illumina®

Local Run Manager
Analysemodul für somatische Varianten

Workflow-Anleitung für MiSeqDx

FÜR IN-VITRO-DIAGNOSTIK

Dokument-Nr. 1000000012183 v01 DEU
August 2021

FÜR IN-VITRO-DIAGNOSTIK

Überblick

Das Analysemodul für somatische Varianten in Local Run Manager ist für die Verwendung mit dem TruSeq Custom Amplicon Kit Dx-Assay von Illumina bestimmt. In Kombination mit dem Modul für somatische Varianten dient der Assay zur Vorbereitung von Bibliotheken für die Sequenzierung von DNA aus formalinfixierten, in Paraffin eingebetteten Gewebeproben. Der Assay erkennt somatische Mutationen bei niedriger Variantenhäufigkeit.

Das Analysemodul untersucht kurze Regionen von amplifizierter DNA oder Amplikons auf Varianten. Das fokussierte Sequenzieren von Amplikons ermöglicht eine hohe Abdeckung bestimmter Regionen über eine große Anzahl von Proben hinweg. Das Analysemodul führt unter Verwendung des Doppelstrang-Ansatzes, der Vorwärts- und Rückwärts-Oligo-Pools beinhaltet, eine Sekundäranalyse von Sequenzierungsläufen durch und generiert Berichte. Weitere Informationen finden Sie in der Packungsbeilage zum *TruSeq Custom Amplicon-Kit Dx (Dokument-Nr. 1000000029772)*.

Das Analysemodul für somatische Varianten benötigt die im MiSeqDx Reagenzien-Kit v3 enthaltenen Verbrauchsmaterialien für die Sequenzierung. Weitere Informationen finden Sie in der Packungsbeilage zum *MiSeqDx Reagenzien-Kit v3 (Dokument-Nr. 1000000030849).*

Über diese Anleitung

Diese Anleitung enthält Anweisungen zum Festlegen der Laufparameter für die Sequenzierung und Analyse für das Analysemodul für somatische Varianten. Informationen zu den Dashboard- und Systemeinstellungen für Local Run Manager finden Sie im *Software-Referenzhandbuch zu Local Run Manager (Dokument-Nr. 1000000011880)*.

Eingeben von Laufinformationen

Festlegen von Parametern

1 Melden Sie sich bei Local Run Manager an.

2 Klicken Sie auf **Create Run** (Lauf erstellen) und wählen Sie **Somatic Variant** (Somatische Varianten).

3 Geben Sie einen Namen ein, mit dem der Lauf von der Sequenzierung bis zur Analyse identifiziert werden kann.
Verwenden Sie alphanumerische Zeichen, Leerzeichen, Unterstriche oder Bindestriche.

4 [Optional] Geben Sie eine Laufbeschreibung ein, die hilft, den Lauf zu identifizieren.
Verwenden Sie alphanumerische Zeichen, Leerzeichen, Unterstriche oder Bindestriche.

5 Wählen Sie die Anzahl der Proben und das Index-Set aus der Dropdown-Liste aus.

Importieren von Manifestdateien für den Lauf

1 Stellen Sie sicher, dass die gewünschten Manifestdateien an einem zugänglichen Netzwerkspeicherort oder auf einem USB-Laufwerk verfügbar sind.

2 Klicken Sie auf **Import Manifests** (Manifeste importieren).

3 Navigieren Sie zur Manifestdatei und wählen Sie die Manifeste aus, die Sie hinzufügen möchten.

HINWEIS

Um Manifestdateien für alle Läufe verfügbar zu machen, die das Analysemodul für somatische Varianten verwenden, fügen Sie die Dateien über die Funktion „Module Settings“ (Moduleinstellungen) hinzu. Für diese Funktion werden Administratorrechte benötigt. Weitere Informationen finden Sie im Software-Referenzhandbuch zu *Local Run Manager (Dokument-Nr. 1000000011880)*.

Angeben der Proben für den Lauf

Wählen Sie eine der folgenden Optionen, um die Proben für den Lauf festzulegen, und führen Sie die entsprechenden Schritte durch.

- **Enter samples manually** (Manuelles Eingeben der Proben): Verwenden Sie die leere Tabelle auf dem Bildschirm „Create Run“ (Lauf erstellen).
- **Import samples** (Importieren von Proben): Navigieren Sie zu einer externen Datei mit kommagetrennten Werten (*.csv). Im Bildschirm „Create Run“ steht eine Vorlage zum Herunterladen zur Verfügung.

Nachdem Sie die Probentabelle ausgefüllt haben, können Sie die Probeninformationen in eine externe Datei exportieren. Diese Datei können Sie bei der Vorbereitung von Bibliotheken als Referenz verwenden oder für einen anderen Lauf importieren.

