



# DRAGEN TruSight Oncology 500 Analysis Software v1.1 (Local)

User Guide

ILLUMINA PROPRIETARY

文書番号：200014849 v00 JPN

2022年4月

**本製品の使用目的は研究に限定されます。診断での使用はできません。**

本文書およびその内容は、Illumina, Inc. およびその関連会社（以下、「イルミナ」という）の所有物であり、本文書に記載された製品の使用に関連して、イルミナの顧客が契約上使用することのみを意図したものであり、その他の目的を意図したものではありません。本文書およびその内容を、イルミナの書面による事前同意を得ずにその他の目的で利用または配布してはならず、また方法を問わず、その他伝達、開示または複製してはなりません。イルミナは、本文書によって、自身の特許、商標、著作権またはコモンロー上の権利に基づくいかなるライセンスも譲渡せず、また第三者の同様の権利も譲渡しないものとします。

本文書に記載された製品の適切かつ安全な使用を徹底するため、資格を有した、適切なトレーニングを受けた担当者が、本文書の指示を厳密かつ明確に遵守しなければなりません。当該製品の使用に先立ち、本文書のすべての内容を熟読し、理解する必要があるものとします。

本文書に含まれるすべての説明を熟読せず、明確に遵守しない場合、製品を損ない、使用者または他者を含む個人に傷害を及ぼし、その他の財産に損害を与える結果となる可能性があり、また本製品に適用される一切の保証は無効になるものとします。

イルミナは、本文書に記載された製品（その部品またはソフトウェアを含む）の不適切な使用から生じる責任、または、顧客による当該製品の取得に関連してイルミナから付与される明示的な書面によるライセンスもしくは許可の範囲外で当該製品が使用されることから生じる責任を一切負わないものとします。

© 2022 Illumina, Inc. All rights reserved.

すべての商標および登録商標は、Illumina, Inc. または各所有者に帰属します。商標および登録商標の詳細は [jp.illumina.com/company/legal.html](http://jp.illumina.com/company/legal.html) をご覧ください。

# 目次

<b>概要</b> .....	<b>1</b>
インストール要件 .....	2
DRAGEN TruSight Oncology 500 Analysis Softwareのインストール .....	3
DRAGEN TruSight Oncology 500 Analysis Softwareのアンインストール .....	4
<b>DRAGEN TruSight Oncology 500 Analysis Softwareの実行</b> .....	<b>5</b>
サンプルシートの要件 .....	5
コマンドラインオプション .....	7
装置ランフォルダーから開始 .....	8
FASTQファイルから開始 .....	10
複数のDRAGENサーバー上での実行 .....	10
<b>解析方法</b> .....	<b>12</b>
DNA解析方法 .....	13
RNA解析方法 .....	20
品質管理 (QC) .....	21
<b>解析の出力</b> .....	<b>23</b>
メトリクスの出力 .....	23
シングルノード解析の出力フォルダーの構成 .....	24
複数ノード解析の出力フォルダーの構成 .....	27
組み合わせバリエーション出力 .....	27
DNA出力 .....	29
RNA出力 .....	31
ブロックリスト .....	36
<b>トラブルシューティング</b> .....	<b>37</b>
DNA拡張メトリクス .....	37
RNA拡張メトリクス .....	38
<b>テクニカルサポート</b> .....	<b>40</b>

# 概要

Illumina® DRAGEN™ TruSight™ Oncology 500 Analysis Software v1.1は、ホルマリン固定パラフィン包埋（FFPE）組織検体から生成されたDNAおよびRNAライブラリーのローカル解析をサポートします。TruSight Oncology 500アッセイは、DNAバイオマーカーの523の遺伝子について、生物学的に関連のあるコーディングエクソンと補助領域にわたり低頻度の体細胞バリエーションで高い感度と特異度が得られるように最適化されています。DNAバイオマーカーには以下が含まれます。

- 1塩基変異（SNV）
- 挿入
- 欠失
- コピー数バリエーション（CNV）
- 複数塩基変異（MNV）

TruSight Oncology 500はDNAの腫瘍変異負荷（TMB）やマイクロサテライト不安定性（MSI）など免疫療法のバイオマーカーも検出します。DNAライブラリー解析の出力には、TMB、スモールバリエーションや複合バリエーションのバリエーションコールファイル、MSI、遺伝子増幅などがあります。融合遺伝子とスプライスバリエーションは55遺伝子のRNAで検出され、RNAライブラリー解析の出力には融合およびスプライスバリエーションコールファイルが含まれます。

対応する領域の詳細はアッセイのマニフェストファイルに記載されています。お住まいの地域のイリミナ担当者にお尋ねください。

DRAGEN TruSight Oncology 500 Analysis Software v1.1は、単一のDRAGENサーバー上で、または複数サーバーに分割して解析できます。

## 互換性

イリミナの[サポートウェブサイト](#)のDRAGEN TruSight Oncology 500 Analysis Software v1.1サポートページでは、イリミナのシーケンスシステムとの互換性に関する情報を提供しています。

DRAGEN TruSight Oncology 500 Analysis Software v1.1のFASTQファイルを生成するにはBCL Convertを使用します。bcl2fastqを使用しても同じ結果は得られないため推奨されません。DRAGEN TruSight Oncology 500 Analysis Software v1.1とBCL Convertを使用するための設定および互換性に関する情報については、イリミナの[サポートウェブサイト](#)のDRAGEN TruSight Oncology 500 Analysis Software v1.1サポートページを参照してください。

## 追加リソース

イリミナの[サポートウェブサイト](#)のDRAGEN TruSight Oncology 500 Analysis Softwareサポートページでは追加のリソースを提供しています。これらのリソースには、ソフトウェア、トレーニング、適合製品、サンプルシート、および以下の添付資料が含まれています。サポートページで常に最新バージョンの資料をご確認ください。

文書	内容説明
『TruSight Oncology 500 Reference Guide』（文書番号：1000000067621）	TruSight Oncology 500 Kit の使用に関する情報を記載

## インストール要件

DRAGEN TruSight Oncology 500 Analysis Softwareは、Illumina DRAGEN Server v3と互換性があります。

### ハードウェア

- DRAGEN TruSight Oncology 500 Analysis Softwareは、DRAGENサーバー上でのみ動作します。
- アッセイパイプラインでは、ネットワークアタッチストレージ（NAS）上でmkfifoが有効である必要があります。

### ソフトウェア

- デフォルトで、Linux CentOS 7.3以降のオペレーティングシステムが提供されます。
- DRAGEN TruSight Oncology 500 Analysis Softwareをインストールする前に、コンテナエンジン Docker v18.09以降が必要です。Dockerの文書でCentOSのインストール手順を参照してください。

### ストレージ要件

最適な性能を発揮するには、DRAGENサーバー上にローカルに格納されたデータで解析を実行します。NAS上に格納されたデータの解析には長時間を要し、性能の面で信頼性が低下する可能性があります。

DRAGENサーバーには、ソフトウェアの出力ディレクトリとして使用する/stagingディレクトリにNVMe SSDが搭載されています。長期間の保管には、ネットワークアタッチストレージが必要です。

DRAGEN TruSight Oncology 500 Analysis Softwareの実行時には、`--analysisFolder`コマンドラインオプションを/stagingのディレクトリに設定し、DRAGENサーバープロセスでNVMe SSDにデータを読み書きできるようにします。

解析を開始する前に、DRAGENサーバーからデータをネットワークアタッチストレージにコピーして移す方策を講じることを推奨します。DRAGENサーバー上にある出力データをできる限り速やかに削除してください。

各シーケンスシステムの101 bpあたりのランおよび解析の出力サイズを以下に示します。

シーケンスシステム	ランフォルダー出力 (Gb)	解析の出力 (Gb)
NovaSeq 6000 SPフローセル	85~100	250~374
NovaSeq 6000 S1フローセル	164~200	360~665
NovaSeq 6000 S2フローセル	290~460	890~1,600
NovaSeq 6000 S4フローセル	800~1,200	2,700~4,100
NextSeq 500/550および550Dx HOフローセル	32~55	82~85

解析の実行時に十分なディスクスペースが利用できない場合、ランが完了しません。

# DRAGEN TruSight Oncology 500 Analysis Softwareのインストール

イルミナでは、Dockerグループにユーザーを追加してから、root以外のユーザーとしてDockerを実行することを推奨します。DRAGEN TruSight Oncology 500 Analysis Softwareをrootとして実行することも可能ですが、この方法は推奨されません。Dockerの許可要件とrootとして実行する代替手段については、Dockerの文書のLinuxのページでインストール後の手順を参照してください。dockerイメージをロードするには、sudo権限または権限昇格が必要となります。

DRAGEN TruSight Oncology 500 Analysis Softwareのインストールには、root権限が必要です。インストールスクリプトを実行すると、サーバー上の既存のDRAGENソフトウェアがアンインストールされます。別のDRAGENパイプラインを使用するには、DRAGEN TruSight Oncology 500 Analysis Softwareをアンインストールして、Illumina DRAGEN Bio-IT PlatformサポートページからDRAGENソフトウェアのインストールパッケージをダウンロードします。DRAGEN TruSight Oncology 500 Analysis Softwareをアンインストールするには、[4ページの「DRAGEN TruSight Oncology 500 Analysis Softwareのアンインストール」](#)を参照してください。

## インストール手順

1. イルミナカスタマーサポートに連絡して、DRAGEN TruSight Oncology 500 Analysis Softwareのインストールパッケージを入手します。
2. Dockerの文書でCentOSのインストール手順を参照して、Docker 18.09以降をインストールします。
3. イルミナから届く電子メールに従ってDRAGEN TSO 500インストールスクリプトをダウンロードします。リンクの有効期限は7日間です。
4. インストールスクリプトを/stagingディレクトリにコピーしてその場所に保存します。

5. 以下のコマンドを用いてランスクリプトの権限を更新します。

```
sudo chmod +x /staging/install_DRAGEN_TSO500-<version>.run
```

6. 以下のコマンドを用いてインストールスクリプトを実行します。所要時間は約10分です。

```
sudo TMPDIR=/staging /staging/install_DRAGEN_TSO500-<version>.run
```

スクリプトを実行すると、インストール済みのDRAGENサーバーソフトウェアがアンインストールされます。インストールの過程でシステムをリブートまたはパワーサイクルするようメッセージが表示されることがあります。これはDRAGENサーバーのFPGAハードウェアのインストールを完了させるために必要です。システムをパワーサイクルするには、サーバーをシャットダウンして再起動する必要があります。

7. 以下のコマンドを用いてDRAGENサーバーのライセンスをインストールします。

```
sudo /opt/edico/bin/dragen_lic -i auto
```

8. 以下のコマンドを用いてDRAGENサーバーのハッシュテーブルをビルドします。所要時間は約20分です。

```
sudo /usr/local/bin/build-hashtable_DRAGEN_TSO500-<version>.sh
```

## システムチェックの実行

以下のコマンドを用いてシステムが適切に機能することを確認します。

```
sudo /usr/local/bin/check_DRAGEN_TSO500-<version>.sh
```

