenen Reihenfolge. • DP: Ungefähre Read-Tiefe (Reads mit MQ = 255 oder schlechten Mates werden gefiltert). • F1R2: Anzahl der Reads in F1R2-Paarorientierung, die jedes Allel unterstützen. • F2R1: Anzahl der Reads in F2R1-Paarorientierung, die jedes Allel unterstützen. • GP: Phred-skalierte Posterior-Wahrscheinlichkeiten für Genotypen gemäß der Definition in der VCF-Spezifikation. • GQ: Genotypqualität. • GT: Genotyp. • ICNT: Anzahl der informativen INDEL-Reads basierend auf dem Referenz-Konfidenzmodell. • MB: Komponentenstatistiken pro Probe zur Erkennung von Mate-Verzerrungen. • MIN_DP: Mindesttiefe, die innerhalb des gVCF-Blocks beobachtet wird. • PL: Normalisierte, Phred-skalierte Wahrscheinlichkeiten für Genotypen gemäß der Definition in der VCF-Spezifikation. • PRI: Phred-skalierte Prior-Wahrscheinlichkeiten für Genotypen. • PS: Physische Phasen-ID-Informationen, bei denen jede eindeutige ID innerhalb einer bestimmten Probe (aber nicht über Proben hinweg) Datensätze innerhalb einer Phasengruppe verbindet. • SB: Komponentenstatistiken pro Probe mit dem exakten Fisher-Test zur Erkennung von Strang-Verzerrungen. • SPL: Normalisierte, Phred-skalierte Wahrscheinlichkeiten für SNPs basierend auf dem Referenz-Konfidenzmodell. • SQ: Somatische Qualität.
SAMPLE	Die Probenspalte enthält die Werte, die in der Spalte „FORMAT“ angegeben sind.

Copy Number Variant-VCF

Die Target-Zählung ist die erste Verarbeitungsphase in der DRAGEN CNV-Pipeline, die `<Sample_ID>.target.counts.gz` erzeugt; dann wird eine GC-Biaskorrektur durchgeführt, die eine `*.target.counts.gc-korrigierte.gz`-Datei erzeugt. In der Normalisierungsphase wird eine `*.tn.tsv.gz`-Datei erzeugt. Die DRAGEN Host Software generiert viele intermediäre Dateien. `*.seg.called.merged` ist die endgültige Call-Datei, die Amplifikations- und Deletionsereignisse enthält.

Zusätzlich zu der Segmentdatei gibt DRAGEN die Calls im VCF-Standardformat aus. Die Ausgabe wird unter `<Sample_ID>.cnv.vcf.gz` im Verzeichnis `<Sample_ID>/Analysis` gespeichert.

Definitionen der für CNV-Caller spezifischen Spalten:

Die POS-Spalte ist die Startposition der Variante. Gemäß der VCF-Spezifikation gilt Folgendes: Wenn eines der ALT-Allele ein symbolisches Allel ist, wie z. B. ``, ist die Padding-Base erforderlich, und POS bezeichnet dann die Koordinate der Base vor dem Polymorphismus. Alle Koordinaten im VCF sind 1-Basen-Koordinaten.

Die Spalte `ID` wird zur Darstellung des Ereignisses verwendet. Das Feld `ID` codiert den Ereignistyp und die Koordinaten des Ereignisses.

In der Spalte `REF` sind alle CNV-Ereignisse mit einem N aufgeführt.

In der Spalte `ALT` wird die Art des CNV-Ereignisses angegeben. Da in der VCF-Datei ausschließlich CNV-Ereignisse angegeben sind, wird nur der Eintrag in `DEL` oder `DUP` verwendet.

Die Spalte `QUAL` enthält einen Qualitäts-Schätzwert für das CNV-Ereignis, der für die harte Filterung verwendet wird.

In der Spalte `FILTER` wird `PASS` (Bestanden) angegeben, wenn das CNV-Ereignis alle Filter bestanden hat. Andernfalls enthält diese Spalte den Namen des fehlgeschlagenen Filters.

Die Spalte `INFO` enthält Informationen zum Ereignis selbst. Der Eintrag `REFLEN` gibt die Ereignislänge wieder. Der Eintrag `SVTYPE` lautet stets CNV. Der Eintrag `END` gibt die Endposition des Ereignisses wieder.

Die FORMAT-Felder werden in der Kopfzeile beschrieben.