Manuelles Eingeben der Proben

1. Geben Sie im Feld „Sample Name“ (Probenname) einen eindeutigen Probennamen ein.
   Verwenden Sie alphanumerische Zeichen, Bindestriche oder Unterstriche.
   Der Probenname wird automatisch für den entsprechenden Well in dem anderen Pool übernommen.
2. [Optional] Bei positiven oder negativen Kontrollproben klicken Sie mit der rechten Maustaste und wählen Sie den Kontrolltyp aus.
   Die Kontrollprobe in einem Proben-Well wird automatisch für den entsprechenden Well in dem anderen Pool übernommen.
3. [Optional] Geben Sie im Feld „Sample Description“ (Probenbeschreibung) eine Probenbeschreibung ein.
   Verwenden Sie alphanumerische Zeichen, Bindestriche oder Unterstriche.
   Die Probenbeschreibung wird automatisch für den entsprechenden Well in dem anderen Pool übernommen.
   Jede Probenbeschreibung ist einer Proben-ID zugeordnet. Wenn Sie dieselbe Proben-ID für einen nachfolgenden Lauf verwenden, wird die entsprechende Probenbeschreibung überschrieben.
4. Wählen Sie einen Index-1-Adapter aus der Dropdown-Liste „Index 1 (i7)“ aus.
5. Wählen Sie einen Index-2-Adapter aus der Dropdown-Liste „Index 2 (i5)“ aus.
6. Wählen Sie eine Manifestdatei aus der Dropdown-Liste „Manifest“ aus.
   Proben im Pool A erfordern ein anderes Manifest als Proben im Pool B.
7. Wählen Sie eine der folgenden Optionen zum Anzeigen, Drucken oder Speichern des Plattenlayouts, sodass es bei der Vorbereitung von Bibliotheken als Referenz zur Verfügung steht:
   - Klicken Sie auf das Symbol **Print** (Drucken), um das Plattenlayout anzuzeigen. Wählen Sie **Print** (Drucken), um es auszudrucken.
   - Klicken Sie auf **Export** (Exportieren), um Probeninformationen in eine externe Datei zu exportieren.

   Vergewissern Sie sich, dass die Manifest- und Probeninformationen korrekt sind. Fehlerhafte Informationen können die Ergebnisse verfälschen.
8. Klicken Sie auf **Save Run** (Lauf speichern).

Importieren von Proben

1 Klicken Sie auf **Import Samples** (Proben importieren) und navigieren Sie zum Speicherort der Datei mit den Probeninformationen. Sie können zwei Dateitypen importieren.
   - Klicken Sie auf dem Bildschirm „Create Run“ (Lauf erstellen) auf **Template** (Vorlage), um ein neues Plattenlayout zu erstellen. Die Vorlagendatei enthält die korrekten Spaltenüberschriften für den Import. Tragen Sie in jeder Spalte Informationen zu den Proben im Lauf ein. Löschen Sie die Probeninformationen in nicht genutzten Zellen und speichern Sie anschließend die Datei.
   - Verwenden Sie eine Datei mit Probeninformationen, die mittels der Export-Funktion aus dem Analysemodul für somatische Varianten exportiert wurde.

2 Klicken Sie auf das Symbol **Print** (Drucken), um das Plattenlayout anzuzeigen.

3 Wählen Sie **Print** (Drucken), um das Plattenlayout auszudrucken, damit es bei der Vorbereitung von Bibliotheken als Referenz zur Verfügung steht.

4 [Optional] Klicken Sie auf **Export** (Exportieren), um Probeninformationen in eine externe Datei zu exportieren.
   Vergewissern Sie sich, dass die Manifest- und Probeninformationen korrekt sind. Fehlerhafte Informationen können die Ergebnisse verfälschen.

5 Klicken Sie auf **Save Run** (Lauf speichern).

Bearbeiten eines Laufs

Anweisungen für das Bearbeiten der Informationen in Ihrem Lauf vor der Sequenzierung finden Sie im *Software-Referenzhandbuch zu Local Run Manager (Dokument-Nr. 1000000011880)*.

Analysemethoden

Das Analysemodul für somatische Varianten führt die nachfolgenden Analyseschritte durch und generiert anschließend Analyse-Ausgabedateien im Alignment-Ordner.

- Demultiplexieren von Index-Reads
- Generieren von FASTQ-Dateien
- Alignieren mit einer Referenz
- Identifizieren von Varianten

Demultiplexierung

Bei der Demultiplexierung wird jede Index-Read-Sequenz mit den für den Lauf angegebenen Indexsequenzen verglichen. In diesem Schritt werden keine Qualitätswerte berücksichtigt.

Index-Reads werden wie folgt identifiziert:

- Die Proben sind gemäß der Reihenfolge, in der sie für den Lauf aufgelistet sind, und mit 1 beginnend durchnummeriert.
- Die Probennummer 0 ist für Cluster reserviert, die keiner Probe zugeordnet wurden.
- Cluster werden einer Probe zugewiesen, wenn die Indexsequenz genau übereinstimmt bzw. je Index-Read maximal eine Nichtübereinstimmung festgestellt wird.

Generieren von FASTQ-Dateien

Nach der Demultiplexierung generiert die Software temporäre Analysedateien im FASTQ-Format, dem Textformat für die Darstellung von Sequenzen. FASTQ-Dateien enthalten die Reads jeder Probe und die entsprechenden Qualitäts-Scores. Cluster, die den Filter nicht passiert haben, werden nicht aufgenommen.

Jede FASTQ-Datei enthält nur die Reads einer Probe. Der Name der Probe ist Bestandteil des FASTQ-Dateinamens. FASTQ-Dateien sind die primären Eingabedateien für das Alignment. Für jede Probe werden vier FASTQ-Dateien erzeugt, zwei aus Pool A und zwei aus Pool B.

