



DRAGEN TruSight Oncology 500 Analysis Software v2.1 (Local)

User Guide

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。

ILLUMINA PROPRIETARY

文書番号：200019138 v01 JPN

2022 年 8 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き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負わせ、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 をご覧ください。

目次

概要.....	1
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.....	6
DRAGEN TruSight Oncology 500 Analysis Software の実行	7
サンプルシートの要件.....	7
コマンドラインオプション	10
BCL ファイルから開始.....	12
FASTQ ファイルから開始	13
複数の DRAGEN サーバー上での実行	13
解析方法	15
FASTQ の生成	15
DNA 解析方法	16
RNA 解析方法.....	20
品質管理 (QC).....	21
解析の出力.....	24
メトリクス出力.....	24
シングルノード解析の出力フォルダーの構成	25
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	28
DNA 出力	29
RNA 出力	32
ブロックリスト	37
解析パフォーマンステスト	37
in silico 特性化法	38
トラブルシューティング	38
DNA 拡張メトリクス	39
RNA 拡張メトリクス.....	40
テクニカルサポート	41
リソースと参考文献	42
改訂履歴	42

概要

Illumina® DRAGEN™ TruSight™ Oncology 500 Analysis Software v2.1 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）
- エクソンレベル 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。TruSight Oncology 500 Homologous Recombination Deficiency（HRD）では、ゲノム不安定性スコア（GIS）も判定します。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遺伝子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ルミナ担当者にお尋ねください。

DRAGEN TruSight Oncology 500 Analysis Software v2.1 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

イルミナの[サポートウェブサイト](#)の DRAGEN TruSight Oncology 500 Analysis Software v2.1 サポートページでは、イル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 v2.1 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 v2.1 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、イルミナの[サポートウェブサイト](#)の DRAGEN TruSight Oncology 500 Analysis Software v2.1 サポートページを参照してください。

追加リソース

イルミナの[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、および以下の文書が含まれています。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 または v4 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。
- TSO 500 HT 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 7.9 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 v20.10 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 を用意しています。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、デフォルト設定を使用するか、`--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て、DRAGEN サーバードメインで 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)	最小ディスクスペース (Gb)
NextSeq 500/550 および 550Dx HO フローセル	32 ~ 55	82 ~ 85	150
NovaSeq 6000 SP フローセル	85 ~ 100	250 ~ 374	300
NovaSeq 6000 S1 フローセル	164 ~ 200	360 ~ 665	800
NovaSeq 6000 S2 フローセル	290 ~ 460	890 ~ 1,600	1,500
NovaSeq 6000 S4 フローセル	800 ~ 1,200	2,700 ~ 4,100	3,000

解析を開始すると、必要な最小ディスクスペースが利用可能かどうかをソフトウェアがチェックします。最小ディスクスペースが利用できない場合、エラーメッセージが表示され、解析は開始されません。実行中にディスクスペースが不足した場合、エラーが表示され、解析が停止します。

許可される操作

イルミナでは、ソフトウェアのインストールと実行について以下の操作を許可することを推奨しています。

- root 以外のユーザーは、Docker グループのメンバーである場合のみ Docker を実行できます。Docker の許可要件と root として実行する代替手段の詳細については、Docker のウェブサイトで購入できる Docker の文書を参照してください。
- DRAGEN TruSight Oncology 500 Analysis Software のインストールとアンインストール、およびシステムチェックの実行には、root 権限が必要です。
- root ユーザーとしてログインせずに、DRAGEN TruSight Oncology 500 Analysis Software を実行します。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ず、また必須でもありません。

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

インストールスクリプトでは、DRAGEN TruSight Oncology 500 Analysis Software に加えて、必要な DRAGEN サーバーソフトウェアの依存関係もインストールされます (CentOS 7.9 向け DRAGEN 3.10.9)。インストールスクリプトを実行すると、既存の DRAGEN サーバーソフトウェアバージョンがアンインストールされます。DRAGEN TruSight Oncology 500 Analysis Software をアンインストールするには、[6 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。

DRAGEN v3 サーバーで製品を正常に実行するには、十分なストレージ容量がある DRAGEN サーバーの /staging ディレクトリにイメージなどの作業ファイルを保存するように、Docker 構成を更新する必要があります。Docker 構成を更新するには、Docker を再構成するための root 権限が必要です。

インストール手順

1. イルミナカスタマーケアに連絡して、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. イルミナから届く電子メールに従ってDRAGEN TSO 500インストールパッケージをダウンロードします。リンクの有効期限は7日間です。
3. 他の解析が実行されていないことを確認します。
他の解析の実行中にDRAGEN TruSight Oncology 500 Analysis Softwareをインストールすると、インストーラーの動作と実行中のDockerコンテナ間で干渉が発生する可能性があります。
4. Dockerの文書でCentOSのインストール手順を参照して、Docker v20.10以降をインストールします。
5. 既存のdockerデータを/var/lib/dockerから/staging/dockerに移動するには、以下のコマンドを入力します。

```
rsync -aqxP /var/lib/docker /staging/docker
```

6. Docker構成ファイルの出力ディレクトリを変更します。

構成ファイルのデフォルトの場所は /etc/docker/daemon.json です。-config-flag パラメーターを使用すると、Docker のインストール中にデフォルトの場所を変更できます。