- `GT`: Genotyp
- `SM`: Lineares Kopienverhältnis des Segmentmittelwerts
- `CN`: Geschätzte Kopienzahl
- `BC`: Anzahl der Klassen in der Region
- `PE`: Anzahl nicht zusammengehöriger Paired-End-Reads an Start- und Stopp-Unterbrechungspunkten

Repeats-VCF

Der ExpansionHunter führt das Realignment der Reads in und um jedes Target-Repeat anhand eines Sequenzdiagramms durch. Anschließend erfolgt im ExpansionHunter die Genotypisierung der Repeat-Länge in den einzelnen Allelen anhand dieses Diagramm-Alignments.

Weitere Informationen und Analysen sind folgenden Dokumenten zu ExpansionHunter zu entnehmen:

- Dolzhenko et al., *Detection of long repeat expansions from PCR-free whole-genome sequence data* 2017
- Dolzhenko et al., *ExpansionHunter: A sequence-graph based tool to analyze variation in short tandem repeat regions* 2019

Der TruSight Whole Genome Analysis Application STR-Variantenkatalog enthält Spezifikationen zu krankheitsverursachenden Repeats in den Genen AFF2, AR, ATN1, ATXN1, ATXN10, ATXN2, ATXN3, ATXN7, ATXN8OS, C9ORF72, CACNA1A, CBL, CNBP, CSTB, DIP2B, DMPK, FMR1, FXN, GLS, HTT, JPH3, NIPA1, NOP56, NOTCH2NL, PABPN1, PHOX2B, PPP2R2B und TBP.

Die Ergebnisse von Repeat-Genotypisierungen werden als einzelne VCF-Dateien ausgegeben, die die Länge eines jeden Allels bei jedem Call-fähigen Repeat gemäß Definition in der Katalogdatei „repeat-specification“ angibt. Der Dateiname lautet `<Sample_ID>.repeats.gz` und befindet sich im Verzeichnis `<Sample_ID>/Analysis`.

Einige Spalten sind spezifisch für Repeat-Expansion-Caller:

Tabelle 9 VCF-Kernfelder

Feld	Beschreibung
CHROM	Bezeichner für Chromosomen
POS	Position der ersten Base vor der Repeat-Region in der Referenz
ID	Immer <code>.</code>
REF	Die Referenzbase mit der Position POS
ALT	Liste der Repeat-Allele im Format <code><STR_n></code> . N ist die Anzahl der Repeat-Einheiten.
QUAL	Immer <code>.</code>
FILTER	Der LowDepth-Filter wird angewendet, wenn die Locus-Gesamttiefe unter dem 10-fachen liegt oder die Anzahl der Reads, die sich über ein oder beide Bruchenden erstrecken, unter 5 liegt.

Tabelle 10 Zusätzliche INFO Felder

Feld	Beschreibung
END	Position der letzten Base der Repeat-Region in der Referenz
REF	Referenz-Kopienzahl
REPID	Varianten-ID aus dem Variantenkatalog
RL	Referenzlänge in bp
RU	Repeat-Einheit in der Referenzausrichtung
VARID	Varianten-ID aus dem Variantenkatalog

Tabelle 11 GENOTYP-Felder (je Probe)

Feld	Beschreibung
AD	Alleltiefen für die Ref- und Alt-Allele in der angegebenen Reihenfolge
ADFL	Anzahl an angrenzenden Reads, die dem Allel entsprechen

Feld	Beschreibung
ADIR	Anzahl an Reads innerhalb des Repeats, die dem Allel entsprechen
ADSP	Anzahl an umspannenden Reads, die dem Allel entsprechen
DST	Ergebnisse (+ erkannt, - nicht erkannt, ? unbestimmt) des Tests, dargestellt nach Variante
GT	Genotyp
LC	Locus-Coverage
REPCI	Konfidenzintervall für REPCN
REPCN	Anzahl der vom Allel umspannten Repeat-Einheiten
RPL	Log-Likelihood-Verhältnis für das Vorhandensein des Referenzallels
SO	Read-Typ, der das Allel unterstützt. Die Werte können <i>SPANNING</i> (umspannend), <i>FLANKING</i> (angrenzend) oder <i>INREPEAT</i> (innerhalb eines Repeats) sein. Diese Werte zeigen an, ob die Reads umspannend, angrenzend oder gänzlich im Repeat enthalten sind.