Alignment

Beim Alignment werden Cluster von jeder Probe mithilfe eines beschränkten („banded“) Smith-Waterman-Algorithmus an den in der Manifestdatei angegebenen Amplikon-Sequenzen ausgerichtet.

Der „banded“ Smith-Waterman-Algorithmus führt semiglobale Sequenz-Alignments durch, um ähnliche Regionen zwischen zwei Sequenzen zu ermitteln. Statt die gesamte Sequenz zu betrachten, vergleicht der Smith-Waterman-Algorithmus Segmente aller möglichen Längen.

Jeder Paired-End-Read wird in Bezug auf sein Alignment mit den entsprechenden Sondensequenzen dieses Reads untersucht.

- Read 1 wird anhand des umgekehrten Komplements der stromabwärts gelegenen lokusspezifischen Oligos (Downstream Locus-Specific Oligos, DLSO) beurteilt.
- Read 2 wird anhand der stromaufwärts gelegenen lokusspezifischen Oligos (Upstream Locus-Specific Oligos, ULSO) bewertet.
- Wenn der Beginn eines Reads mit einer Sondensequenz übereinstimmt, d. h., wenn er maximal drei Unterschiede (Nichtübereinstimmungen oder Verschiebungen aufgrund vorangehender Indels) aufweist, wird die volle Länge des Reads mit dem Amplikonziel dieser Sequenz aligniert.
- Indels innerhalb der DLSO und ULSO werden aufgrund der Assay-Chemie nicht beobachtet.

Die Alignments werden nach den Nichtübereinstimmungsraten über die Region von Interesse oder das gesamte Amplikon hinweg – je nach Länge des Amplikons – aus den Alignment-Ergebnissen herausgefiltert. Die herausgefilterten Alignments werden als „nicht aligniert“ in Alignment-Dateien protokolliert und beim Varianten-Calling nicht berücksichtigt.

Varianten-Calling

Der von Illumina entwickelte Pisces Variant Caller (Pisces-Varianten-Caller) identifiziert Varianten, die mit geringer Häufigkeit in DNA-Proben vorhanden sind.

Der Pisces-Varianten-Caller identifiziert SNPs in drei Schritten:

- Separates Betrachten jeder Position im Referenzgenom.
- Zählen der Basen an der angegebenen Position bei alignierten Reads mit einer Überlappung an der Position.
- Berechnen eines Varianten-Scores, der die Qualität des Calls nach dem Poisson-Modell bewertet. Varianten mit einem Qualitäts-Score unter Q30 werden ausgeschlossen.

Zunächst werden separate Varianten-Calls für jeden Pool durchgeführt. Anschließend werden die Varianten aus jedem Pool verglichen und zusammen in einer einzelnen Ausgabedatei ausgegeben. Wenn eine Variante folgende Kriterien erfüllt, wird sie in der Varianten-Call-Datei (VCF) mit „PASS“ (BESTANDEN) markiert:

- Die Variante ist in beiden Pools vorhanden.
- Sie hat eine 900-fache kumulative Abdeckungstiefe (mindestens 450-fach pro Pool).

- Sie hat eine Variantenhäufigkeit von ≥ 2,6 %, wie in der zusammengeführten VCF-Datei festgehalten.

Pisces (Somatischer Varianten-Caller)

Pisces führt Calls von somatischen Varianten durch, um seltene Varianten in DNA-Proben zu identifizieren. Die Anwendung kann für alle Proben eingesetzt werden und erzeugt VCF- und gVCF-Dateien.

Weitere Informationen finden Sie unter github.com/Illumina/Pisces/wiki.

Anzeigen von Lauf- und Probendaten

1 Klicken Sie im Dashboard von Local Run Manager auf den Laufnamen.

2 Prüfen Sie in der Registerkarte „Run Overview“ (Laufübersicht) die Kennzahlen des Sequenzierungslaufs.

3 [Optional] Klicken Sie auf das Symbol **Copy to Clipboard** (In Zwischenablage kopieren), um den Pfad des Ausgabeordners des Laufs zu kopieren.

4 Klicken Sie auf die Registerkarte „Sequencing Information“ (Sequenzierungsinformationen), um die Informationen zu Laufparametern und Verbrauchsmaterialien zu prüfen.

5 Klicken Sie auf die Registerkarte „Samples and Results“ (Proben und Ergebnisse), um den Speicherort des Analyseordners zu sehen.
   - Falls die Analyse wiederholt wurde, erweitern Sie die Dropdown-Liste „Select Analysis“ (Analyse auswählen) und wählen Sie die entsprechende Analyse aus.

6 Klicken Sie auf das Symbol **Copy to Clipboard** (In Zwischenablage kopieren), um den Pfad des Analyseordners zu kopieren.

Weitere Informationen zu den Registerkarten „Run Overview“ (Laufübersicht) und „Sequencing Information“ (Sequenzierungsinformationen) sowie dazu, wie Analysen erneut in die Warteschlange gestellt werden, finden Sie im *Software-Referenzhandbuch zu Local Run Manager (Dokument-Nr. 1000000011880)*.

Analysebericht

Eine Zusammenfassung der Analyseergebnisse finden Sie auf der Registerkarte „Samples and Results“ (Proben und Ergebnisse) und in Form eines Übersichtsberichts im Alignment-Ordner. Ein Bericht zu jeder Probe im PDF-Format steht ebenfalls zur Verfügung.