構成ファイルをデフォルト以外の場所に保存するように Docker を構成していない場合は、以下の手順に従います。

- a. 以下のコマンドを用いて、構成ファイルが存在するかどうかを確認します。

```
/etc/docker/daemon.json
```

- b. 構成ファイルが存在しない場合は、以下のコマンドを用いて構成ファイルを作成します。

```
touch /etc/docker/daemon.json
```

- c. /etc/docker/daemon.json ファイルに以下のテキストペアが存在する場合は、以下のテキストに一致するように更新します。存在しない場合は、ファイルに追加します。

```
{"data-root": "/staging/docker/",  
  "experimental": true}
```

構成ファイルをデフォルト以外の場所に置くように Docker を明示的に構成している場合は、次の手順に従います。

- a. Docker構成ファイルを開きます。
- b. 構成ファイルに以下のテキストペアが存在する場合は、以下のテキストに一致するように更新します。存在しない場合は、以下のように追加します。

```
{"data-root": "/staging/docker/",  
  "experimental": true}
```

7. Dockerが再構成されたら、以下のコマンドを用いてDockerを再起動します。

```
systemctl enable docker.service
```

8. 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、スクリプトをこのディレクトリに保存します。
9. 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。

```
chmod +x /staging/install_DRAGEN_TSO500-2.1.0.run
```

10. 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。

```
TMPDIR=/staging /staging/install_DRAGEN_TSO500-2.1.0.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、サーバをシャットダウンして再起動する必要があります。

11. 以下のコマンドを用いてDRAGENサーバのハッシュテーブルをビルドします。所要時間は約60分です。

```
/usr/local/bin/build-hashtable_DRAGEN_TSO500-2.1.0.sh
```

12. 以下のようにDRAGENサーバのライセンスをインストールします。

- サーバがインターネットに接続されている場合は、以下のように DRAGEN TSO 500 のソフトウェアライセンスをアクティブにします。
 - a. サーバがインターネットに接続されていることをテストして確認します。
例: ping www.illumina.com
 - b. ライセンスをアクティブにするには、次のように入力します。
/opt/edico/bin/dragen_lic -i auto
- サーバがインターネットに**接続されていない**場合は、ライセンス情報についてイルミナカスタマーサービスにお問い合わせください。

13. DRAGENサーバのライセンスをインストールしたら、以下のコマンドを実行して、インストールされているDRAGENサーバのライセンスのリストを生成します。

```
/opt/edico/bin/dragen_lic
```

ライセンスのインストールが正常に実行されると、リストには TSOCombined が表示されます。HRD のライセンスを所有している場合は、リストに TSO500_HRD と Genome が表示されます。適切なライセンスがインストールされていない場合は、イルミナカスタマーケアにお問い合わせください。

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。

```
/usr/local/bin/check_DRAGEN_TSO500-2.1.0.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 Analysis Software 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。

MacOS を使用している場合、ローカル設定が英語でないとエラーが発生することがあります。このエラーを解消するには、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。