Die Datei `<Sample_ID>.repeats.vcf.gz` enthält die SMN-Ausgabe sowie alle Target-Repeats. Die SMN-Ausgabe wird als ein einzelner SNV-Call an der das Spleißen beeinflussenden Position in SMN1 (NM_000344.3:c.840C/T) dargestellt, wobei der Status der Spinalen Muskelatrophie (SMA) in den folgenden benutzerdefinierten Feldern enthalten ist:

Tabelle 12 SMA-Ergebnis in Ausgabedatei repeats.vcf

Feld	Beschreibung
VARID	SMN kennzeichnet den SMN-Call.
GT	Genotyp-Call an dieser Position unter Verwendung eines normalen (diploiden) Genotypmodells.
DST	SMA-Status-Call: + steht für „erkannt“ - steht für „nicht erkannt“ ? steht für „unbestimmt“
AD	Gesamtzahl der Reads, die das C- und T-Allel unterstützen.
RPL	Log10-Likelihood-Quotient zwischen den betroffenen und nicht betroffenen Modellen. Positive Werte deuten darauf hin, dass das nicht betroffene Modell wahrscheinlicher ist.

ROH-BED

Homozygotie-Regionen (ROH) werden als Teil des Callers für kleine Varianten erkannt. Der Caller erkennt die Homozygotie-Läufe von Gesamtgenom-Calls in autosomalen menschlichen Chromosomen und gibt diese aus. Geschlechtschromosomen werden ignoriert, es sei denn, der Proben-Geschlechtskaryotyp ist XX, wie vom Ploidy Estimator bestimmt. Die ROH-Ausgabe ermöglicht nachgeschalteten Tools, ein Screening auf und eine Prognose für Blutsverwandtschaft zwischen den Elternteilen des Probanden durchzuführen bzw. abzugeben.

Eine Region ist definiert als aufeinanderfolgende Varianten-Calls auf dem Chromosom ohne große Lücke zwischen diesen Varianten. Anders formuliert werden Regionen nach Chromosom oder nach großen Lücken ohne SNV-Calls aufgeschlüsselt. Die Lückengröße ist auf 3 Mbasen festgelegt.

Der ROH-Caller erstellt eine ROH-Ausgabedatei mit dem Namen `<Sample_ID>.roh.bed` im Verzeichnis `<Sample_ID>/Analysis`. Jede Zeile steht für eine Homozygotie-Region. Die BED-Datei enthält die folgenden Spalten:

```
Chromosome Start End Score #Homozygous #Heterozygous
```

Wobei gilt:

- Der Score ist von der Anzahl der homozygoten und heterozygoten Varianten abhängig, wobei gilt, dass jede homozygote Variante den Score um 0,025 erhöht und jede heterozygote Variante den Score um 0,975 verringert.
- Start- und Endpositionen sind 0-basierte, halboffene Intervalle.
- #Homozygous gibt die Anzahl der homozygoten Varianten in der Region an.
- #Heterozygous gibt die Anzahl der heterozygoten Varianten in der Region an.

Der Caller generiert auch eine Metrikdatei mit dem Namen `<Sample_ID>.roh_metrics.csv`, in der die Anzahl großer ROH und der Prozentsatz von SNPs in großen ROH (> 3 MB) aufgeführt sind.

Ploidy-Estimator-Metriken

Der Ploidy Estimator wird standardmäßig ausgeführt. Der Ploidy Estimator verwendet Reads vom Mapper/Aligner zur Berechnung der Coverage-Tiefe der Sequenzierung für jedes Autosom und Gonosom im Humangenom. Der Geschlechtskaryotyp der Probe wird anschließend anhand des Verhältnisses der mittleren Geschlechtschromosomen-Coverages zum Median der Autosomen-Coverage bestimmt. Im konsolidierten Bericht werden XX oder XY und CONCORDANT (konkordant), DISCORDANT (diskordant) oder ND (nicht bestimmt) im Vergleich zu den bereitgestellten Probandaten angegeben. Die detaillierten Ergebnisse, einschließlich jeder normalisierten medianen Coverage werden in der Datei `<Sample_ID>.ploidy_estimation_metrics.csv` berichtet.

FASTQ-Dateien

FASTQ (*.fastq.gz, *.fastq.ora) ist ein textbasiertes Dateiformat mit den Base-Calls und Qualitätswerten pro Read. Jeder Eintrag enthält die folgenden Informationen:

- den Probenbezeichner

- die Sequenz
- ein Pluszeichen (+)
- die Phred-Qualitäts-Scores in einem ASCII + 33-codierten Format

Die Software generiert eine FASTQ-Datei für jede Probe, jedes Read und jede Lane. Beispielsweise generiert die Software für jede Probe in einem Paired-End-Lauf zwei FASTQ-Dateien: eine für Read 1 und eine für Read 2. Zusätzlich zu diesen FASTQ-Probedateien generiert die Software zwei FASTQ-Dateien pro Lane, die alle unbekanntes Proben enthalten. FASTQ-Dateien für Index Read 1 und Index Read 2 werden nicht generiert, da die Sequenz in der Kopfzeile jedes FASTQ-Eintrags enthalten ist. Das Dateinamenformat wird aus den im Probenblatt angegebenen Feldern erstellt und es wird das Dateinamenformat `<Sample_ID>_S#_L00#_R#_001.fastq.gz`