このスクリプトでは以下の機能を確認します。

- 要求されたすべてのサービスが動作しているか
- 正しいDockerイメージがインストールされているか
- Illumina DRAGEN TruSight Oncology 500パイプラインがテストデータセットで正常に動作するか

このセルフテストには約30分かかります。セルフテストでエラーメッセージが出力される場合は、イルミナテクニカルサポートに連絡し、/staging/check\_DRAGEN\_TSO500\_<timestamp>.tgzの出力ファイルを提供してください。

Secure Shell Protocol (SSH) を介してDRAGENサーバー上で解析を実行するとき、解析実行コマンドの前にハングアップを無視するnohupコマンドを指定します。このコマンドはDRAGENサーバーとの接続が切断された場合でも解析が終了しないようにします。解析コマンドからの出力はすべて、そのディレクトリ内のnohup.outにリダイレクトされます。

MacOSを使用する場合は、以下のエラーが起こらないように、ターミナル設定で環境変数を自動的に設定する機能を無効にします。

```
ERROR: locale::facet_S_create_c_locale name not valid
```

## DRAGEN TruSight Oncology 500 Analysis Softwareのアンインストール

DRAGEN TruSight Oncology 500 Analysis Softwareのインストールには、uninstall\_DRAGEN\_TSO500-<VERSION>.shというアンインストールスクリプトが含まれています。このスクリプトは/usr/local/binにインストールされています。

アンインストールスクリプトを実行すると、以下のアセットが削除されます。

- すべてのスクリプト (DRAGEN\_TSO500.sh、test\_DRAGEN\_TSO500-<VERSION>.sh、およびuninstall\_DRAGEN\_TSO500-<VERSION>.sh)
- staging/illumina/DRAGEN\_TSO500に存在するリソース
- dragen\_tso500:<VERSION>のDockerイメージ

以下のコマンドをrootユーザーとして実行し、DRAGEN TruSight Oncology 500 Analysis Softwareをアンインストールします。

```
uninstall_DRAGEN_TSO500-<version>.sh
```

DockerやDRAGENをアンインストールしないでください。DockerやDRAGENは、関連のRPMパッケージを削除することで削除できます。

# DRAGEN TruSight Oncology 500 Analysis Softwareの実行

DRAGEN TruSight Oncology 500 Analysis Softwareは、`/usr/local/bin`ディレクトリにインストールされている`DRAGEN_TSO500.sh`というBashスクリプトにより開始します。Bashスクリプトはコマンドラインで実行され、Dockerによりソフトウェアを動作させます。

引数については、[7ページの「コマンドラインオプション」](#)を参照してください。BCLファイルから、またはBCL Convertiによって作成されるFASTQフォルダーから開始できます。いずれの方法にも以下の要件が適用されます。

- シーケンスランまたはFASTQフォルダーへのパス。ランまたはFASTQフォルダーをDRAGENサーバーへ、推奨される編成`/staging/runs/{RunID}`のstagingフォルダー内にコピーします。ランフォルダーのDRAGENサーバーへのコピーは、`rsync`などのLinuxコマンドで実行できます。コマンドラインで別の指定をしない限り、ランフォルダー内のサンプルシートが使用されます。
- リソースファイルフォルダーへのパス。このフォルダーの場所はインストール中に事前設定されます。このフォルダー内のファイルを変更すると実行時にエラーが発生します。
- ハッシュテーブルへのパス。このファイルの場所はインストール中に事前設定されます。
- 解析の出力フォルダーへのパス。このフォルダーが作成され、解析の出力ファイルが格納されます。

## サンプルシートの要件

DRAGEN TruSight Oncology 500 Analysis Softwareのサンプルシートは各解析につき1つが必要です。サンプルシートは、シーケンスランのセットアップと解析のための情報を格納するCSVファイル (\*.csv) です。サンプルシートは、サンプルのリストと各サンプルのインデックスシーケンスで構成されています。DRAGEN TruSight Oncology 500 Analysis Softwareはv1とv2のサンプルシートをサポートします。バージョン間の相違点の詳細については、[イルミナサポートサイト](#)を参照してください。

シーケンスシステムのランセットアップのステップ中にサンプルシートを用意します。ランに合わせた適切なサンプルシートのテンプレートについては、[イルミナサポートサイト](#)を参照してください。

サンプルシートは、サンプルのリストと各サンプルのインデックスシーケンスで構成されています。必要のないインデックスIDは削除してください。さまざまな種類のシーケンスランで、異なるインデックスアダプターを使用することがあります。DRAGEN TruSight Oncology 500 Analysis Softwareのリソースセットに記載されたインデックスIDを使用してください。

### サンプルシートの作成

以下のステップに従って、TruSight Oncology 500のサンプルシートを作成してください。

- [イルミナサポートサイト](#)のTruSight Oncology 500ページから適切なサンプルシートのテンプレートをダウンロードします。

2. [Data] セクションに、各サンプルの必須パラメーターを入力します。次の表に必須パラメーターとオプションのパラメーターを示します。

サンプルのパラメーター	必須	詳細
Sample_ID	はい	Sample_IDは出力ファイル名の一部になる。サンプルIDでは、大文字と小文字が区別されない。サンプルIDは以下の特徴を備える必要がある <ul style="list-style-type: none"> <li>各ランに固有のもの</li> <li>スペースを含まない</li> <li>英数字、アンダースコア、ダッシュ。アンダースコアまたはダッシュを使用する場合は、アンダースコアまたはダッシュの前後に英数字を入れる。例：Sample1-T5B1_022515</li> <li>all、default、none、unknown、undetermined、stats、またはreportsという名称にすることはできない</li> <li>ペアIDに基づくサンプルIDを推奨する。例： &lt;PairID&gt;-DNA、&lt;PairID&gt;-RNA</li> </ul>
Sample_Type	はい	DNAまたはRNAと入力する
Sample_Name	いいえ	サンプルの名前
Sample_Plate	いいえ	サンプルプレートの名前
Sample_Well	いいえ	サンプルウェルの名前
Pair_ID	はい	同じ個体のDNAサンプルとRNAサンプルを対にするために使用する。2つのサンプルを関連付けるために共通のペアIDを使用する
Index adapter sequences	はい	サンプルに対してTruSight Oncology 500に互換性のあるインデックスアダプターシーケンスを入力する。各サンプルに固有のインデックスペアを選択する。インデックスの例をSampleSheet.csvに示す
Index	はい	Index 1シーケンス
Index2	はい	Index 2シーケンス
I7_Index_ID	いいえ	I7インデックスID
I5_Index_ID	いいえ	I5インデックスID
Description	いいえ	サンプルの説明
Lane	いいえ	NovaSeq 6000 XPワークフローのみに使用する。どのレーンが特定のサンプルに対応しているかを示す。行ごとに1つの数値を入力する。ヘッダーがある場合、空欄にすることはできない

3. 以下のいずれかの方法でシーケンスランフォルダーにサンプルシートを保存します。
- サンプルシートをSampleSheet.csvという名前で保存します。
  - サンプルシートに任意の名前を付け、コマンドラインオプションにサンプルシートへのパスを指定します。

## コマンドラインオプション

DRAGEN TruSight Oncology 500 Analysis Softwareでは、以下のコマンドラインオプションを使用できます。例えば、表 1を参照してください。

入力要件の詳細を確認する場合は、--helpコマンドラインオプションを使用します。

オプション	必須	内容説明
--help	いいえ	使用可能なオプションをヘルプ画面に表示する
--analysisFolder	はい	ローカルの解析フォルダーへのフルパスを指定する。フォルダーには十分なスペースがあり、NVMe SSDドライブ上になければならない <ul style="list-style-type: none"> <li>• NovaSeqは3,000 GB以上の空きを必要とする</li> <li>• NextSeqは500 GB以上の空きを必要とする</li> </ul>
--resourcesFolder	はい	ローカルのリソースフォルダーへのフルパスを指定する
--runFolder	はい	--fastqFolderが指定されていない場合は必須。ローカルのランフォルダーへのフルパスを指定する
--fastqFolder	はい	--runFolderが指定されていない場合は必須。ローカルのFASTQフォルダーへのフルパスを指定する。この場所から解析が開始される
--user	いいえ	dockerではオプション。Dockerコンテナ内で使用するユーザーIDを指定する
--remove	いいえ	dockerではオプション。後のコンテナを実行後に削除するには、--rmオプションにコンテナを渡す
--version	いいえ	ソフトウェアのバージョンを表示する
--sampleSheet	いいえ	ランフォルダー内のSampleSheet.csvとして指定しない場合は、ファイル名を含むフルパスを指定する
--sampleOrPairIDs	いいえ	このノードで処理する必要があるカンマ区切りのサンプルIDまたはペアIDを指定する。例： Pair_1, Pair_2, Sample_1
--demultiplexOnly	いいえ	デマルチプレックスを実施し、追加の解析なしでFASTQのみを生成する
--gather	いいえ	このオプションの後に、結果を1つのResultsフォルダーに集約する任意のディレクトリを指定する
--hashtableFolder	いいえ	ローカルのDRAGENハッシュテーブルへのフルパスを指定する

## 装置ランフォルダーから開始

その他のコマンドについては、[7ページの「コマンドラインオプション」](#)を参照してください。

コマンドラインにファイルパスを指定するときは、フルパスを使用します。&、\*、#、スペースなど特殊文字の使用は避けてください。

NovaSeq 6000システムを使用してDRAGEN TruSight Oncology 500 Analysis Softwareを実行する場合は、以下のコマンドのいずれかに--isNovaSeqを追加します。

1. 新たな解析を開始する前に、実行中のDRAGEN TruSight Oncology 500 Analysis Softwareコンテナが完了するまで待ちます。実行中のコンテナのリストを生成するために、以下のコマンドを実行します。

```
sudo docker ps
```

2. 次のオプションのいずれかを選択します。

- ランフォルダー内のBCLファイルとランフォルダー内のサンプルシートから開始します。

```
DRAGEN_TSO500.sh \
--resourcesFolder /staging/illumina/DRAGEN_TSO500/resources \
--hashtableFolder /staging/illumina/DRAGEN_TSO500/ref_hashtable \
--runFolder /staging/{RunFolderName} \
--analysisFolder /staging/{AnalysisFolderName}
```

- ランフォルダー内のBCLファイルから、別のサンプルシートを指定して開始します。

```
DRAGEN_TSO500.sh \
--resourcesFolder /staging/illumina/DRAGEN_TSO500/resources \
--hashtableFolder /staging/illumina/DRAGEN_TSO500/ref_hashtable \
--runFolder /staging/{RunFolderName} \
--analysisFolder /staging/{AnalysisFolderName} \
--sampleSheet /staging/{SampleSheetName}.csv
```

- ランフォルダー内のBCLファイルから、別のサンプルシートとデマルチプレックスのみを指定して開始します。

```
DRAGEN_TSO500.sh \
--resourcesFolder /staging/illumina/DRAGEN_TSO500/resources \
--hashtableFolder /staging/illumina/DRAGEN_TSO500/ref_hashtable \
--runFolder /staging/{RunFolderName} \
--analysisFolder /staging/{AnalysisFolderName} \
--sampleSheet /staging/{SampleSheetName}.csv \
--demultiplexOnly
```