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_TSO500-2.1.0.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 (build-hashtable_DRAGEN_TSO500-2.1.0.sh、check_DRAGEN_TSO500-2.1.0.sh、DRAGEN_TSO500.sh、uninstall_DRAGEN_TSO500-2.1.0.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_TSO500-2.1.0.sh
```

Docker や DRAGEN サーバーソフトウェア (3.10.9) をアンインストールする必要はありません。Docker を削除するには、Docker の文書で CentOS のインストール手順を確認してください。

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_TS0500.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、[10 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 Convert によって作成される FASTQ フォルダから開始できます。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 フォルダへのパス。ランまたは FASTQ フォルダを DRAGEN サーバーへ、推奨される編成 `/staging/runs/{RunID}` の staging フォルダ内にコピーします。ランフォルダの 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、ランフォルダ内のサンプルシートが使用されます。
- ランフォルダは変更しないでください。入力要件については、[12 ページの「BCL ファイルから開始」](#) を参照してください。
- 解析の出力フォルダのパスがデフォルト設定と異なる場合は、そのパスを指定します。[10 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。

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。DRAGEN TruSight Oncology 500 Analysis Software はサンプルシート v2 をサポートします。

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。互換性のあるサンプルシートのテンプレートは、ソフトウェアと一緒にインストールされたリソースファイルのフォルダにも用意されています。

サンプルシートには、サンプルのリスト、各サンプルのインデックスシーケンス、オプションのサンプル情報が記載されています。TruSight Oncology 500 HRD を使用して濃縮された DNA サンプルは、サンプルシートの [Sample Feature] 列に示す必要があります。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。

- サンプルシート v2 テンプレートをダウンロードします。以下の 2 箇所から入手できます。
 - [イルミナサポートサイト](#) の TruSight Oncology 500 ページ
 - ソフトウェアと一緒にインストールされたリソースセット (`/staging/illumina/DRAGEN_TS0500/resources/sampleSheet/`)

2. [BCL Convert Settings] セクションに、以下の必須パラメーターを入力します。

サンプルのパラメーター	必須	詳細
SoftwareVersion	はい	2.1 と入力する
AdapterRead1	はい	8 bp インデックスを使用している場合： AGATCGGAAGAGCACACGTCTGAACTCCAGTCA 10 bp インデックスを使用している場合： CTGTCTCTTATACACATCTCCGAGCCCACGAGAC
AdapterRead2	はい	8 bp インデックスを使用している場合： AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT 10 bp インデックスを使用している場合： CTGTCTCTTATACACATCTGACGCTGCCGACGA
AdapterBehavior	はい	trim と入力する。これにより、BCL Convert ソフトウェアが各リードから指定されたアダプターシーケンスをトリミングすることを示す
MinimumTrimmedReadLength	はい	35 と入力する。このポイントより下の長さでトリミングされたリードがマスクされる
MaskShortReads	はい	35 と入力する。このポイントより下の長さでトリミングされたリードがマスクされる

3. [BCL Convert Data] セクションに、各サンプルの以下のパラメーターを入力します。

サンプルのパラメーター	必須	詳細
Sample_ID	はい	[TSO500S_Data] セクションに記載されている Sample_ID に一致する必要がある
Index	はい	[TSO500S_Data] セクションの一致する Sample_ID に割り当てられた Index_ID に有効な Index 1 シーケンス
Index2	はい	[TSO500S_Data] セクションの一致する Sample_ID に割り当てられた Index_ID に有効な Index 2 シーケンス
Lane	NovaSeq 6000 XP ワークフローのみ	どのレーンが特定のサンプルに対応しているのかを示す。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。レーン値が指定されている場合、指定した Index ID はレーンごとに一意でなければならない

4. [TSO 500 Data] セクションに、各サンプルの以下のパラメーターを入力します。

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 - 各ランに固有のもの 1～40 文字 - スペースを含まない - 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 - 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 - ペア ID に基づくサンプル ID にすることを推奨する。 例：<PairID>-DNA、<PairID>-RNA
Index_ID	はい	サンプルに使用したインデックスアダプター ID。サンプルが異なるレーンに割り当てられている場合を除き、サンプルごとに一意にしなければならない DNA サンプルでは、インデックス UP、CP、および UDP が有効であるが、RNA サンプルに使用できるのはインデックス UP および UDP のみである
Sample_Type	はい	DNA または RNA と入力する
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
Sample_Feature	いいえ	HRD 濃縮サンプルにのみ必須。 HRD 濃縮されている DNA サンプルの場合は、サンプルシートのこの列に HRD と入力する。サンプルが HRD 濃縮されていない場合は、このフィールドは空欄にする
Sample_Description	いいえ	サンプルの説明は以下の要件を満たす必要がある 1～50 文字 - 英数字、アンダースコア、ダッシュ、スペース。アンダースコア、ダッシュ、またはスペースを入力する場合は、その前後に英数字を入力する。例：Solide-FFPE_213

5. サンプルシートの下部に空白行が含まれていないことを確認します。空白行が入っていると、解析が正常に終了しません。
6. 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	詳細
<code>--help</code>	いいえ	使用可能なオプションをヘルプ画面に表示する
<code>--analysisFolder</code>	いいえ	ローカルの解析フォルダーへのパス。デフォルトの場所は <code>/staging/DRAGEN_TSO500_Analysis_{timestamp}</code> 。デフォルトの場所を使用しない場合は、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。例: DRAGEN サーバーの <code>/staging</code> ディレクトリ 最小ディスクスペースの要件については、 2 ページの「インストール要件」 の表を参照すること
<code>--resourcesFolder</code>	いいえ	リソースフォルダーへのパス。デフォルトの場所は <code>/staging/illumina/DRAGEN_TSO500/resources</code> 。デフォルトの場所を使用しない場合は、リソースフォルダーへのフルパスを入力する
<code>--runFolder</code>	はい	<code>--fastqFolder</code> 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
<code>--fastqFolder</code>	はい	<code>--runFolder</code> 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
<code>--user</code>	いいえ	<code>docker</code> のオプション。Docker コンテナ内で使用するユーザー ID を指定する
<code>--version</code>	いいえ	ソフトウェアのバージョンを表示する

オプション	必須	詳細
--sampleSheet	いいえ	ランフォルダー内の SampleSheet.csv を使用しない場合は、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のあるカンマ区切りのサンプル ID またはペア ID をスペースなしで指定する。 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	インストール時に作成された DRAGEN ハッシュテーブルの場所がデフォルトになる。ユーザーがこの場所を変更した場合のみ指定する必要がある

その他のコマンドについては、[10 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。BCL ファイルから開始する場合、指定する必要があるのはランフォルダーのみです。BCL ファイルを含む直接の親ディレクトリを指定する必要はありません。

SSH を使用して解析ソフトウェアを実行する場合、イルミナでは、解析が予期せずに終了することを防ぐために追加のソフトウェアを使用することを推奨します。イルミナでは screen と tmux を推奨しています。

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。