FASTQ-Dateien werden im Verzeichnis `<Sample_ID>/Conversion` gespeichert. Im FASTQ-Verzeichnis des Analyseordners finden Sie das Verzeichnisprotokoll mit BCL-zu-FASTQ-Konvertierungsprotokollen und das Verzeichnis Reports (Berichte), das verschiedene gelesene Metrikdateien enthält, sowie `SampleSheet.csv`, das für die FASTQ-Konvertierung verwendet wird. FASTQ-Dateien von unbestimmten Reads finden Sie im Verzeichnis `Undetermined/Conversion` (unbestimmt/Konversion) des Analyseordners.

Der Probenbezeichner ist so formatiert:

```
@Instrument:RunID:FlowCellID:Lane:Tile:X:Y ReadNum:FilterFlag:0:SampleNumber
```

Beispiel:

```
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
```

```
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
```

```
+
```

```
<>;##=><9=AAAAAAAAA9#:<#<;<<<????#=#
```

CRAM-Dateien

Komprimierte referenzorientierte Alignment Map- bzw. CRAM-Dateien (*.cram) werden im Verzeichnis `<Sample_ID>/Analysis` gespeichert und enthalten Kopfzeilen und Alignment-Datensätze im Verhältnis zur genomischen Referenzdatei, die während des Alignments verwendet wird. Der Pfad zur Referenzdatei ist in der Datei `<Sample_ID>/Analysis/<Sample_ID>-replay.json` als `--ht-reference`-Parameter aufgeführt, der standardmäßig auf `hg38.fa` eingestellt ist.

CRAM-Dateien verfügen über einen Kopfzeilen- und einen Alignment-Abschnitt:

- **Kopfzeile:** Enthält Informationen über die gesamte Datei, z. B. den Probenamen, die Probenlänge und die Alignment-Methode. Alignments im Abschnitt „Alignments“ (Ausrichtungen) sind bestimmten Informationen im Abschnitt „Header“ (Kopfzeile) zugeordnet.
- **Ausrichtungen:** Enthält Read-Name, Read-Sequenz, Read-Qualität, Informationen zu den Alignments und benutzerdefinierte Tags. Der Read-Name enthält das Chromosom, die Startkoordinate, die Alignment-Qualität und den Match-Deskriptor-String.

Im Abschnitt „Alignments“ (Ausrichtungen) sind diese Informationen für jeden Read oder jedes Read-Paar enthalten:

- AS: Paired-End-Alignment-Qualität.
- RG: Read-Gruppe, die die Anzahl der Reads für eine bestimmte Probe angibt.
- BC: Barcode-Tag, der die demultiplexierte Proben-ID angibt, die dem Read zugeordnet ist.
- SM: Single-End-Alignment-Qualität.
- XC: Match-Deskriptor-String.
- XN: Amplikon-Namen-Tag, das die dem Read zugeordnete Amplikon-ID aufzeichnet.

Um Alignment-Datensätze anzuzeigen, können `samtools` als `samtools view --reference <path_to_reference_folder>/hg38.fa <Sample_ID>.cram` verwendet werden.

Es werden auch eine Indexdatei und eine Prüfsummendatei generiert.

Technische Unterstützung

Wenn Sie technische Unterstützung benötigen, wenden Sie sich an den technischen Support von Illumina.

Website: www.illumina.com

E-Mail: techsupport@illumina.com

Sicherheitsdatenblätter (SDS) sind auf der Illumina-Website unter support.illumina.com/sds.html verfügbar.

Die **Produktdokumentation** steht unter support.illumina.com zum Herunterladen zur Verfügung.



Illumina, Inc.
5200 Illumina Way
92122 San Diego, Kalifornien, USA
+1.800.809.ILMN (4566)
+1.858.202.4566 (außerhalb von Nordamerika)
techsupport@illumina.com
www.illumina.com



Illumina Netherlands B.V.
Steenoven 19
5626 DK Eindhoven
The Netherlands

Australischer Sponsor

Illumina Australia Pty Ltd
Nursing Association Building
Level 3, 535 Elizabeth Street
3000 Melbourne, VIC
Australien

FÜR IN-VITRO-DIAGNOSTIK.

© 2024 Illumina, Inc. Alle Rechte vorbehalten.

illumina[®]