- FASTQフォルダーとFASTQフォルダー内のサンプルシートから開始します。

```
DRAGEN_TSO500.sh \
--resourcesFolder /staging/illumina/DRAGEN_TSO500/resources \
--hashtableFolder /staging/illumina/DRAGEN_TSO500/ref_hashtable \
--fastqFolder /staging/{FastqFolderName} \
--analysisFolder /staging/{AnalysisFolderName}
```

- FASTQフォルダーとFASTQフォルダー内のサンプルシートおよびサンプルまたはペアのサブセットから開始します。

```
DRAGEN_TSO500.sh \
--resourcesFolder /staging/illumina/DRAGEN_TSO500/resources \
--hashtableFolder /staging/illumina/DRAGEN_TSO500/ref_hashtable \
--fastqFolder /staging/{FastqFolderName} \
--analysisFolder /staging/{AnalysisFolderName} \
--samplePairIDs "Pair_1, Pair2"
```

## BCLファイルから開始

BCL (\*.bcl) ファイルから開始する場合、DRAGEN TruSight Oncology 500 Analysis Softwareでは、特定のファイルとフォルダーが含まれるランフォルダーが必要です。これらの入力はDockerに必要となります。

ランフォルダーにはシーケンスランのデータが含まれます。ランフォルダーから開始する場合、フォルダーに以下のファイルが含まれていることを確認してください。

フォルダー/ファイル	内容説明
Configフォルダー	構成ファイル
Dataフォルダー	*.bclファイル
Imagesフォルダー	(オプション) 生のシーケンス画像ファイル
InterOpフォルダー	Interopメトリクスファイル
Logsフォルダー	(オプション) シーケンスシステムのログファイル
RTALogsフォルダー	Real Time Analysis (RTA) のログファイル
RunInfo.xmlファイル	ラン情報
RunParameters.xmlファイル	ランパラメーター
SampleSheet.csvファイル	サンプル情報。ランフォルダー内にはないサンプルシート、またはSampleSheet.csvという名前ではないサンプルシートを使用する場合はフルパスを指定する

## FASTQファイルから開始

FASTQ (\*.fastq) ファイルを使用してDRAGEN TruSight Oncology 500 Analysis Softwareを実行するためには、以下の入力が必要です。この要件はDockerに適用されます。

- 既存のFASTQフォルダーへのフルパス
- FASTQフォルダーの構成が10ページの「FASTQファイル編成」のフォルダー構成と一致していること
- サンプルシートがFASTQフォルダーのパスに存在するか、--sampleSheet上書きコマンドでサンプルシートへのパスを設定できること

解析を完了するのに十分なディスクスペースがあることを確認します。ディスクスペースの要件については、--helpコマンドライン引数の詳細を参照してください。

### FASTQファイル編成

特定のサンプルIDに対応する個々のサブフォルダー内にFASTQファイルを保存します。ファイルペアを同一フォルダーに一緒に入れます。

DRAGEN TruSight Oncology 500 Analysis Softwareでは、サンプルごとに別々のFASTQファイルが必要です。FASTQファイルを統合しないでください。

装置では各フローセルレーンに対し2個のFASTQファイルが生成されるため、1サンプルあたり8個のFASTQファイルが得られます。

```
Sample1_S1_L001_R1_001.fastq.gz
```

- Sample1が、サンプルIDに相当します。
- S1のSはサンプルを意味し、S1の1はサンプルシート内のサンプルの順番に基づいています。そのため、S1は1番目のサンプルになります。
- L001はフローセルレーン番号に相当します。
- R1のRはリードを意味するため、R1はRead 1を表します。

## 複数のDRAGENサーバー上での実行

DRAGEN TruSight Oncology 500 Analysis Softwareは、処理時間を短縮するためにサンプルのサブセットを異なるDRAGENサーバー上で実行できます。これはスキャッター/ギャザーという3段階のプロセスによって可能になります。このプロセスはデマルチプレックス、解析、結果のギャザリングで構成されます。

第1段階はデマルチプレックスです。デマルチプレックスはランフォルダー全体で1回実行され、その実行においてサンプルごとにFASTQファイルを生成した後、サンプルファイルを該当のフォルダーに分割します。完了後、出力ディレクトリにはFASTQファイルを保持するサンプルディレクトリが含まれます。

複数のDRAGENサーバーに解析を分散させるプロセスを以下に示します。

1. 実行に使用可能なDRAGENサーバーの数を決定します。
2. 単一のDRAGENサーバー上でデマルチプレックスを実行します。
  - i** XP ワークフローを使用して複数の DRAGEN サーバーでランを順番に処理するには、サンプルシートを変更してレーンのサブセットを指定します。例えば、S2 フローセルの場合、2 つの変更したサンプルシートを作成し、1 つはレーン 1 のサンプル、もう 1 つはレーン 2 のサンプルとすることができます。これにより、サーバー間でファイルをコピーする代わりにサンプルシートを変更するだけで済みます。この方法では、`--demultiplexOnly` オプションを指定することなく `Run Folder` コマンドから開始します。デマルチプレックスはサーバーごとに 1 回実行されるため、ランフォルダー全体を各解析サーバーにコピーする必要があります。
3. 元のDRAGENサーバーから追加のサーバーに、FASTQフォルダーの出力を転送します。Logs\_Intermediates/FastqGeneration。
4. 元のDRAGENサーバーと追加のDRAGENサーバーの両方に対して、`--fastqFolder`オプションを指定して解析ソフトウェアを実行します。
  - オプション1：元のSampleSheet.csvを各サーバーにコピーします。次に、各DRAGENサーバー上のBashスクリプトに対してサブセットに分割したリストを渡して、実行するサンプル/ペアを指定します。
  - オプション2：各DRAGENサーバーにSampleSheet.csvをコピーして、実行するサンプル/ペアのリストのみが含まれるように変更します。

解析の開始時に`--samplePairsIDs`コマンドラインオプションが指定されていない限り、ソフトウェアでは、サンプルシート内のすべてのサンプルがFASTQフォルダー内に含まれることを検証します。これらのチェックに失敗するとエラーが発生します。

表 1 マルチノード解析用のコマンド

ステップ	コマンド
デマルチプレックス	<pre>DRAGEN_TSO500.sh --resourcesFolder /staging/illumina/DRAGEN_TSO500/resources -- hashtableFolder /staging/illumina/DRAGEN_ TSO500/ref_hashtable --runFolder /staging/ {RunFolderName} --analysisFolder /staging/ {DemultiplexAnalysisFolderName} --demultiplexOnly --sampleSheet /staging/illumina/{SampleSheetName}</pre>
解析 (単一サーバー)	<pre>DRAGEN_TSO500.sh --resourcesFolder /staging/illumina/DRAGEN_TSO500/resources -- hashtableFolder /staging/illumina/DRAGEN_ TSO500/ref_hashtable --fastqFolder /staging/ {DemultiplexAnalysisFolderName}/Logs_ Intermediates/FastqGeneration/ --analysisFolder /staging/{Node1AnalysisFolderName} --sampleSheet /staging/illumina/{SampleSheetName} -- samplePairIDs Pair_1,Pair_2</pre>

ステップ	コマンド
解析（追加サーバー）	<pre>DRAGEN_TSO500.sh --resourcesFolder /staging/illumina/DRAGEN_TSO500/resources -- hashtableFolder /staging/illumina/DRAGEN_ TSO500/ref_hashtable --fastqFolder /staging/ {DemultiplexAnalysisFolderName}/Logs_ Intermediates/FastqGeneration/ --analysisFolder /staging/{Node1AnalysisFolderName} --sampleSheet /staging/illumina/{SampleSheetName} -- samplePairIDs Pair_3</pre>
ギャザー	<pre>DRAGEN_TSO500.sh --analysisFolder /Gathered_ Results --resourcesFolder \${RESOURCES} --runFolder \${RUN_FOLDER} --sampleSheet \${SAMPLE_SHEET} -- gather /Demultiplex_Output /Node1_Output /Node2_ Output</pre>

## 解析方法

DRAGEN TruSight Oncology 500 Analysis Softwareのワークフローでは、以下の解析ステップを実施して、指定したフォルダーに解析出力ファイルを書き込みます。

- FASTQの生成
- 以下の方法を用いたDNA解析：
  - DNAアライメントと再アライメント
  - リードCollapsing
  - Indel再アライメントとリードスティッチング
  - スモールバリエントコール
  - スモールバリエントフィルタリング
  - コピー数バリエントコール
  - フェーズドバリエントコール
  - バリエントのマージ
  - アノテーション
  - 腫瘍変異負荷
  - マイクロサテライト不安定性ステータス

- 以下の方法を用いるRNA解析：
  - ダウンサンプリング
  - リードトリミング
  - アライメント
  - 重複マーキング
  - 融合遺伝子コール
  - RNA融合遺伝子フィルタリング
  - スプライスバリエーションコール
  - アノテーション
  - 融合遺伝子のマージ
- 品質管理 (QC)
  - ランQC
  - DNAサンプルQC
  - RNAサンプルQC
  - コンタミネーションの検出

## DNA解析方法

### DNAアライメントと再アライメント

アライメントステップでは、SAM Toolsユーティリティを搭載したBurrows-Wheeler Aligner (BWA-MEM) を使用してFASTQファイル内のDNA配列とhg19ゲノムをアライメントします。このアライメントステップでは、BAMファイル (\*.bam) とBAMインデックスファイル (\*.bam.bai) が生成され、DnaAlignmentフォルダーに保存されます。BAMファイルは、アライメントした配列の表示に使用するSAMファイルの圧縮バイナリバージョンです。

このソフトウェアでは、リードCollapsingステップの完了後にFASTQファイルに2回目のアライメントを実行します。2回目のアライメントによって、分子バーコード (UMI) でCollapsingしたリードのみを使用したサンプルリードの再アライメントが可能になります。

BWA-MEMの詳細については、Burrows-Wheeler Alignerのウェブサイトを参照してください。SAM/BAM形式のファイルの詳細は、GitHubでSequence Alignment/Mapの仕様のページを参照してください。

## リードCollapsing

リードCollapsing解析のステップでは、ゲノム上の位置が類似するリードのセット（ファミリー）をUMIタグを用いて代表配列にCollapsingするアルゴリズムを実行します。この処理によって、非常に頻度の低い配列バリエーションのシグナルが失われることなく、重複するリードを正確に除去できます。さらに、UMI Collapsingでは相補鎖からの情報を組み合わせる重複除去を用いてFFPEの脱アミノ化アーチファクトを減少させます。リードCollapsingのステップでは、FASTQファイルが生成されて、CollapsedReads出力フォルダーのメトリクスファイルと関連付けられます。リードCollapsingでは以下のBAMタグを追加します。

- **RX/XU** : UMI
- **XV** : ファミリー内にあるリード数
- **XW** : 重複ファミリー内のリード数、または0（重複ファミリーでない場合）