```
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 \
--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 \
--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 \
--runFolder /staging/{RunFolderName} \
--analysisFolder /staging/{AnalysisFolderName} \
--sampleSheet /staging/{SampleSheetName}.csv \
--demultiplexOnly
```

- FASTQ フォルダー内のサンプルシートで、異なるリソースとハッシュテーブルフォルダーを指定して FASTQ 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 \
--fastqFolder /staging/{FastqFolderName} \
--analysisFolder /staging/{AnalysisFolderName} \
--sampleOrPairIDs "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。これらの入力には 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。BCL ファイルから開始する場合、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 フォルダへのフルパス
- FASTQ フォルダの構成が [13 ページの「FASTQ ファイル編成」](#) のフォルダ構成と一致していること
- サンプルシートが FASTQ フォルダ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。

 BCL Convert は、UMI シーケンスを FASTQ ファイルのリードヘッダーに書き込むように設定されています。

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 `--sampleOrPairIDs` 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。

5. デマルチプレックスと各解析ランの結果を単一のサーバーにコピーし、最終の `/Results` ディレクトリを生成して、ここに集計結果を入れます。これを行うには、`--gather` コマンドの後に、デマルチプレックスステップと個々の解析ランの出力ディレクトリを指定して使用します。

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>

ステップ	コマンド
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} -- sampleOrPairIDs Pair_1,Pair_2</pre>
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} -- sampleOrPairIDs 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

収集されたシーケンスデータは、DRAGEN TruSight Oncology 500 Analysis Software モジュールによって処理され、品質管理、バリエーションの検出のほか、腫瘍変異負荷（TMB）、マイクロサテライト不安定性（MSI）ステータス、ゲノム不安定性スコア（GIS）の決定、結果のレポートが行われます。以降のセクションで解析方法を説明します。

FASTQ の生成

BCL 形式で保存されたシーケンスデータは、各サンプルに固有のインデックスシーケンスを使用してクラスターを元のライブラリーに割り当てるプロセスによってデマルチプレックスされます。各クラスターには 2 つのインデックス（i7 シーケンスと i5 シーケンスが、ライブラリーフラグメントの各端に 1 つずつ）が含まれていて、これらのインデックスシーケンスの組み合わせが、プールされたライブラリーのデマルチプレックスに使用されます。

デマルチプレックス後に、このプロセスは FASTQ ファイルを生成します。このファイルには、個々のサンプルライブラリーのシーケンスリードと、各ベースコールの関連するクオリティスコアが含まれます。ただしフィルターを通過しなかったクラスターのリードは除外されます。

DNA 解析方法

DNA アライメントとエラー補正

DNA アライメントとエラー補正では、DNA ライブラリーから抽出したシーケンスリードとリファレンスゲノムをアライメントし、バリエーションコール前にシーケンスリードのエラーを補正します。

DRAGEN UMI のエラー補正は、3 つの主な手順で構成されます。

1. DRAGEN UMIは、ハードウェアで高速化されたマッピング機能（ハッシュテーブル実装に基づく）を使用して、FASTQファイル内のDNA配列とhg19リファレンスゲノムをアライメントします。これらのアライメントはBAMに書き込まれません。
2. 生のアライメントを処理して、FFPE保存中、PCR増幅中、シーケンシング中などに取り込まれるエラーを除去します。元のDNA分子が同じリードは、ライブラリー調製時に、同じ分子バーコード（UMI）でタグ付けされます。DRAGENはUMIを使用することで、関連するリードを比較して、外れ値のシグナルを除去し、複数のリードを1つの高品質のシーケンスにCollapsingできます。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）
3. DRAGENは、UMIでCollapsingしたリードに最終アライメントステップを実施します。これらの最終アライメントはその後BAMファイルに書き込まれ、対応するBAMインデックスファイルが作成されます。

DRAGEN は引き続きこれらの最終アライメントを、遺伝子増幅（コピー数）コール、スモールバリエーションコール（SNV、INDEL、MNV、DELIN）、マイクロサテライト不安定性（MSI）ステータスの決定、DNA ライブラリーの品質管理の入力として使用します。

スモールバリエーションコールとフィルタリング

DRAGEN では、腫瘍サンプルのマッピングおよびアライメントした DNA リードを入力として使用することで、Tumor-only サンプルの SNV、挿入、欠失、MNV、DELIN のコールをサポートします。バリエーションは、列ごとのパイルアップ解析と、ハプロタイプのローカルな de novo アセンブルの両方によって検出されます。de novo ハプロタイプを使用することで、列ごとのパイルアップ解析のみの場合と比べて、はるかに大きな挿入と欠失を検出できます。DRAGEN の挿入と欠失は、少なくとも 0 ~ 25 bp の長さで検証されます。さらに、DRAGEN では de novo アセンブルを使用して、同時にフェージングされ、同じハプロタイプの一部である SNV、挿入、欠失を検出します。15 bp のウィンドウ内にあるこうした同時にフェージングされたバリエーションは、その後、複合バリエーション（MNV と DELIN）に再構築できます。Tumor-only パイプラインは、生殖細胞系バリエーションと体細胞バリエーションの両方を含む VCF ファイルを生成します。これをさらに解析することで、腫瘍変異を同定できます。パイプラインは倍数性の推測を行わず、低頻度アレルの検出を可能にします。

DRAGEN のスモールバリエントコールには、以下のステップが含まれます。

1. 十分なリードカバレッジのある領域を検出する（コール可能な領域）
2. リードがリファレンスから逸脱していて、生殖細胞コールまたは体細胞コールの可能性がある領域を検出する（アクティブ領域）
3. de novo グラフハプロタイプをリードからアセンブルする（ハプロタイプアセンブル）