Informationen auf der Registerkarte „Sample and Details“ (Probe und Details)

1 Klicken Sie auf eine Probe in der Liste, um den Probenbericht anzusehen.

Tabelle 1 Tabelle „Run Information“ (Laufinformationen)

Spaltenüberschrift	Beschreibung
Run Status (Laufstatus)	Gibt an, ob der Sequenzierungslauf erfolgreich war oder nicht.
Total Yield (GB) (Gesamtergebnis, GB)	Anzahl der im Sequenzierungslauf durchgeführten Base-Calls. Zeigt den Grenzwert sowie den Status „Pass“ (Bestanden) bzw. „Fail“ (Fehlgeschlagen) an.
% ≥ Q30	Prozentsatz der Reads im Sequenzierungslauf mit einem Qualitäts-Score von 30 (Q30) oder höher. Zeigt den Grenzwert sowie den Status „Pass“ (Bestanden) bzw. „Fail“ (Fehlgeschlagen) an.

Spaltenüberschrift	Beschreibung
Sample Name (Probenname)	Probenname, der bei der Erstellung des Laufs angegeben wurde.
Total PF Reads (PF-Reads ingesamt)	Gesamtzahl der Reads nach Filterung.
Read 1% ≥ Q30	Prozentsatz der Reads in Read 1 der Probe mit einem Qualitäts-Score von 30 (Q30) oder höher.
Read 2% ≥ Q30	Prozentsatz der Reads in Read 2 der Probe mit einem Qualitäts-Score von 30 (Q30) oder höher.
Autosome call rate (Call-Rate für Autosomen)	Anzahl der genomischen Positionen in den Autosomen (Chromosomen 1 bis 22), die einen vordefinierten Konfidenzgrenzwert erreichen, geteilt durch die Gesamtzahl der untersuchten genomischen Positionen in Autosomen. Die Call-Rate wird auf einer Pro-Proben-Basis und als Prozentsatz angegeben, der wie folgt berechnet wird: 1 minus Anzahl der Autosomen-Positionen mit unvollständigen Calls geteilt durch die Gesamtzahl der sequenzierten Autosomen-Positionen.

Tabelle 2 Tabelle „Sample Reports“ (Probenberichte)

Spaltenüberschrift	Beschreibung
Sample (Probe)	Probenname, der bei der Erstellung des Laufs angegeben wurde.
Report Date (Datum des Berichts)	Zeitpunkt, zu dem der Bericht generiert wurde.
Sample Information (Probeninformationen)	Proben-ID, die bei der Erstellung des Laufs angegeben wurde, und Gesamtzahl der Reads der Probe, die den Filter passiert haben, sowie Prozentsatz der Reads der Probe mit einem Qualitäts-Score von 30 (Q30) oder höher und Call-Rate der Autosomen.
Amplicon Summary (Amplikon-Zusammenfassung)	Gesamtzahl der sequenzierten Amplikonregionen und Gesamtlänge in Basenpaaren der sequenzierten Amplikons in den Zielregionen für die Probe in Pool A und Pool B und die für jeden Pool verwendete Manifestdatei. Die Manifestdatei gibt das Referenzgenom und die Ziel-Referenzregionen an, die im Alignment-Schritt verwendet werden.
Read Level Statistics (Statistikwerte auf Read-Ebene)	Anzahl und Prozentsatz der Reads der Probe, die jede Position in der Referenz abdecken, für Read 1 und für Read 2 in Pool A und Pool B.
Variants Summary (Varianten-Zusammenfassung)	Anzahl der in der Probe nachgewiesenen SNVs, Insertionen und Deletionen, die vordefinierte Werte zur Feststellung, ob die Qualität der Ergebnisse innerhalb eines zulässigen Bereichs liegt, erfüllen.
Coverage Summary (Abdeckungszusammenfassung)	Die Gesamtzahl der alignierten Basen geteilt durch die Größe der Zielregion sowie der Prozentsatz der Amplikonregionen mit höheren Abdeckungswerten als der Grenzwert für eine geringe Abdeckung von 0,2-facher mittlerer Amplikonabdeckung für die Probe in Pool A und Pool B.
Coverage Plots (Abdeckungsdiagramme)	Das Diagramm „Coverage by Amplicon Region“ (Abdeckung nach Amplikonregion) zeigt die Abdeckung für die Probe über die Amplikonregionen hinweg. Regionen mit Abdeckungswerten unter dem Abdeckungsgrenzwert werden rot hervorgehoben. Der Durchschnitt aller Werte wird mit einer orangefarbenen Linie dargestellt. Die Abdeckung von Pool A und Pool B wird in einem Diagramm dargestellt.
Software Versions (Software-Versionen)	Versionen der Software-Programme, mit denen die Probe sequenziert wurde. Hierzu gehören die Versionen der MiSeq Operating Software, der Local Run Manager-Software, der RTA-Software und die des Analysemoduls für somatische Varianten.

Analyse-Ausgabedateien

Die folgenden Analyse-Ausgabedateien werden für das Analysemodul für somatische Varianten generiert. Sie enthalten Analyseergebnisse für das Alignment und das Varianten-Calling. Die Analyse-Ausgabedateien befinden sich im Alignment-Ordner.