## Indel再アライメントとリードスティッチング

スモールバリエーションコールの結果を改善するために、Geminiソフトウェアでは、ローカルIndel再アライメント、ペアリードスティッチング、およびリードのフィルターを実施します。スティッチしたリードは、ペアとなるリードから組み合わせたシングルリードです。検出されたIndel近辺のリードを再アライメントして、アライメントのアーチファクトを除去します。このソフトウェアでは、1個のBAMファイルとアライメントに使用したゲノムFASTAを入力すると、対応するBAMファイル1個が出力されます。この出力ファイルには、スティッチし、ペアとして再アライメントしたリードが含まれます。マップクオリティが低いリードペアや、入力BAMに由来する補足的および二次的アライメントを含むリードペアは無視されます。

正常にスティッチしたリードの場合、Geminiは以下のBAMタグを追加します。

- **XD** : 方向性をサポートするストリングで、順方向、逆方向、スティッチした位置を示す
- **XR** : ペアの方向（FRまたはRF）

## スモールバリエーションコール

Piscesソフトウェアでは、体細胞バリエーションコールを実施してDNAサンプル内の頻度の低いバリエーションを同定します。Piscesは、StitchedRealignedの解析ステップで生成されたBAMファイル内からスモールバリエーションコールを実施します。

Piscesでは、バリエーション候補ごとに、genome.vcfファイルの [Format] 列の下にUSフィールドを追加します。このファイルには以下のリードタイプカウントの変異体サポートに対する値が含まれます。

- Duplex stitched
- Duplex nonstitched
- Simplex forward stitched
- Simplex forward nonstitched
- Simplex reverse stitched
- Simplex reverse nonstitched

この後に同じリードタイプカウントの合計サポートが続きます。

このステップのsmallバリエントコールgenome.vcfは候補を収集するだけであり、その後に対応するリードサポート情報を出力します。最終的なバリエントコールは、以下の後処理ステップで判定します。

ソフトウェアコンポーネントPsaraは、パネルマニフェストに基づくgVCFのトリミングに使用されます。バリエントが含まれるのは、マニフェストと重複する場合または重複するIndel内に含まれる場合です。smallバリエントフィルタリングが、最終的なバリエントコールを決定します。

詳細は、GitHubでPiscesプロジェクトのPisces設計文書を参照してください。

## smallバリエントフィルタリング

ソフトウェアコンポーネントのPepeでは、バックグラウンドを滑らかにしてクオリティスコアを補正するために、smallバリエントコールのゲノムVCFで後処理を実施します。エラー率がクオリティの閾値に達しない場合、ソフトウェアによってバリエントが除去されます。この解析ステップでは、ゲノムVCFファイルと関連するエラー率ファイルを生成します。リファレンスコールの最小リード深度は100です。VAFの検出下限は、最小リード深度で5%です。

Pepeでは、以下の条件に応じて動的に補正した2つのクオリティスコアを候補ごとに算出します。

- バックグラウンドノイズ
- トリヌクレオチドの変化
- リードサポートタイプ

バリエント候補ごとに、同一位置でのバックグラウンドノイズを、クオリティの異なる正常なベースラインサンプルから推定します。観察された変異体の深度、合計深度、バックグラウンドノイズを用いて、二項分布によりp値を算出します。次に、このp値をバリエントクオリティスコア (AQ) に変換します。各トリヌクレオチド変化のサンプル特異的なエラー率を、アリル頻度が1%未満の全位置を用いることで、各サンプルのさまざまなサポートカテゴリから推定します。バリエント候補ごとの尤度比スコア (LQ) を、観察された総リードと変異体リードの対応するエラー率別に算出します。バイアスコア (BFQ) を各バリエント候補で算出して、異なるサポートグループ間で変異体サポートリードと総サポートリードを比較した場合の不均衡状態を評価します。

Catalogue of Somatic Mutations in Cancer (COSMIC) のカウントが50を超えるバリエントの場合、LQおよびAQの閾値を20にし、残りの位置では60にします。Indelの場合、少なくとも1個のスティッチした変異体サポートが必要です。COSMICに未登録のバリエントでは、BFQの閾値が20未満です。さらに、フラグメント内にある変異体およびWTアリルの位置情報がバリエント候補ごとに抽出されます。コルモゴロフ-スミルノフ検定を適用して、変異体位置とWT位置間でp値を算出します。p値が0.05未満で中央値差が0.5以上のバリエントをフィルターし、VarBiasのラベルを付けます。リードCollapsingおよびバリエントフィルタリングの正味の影響により、偽陽性が有意に減少します。例えば、標準的なセルフリーDNAサンプルの偽陽性が、約1,500/Mbから5/Mb未満に減少しました。

クオリティスコアの評価に加えて、製品マニフェストで対応する特定の領域は高いバックグラウンドノイズのために解析から除外されます。除外されるバリエントはすべてVCF内でフラグ付けされています。除外する位置のブロックリストはお住まいの地域のイルミナ担当者にお尋ねください。

一部の領域はシーケンスが難しいことが知られています。このような領域の一例はTERTプロモーター領域です。TERTプロモーター領域に対してシーケンスを実行できますが、シーケンス領域のGCリッチなコンテンツのため、この場所はカバレッジが低くなる可能性があります。

## コピー数バリエーションコール

CRAFTコピー数バリエーションコーラーでは、アッセイ内でターゲットCNV遺伝子に対し増幅、リファレンス、欠失コールを実施します。CRAFTソフトウェアコンポーネントでは、パネル上の各ターゲット区間のカバレッジを計算、ノーマライゼーションを実施、各遺伝子の倍率変化値を算出し、各CNVターゲット遺伝子のCNVステータスを決定します。ノーマライゼーションのステップ中に、潜在変数であるシーケンス深度、ターゲットサイズ、PCRデュプリケート、プローブ効率、GCの偏り、DNAタイプなどを用いてカバレッジの偏りが補正されます。正常なFFPEとゲノムDNAサンプルのコレクションを使用して、これらの変数の一部を補正します。ターゲットCNV遺伝子ごとに、in silicoデータのトレーニングを行って遺伝子固有の増幅と欠失の閾値を判定します。入力はBAM形式内のCollapsingしたリードで、出力はVCFファイルです。増幅は、VCFファイル内にDUPとしてアノテーションされます。欠失ステータス (DEL) は情報のためのみに提供され、VCFファイルで常にLowValidationとしてラベルされます。

## フェーズドバリエーションコール

Scyllaは、所定のサンプル中の複数塩基変異 (MNV) をすばやく検出します。このソフトウェアでは、Scyllaを使用して、具体的に、臨床的に重要なEGFRエクソン19の変異体を検出します。これは、それ以外の方法ではバリエーションコーラーの範囲外になる変異体です。Psaraでは、EGFRのエクソン19内にある小領域にスモールバリエーションgVCFのフィルターをかけます。このgVCFサブセットに由来するSNP、MNV、Indelの候補が、GeminiのBAM出力と共にScyllaに送られます。Scyllaでは、元のBAMを使用して、これらのスモールバリエーション中のどれを、より長いMNVにまとめてフェージングする必要があるのかを判定します。

Scyllaでは、高いレベルで、入力gVCF中のフェージング候補となるバリエーションを識別して、そのバリエーションを近傍領域に配置します。その後、ScyllaではサンプルBAMファイルを検索して、これらのスモールバリエーションが互いに同一クローンのサブ集団で生じているというエビデンスを取り出します。これは、同じバリエーションを含むクラスターの最小セットに近傍領域の重複するリードをクラスタリングすることで達成できます。

## バリエーションのマージ

ソフトウェアで、フェージングしたバリエーションと、スモールバリエーションフィルタリングのステップで生成した他のスモールバリエーションを統合し、gVCFファイルを生成します。このプロセスでは、染色体、位置、リファレンスアレルおよび代替アレルと一致する完全な重複が除去されます。フェーズドバリエーションコールから検出された場合、以下の上皮成長因子受容体 (EGFR) バリエーションが追加されます。他のすべてのEGFRバリエーションは、バリエーションのマージ中に除去されます。

表 2 EGFR バリエーション

染色体	位置	リファレンスアレル	代替アレル
chr7	55242463	AAGGAATTAAGAGAAG	A
chr7	55242464	AGGAATTAAGAGA	A
chr7	55242464	AGGAATTAAGAGAAGC	A
chr7	55242465	GGAATTAAGAGAAGCA	G
chr7	55242467	AATTAAGAGAAGCAAC	A
chr7	55242469	TTAAGAGAAGCAACATCTC	T
chr7	55242468	ATTAAGAGAAGCAACATCT	A
chr7	55242466	GAATTAAGAGAAGCAA	G
chr7	55242465	GGAATTAAGA	G
chr7	55242469	TTAAGAGAAGCAA	T
chr7	55242462	CAAGGAATTAAGAGAA	C
chr7	55242466	GAATTAAGAGAAGCAACAT	G
chr7	55242482	CATCTCCGAAAGCCAACAAGGAAAT	C
chr7	55242465	GGAATTAAGAGAAGCAACA	G
chr7	55242467	AATTAAGAGAAGCAACATC	A
chr7	55242469	TTAAGAGAAG	C
chr7	55242467	AATTAAGAGAAGCAACATC	T
chr7	55242469	TTAAGAGAAGCAA	C
chr7	55242467	TTAAGAGAAGCAA	TTGCT
chr7	55242468	ATTAAGAGAAG	GC
chr7	55242469	TTAAGAGAAGCAACATCTCC	CA
chr7	55242465	GGAATTAAGAGAAG	AATTC
chr7	55242465	GGAATTAAGAGAAGCAAC	AAT
chr7	55242467	AATTAAGAGAAGCAAC	T
chr7	55242467	AATTAAGAGAAGCAACATCTC	TCT
chr7	55242469	TTAAGAGAAGCAACATCT	CAA
chr7	55242465	GGAATTAAGAGAAGCAA	AATTC
chr7	55242465	GGAATTAAGAGAAGCAACATC	AAT
chr7	55242468	ATTAAGAGAAGCAAC	GCA

## アノテーション

Illumina Annotation Engine Nirvanaでスモールバリエーションのアノテーションを実施します。入力はgVCFファイルであり、出力はアノテーションしたJSONファイルです。

Nirvanaによって処理される各バリエーションの入力は、データベースから入手可能なdbSNP、gnomAD genome and exome、1000 genomes、ClinVar、COSMIC、RefSeq、Ensemblなどの情報でアノテーションされます。バージョン情報と一般的な詳細情報はヘッダーから取得できます。アノテーションされた各バリエーションは、ヘッダーに続く個別の行に入れ子式の辞書構造として格納されます。各アノテーションデータベースのバージョン情報を次の表に示します。

データベース	バージョン
gnomAD	2.1
COSMIC	v84
ClinVar	2019-02-04
dbSNP	v151
1000 Genomes Project	Phase 3 v5a
RefSeq	NCBI Homo sapiens Annotation Release 105.20201022
Ensembl	VEP build 91

## 腫瘍変異負荷

腫瘍変異負荷（TMB）解析ステップでは、アノテーションしたスモールバリエーションのJSONファイルとスモールバリエーションのフィルター解析ステップで生成されたgVCFファイルから、TMBメトリクスを生成します。アノテーションしたJSONファイルを使用して、公開データベースのアリルのカウントや転写レベルでの結果など、個々のバリエーションに関する情報を取得できます。gVCFファイルを使用して有効なパネルサイズの名分母を評価します。