4. 列ごとのパイルアップ解析から、体細胞コールまたは生殖細胞コールの可能性があるものを抽出する
5. FFPE ノイズを考慮して、リードベースのクオリティを調整する
6. 各リード/ハプロタイプペアのリードの尤度を計算する
7. すべてのリード/ハプロタイプペアで遺伝型の確度を合計することで、変異コールを実施する
8. 系統ノイズファイルを使用するなど、追加でフィルタリングを実施して、バリエントコールの精度を向上させる。系統ノイズファイルは、ゲノムの特定の位置におけるノイズの統計的確度を示します。このノイズファイルは、クリーンな（正常な）サンプルを使用して構成されています。ノイズがよく見られる（領域のマッピングが困難な）領域では、ノイズ値が高くなります。スモールバリエントコーラーは、これらの領域にペナルティーを科すことで、コールの偽陽性の確率を減少させます。TSO 500 で使用されるフィルターは、「DNA 出力」セクションに記載されています。

コピー数バリエントコール

DRAGEN では、Tumor-only サンプルで遺伝子増幅を同定できます。

DRAGEN の遺伝子増幅コールには、以下のステップが含まれます。

1. パネルの対象となる領域ごとにターゲットをカウントする
2. 事前処理された正常なサンプルのパネルを利用して、ターゲットカウントをノーマライズする
3. GC バイアスを原因とするカバレッジにおける差を補正する
4. 統計モデルを適用して、観察したカバレッジの CNV イベントのスコアを計算し、コピー数コールを実施する

エクソンレベルのコピー数バリエントコール

BRCA の大規模な再編成ステップにより、BAM ファイルからエクソンレベルの CNV を検出することを目的に、BRCA1 遺伝子と BRCA2 遺伝子のセグメンテーションが生成されます。この大規模な再編成コンポーネントは、CNV コールと同じ方法を使用して、パネルの各ターゲット区間のカバレッジを計算し、ノーマライゼーションを行い、BRCA 遺伝子全体でプローブごとの倍率変化値を算出します。ノーマライゼーションでは、正常な FFPE とゲノム DNA サンプルのコレクションを使用した、GC の偏り補正、シーケンス深度、プローブ効率などが対象になります。最初のセグメンテーションは、循環バイナリセグメンテーションを使用して遺伝子ごとに実施されます。次に、in silico データで確立された閾値を使用して、隣接するセグメントの振幅、ノイズ、分散により、セグメントのマージが決定されます。複数のセグメントを持つ遺伝子では、大規模な再編成がレポートされます。各 BRCA 遺伝子セグメントのエクソンレベルの CNV と対数平均倍率変化の座標は、*_DragenExonCNV.json ファイルで確認できます。

アノテーション

Illumina Annotation Engine で、スモールバリエーション、CNV、エクソンレベルの CNV 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

腫瘍変異負荷

DRAGEN は、十分なカバレッジがあるコーディング領域で、腫瘍変異負荷（TMB）の計算に使用されます。

以下のバリエーションは、TMB の計算から除外されます。

- フィルターをパスしなかったバリエーション
- ミトコンドリアバリエーション
- MNV
- 最小深度の閾値に達しないバリエーション
- 最小バリエーションアレルの閾値に達しないバリエーション
- 適格領域から外れるバリエーション
- 腫瘍発生に関与する遺伝子変異。腫瘍発生に関与する遺伝子変異として扱われる集団アレルカウントが 50 以上のバリエーション

生殖細胞系列バリエーションは TMB にカウントされません。バリエーションは、データベースとプロキシフィルターに基づいて生殖細胞系列と決定されます。

1000 Genomes データベースまたは gnomAD データベースのいずれかで観察された、集団アレルカウントが 50 以上のバリエーションが生殖細胞系列としてマークされます。プロキシフィルターが特定のバリエーションの周囲のバリエーションをスキャンし、同様のバリエーションアレル頻度（VAF）を持つバリエーションを特定します。同様の VAF を持つ周囲のバリエーションの大部分が生殖細胞系列の場合、そのバリエーションも生殖細胞系列としてマークされます。

TMB の計算の式は以下のようになります。

- $TMB = \text{フィルターしたバリエーション} / \text{適格な領域サイズ (Mbp)}$
- $\text{非同義 TMB} = \text{フィルターした非同義バリエーション} / \text{適格な領域サイズ (Mbp)}$

出力は、TMB の計算に使用したバリエーションの情報を含む *_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

DRAGEN はサンプルの MSI ステータスの決定に使用されます。DRAGEN では、正常なサンプルセットから作成された、正常なリファレンスファイルを使用します。シーケンス中に、各マイクロサテライト部位のリードカウントを集計することで、正常なリファレンスファイルが生成されます。正常なファイルには、各マイクロサテライトのリードカウントの分布が記載されます。

Tumor-only サンプルに対する MSI コールは、最初に各マイクロサテライト部位のリードアライメントから腫瘍カウントを集計することにより実行されます。次に、腫瘍と正常なベースラインサンプルの各ペアの Jensen-Shannon distance (JSD) が計算されます。DRAGEN は、腫瘍の JSD 分布と正常な JSD 分布のカイ二乗検定を実行して、不安定な部位を特定します。2 つの JSD 分布の平均距離差が距離閾値以上で、カイ二乗 p 値が p 値閾値以下の場合、不安定な部位としてコールされます。最後に、DRAGEN は、評価された部位のカウント、不安定な部位のカウント、すべての評価された部位における不安定な部位の割合、すべての不安定な部位の Jensen-Shannon distance の合計を考慮して、MSI ステータスを生成します。

ゲノム不安定性スコア

ゲノム不安定性スコア (GIS) は、相同組換え欠損の全ゲノムサインです。GIS は、ヘテロ接合性欠失、テロメアアリル不均衡、大規模な状態遷移という 3 つの要素の合計で構成されています。これらの要素は、Myriad Genetics から契約で得た GIS アルゴリズムを使用して推定されます。GIS アルゴリズムでは、ゲノム全体の 1 塩基パネルで b アリル頻度とカバレッジの入力を使用します。正常なサンプルのパネルは、GIS の推定の前に、バイアス低減とノーマライゼーションの両方で使用されます。GIS の最終結果は、*.gis.json ファイルに記載されています。

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、スモールバリエーションコール中に生成された SNP エラーファイルとパイルアップファイル、および TMB トレースファイルを使用して、外来のヒト DNA のコンタミネーションを検出します。ソフトウェアは、コンタミネーションスコアを使用して、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のあるサンプルでは、SNP のバリエーションアリル頻度が 0%、50% または 100% の期待値から変化します。アルゴリズムでは、バリエーションアリル頻度が 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、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。コンタミネーションスコアは、サンプルの最小アリル頻度が 25% 未満で、CNV イベントが原因である可能性が低い、既定の SNP 位置にわたるすべての対数尤度スコアの合計です。