Dateiname	Beschreibung
Demultiplexierung (*.txt)	Temporäre Dateien mit zusammengefassten Ergebnissen der Demultiplexierung.
FASTQ (*.fastq.gz)	Temporäre Dateien mit hinsichtlich ihrer Qualität ausgewerteten Base-Calls. FASTQ-Dateien sind die primären Eingabedateien für den Alignment-Schritt.
Alignment-Dateien im BAM-Format (*.bam)	Enthalten alignierte Reads für eine bestimmte Probe.
Varianten-Call-Dateien im VCF-Format (*.vcf.gz) pro Pool	Enthält Varianten-Calls für jede Position entweder aus dem Vorwärts- oder dem Rückwärts-Pool.
Varianten-Call-Dateien im Genom-VCF-Format (*.genome.vcf.gz)	Enthalten den Genotyp für jede Position, als Varianten- oder als Referenz-Call.
Konsensus-Varianten-Call-Dateien im VCF-Format (*.vcf.gz)	Enthält Varianten-Calls an jeder Position aus beiden Pools.
AmpliconCoverage_M1.tsv	Enthält Informationen zur Abdeckung pro Amplikon und Probe für jedes angegebene Manifest. M# steht für die Manifestnummer.

Demultiplexierungsdateiformat

Bei der Demultiplexierung werden die den Clustern zugeordneten Indexsequenzen gelesen, um festzustellen, aus welcher Probe der Cluster stammt. Die Zuordnung zwischen den Clustern und Probennummern wird für jede Platte der Fließzelle in eine Demultiplexierungsdatei (*.demux) geschrieben.

Die Dateinamen der Demultiplexierungsdateien haben das Format s_1_X.demux, wobei X der Platzhalter für die Plattennummer ist.

Demultiplexierungsdateien beginnen mit einem Dateivorspann mit folgenden Daten:

- Version (4-Byte-Integer), derzeit 1
- Clusterzahl (4-Byte-Integer)

Der Rest der Datei besteht aus Probennummern für jeden Cluster der Platte.

Nach Abschluss des Demultiplexierungsschritts generiert die Software eine Demultiplexierungsdatei mit dem Namen DemultiplexSummaryF1L1.txt.

- F1 im Dateinamen ist der Platzhalter für die Fließzellennummer.
- L1 im Dateinamen ist der Platzhalter für die Plattennummer.
- Die Demultiplexierungsergebnisse werden in einer Tabelle mit einer Zeile für jede Platte und einer Spalte für jede Probe, einschließlich Probe 0, aufgeführt.
- Die in Index-Reads am häufigsten vorkommenden Sequenzen.

FASTQ-Dateiformat

FASTQ ist ein textbasiertes Dateiformat mit Base-Calls und Qualitätswerten pro Read. Jeder Datensatz besteht aus vier Zeilen mit:

- dem Bezeichner
- der Sequenz
- einem Pluszeichen (+)
- den Phred-Qualitäts-Scores in einem ASCII-33-codierten Format

Der Bezeichner hat folgendes Format:

@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=AAAAAAAAAA9#:<#<;<<<????#=
```

BAM-Dateiformat

Eine BAM-Datei (*.bam) ist die komprimierte Binärversion einer SAM-Datei, die für die Darstellung alignierter Sequenzen von bis zu 128 Mb verwendet wird. Eine ausführliche Beschreibung des SAM- und des BAM-Formats finden Sie unter samtools.github.io/hts-specs/SAMv1.pdf.

BAM-Dateien verwenden das Dateinamensformat Probenname_S#.bam, wobei # für die Probennummer steht, die sich von der Reihenfolge ableitet, in der die Proben für den Lauf aufgeführt sind.

BAM-Dateien enthalten einen Dateivorspann und einen Alignment-Bereich:

- **Dateivorspann:** Enthält Informationen über die gesamte Datei, z. B. den Namen der Probe, die Probenlänge und die Alignment-Methode. Alignments im Alignment-Bereich sind mit spezifischen Informationen im Kopfbereich verbunden.
- **Alignment-Bereich**: Enthält den Read-Namen, die Read-Sequenz, die Read-Qualität, Alignment-Informationen und anwendungsspezifische Tags. Der Read-Name enthält das Chromosom, die Startkoordinate, die Alignment-Qualität und die Übereinstimmungsdeskriptor-Zeichenfolge.

Im Alignment-Bereich sind für jeden Read bzw. jedes Read-Paar folgende Informationen aufgeführt:

- **AS:** Paired-End-Alignment-Qualität.
- **BC:** Barcode-Tag, das die dem Read zugeordnete demultiplexierte Proben-ID angibt.
- **SM:** Single-End-Alignment-Qualität.
- **XC:** Übereinstimmungsdeskriptor-Zeichenfolge.
- **XN:** Amplikon-Namens-Tag, das die dem Read zugeordnete Amplikon-ID aufzeichnet.

BAM-Indexdateien (*.bam.bai) zeigen einen Index der entsprechenden BAM-Datei.

VCF-Dateiformat

Das Variant Call Format (VCF) ist ein gängiges Dateiformat, das von der Genomik-Wissenschaftsgemeinde entwickelt wurde. Es enthält Informationen über Varianten, die an spezifischen Positionen in einem Referenzgenom gefunden wurden. VCF-Dateien haben die Dateierweiterung „.vcf“.

Der Dateivorspann der VCF-Datei enthält die Version des VCF-Dateiformats und des Varianten-Callers sowie eine Liste der im Rest der Datei verwendeten Annotationen. Im VCF-Dateivorspann sind außerdem die Referenzgenomdatei und die BAM-Datei angegeben. Die letzte Zeile im Dateivorspann enthält die Spaltenüberschriften für die Datenzeilen. Jede Datenzeile in der VCF-Datei enthält Informationen zu einer Variante.

Überschriften der VCF-Datei

Überschrift	Beschreibung
CHROM	Das Chromosom des Referenzgenoms. Chromosomen werden in derselben Reihenfolge wie in der FASTQ-Referenzdatei aufgeführt.
POS	Die Einzelbasenposition der Variante im Referenzchromosom. Bei SNPs ist diese Position die Referenzbase mit der Variante. Bei Indels oder Deletionen ist diese Position die Referenzbase unmittelbar vor der Variante.
ID	Die rs-Nummer für die Variante aus dbSNP.txt, falls vorhanden. Wenn es mehrere rs-Nummern an dieser Position gibt, wird die Liste durch Semikola getrennt. Wenn an dieser Position kein dbSNP-Eintrag existiert, wird eine Kennung für einen fehlenden Wert ('.') verwendet.
REF	Der Referenz-Genotyp. Beispielsweise wird die Deletion eines einzelnen T als Referenz-TT und alternatives T dargestellt. Eine Einzelnukleotid-Variante A bis T wird als Referenz-A und alternatives T dargestellt.
ALT	Die Allele, die sich vom Referenz-Read unterscheiden. Beispielsweise wird die Insertion eines einzelnen T als Referenz-A und alternatives AT dargestellt. Eine Einzelnukleotid-Variante A bis T wird als Referenz-A und alternatives T dargestellt.
QUAL	Ein vom Varianten-Caller zugewiesener Phred-skalierter 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 in Relation zur beobachteten Fehlerrate hoch sind.

Anmerkungen in der VCF-Datei

Überschrift	Beschreibung
FILTER	Wenn alle Filter passiert werden, wird PASS in die Filterspalte geschrieben. • LowDP: Wird auf Positionen angewendet, deren Abdeckungstiefe geringer als 450-fach in jedem Pool ist. Bei Amplikonpositionen, die vom Vorwärts- und vom Rückwärtsread abgedeckt werden, entspricht dies einer 900-fachen Single-Read-Abdeckung. • LowGQ: Die Genotypisierungsqualität (GQ) liegt unter dem Cutoff. • q30: Qualitäts-Score < 30. • LowVariantFreq: Die Variantenhäufigkeit liegt unter dem angegebenen Grenzwert. • PB: Probe Pool Bias (Sondenpool-Verzerrung). Die Variante wurde nicht oder mit geringer Häufigkeit in einem oder zwei Sondenpools gefunden. • R3x6: Anzahl der an die Varianten-Calls angrenzenden Wiederholungen (Länge von einem bis drei Basenpaaren) ≥ 6. • SB: Der Strand Bias (die Strangverzerrung) liegt über dem angegebenen Grenzwert.