TMBの計算から生殖細胞系列バリエーションを除外するために、ソフトウェアでは公開データベースのフィルタリングと、アリル頻度情報および近接した位置にあるバリエーションを使用する事後データベースフィルタリング法を併用しています。

最初に、コンポーネントはGnomAD exome、genome、1000 genomesデータベースのいずれかで観察されたアリルカウントが10以上のバリエーションを除外します。データベースで観察されない生殖細胞系列バリエーションをフィルタリングするために、ソフトウェアは同じ染色体上にある一定範囲のアリル頻度のバリエーションを同定します。特定のバリエーションがデータベースの出現頻度に基づいて除外されない場合、同じ染色体上にある類似のアリル頻度のバリエーションをグループ化します。5以上の類似バリエーションがフィルタリングされる場合、対象バリエーションはTMBの計算から除外されます。さらに、アリル頻度が90%以上のバリエーションもTMBの計算から除外されます。TMBは以下の式を用いて算出します。

$$\text{TMB} = \text{適格なバリエーション} / \text{有効なパネルサイズ}$$

**適格なバリエント**

- フィルタリングによって除かれなかったバリエント
- コーディング領域のバリエント (RefSeq Cds)
- バリエント頻度が5%以上
- カバレッジが50X以上
- SNVおよびIndel (MNVを除く)
- 非同義バリエントと同義バリエント
- COSMICのカウントが50以上のバリエントを除外

**有効なパネルサイズ**

- カバレッジが50Xを超える総コーディング領域
- バリエントコールの対象とならない低信頼度の領域を除外

出力は、TMBの計算に使用したバリエントの情報を含む\*\_TMB\_Trace.tsvファイルと、TMBスコアの計算および構成の詳細を記載した\*.tmb.jsonファイルに記録されます。

**マイクロサテライト不安定性ステータス**

マイクロサテライト不安定性ステータス (MSI) のステップでは、リードスティッチングの解析ステップで生成されたBAMファイルからマイクロサテライト不安定性を判定し、MSIメトリクスファイルを生成します。ソフトウェアは、情報エントロピーメトリクスに基づくベースライン正常サンプルのセットを基準にして、マイクロサテライト部位に不安定性のエビデンスがあるか評価します。評価した全MSI Siteに対する不安定性MSI Siteの割合を、サンプルレベルのマイクロサテライトスコアとしてレポートします。

**コンタミネーションの検出**

コンタミネーション解析ステップでは、スモールバリエントフィルタリングステップで生成されたVCFファイル内にある外来のDNAによるコンタミネーションを検出します。ソフトウェアで、コンタミネーションp値 (pスコア) とコンタミネーションスコアを組み合わせることで、サンプルに外来のDNAが存在するか判定します。

コンタミネーションスコアは、全位置にわたるすべての対数尤度スコアを合計した値です。p値はSNPが染色体全体に不均一に分布することの有意性を表します。これは大規模に再編成されたゲノムを示し、コンタミネーションによる偽陽性を引き起こす可能性があります。

コンタミネーションのあるサンプルでは、SNPのバリエントアリル頻度が0%、50%または100%から外れています。アルゴリズムでは、バリエントアリル頻度が25%未満または75%超となるコモンSNPと重複するすべての位置を収集します。その後、以下の適格性評価により、その位置がエラーである尤度、または実際に変異である尤度を算出します。

- サンプルあたりのエラー率の推定
- 変異サポート
- 選択された各位置の合計深度

## RNA解析方法

### ダウンサンプリング

各サンプルを3,000万RNAリードにダウンサンプリングします。この数は、シングルリードの総数を表します（R1 + R2、全レーンから）。推奨シーケンス構成または推奨プレックスを使用すると、サンプルをダウンサンプリングの限界よりも少ないリード数とすることができます。このような場合、FASTQファイルはそのままの状態です。

### リードトリミング

後続の処理のためにリードを76塩基対にトリミングします。

### DRAGENサーバーのRNAアライメントおよび遺伝子融合検出

DRAGENサーバーは、定位置不明のコンティグ（chrUn\_gl領域）を含むヒトhg19ゲノムを用いて transcript-awareモードでRNAリードのアライメントを行い、GENCODEv19転写アノテーションを使用してスプライスサイトを同定します。DRAGENサーバーは、アライメントの開始および終了座標を用いて、重複するリードのアライメントを同定してラベルします（soft-clipリードで補正）。

融合遺伝子とスプライスバリエーションコールは、バリエーションのスコアリングに重複を除去したフラグメントのみを使用します。DRAGENサーバーは、複数の遺伝子に対してキメラスプリットリードのアライメント（一次アライメントと補足的アライメントのペア）を用いて融合遺伝子候補を同定します。DRAGENサーバーは、サポートリード数、サポートリードのマッピングクオリティ、親遺伝子間のシーケンスの相同性など、各候補のさまざまな特徴に基づいてスコア付けとフィルタリングを行います。

DRAGENサーバーへの入力FASTQ形式のトリミングされたリードです。出力は、重複マーキングのあるリードアライメントを含むBAMファイル、アノテーションのないスプライスジャンクションを含むSJ.out.tabファイル、融合遺伝子候補を含むCSVファイルなどです。

### スプライスバリエーションコール

スプライスバリエーションコールは、内部で開発されたソフトウェアを使用して実行されます。入力はDRAGENサーバーのBAMファイルとSJ.out.tabファイルです。SJ.out.tabのジャンクションは、最初にGENCODEv19のスプライスアノテーションを用いてフィルタリングされ、次にさまざまな組織タイプの非腫瘍FFPEサンプルコホートのベースラインを用いてさらにフィルタリングされます。許可リスト（一部のファイルではホワイトリストという）にあるスプライスジャンクションはフィルタリングされません。許可リストには、ARv7、METエクソン14スキッピング、EGFRvIIIが含まれます。出力は、最終出力であるVCFファイルと、マージに使用される遺伝子間バリエーションを含むTSVファイルです。次の表に示すように、スプライスバリエーションは0～10にスコア付けされます。

表 3 スプライスバリエーションコーラーでスコア付けされる特徴

スコアコンポーネント	スプライスの特徴	スコア付けの範囲	係数
スプリットリード	split_unique_reads_alt	0～10	1

### アノテーション

Illumina Annotation Engineでスプライスバリエーションのアノテーションを実施します。入力と出力はVCFファイルです。

## RNAのマージ

融合遺伝子イベントの大部分は、補足的アライメントとsoft-clipアライメントを用いて融合遺伝子ブレイクポイントを検出するDRAGEN融合遺伝子コーラーによってコールされます。場合によっては、融合遺伝子ブレイクポイントが十分に近接しているため、融合遺伝子をサポートするリードがアライメントの間隙（CIGAR列内のN）を作ります。このような融合遺伝子イベントは、スプライスバリエーションコール（TSO 500ワークフローでは独立したステップ）中に検出されます。RNA融合遺伝子のマージステップでは、DRAGEN融合遺伝子コールおよびスプライスバリエーションコール中に、検出された融合遺伝子を組み合わせ、検出されたすべての融合遺伝子を含む1つの出力にします。

## 品質管理（QC）

DRAGEN TruSight Oncology 500 Analysis Software v1.1には、数種類のQCが含まれます。

### ランQC

ランメトリクスレポートでは、シーケンスランフォルダーのInterOpファイルを使用して、ラン品質の結果が許容基準を満たす範囲内であるか判断するための推奨値を示します。クオリティ閾値はシステム間で異なり、シーケンスランフォルダーに基づいて自動的に検出されます。システムごとのランメトリクスとクオリティ閾値の情報を以降の表に示します。

表 4 ハイスループットシステム（NovaSeq 6000 システム）

メトリクス	内容説明	推奨されるガイドラインのクオリティ閾値	バリエーションクラス
PCT_PF_READS (%)	パスフィルターリードの全体に占める割合	≥55.0	すべて
PCT_Q30_R1 (%)	クオリティスコアが30以上のRead 1リードの割合	≥80.0	すべて
PCT_Q30_R2 (%)	クオリティスコアが30以上のRead 2リードの割合	≥80.0	すべて

表 5 ロースループットシステム（NextSeq 500/550 システム）

メトリクス	内容説明	推奨されるガイドラインのクオリティ閾値	バリエーションクラス
PCT_PF_READS (%)	パスフィルターリードの全体に占める割合	≥80.0	すべて
PCT_Q30_R1 (%)	クオリティスコアが30以上のRead 1リードの割合	≥80.0	すべて
PCT_Q30_R2 (%)	クオリティスコアが30以上のRead 2リードの割合	≥80.0	すべて

## サンプルのDNAライブラリーQCメトリクス

DNAライブラリーQCメトリクスの入力は、DNAアライメント、リードCollapsingのBAM、Indel再アライメント、リードスティッチングのBAM、およびCRAFTでノーマライズしたBinCount.tsvファイルです。メトリクスとガイドラインの閾値はMetricsOutput.tsvファイルに記載されています。

メトリクス	内容説明	推奨されるガイドラインのクオリティ閾値	バリエーションクラス
CONTAMINATION_SCORE および CONTAMINATION_P_VALUE	SNPのVAF分布に基づくコンタミネーションスコア。コンタミネーションp値は大規模に再編成されたゲノムを評価するために使用し、コンタミネーションスコアがUSLを上回る場合にのみ必要である。0.05未満のp値は、そのサンプルに大規模な再編成があり、実際にサンプルのコンタミネーションがなくても高いコンタミネーションスコアを生じる可能性があることを示唆する	コンタミネーションスコア $\leq$ 3106 または コンタミネーションスコア $>$ 3106 および コンタミネーションp値 $\leq$ 0.049	すべて
MEDIAN_EXON_COVERAGE	全エクソン塩基にわたるエクソンフラグメントカバレッジの中央値	$\geq$ 150	スモールバリエーション、TMB
PCT_EXON_50X	50Xフラグメントカバレッジを持つエクソン塩基の割合	$\geq$ 90.0	スモールバリエーション、TMB
MEDIAN_INSERT_SIZE	サンプルのフラグメント長の中央値	$\geq$ 70	スモールバリエーション、TMB
USABLE_MSI_SITES	MSIコールに使用可能なMSI Siteの数	$\geq$ 40	MSI
COVERAGE_MAD	絶対偏差の中央値。CNVコールに使用する全領域にわたるノーマライズした偏差の中央値を表す	$\leq$ 0.210	CNV
MEDIAN_BIN_COUNT_CNV_TARGET	CNVターゲットあたりの生のbinカウント中央値	$\geq$ 1.0	CNV

## RNAライブラリーQC

RNAライブラリーQCの入力はRNAアライメントです。メトリクスとガイドラインの閾値はMetricsOutput.tsvファイルに記載されています。

メトリクス	内容説明	推奨されるガイドラインのクオリティ閾値	バリエーションクラス
MEDIAN_ CV_ GENE_ 500X	カバレッジの中央値が500xを超えるすべての遺伝子のCV中央値。カバレッジの中央値が500xを超える遺伝子が高発現する可能性がある。500xを超える高いCV中央値は、ライブラリー調製の問題を示唆している（サンプルインプット量の不足またはプローブでの濃縮時のプルダウンの問題）	$\leq 93$	融合遺伝子のスプライス
MEDIAN_ INSERT_ SIZE	サンプルのフラグメント長の中央値	$\geq 80$	融合遺伝子のスプライス
TOTAL_ ON_ TARGET_ READS	ターゲット領域にマッピングするリード総数	$\geq 9000000$	融合遺伝子のスプライス