コンタミネーションスコアが大きくなるほど、外来の DNA によるコンタミネーションである可能性が高くなります。コンタミネーションスコアが既定のクオリティ閾値を超えた場合、サンプルにはコンタミネーションがあると見なされます。コンタミネーションスコアは、ゲノムが大規模に再編成されたサンプルや HRD サンプルで高いことがわかっています。HRD サンプルのうち、1% では、実際のコンタミネーションのエビデンスがない状態で、閾値を超えているものがありました。

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（すなわち、全レーンからの R1+R2）。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。

RNA アライメントおよび融合遺伝子検出

RNA アライメントおよび融合遺伝子検出は、FASTQ 形式のトリミングされたリードを入力として使用します。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。

DRAGEN 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 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 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 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 の RNA 融合遺伝子コーラーは、2 つの異なる親遺伝子にまたがるキメラリードを探すことで、遺伝子融合を同定します。DRAGEN はまず、キメラリードに基づいて融合遺伝子の候補のリストを作成し、次に候補にスコアを付け、候補プールから信頼度の高い融合遺伝子コールのリストを作成します。

DRAGEN の RNA 融合遺伝子コーラーは、以下のステップを実施します。

1. スプリットリードアライメントに基づいて融合遺伝子の候補を生成する
2. 融合遺伝子の不一致のサポートリードペアと soft-clip リードから追加のエビデンスを収集する
3. 遺伝子カバレッジ、リードマッピングクオリティ、代替アレル頻度、遺伝子相同性、アライメントアンカー長、エクソン境界からのブレイクポイント距離などの融合遺伝子候補の特徴を算出する

4. ロジスティック回帰モデルを使用して融合遺伝子候補にスコアを付け、ランク付けする
5. スコアや他のフィルター（サポートリードの数、固有のリードアライメントカウント、Read-through transcript、濃縮された領域に一致する融合遺伝子など）に基づいて、融合遺伝子コールの最終リストを選択する

スプライスバリエーションコール

RNA スプライスバリエーションコールは、RNA サンプルライブラリーに対して実行されます。RNA アライメントのスプライスバリエーション候補（ジャンクション）が、既知の転写産物のデータベースと、さまざまな組織タイプの正常な FFPE サンプルから生成された非腫瘍ジャンクションのスプライスバリエーションベースラインと比較されます。データベースまたはベースラインと一致するスプライスバリエーションは、腫瘍学的機能が知られている一連のジャンクションに含まれていない限り、除外されます。十分なリードサポートがある場合、スプライスバリエーション候補は保持されます。このプロセスで、RNA 融合遺伝子候補も同定されます。

RNA 融合遺伝子のマージ

RNA 融合遺伝子コール中に同定された融合遺伝子は、RNA スプライスバリエーションコール中に同定された隣接遺伝子の融合遺伝子とマージされます。これらは、転写産物の静的データベース（GENCODE v19）に関して、遺伝子記号または名前がアノテーションされます。このプロセスの結果として、一連の融合遺伝子コールがレポート作成に適切になります。

RNA スプライスバリエーションのアノテーション

Illumina Annotation Engine は、RefSeq に関して、検出された RNA スプライスバリエーションコールに転写レベルの変化（遺伝子の転写においてエクソンが影響を受けた）をアノテーションします。この RefSeq データベースは、スモールバリエーションアノテーションプロセスで使用する RefSeq データベースと同じです。

品質管理（QC）

DRAGEN TruSight Oncology 500 Analysis Software v2.1 には、数種類の QC が含まれます。

ラン QC

メトリクスの出力レポートの [Run Metrics] セクションでは、シーケンスランのクオリティメトリクスとともに、それらが許容基準を満たす範囲内であるかを判断するための推奨値を示します。パスフィルターリードの全体的な割合が最小閾値と比較されます。Read 1 と Read 2 に関して、不正確なベースコールの確度の予測値（Q スコア）を示す、Q30 以上の塩基の平均割合も最小閾値と比較されます。以降の表に、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示します。

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

表2 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3 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

DNA サンプル QC

DRAGEN TruSight Oncology 500 Analysis Software v2.1 では、QC メトリクスを使用して、コンタミネーションの品質管理をパスする DNA ライブラリーのスモールバリエーションコール、TMB、MSI、および遺伝子増幅の有効性を評価します。ライブラリーが 1 つまたは複数のクオリティメトリクスをパスしなかった場合、対応するバリエーションタイプまたはバイオマーカーはレポートされず、レポートヘッダーの関連する QC カテゴリに FAIL と表示されます。さらに、以下の QC カテゴリの 1 つまたは複数での QC のパスを必要としているコンパニオン診断結果は利用できません。

DNA ライブラリーの QC の結果は、MetricsOutput.tsv ファイルに記載されます。詳細については、[24 ページの「メトリクスの出力」](#)を参照してください。

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	コンタミネーションスコアは SNP の VAF 分布に基づく	コンタミネーションスコア ≤ 1457	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	≥ 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	≥ 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	≥ 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 部位の数	≥ 40	MSI
GENE_SCALED_MAD	欠失の可能性のある遺伝子を除外し、各 CNV ターゲット領域の各遺伝子についてノーマライズしたカウントの中央値を補正した後の、ノーマライズしたカウントの中央値からの絶対偏差の中央値 (MAD)	≤ 0.134	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	≥ 1.0	CNV
PCT_TARGET_HRD_50X	(HRD) 50X フラグメントカバレッジを持つ HRD プローブの割合。HRD サンプルのみレポート	≥ 50.0	GIS

RNA サンプル 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。詳細については、[24 ページの「メトリクスの出力」](#) を参照してください。