INFO	Die Spalte „INFO“ kann folgende Angaben enthalten: • AC: Allelanzahl in Genotypen für jedes ALT-Allel, in derselben Reihenfolge wie aufgeführt. • AF: Allelhäufigkeit für jedes ALT-Allel, in derselben Reihenfolge wie aufgeführt. • AN: Gesamtzahl der Allele in aufgerufenen Genotypen. • CD: Markierung, die angibt, dass der SNP in der codierenden Region von mindestens einem RefGene-Eintrag vorkommt. • DP: Tiefe (Anzahl der Base-Calls, die an einer Position ausgerichtet sind und beim Varianten-Calling verwendet werden). • Exon: Eine kommagetrennte Liste der Exon-Regionen, die aus dem RefGene gelesen wurden. • FC: Functional Consequence (funktionale Konsequenz). • GI: Eine kommagetrennte Liste der Gen-IDs, die aus dem RefGene gelesen wurden. • QD: Variantenzuverlässigkeit/-qualität nach Tiefe. • TI: Eine kommagetrennte Liste der Transkript-IDs, die aus dem RefGene gelesen wurden.
FORMAT	Die Spalte „Format“ enthält durch Doppelpunkt getrennte Felder. Beispiel: GT:GQ. Welche Felder in der Liste aufgeführt werden, hängt vom verwendeten Varianten-Caller ab. Folgende Felder sind verfügbar: • AD: Eintrag im Format X,Y, wobei X für die Anzahl der Referenz-Calls und Y für die Anzahl der alternativen Calls steht. • DP: Ungefähre Read-Tiefe. Reads mit MQ = 255 oder mit fehlerhaften „Mates“ werden herausgefiltert. • GQ: Genotypqualität. • GQX: Genotypqualität. GQX ist das Minimum des GQ-Werts und der QUAL-Spalte. In der Regel sind diese Werte ähnlich. Als Minimum ist GQX der konservativere Messwert für die Genotypqualität. • GT: Genotyp. 0 entspricht der Referenzbase, 1 entspricht dem ersten Eintrag in der ALT-Spalte usw. Der Schrägstrich (/) gibt an, dass keine Phasierungsinformationen vorhanden sind. • NC: Anteil der Basen, bei denen kein Call erfolgte oder deren Base-Call-Qualität unter dem Mindestwert lag. • NL: Noise Level (Rauschpegel) – geschätzte Rauschstärke beim Base-Calling an dieser Position. • PB: Probe Pool Bias (Sondenpool-Verzerrung). Werte gegen 0 weisen auf eine größere Verzerrung eines Sondenpools und auf die geringere Zuverlässigkeit eines Varianten-Calls hin. • SB: Strand Bias (Strangverzerrung) an dieser Position. Höhere negative Werte geben eine geringere Verzerrung und Werte nahe 0 eine höhere Verzerrung an. • VF: Variant frequency (Variantenhäufigkeit). Der Prozentsatz der Reads, die das alternative Allel unterstützen.
SAMPLE (PROBE)	Die Probenspalte enthält die Werte, die in der Spalte „FORMAT“ angegeben sind.

Genom-VCF-Dateien

Genom-VCF-Dateien (gVCF) sind Dateien im Format VCF v4.1, die mehrere Konventionen befolgen, um alle Bereiche im Genom in einem angemessenen kompakten Format darzustellen. gVCF-Dateien (*.genome.vcf.gz) erfassen für jede Probe alle Bereiche der Region von Interesse in einer einzigen Datei.

Die gVCF-Datei weist No-Calls an Positionen aus, die nicht alle Filter passieren. Das Genotyp-Tag (GT) **./.** gibt einen No-Call an.

Weitere Informationen finden Sie unter sites.google.com/site/gvcftools/home/about-gvcf.

VCF-Dateien pro Pool und Konsensus-VCF-Dateien

Der Workflow für somatische Varianten erzeugt zwei Sets von Varianten-Call-Dateien.

- **Pro-Pool-VCF-Dateien:** Enthalten Varianten-Calls entweder aus dem Vorwärts- oder aus dem Rückwärts-Pool. Die Pro-Pool-Dateien werden im Ordner „VariantCallingLogs“ gespeichert.
- **Konsensus-VCF-Dateien:** Enthalten Varianten-Calls aus beiden Pools. Konsensus-Dateien werden im Alignment-Ordner gespeichert.

Zu den Pro-Pool-VCF-Dateien und den Konsensus-VCF-Dateien gehören sowohl VCF-Dateien (*.vcf) als auch gVCF-Dateien (*.genome.vcf). Es gelten folgende Namenskonventionen, wobei P# die Reihenfolge repräsentiert, in der die Proben für den Lauf aufgeführt sind:

- **Berichte für alle Stellen:** ProbenName_P#.genome.vcf
- **Berichte nur für Varianten:** ProbenName_P#.vcf

Die Software vergleicht die Pro-Pool-VCF-Dateien und kombiniert die Daten an jeder Position, um eine Konsensus-VCF-Datei für die Probe zu erzeugen.

Die Varianten-Calls aus jedem Pool werden unter Verwendung der folgenden Kriterien in Konsensus-VCF-Dateien zusammengeführt.

Kriterien	Result (Ergebnis)
Ein Referenz-Call in jedem Pool	Referenz-Call
Ein Referenz-Call in einem Pool und ein Varianten-Call im anderen Pool	Gefilterter Varianten-Call
Übereinstimmende Varianten-Calls mit ähnlicher Häufigkeit in jedem Pool	Varianten-Call
Übereinstimmende Varianten-Calls mit signifikant unterschiedlicher Häufigkeit in jedem Pool	Gefilterter Varianten-Call
Nicht übereinstimmende Varianten-Calls in jedem Pool	Gefilterter Varianten-Call

Kennzahlen aus jedem Pool werden unter Verwendung folgender Werte zusammengeführt.

Kennzahl	Wert
Tiefe	Addition der Tiefe aus beiden Pools
Variantenhäufigkeit	Gesamtzahl der Varianten geteilt durch die Gesamtabdeckungstiefe
Q-Score	Mindestwert der beiden Pools

Amplikon-Abdeckungsdatei

Für jede Manifestdatei wird eine Amplikon-Abdeckungsdatei generiert. „M#“ im Dateinamen steht für die Manifestnummer, die in der Probentabelle für den Lauf aufgeführt ist.

Jede Datei beginnt mit einem Dateivorspann, der die dem Manifest zugeordneten Proben-IDs enthält. In den drei Spalten unter dem Dateivorspann sind folgende Informationen aufgeführt:

- Target-ID, wie im Manifest aufgelistet.
- Abdeckungstiefe der Reads nach Filterung.
- Gesamtabdeckungstiefe.

Ergänzende Ausgabedateien

Die folgenden Ausgabedateien bieten ergänzende Informationen oder eine Zusammenfassung von Laufergebnissen und Analysefehlern. Zwar sind diese Dateien für die Beurteilung der Analyseergebnisse nicht erforderlich, sie können jedoch für die Fehlerbehebung verwendet werden. Die Dateien befinden sich im Alignment-Ordner, sofern nicht anders angegeben.

Dateiname	Beschreibung
AnalysisLog.txt	Verarbeitungsprotokoll, in dem jeder Schritt während der Analyse des aktuellen Laufordners beschrieben ist. Diese Datei enthält keine Fehlermeldungen. Die Datei befindet sich im Ordner „Alignment“.
AnalysisError.txt	Verarbeitungsprotokoll, in dem alle während der Analyse aufgetretenen Fehler aufgeführt sind. Wenn keine Fehler auftreten, enthält die Datei keine Einträge. Die Datei befindet sich im Ordner „Alignment“.
DemultiplexSummaryF1L1#.txt	Gibt die Demultiplexierungsergebnisse in einer Tabelle mit einer Zeile für jede Platte und einer Spalte für jede Probe an. „#“ ist der Platzhalter für Lane 1, 2, 3 oder 4 der Fließzelle. Die Datei befindet sich im Ordner „Alignment“.
AmpliconRunStatistics.xml	Enthält eine laufspezifische Zusammenfassungsstatistik. Die Datei befindet sich im Ordner „Alignment“.

Analyseordner

Im Analyseordner werden die von der Local Run Manager-Software generierten Dateien gespeichert.

Das Zusammenspiel zwischen dem Ausgabeordner und dem Analyseordner lässt sich wie folgt zusammenfassen:

- Während der Sequenzierung stellt die Echtzeitanalyse (RTA) die bei der Bildanalyse, dem Base-Calling und der Qualitätsbewertung generierten Dateien in den Ausgabeordner.
- Die Echtzeitanalyse kopiert Dateien in Echtzeit in den Analyseordner. Nachdem die Echtzeitanalyse jeder Base für jeden Zyklus einen Qualitäts-Score zugeordnet hat, legt sie in beiden Ordnern die Datei RTAComplete.txt an.
- Wenn die Datei RTAComplete.txt vorhanden ist, wird die Analyse gestartet.
- Während der Ausführung der Analyse erstellt Local Run Manager Ausgabedateien im Analyseordner und kopiert diese anschließend zurück in den Ausgabeordner.

Alignment-Ordner

Jedes Mal, wenn die Analyse erneut in die Warteschlange gestellt wird, erstellt Local Run Manager einen Alignment-Ordner mit dem Namen **Alignment_N**, wobei N eine fortlaufende Zahl ist.

Ordnerstruktur

- **Alignment:** Enthält *.bam-, *.vcf- und FASTQ-Dateien sowie analysemodulspezifische Dateien.
  - **Date and Time Stamp:** Datum und Uhrzeit der Analyse in folgendem Format: JJJJMMTT_HHMMSS

- AnalysisError.txt
- AnalysisLog.txt
- AmpliconRunStatistics.xml
- Sample1.genome.vcf.gz
- Sample1.coverage.csv
- Sample1.report.pdf
- Sample1.summary.csv
- Sample1.vcf.gz
- Sample1.bam
- FASTQ
  - Sample1
  - Stats
    - DemuxSummaryF1L1.txt
    - FastqSummaryF1L1.txt

Data

- Intensities
  - BaseCalls
    - L001: Enthält einen Unterordner pro Zyklus mit *.bcl-Dateien.
  - L001: Enthält *.locs-Dateien, eine Datei pro Platte.
- RTA Logs: Enthält Protokolldateien von der RTA-Software-Analyse.

InterOp: Enthält Binärdateien, die zur Übermittlung der Sequenzierungslaufkennzahlen verwendet werden.

Logs: Enthält Protokolldateien, in denen die bei der Sequenzierung durchgeführten Schritte aufgeführt sind.

RTAComplete.txt

RunInfo.xml

runParameters.xml

Technische Unterstützung

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

Website: www.illumina.com
E-Mail: techsupport@illumina.com

Telefonnummern des Illumina-Kundendiensts

Region	Gebührenfrei	Regional
Nordamerika	+1.800.809.4566	
Australien	+1.800.775.688	
Belgien	+32 80077160	+32 34002973
China	400.635.9898	
Deutschland	+49 8001014940	+49 8938035677
Dänemark	+45 80820183	+45 89871156
Finnland	+358 800918363	+358 974790110
Frankreich	+33 805102193	+33 170770446
Großbritannien	+44 8000126019	+44 2073057197
Hongkong	800960230	
Irland	+353 1800936608	+353 016950506
Italien	+39 800985513	+39 236003759
Japan	0800.111.5011	
Neuseeland	0800.451.650	
Niederlande	+31 8000222493	+31 207132960
Norwegen	+47 800 16836	+47 21939693
Schweden	+46 850619671	+46 200883979
Schweiz	+41 565800000	+41 800200442
Singapur	+1.800.579.2745	
Spanien	+34 911899417	+34 800300143
Taiwan	00806651752	
Österreich	+43 800006249	+43 19286540
Andere Länder	+44.1799.534000	

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

Die **Produktdokumentation** steht auf der Illumina-Website im PDF-Format zum Herunterladen zur Verfügung. Gehen Sie zu support.illumina.com, wählen Sie ein Produkt und wählen Sie anschließend **Documentation & Literature** (Dokumentation und Literatur).

Dokument-Nr. 1000000012183 v01 DEU

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

IVD

EC REP

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
Melbourne, VIC 3000
Australien

FÜR IN-VITRO-DIAGNOSTIK

illumina®