## 解析の出力

解析が完了すると、DRAGEN TruSight Oncology 500 Analysis Softwareでは、指定の場所に解析の出力フォルダーが生成されます。

解析の出力を表示するには、解析の出力フォルダーに移動して、表示するファイルを選択します。

## メトリクスの出力

MetricsOutput.tsvファイルには、以下に示す全サンプルの品質管理メトリクスが含まれます。

- 以下のメトリクスのQCメトリクス：
  - スモールバリエーションコール (SVC)
  - TMB
  - MSI
  - CNV
  - 融合遺伝子
  - スプライスバリエーション
- ランQCメトリクス、解析ステータス、およびコンタミネーション

このTSVファイルには総リード、Collapsingしたリード、キメラリード、オンターゲットリードに基づく、サンプルごとの拡張DNAライブラリーQCメトリクスも含まれます。RNAサンプルを用いる解析では、RNAライブラリーQCメトリクスと、総リードとカバレッジに基づくサンプルごとの拡張RNAライブラリーQCメトリクスも生成されます。

MetricsOutput.tsvファイルは、\*.tsvファイルに、サンプルステータス、主要な解析メトリクス、およびメタデータを組み入れた最終の統合メトリクスレポートです。レポート内のサンプルメトリクスは、ガイドライン推奨の下限値（LSL）と上限値（USL）を、ランにおけるサンプルごとに示しています。

トラブルシューティングの情報については、[37ページの「トラブルシューティング」](#)を参照してください。

## ランメトリクス

解析モジュールのランメトリクスは、シーケンスランのクオリティを示します。

ランのデータクオリティを評価するために、以下のメトリクスを確認します。

メトリクス	内容説明	推奨閾値
PCT_PF_READS	フィルターをパスした、シーケンスフローセル上のリードの割合	≥ 55.0
PCT_Q30_R1	Read 1に由来するクオリティスコアが30以上の塩基の割合	≥ 80.0
PCT_Q30_R2	Read 2に由来するクオリティスコアが30以上の塩基の割合	≥ 80.0

以下の状況で、[Run Metrics] セクションの値がNAと表示されます。

- 解析をFASTQファイルから開始した場合
- 解析をBCLファイルから開始し、かつInterOpファイルが紛失または破損している場合

## シングルノード解析の出力フォルダーの構成

このセクションでは、シングルノードの解析ランから生成された出力フォルダーのコンテンツを示します。

シングル出力フォルダーの構成は以下のとおりです。

- 📁 Logs\_Intermediates
  - 📁 AlignmentCollapser
    - 📁 UMI CollapsingのBAMファイルと生のDRAGENメトリクスを含むDNAサンプルIDごとのサブフォルダー
  - 📁 Annotation : アノテーション出力ログが含まれます。
  - 📁 Cleanup : 出力クリーンアップログが含まれます。
  - 📁 CnvCaller
  - 📁 CombinedVariantOutput : CombinedVariantの出力ログ
  - 📁 Contamination
    - 📁 コンタミネーションメトリクスのJSONを含むサンプルIDごとのサブフォルダー
    - 📁 コンタミネーション出力ログ
  - 📁 DnaFastqValidation
- 📁 DnaQCMetrics

- 📁 アライメント、Collapsing、ステッチメトリクスのJSONファイルを含むサンプルIDごとのサブフォルダー
- 📁 DnaQCMetricsの出力ログ
- 📁 FastqDownsample
  - 📁 FASTQファイルを含むサンプルIDごとのサブフォルダー
  - 📁 FastqGenerationの出力ログ
- 📁 FastqGeneration
  - 📁 FASTQファイルを含むサンプルIDごとのサブフォルダー
  - 📁 FastqGenerationの出力ログ
- 📁 MergedAnnotation
- 📁 MetricsOutput
- 📁 Msi
  - 📁 MSIメトリクスのJSONを含むサンプルIDごとのサブフォルダー
  - 📁 Msiの出力ログ
- 📁 PhasedVariants : この名前のフォルダーが出力フォルダー構成内に複数存在する可能性があります。
  - 📁 フェージングしたバリエーションのメトリクスのJSONを含むサンプルIDごとのサブフォルダー
  - 📁 Psaraおよび Scyllaの出力ログ
- 📁 ResourceVerification
- 📁 RnaAlignmentFusionCaller
  - 📁 重複マーキングのあるアライメントされたBAMファイルおよびインデックスファイルを含むサンプルIDごとのサブフォルダー
  - 📁 RnaAlignmentFusionCallerの出力ログ
- 📁 RnaAnnotation
  - 📁 アノテーションされたVCFファイルを含むサンプルIDごとのサブフォルダー
  - 📄 [ dsdm json]
- 📁 RnaFastqValidation
- 📁 RnaFusionMerge
- 📁 RnaQCMetrics
  - 📁 アライメント、Collapsing、ステッチメトリクスのJSONファイルを含むサンプルIDごとのサブフォルダー
  - 📁 RnaQCMetricsの出力ログ
- 📁 RnaSpliceVariantCalling
  - 📁 スプライスバリエーションVCFを含むサンプルIDごとのサブフォルダー
  - 📄 [ dsdm JSON]

- 📁 RunQc
  - 📄 RunQC Metrics JSONファイル
  - 📁 RunQCの出力ログ
- 📁 SampleAnalysisResults
- 📁 SamplesheetValidation : サンプルシートのバリデーション出力ログを含む
- 📁 SmallVariantFilter
  - 📁 エラー率表を含むサンプルIDごとのサブフォルダー
  - 📁 SmallVariantFilterの出力ログ
- 📁 StitchedRealigned
  - 📁 スティッチと再アライメントしたBAMファイルおよびインデックスファイルを含むサンプルIDごとのサブフォルダー。StitchedRealigned BAMは、最終的に出力されるDNAのBAM
  - 📁 StitchedRealignedの出力ログ
- 📁 Tmb
  - 📁 TMBメトリクスのJSONを含むサンプルIDごとのサブフォルダー
  - 📁 TMBの出力ログ
- 📁 TrimFastq
  - 📁 FASTQファイルを含むサンプルIDごとのサブフォルダー
  - 📄 dsdm JSON
- 📁 VariantCaller
  - 📁 フィルタリングされていないゲノムVCFファイルを含むサンプルIDごとのサブフォルダー
  - 📁 VariantCallerの出力ログ
- 📁 VariantMatching
- 📁 Logs\_Intermediates
  - メトリクスの出力TSV
  - 📁 Sample ID : サンプルごとに以下の出力が生成される
    - 📄 組み合わせバリエーション出力TSV
    - 📄 TMBトレースTSV
    - 📄 スモールバリエーションゲノムVCF
    - 📄 スモールバリエーションゲノムのアノテーションJSON
    - 📄 コピー数バリエーションVCF
    - 📄 すべての融合遺伝子CSV
    - 📄 スプライスバリエーションVCF
    - 📄 スプライスバリエーションのアノテーションJSON

## 複数ノード解析の出力フォルダーの構成

このセクションでは、解析から生成された出力フォルダーのコンテンツを示します。複数ノードを用いる解析用の解析出力フォルダーの構成は以下のとおりです。

- 📁 Demultiplex\_Output
  - 📁 Logs\_Intermediates：サンプルごとのFASTQファイルを含む
- 📁 Node1\_Output：使用するノードごとに以下の出力が生成される
  - 📁 Logs\_Intermediates
  - 📁 Results：そのノードで実行されるサンプルの結果のみを含む
- 📁 Gathered\_Results
  - 📁 Logs\_Intermediates
- 📁 Results：使用するすべてのノードのすべてのサンプルの結果を含む

## 組み合わせバリエーション出力

ファイル名：{SampleID}\_CombinedVariantOutput.tsv

組み合わせバリエーション出力ファイルでは、1サンプルを基に1つのファイルにまとめてバリエーションとバイオマーカーが記載されています。ペアIDを使用する場合、ファイルは対となる（同じ個体の）DNAサンプルとRNAサンプルに基づきます。出力には、以下のバリエーションとバイオマーカーが含まれます。

- スモールバリエーション（EGFR複合バリエーションを含む）
- 遺伝子増幅
- TMB
- MSI
- 融合遺伝子
- スプライスバリエーション

組み合わせバリエーション出力ファイルには、解析の詳細やシーケンスランの詳細のセクションがあります。それぞれの詳細を次の表に示します。

解析の詳細	シーケンスランの詳細
<ul style="list-style-type: none"> <li>• ペアID</li> <li>• DNAサンプルID (DNAの実行の場合)</li> <li>• RNAサンプルID (RNAの実行の場合)</li> <li>• 出力日</li> <li>• 出力時間</li> <li>• モジュールのバージョン</li> <li>• パイプラインのバージョン (Dockerイメージのバージョン番号)</li> </ul>	<ul style="list-style-type: none"> <li>• ラン名</li> <li>• ラン日</li> <li>• DNAサンプルインデックスID (DNAの実行の場合)</li> <li>• RNAサンプルインデックスID (RNAの実行の場合)</li> <li>• 装置ID</li> <li>• 装置制御ソフトウェアのバージョン</li> <li>• 装置タイプ</li> <li>• RTAのバージョン</li> <li>• 試薬カートリッジのロット番号</li> </ul>

組み合わせバリエーション出力では、以下のいずれかの状況で、空白フィールドを伴うスモールバリエーションを作成します。

- TruSight Oncology 500でターゲット化されていない重複遺伝子において、バリエーションがカノニカルなRefSeq transcriptと一致する場合。
- バリエーションが、リソースフォルダー内にあるTST500\_Manifest.bedファイルのiSNP、indel、またはFlankingに指定された領域に位置する場合。

### バリエーションのフィルターの規則

- **スモールバリエーション**：マージしたゲノムのVCFで [FILTER] フィールドにPASSと記されたすべてのバリエーションは、組み合わせバリエーション出力に表示されます。
  - 遺伝子情報は、Gene Allow List-スモールバリエーション内にあるカノニカルなtranscriptに属するバリエーションにのみ存在します。
  - 転写産物情報は、Gene Allow List-スモールバリエーション内にあるカノニカルなtranscriptに属するバリエーションにのみ存在します。
- **コピー数バリエーション**：コピー数バリエーションは以下の条件を必ず満たす必要があります。
  - [FILTER] フィールドがPASS。
  - [ALT] フィールドが<DUP>。
- **融合遺伝子バリエーション**：融合遺伝子バリエーションは以下の条件を必ず満たす必要があります。
  - フィルターをパスしたバリエーションコール ([KeepFusion] フィールドが true)。
  - 融合遺伝子の許可リストにある遺伝子を少なくとも1つ含みます。
  - ダッシュ (-) で分かれた遺伝子は、融合遺伝子の方向が判定可能なことを示します。スラッシュ (/) で分かれた遺伝子は、融合遺伝子の方向を判定できないことを示します。
- **バイオマーカーTMB/MSI**：DNAサンプルを処理した場合に必ず存在します。
- **スプライスバリエーション**：遺伝子EGFR、MET、およびARに含まれる、フィルターをパスしたスプライスバリエーション。