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 を超える遺伝子は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	≥ 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	≥ 9000000	融合遺伝子のスプライス
GENE_MEDIAN_COVERAGE	RNA パネル内のすべての遺伝子の重複を除去したカバレッジの中央値（55 遺伝子）	該当なし *	融合遺伝子のスプライス

* RNA サンプルが不必要に失敗するのを避けるため、イルミナでは、汎用的な閾値による RNA サンプルクオリティの決定を推奨していません。RNA の発現は、組織タイプと小さなパネルサイズ（55 遺伝子）によって大きく異なるため、ノーマライゼーションが困難になります。ノーマライゼーションでは、組織固有の閾値を考慮できます。

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 DNA ライブラリー QC メトリクス：
 - スモールバリエーションコール
 - TMB
 - MSI
 - CNV
 - GIS（TruSight Oncology 500 HRD が動作している場合）

- RNA ライブラリー QC メトリクス
- 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には、ランの各サンプルに推奨される下限値 (LSL) と上限値 (USL) が含まれています。

トラブルシューティングの情報については、[38 ページの「トラブルシューティング」](#)を参照してください。

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

AdditionalSarjMetrics : PCT_TARGET_250X メトリクスを裏付けるペア ID ごとの計算を含む

Annotation : スモールバリアントアノテーションの出力を含む

 サンプル ID ごとのサブフォルダー : アライメントしたスモールバリアントの JSON を含む

CombinedVariantOutput

 ペア ID ごとのサブフォルダー : 組み合わせバリアント出力 TSV ファイルを含む

 組み合わせ出力ログファイル

Contamination

 DNA サンプル ID ごとのサブフォルダー : コンタミネーションメトリクスの JSON と出力ログを含む

DnaDragenCaller

 サンプル ID ごとのサブフォルダー : アライメントされた BAM とインデックスファイル、スモールバリアントの VCF と gVCF、コピー数バリアントの VCF、MSI JSON、および CSV 形式での QC 出力を含む

DnaDragenExonCNVCaller

 DNA サンプル ID ごとのサブフォルダー : エクソンレベルの CNV JSON と裏付けとなる計算、および QC ファイルを含む

DnaFastqValidation : DNA サンプルの FASTQ バリデーション出力ログを含む

FastqDownsample

- 📁 RNA サンプル ID ごとのサブフォルダー：FASTQ ファイルと出力ログを含む
 - 📁 FastqDownsample の出力
- 📁 Gis：HRD サンプルの GIS 関連ファイルを含む
 - 📁 HRD サンプル ID ごとのサブフォルダー：GIS JSON、裏付けとなる計算、および QC ファイルを含む
- 📁 LrAnnotation
 - 📁 DNA サンプル ID ごとのサブフォルダー：アノテーションされたエクソンレベルの CNV JSON を含む
- 📁 LrCalculator
 - 📁 DNA サンプル ID ごとのサブフォルダー：エクソンレベルの CNV VCF を含む
- 📁 MetricsOutput
 - 📁 ペア ID ごとのサブフォルダー：メトリクス出力 TSV ファイルを含む
 - 📁 組み合わせ出力ログファイル
- 📁 ResourceVerification：リソースファイルのチェックサム検証ログを含む
- 📁 RnaAnnotation
 - 📁 RNA サンプル ID ごとのサブフォルダー：アノテーションされたスプライスバリエーションの JSON を含む
- 📁 RnaDragenCaller
 - 📁 サンプル ID ごとのサブフォルダー：アライメントされた BAM、融合遺伝子候補 CSV、および CSV 形式での QC 出力を含む
- 📁 RnaFastqValidation：RNA サンプルの FASTQ バリデーション出力ログを含む
- 📁 RnaFusion
 - 📁 RNA サンプル ID ごとのサブフォルダー：すべての融合遺伝子 CSV と融合遺伝子プロセッサログを含む
- 📁 RnaQcMetrics
 - 📁 RNA サンプル ID ごとのサブフォルダー：RNA QC メトリクスの JSON を含む
- 📁 RnaSpliceVariantCalling
 - 📁 RNA サンプル ID ごとのサブフォルダー：スプライスバリエーション VCF を含む
- 📁 Run QC：ラン QC メトリクスの JSON、中間ラン QC メトリクスの JSON、ログファイルを含む
- 📁 SampleAnalysisResults
 - 📁 ペア ID ごとのサブフォルダー：サンプル解析結果の JSON と詳細なログファイルを含む
- 📁 SampleSheetValidation：中間サンプルシートとバリデーションログを含む
- 📁 Tmb
 - 📁 DNA サンプル ID ごとのサブフォルダー：TMB メトリクス CSV、TMB トレース TSV、関連ファイルおよびログを含む
 - passing_sample_steps.json：各サンプル ID のパスしたステップを含む

pipeline_trace.txt：実行された各 Nextflow タスクとステータス（COMPLETED、FAILED など）を記載したサマリーおよびトラブルシューティングファイルを含む

run.log：Nextflow パイプライン実行を説明する完全なトレースレベルのログファイルを含む

run_report.html：高レベルのラン統計（性能、使用状況など）を含む

run_timeline.html：解析ランのタイムライン関連の情報を含む

Results

Metrics Output TSV（すべてのペア ID）

Pair ID：サンプルごとに以下の出力が生成される

- Combined Variant Output TSV
- Metrics Output TSV
- TMB Trace TSV
- Small Variant Genome VCF
- Small Variant Genome Annotated JSON
- Copy Number Variant VCF
- GIS JSON
- Exon-level CNV VCF
- Exon-level CNV Annotated JSON
- All Fusion CSV
- Splice Variant VCF
- Splice Variant Annotated 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

NodeX_Output：使用した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たサンプルの結果のみを含む

Results：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: {Pair_ID}_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、ファイルは同じ個体の対となる 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
- (HRD) GIS
- エクソンレベル CNV

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	内容説明