## DNA出力

### マージしたスモールバリエントgVCF

ファイル名: {SAMPLE\_ID}\_MergedSmallVariants.genome.vcf

マージしたバリエントゲノムのバリエントコールファイルは、スモールバリエントゲノムVCF（バリエントフィルタリングの出力）とフェージングしたバリエントコールからの臨床的に重要なEGFRエクソン19のバリエントを統合します。これには評価対象のすべてのスモールバリエント候補の情報が含まれます。バリエントステータスは、以下のゲノムVCFの [FILTER] 列から判定します。

ALT	FILTER	注記
.	PASS	WT
., A, C, G, etc.	LowDP	コールなし (DP < 100X、VAF ≥ 5% で確信をもってバリエントを検出するには不十分な深度)
A, C, G, etc.	PASS	PASSバリエント
A, C, G, etc.	LowSupport	フィルターしたバリエント候補 : <ul style="list-style-type: none"> <li>• AQまたはLQ失敗</li> <li>• ホモポリマーコンテキストのIndelまたはバリエントに対するスティッチしたサポートが0</li> </ul>
A, C, G, etc.	Blocklist	座位の除外リスト（一部のファイルではブラックリストという）。詳細については、 <a href="#">15ページの「スモールバリエントフィルタリング」</a> を参照
A, C, G, etc.	LowVarSupport	変異体サポート < 1のフィルターしたバリエント候補

### マージしたスモールバリエントのアノテーションJSON

ファイル名: {SAMPLE\_ID}\_MergedSmallVariantsAnnotated.json.gz

マージしたスモールバリエントのアノテーションファイルは、フィルターをパスしたバリエントやパスしなかったバリエントなど、マージしたゲノムVCF内にあるリファレンス以外のすべて位置に関するバリエントのアノテーション情報を示します。

### TMBトレース

TMBトレースファイルは、特定のサンプルのTMB値を計算する方法に関する包括的な情報を示します。スモールバリエントのフィルターステップでフィルターをパスしたすべてのスモールバリエントが、このファイルに記載されます。TMB JSON内のTmbPerMb値の分子を計算するためには、TSVファイルフィルターを設定して、値がTrueであるIncludedInTMBNumeratorを使用します。

TMBトレースファイルは、バリエントを解釈する目的で使用するものではありません。フィルタリングステータスは、TMBを計算するための専用の設定です。フィルターを設定しても、バリエントの分類が体細胞または生殖細胞に変わることはありません。

列	内容説明
Chromosome	染色体
Position	バリエントの位置
RefCall	リファレンス塩基
AltCall	バリエント塩基
VAF	バリエントアリル頻度
Depth	バリエント検出位置のカバレッジ
CytoBand	バリエントのサイトバンド
GeneName	遺伝子名（該当する場合）。複数の遺伝子がある場合、セミコロンで区切られたリストを使用する
VariantType	バリエントのタイプ：SNV、挿入、欠失、MNV
CosmicIDs	Cosmic ID。複数のCosmic IDがある場合は「;」で連結する
MaxCosmicCount	最大COSMIC study数
AlleleCountsGnomadExome	gnomAD exomeデータベース中のバリエントアリルカウント
AlleleCountsGnomadGenome	gnomAD genomeデータベース中のバリエントアリルカウント
AlleleCounts1000Genomes	1000 genomesデータベース中のバリエントアリルカウント
MaxDatabaseAlleleCounts	3つのデータベースにわたる最大バリエントアリルカウント
GermlineFilterDatabase	バリエントがデータベースフィルターによってフィルターされた場合、TRUE
GermlineFilterPRoxi	バリエントがプロキシフィルターによってフィルターされた場合、TRUE
CodingVariant	バリエントがコーディング領域にある場合、TRUE
Nonsynonymous	バリエントに、非同義変異の結果を伴う転写アノテーションがある場合、TRUE
IncludedinTMBNumberator	バリエントがTMBの計算で使用される場合、TRUE

## コピー数バリエントVCF

コピー数VCFファイルには、DRAGEN TruSight Oncology 500 Analysis Software v1.1でターゲット化した増幅遺伝子のDNAライブラリーに対するCNVコールが含まれています。CNVコールでは、リファレンス、欠失、増幅として分類される遺伝子ごとに倍率変化の結果を示します。

VCFの [QUAL] 列にある値は、 $Q = -10 \times \log_{10}(p\text{値})$ の式で表されるp値のPhred変換値です。p値は、遺伝子の倍率変化とゲノムの残りの部分の倍率変化間でt検定して得られます。Qスコアが高いほど、CNVコールの信頼度が高いことを示します。

VCFの表記として、<DUP>は、検出された倍率変化 (FC) が既定の増幅カットオフ値よりも大きいことを示します。<DEL>は、検出されたFCがその遺伝子の既定の欠失カットオフ値よりも小さいことを示します。このカットオフ値は遺伝子によって変わることがあります。

<DEL>コールはin silicoデータセットでのみ検証されています。結果として、すべての<DEL>コールにはVCF内にLowValidationフィルターがあります。

各コピー数バリエーションは、二倍体ゲノム中のノーマライズしたリード深度を基準にした場合の、テストサンプル中のノーマライズしたリード深度の倍率変化としてレポートされます。腫瘍の純度を考慮すると、レポートされた倍数変化からサンプル中の遺伝子の倍数性が推測できます。

腫瘍純度をX%とすると、レポートされた倍率変化Yに関して、以下の式を用いることでコピー数nを算出できます。

$$n = [(200 * Y) - 2 * (100 - X)] / X$$

例えば、腫瘍純度30%で倍率変化が2.2倍のMETは、10コピーのMET DNAが観察されたことを示します。

## RNA出力

### スプライスバリエーションVCF

スプライスバリエーションVCFには、RNA解析パイプラインで識別された解析パネルがターゲットとするスプライスバリエーション候補がすべて含まれます。各バリエーション候補に対して以下のフィルターが適用できます。

ファイル名	内容説明
LowQ	Splice Variant Scoreが、Passing Quality Scoreの閾値1未満である
PASS	Splice Variant Scoreが、Passing Quality Scoreの閾値1以上である
LowUniqueAlignments	すべてのスプライスジャンクションのサポートリードは、2つのスプライスサイトの少なくとも1つの近くにある固有のゲノムの区間にマップする

各列の詳細については、出力のヘッダーを参照してください。

### スプライスバリエーションのアノテーションJSON

可能な場合、各スプライスバリエーションにはIllumina Annotation Engineを用いてアノテーションが実施されます。以下の情報はJSONから得られます。

- HGNC遺伝子
- 転写産物
- エクソン
- イントロン
- カノニカル
- バリエーションの結果

## すべての融合遺伝子CSV

すべての融合遺伝子CSVファイルには、RNA解析パイプラインで識別された融合遺伝子候補がすべて含まれます。ファイルの2つの主要な出力列 [Filter] と [KeepFusion] は、融合遺伝子候補を記載します。

次の表は、[Filter] 列に示すセミコロンで区切られた出力について説明しています。出力は信頼度フィルターまたは示される情報のみです。どの信頼度フィルターにもかからなかった場合は [Filter] 列に出力PASSが記録され、フィルターにかかった場合は出力FAILが記録されます。

表 6 [Filter] 列の出力

Filter	フィルタータイプ	内容説明
DOUBLE_BROKEN_EXON	信頼度フィルター	2つのブレイクポイントのいずれもアノテーションされたエクソン境界から離れている場合、サポートリード数は高い閾値要件（サポートリード数10以上）を満たさない
LOW_MAPQ	信頼度フィルター	ブレイクポイントのいずれかで、融合遺伝子のサポートリードのアライメントがすべてMAPQ 20未満である
LOW_UNIQUE_ALIGNMENTS	信頼度フィルター	すべての融合遺伝子のサポートリードのアライメントは、いずれかのブレイクポイントで固有のゲノムの区間にマップする
LOW_SCORE	信頼度フィルター	融合遺伝子候補は、その候補の特徴によって決定される確率スコアを持つ
MIN_SUPPORT	信頼度フィルター	融合遺伝子候補に、融合遺伝子のサポートリードが非常に少ない（サポートリードペア5未満）
READ_THROUGH	信頼度フィルター	ブレイクポイントがリファレンスゲノム上のシス位の近傍（200 kbp未満）である
ANCHOR_SUPPORT	情報のみ	融合遺伝子のサポートリードのリードアライメントが、2つのブレイクポイントのいずれかで長さが不十分（12 bp）
HOMOLOGOUS	情報のみ	候補は、関与する2つの遺伝子に高い相同性があるために生成された偽陽性の可能性がある
LOW_ALT_TO_REF	情報のみ	融合遺伝子のサポートリード数が、2つのブレイクポイントのいずれかでリファレンス転写産物をサポートするリード数の1%未満である
LOW_GENE_COVERAGE	情報のみ	濃縮された遺伝子の各ブレイクポイントが、非ゼロのリードカバーレッジで125 bpより少ない
NO_COMPLETE_SPLIT_READS	信頼度フィルター	融合遺伝子のサポートスプリットリードごとに、2つのブレイクポイントの端から端にわたりアライメントされた塩基の総数がリード長の60%未満である
UNENRICHED_GENE	信頼度フィルター	2つの親遺伝子のいずれも濃縮パネル内にはない

出力の [KeepFusion] 列は、どの信頼度フィルターにもかからなかった場合にTRUEとなります。

各列の詳細については、出力のヘッダーを参照してください。

表 7 融合遺伝子の列

融合遺伝子のオブジェクトフィールド	ソース
Gene A	融合遺伝子のA側に関連する遺伝子。複数の遺伝子がある場合、セミコロンで区切られたリストを使用する
Gene B	融合遺伝子のB側に関連する遺伝子。複数の遺伝子がある場合、セミコロンで区切られたリストを使用する
Gene A Breakpoint	(情報のみ) 染色体と融合遺伝子の遺伝子A側のオフセット
Gene A Location	遺伝子A内のブレイクポイントの場所： <ul style="list-style-type: none"> <li>• <b>IntactExon</b>：エクソン境界に一致</li> <li>• <b>BrokenExon</b>：エクソン内</li> <li>• <b>Intronic</b>：イントロン内</li> <li>• <b>Intergenic</b>：遺伝子重複なし（現在は除外）</li> </ul> 複数の遺伝子が遺伝子Aにある場合、セミコロンで区切られた場所のリスト。この列は遺伝子を同定するために内部的に使用し、複数の遺伝子が重複する領域でブレイクポイントが生じた場合にレポートする。場合によって、GeneAリストから除外された遺伝子に対する追加の値がリストされる
Gene A Sense	左/右のブレイクポイントの順序が、融合遺伝子の転写物が遺伝子Aと同じセンスであることを示唆するかどうかを示すブール変数。複数の遺伝子が遺伝子Aにある場合、セミコロンで区切られたブール変数のリスト
Gene A Strand	遺伝子Aのストランド、順方向は+、逆方向は-
Gene B Breakpoint	(情報のみ) 染色体と融合遺伝子の遺伝子B側のオフセット
Gene B Location	遺伝子B内のブレイクポイントの場所： <ul style="list-style-type: none"> <li>• <b>IntactExon</b>：エクソン境界に一致</li> <li>• <b>BrokenExon</b>：エクソン内</li> <li>• <b>Intronic</b>：イントロン内</li> <li>• <b>Intergenic</b>：遺伝子重複なし（現在は除外）</li> </ul> 複数の遺伝子が遺伝子Bにある場合、セミコロンで区切られた場所のリスト。この列は遺伝子を同定するために内部的に使用し、複数の遺伝子が重複する領域でブレイクポイントが生じた場合にレポートする。場合によっては、GeneBリストから除外された遺伝子に対する追加の値がリストされる
Gene B Sense	左/右のブレイクポイントの順序が、融合遺伝子の転写物が遺伝子Bと同じセンスであることを示唆するかどうかを示すブール変数。複数の遺伝子が遺伝子Bにある場合、セミコロンで区切られたブール変数のリスト

融合遺伝子のオブジェクトフィールド	ソース
Gene B Strand	遺伝子Bのストランド、順方向は+、逆方向は-
Score	DRAGENサーバーによって判定される融合遺伝子のクオリティ
Filter	各コーラーによって判定される融合遺伝子に関連するフィルター。異なるコーラーからの結果は同等ではない
Ref A Dedup	ジャンクションをまたがるペアのリードやジャンクションで分割されたリードの遺伝子A固有のマッピング。融合遺伝子をサポートしない。重複するリードは含まれない
Ref B Dedup	ジャンクションをまたがるペアのリードやジャンクションで分割されたリードの遺伝子B固有のマッピング。融合遺伝子をサポートしない。重複するリードは含まれない
Alt Split Dedup	ジャンクションで分割されたリードの固有のマッピング。融合遺伝子をサポートする。重複するリードは含まれない
Alt Pair Dedup	ジャンクションをまたがるペアのリードの固有のマッピング。融合遺伝子をサポートする。重複するリードは含まれない
KeepFusion	融合遺伝子を保持するか、融合遺伝子のリストから削除するかの判定
Fusion Directionality Known	融合遺伝子の方向性が既知で、遺伝子の順番どおりかどうか

Microsoft Excelを使用してこのレポートを表示する場合、日付に変換できる遺伝子（例えば、MARCH1）が、Excelによって自動的にdd-mm形式（1-Mar）に変換されます。融合遺伝子の許可リストにある遺伝子を以下に示します。

- ABL1
- AKT3
- ALK
- AR
- AXL
- BCL2
- BRAF
- BRCA1
- BRCA2
- CDK4
- CSF1R
- EGFR

- EML4
- ERBB2
- ERG
- ESR1
- ETS1
- ETV1
- ETV4
- ETV5
- EWSR1
- FGFR1
- FGFR2
- FGFR3
- FGFR4
- FLI1
- FLT1
- FLT3
- JAK2
- KDR
- KIF5B
- KIT
- KMT2A
- MET
- MLLT3
- MSH2
- MYC
- NOTCH1
- NOTCH2
- NOTCH3
- NRG1
- NTRK1
- NTRK2
- NTRK3

- PAX3
- PAX7
- PDGFRA
- PDGFRB
- PIK3CA
- PPARG
- RAF1
- RET
- ROS1
- RPS6KB1
- TMPRSS2

## ブロックリスト

ブロックリストは、偽陽性のバリエーションコールが生成される可能性のあるパネルの高ノイズ領域を表します。結果として、gVCF内のすべての位置がFilter=blacklistとしてラベルされ、バリエーションコールの結果が当該領域で信頼できないことを示します。

ブロックリストには以下の遺伝子が含まれます。

- HLA-A
- HLA-B
- HLA-C
- KMT2B
- KMT2C
- KMT2D
- chrY
- 60ベースラインサンプル中の6サンプル以上に生じるVAFが1%超の任意の位置

# トラブルシューティング

エラーのタイプ	措置
ソフトウェア	ログファイル <code>./{analysisFolder}/Logs_Intermediates/TruSight-Oncology-500-pipeline-&lt;timestamp&gt;.log file</code> を開く。ログファイルには、ソフトウェアによるすべてのコマンド実行と解析ステップごとの終了コードが表示される。ステップが失敗した場合、標準の出力ファイルとフォルダー <code>./{analysisFolder}/Logs_Intermediates/</code> の標準のエラーログファイルをレビューする
サンプル	最終のサンプルバイオマーカーレポートログファイル <code>./{analysisFolder}/Results/SampleID/CombinedVariantOutput.tsv</code> を開く。サンプルの解析ステップが失敗した場合、レポートの [SAMPLE STATUS] セクションの [Failed Steps] フィールドにステップ名が表示される。 <code>./{analysisFolder}/Logs_Intermediates/{FailedStep}/</code> に含まれているそのステップのログファイルをレビューする
サンプル	ログファイル <code>./{analysisFolder}/Logs_Intermediates/TruSight-Oncology-500-pipeline-&lt;timestamp&gt;.log file</code> を開く。 サンプルリミットの50 GBを超える場合、以下のメッセージがタイムスタンプおよびサンプルのファイルパスとともに表示される。 <code>TSO500_SOLID_LIMIT_BASES_PER_SAMPLE - TSO500 Solid exceeded sample limit</code>

## DNA拡張メトリクス

DNA拡張メトリクスは情報のみとして提供されます。トラブルシューティングの情報を得ることができますが、仕様の限界を明示せずに提供されるため、サンプルの品質管理には直接使用できません。追加のガイドラインについては、イルミナのテクニカルサポートにお問い合わせください。

メトリクス	内容説明	単位
TOTAL_PF_READS	パスフィルターリードの合計	カウント
MEAN_FAMILY_SIZE	サポートリードの補正、collapsing、フィルタリングの後にファミリ数で除算したファミリーごとのリードの合計	カウント
MEDIAN_TARGET_COVERAGE	塩基カバレッジの中央値	カウント
PCT_CHIMERIC_READS	キメラリードの割合 (%)	%
PCT_EXON_100X	100X超のカバレッジを持つエクソン塩基の割合 (%)	%

メトリクス	内容説明	単位
PCT_READ_ENRICHMENT	ターゲット領域の任意の部分を横断するリードの総リードに対する割合 (%)	%
PCT_USABLE_UMI_READS	使用可能なUMIを持つリードの割合 (%)	%
MEAN_TARGET_COVERAGE	塩基カバレッジの平均値	カウント
PCT_ALIGNED_READS	リファレンスゲノムにアライメントしたリードの割合 (%)	%
PCT_CONTAMINATION_EST	サンプルのコンタミネーションの割合 (%)	%
PCT_PF_UQ_READS	固有のパスフィルターリードの割合 (%)	%
PCT_TARGET_0.4X_MEAN	平均の0.4X超のターゲットカバレッジを持つターゲット塩基の割合 (%)	%
PCT_TARGET_100X	100X超のカバレッジを持つターゲット塩基の割合 (%)	%
PCT_TARGET_250X	250X超のカバレッジを持つターゲット塩基の割合 (%)	%

## RNA拡張メトリクス

RNA拡張メトリクスは情報のみとして提供されます。トラブルシューティングの情報を得ることができますが、仕様の限界を明示せずに提供されるため、サンプルの品質管理には直接使用できません。追加のガイドラインについては、イルミナのテクニカルサポートにお問い合わせください。

メトリクス	内容説明	単位
PCT_CHIMERIC_READS	ゲノム内の非連続領域にマッピングされる2つのセグメントとしてアライメントされたリードの割合 (%)	%
PCT_ON_TARGET_READS	ターゲット領域の任意の部分を横断するリードの総リードに対する割合 (%)。ターゲット領域に部分的にマッピングされるリードはオンターゲットとしてカウントされる	%
SCALED_MEDIAN_GENE_COVERAGE	長さによって拡大縮小される遺伝子の塩基カバレッジ中央値の中央値。パネル内の遺伝子のカバレッジ深度中央値の表示	カウント
TOTAL_PF_READS	パスフィルターリード総数	カウント

メトリクス	内容説明	単位
GENE_ MEDIAN_ COVERAGE	パネル内のすべての遺伝子のカバレッジ深度中央値	カウント
PER_GENE_ MEDIAN_ COVERAGE	各遺伝子の重複を除去したカバレッジの中央値。このメトリクスはLogs_Intermediates/RnaQCMetricsで確認できる。 RnaQCMetricsフォルダー内には、各サンプルのサブフォルダーがあり、{SampleName}_GeneCoverage.tsvファイルが入っている	カウント

# テクニカルサポート

技術的なサポートについては、イルミナのテクニカルサポートにお問い合わせください。

ウェブサイト：[jp.illumina.com](http://jp.illumina.com)  
 電子メール：[techsupport@illumina.com](mailto:techsupport@illumina.com)

## イルミナテクニカルサポート電話番号

地域	フリーダイヤル	国際
アイルランド	+353 1800 936608	+353 1 695 0506
イタリア	+39 800 985513	+39 236003759
インド	+91 8006500375	
インドネシア		0078036510048
英国	+44 800 012 6019	+44 20 7305 7197
オーストラリア	+61 1800 775 688	
オーストリア	+43 800 006249	+43 1 9286540
オランダ	+31 800 022 2493	+31 20 713 2960
カナダ	+1 800 809 4566	
韓国	+82 80 234 5300	
シンガポール	1 800 5792 745	
スイス	+41 800 200 442	+41 56 580 00 00
スウェーデン	+46 2 00883979	+46 8 50619671
スペイン	+34 800 300 143	+34 911 899 417
タイ	+66 1800 011 304	
台湾 (中国)	+886 8 06651752	
中国		+86 400 066 5835
デンマーク	+45 80 82 01 83	+45 89 87 11 56
ドイツ	+49 800 101 4940	+49 89 3803 5677
日本	+81 0800 111 5011	
ニュージーランド	+64 800 451 650	
ノルウェー	+47 800 16 836	+47 21 93 96 93

地域	フリーダイヤル	国際
フィリピン	+63 180016510798	
フィンランド	+358 800 918 363	+358 9 7479 0110
フランス	+33 8 05 10 21 93	+33 1 70 77 04 46
米国	+1 800 809 4566	+1 858 202 4566
ベトナム	+84 1206 5263	
ベルギー	+32 800 77 160	+32 3 400 29 73
香港 (中国)	+852 800 960 230	
マレーシア	+60 1800 80 6789	

安全データシート (SDS) : イルミナのウェブサイト [jp.support.illumina.com/sds.html](http://jp.support.illumina.com/sds.html) から入手できます。

製品関連文書 : [jp.support.illumina.com](http://jp.support.illumina.com) からダウンロードできます。



イルミナ株式会社  
東京都港区芝5-36-7  
三田ベルジュビル22階  
サポート専用フリーダイヤル  
0800-111-5011  
techsupport@illumina.com  
jp.illumina.com

本製品の使用目的は研究に限定されます。診断での使用はできません。

© 2022 Illumina, Inc. All rights reserved.

illumina®