• ペア ID • DNA サンプル ID (DNA の実行の場合) • RNA サンプル ID (RNA の実行の場合) • 出力日 • 出力時間 • モジュールのバージョン • 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)	• ラン名 • ラン日 • DNA サンプルインデックス ID (DNA の実行の場合) • RNA サンプルインデックス ID (RNA の実行の場合) • サンプルの特徴 (HRD) • 装置 ID • 装置コントロールソフトウェアのバージョン • 装置タイプ • RTA のバージョン • 試薬カートリッジのロット番号

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、iIndel、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**: ハードフィルタリングしたゲノムの 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の 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の 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。
- **バイオマーカー GIS**：TruSight Oncology 500 HRD が動作している場合のみ存在します。
- **エクソンレベル CNV**：エクソンレベル CNV は以下の条件を必ず満たす必要があります。
 - BRCA1 または BRCA2 に少なくとも 1 つの影響を受けたエクソンを含む。
 - [ALT] フィールドが <DUP> または <LOSS>。

DNA 出力

スモールバリエーション gVCF

ファイル名：{SAMPLE_ID}_hard-filtered.gvcf.gz

スモールバリエーションゲノムのバリエーションコールファイルには、TSO 500 パネル全体のフェージングしたバリエーションコールからの最大 15 bp の複合バリエーションを含む、評価したすべてのスモールバリエーション候補に関する情報が記載されています。

バリエーションステータスは、以下のゲノム VCF の [FILTER] 列から判定します。

Filter	注記
PASS	PASS バリエーション
base_quality	この座位での代替リードのベースクオリティの中央値が閾値に達しないためにフィルターした部位
filtered_reads	リードの割合が大きすぎるためにフィルターした部位

fragment_length	この座位での代替リードのフラグメント長中央値とリファレンスリードのフラグメント長の中央値の絶対差が閾値を超えるためにフィルターした部位
low_depth	リード深度が低すぎるためにフィルターした部位
low_frac_info_reads	有益なリードの割合が閾値を下回るためにフィルターした部位
long_indel	Indel 長が長すぎるためにフィルターした部位
mapping_quality	この座位での代替リードのマッピングクオリティの中央値が閾値に達しないためにフィルターした部位
multiallelic	2 つ以上の代替アリルが腫瘍 LOD をパスするためにフィルターした部位
no_reliable_supporting_read	信頼できる体細胞サポートリードが存在しないためにフィルターした部位
read_position	リードの開始 / 終了間の距離の中央値とこの座位が閾値を下回るためにフィルターした部位
str_contraction	代替アリルの反復単位がリファレンスより 1 少なく PCR エラーが疑われたためフィルターした部位
too_few_supporting_reads	腫瘍サンプルのサポートリードが少なすぎるためフィルターした部位
weak_evidence	体細胞バリエーションスコア (SQ) が閾値に達しない
systematic_noise	正常なサンプル中の系統ノイズのエビデンスに基づいてフィルターした部位
excluded_regions	VC で除外された領域ベッドと重複する部位

スモールバリエーションのアノテーション JSON

ファイル名: {SAMPLE_ID}_DNAVariants_Annotated.json.gz

スモールバリエーション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ーションやパスしなかったバリエーションなど、ゲノム VCF 内にあるリファレンス以外のすべての位置に関するバリエーション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。スモールバリエーション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ーションが、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、バリエーション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、バリエーションの分類が体細胞または生殖系列に変わることはありません。

列	内容説明
Chromosome	染色体
Position	バリアントの位置
RefCall	リファレンス塩基
AltCall	バリアント塩基
VAF	バリアントアリル頻度
Depth	バリアント検出位置のカバレッジ
CytoBand	バリア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	バリアントのタイプ：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アントアリルカウント
AlleleCountsGnomadGenome	gnomAD genome データベース中のバリアントアリルカウント
AlleleCounts1000Genomes	1000 genomes データベース中のバリア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アントアリルカウント
GermlineFilterDatabase	バリア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	バリア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	バリアントがコーディング領域にある場合、TRUE
Nonsynonymous	バリアントに、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	バリアントが TMB の計算で使用される場合、TRUE

コピー数 VCF

コピー数 VCF ファイルには、DRAGEN TruSight Oncology 500 Analysis Software v2.1 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。 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。

 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 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。

フィルター名	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、DRAGEN RNA 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4 [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5 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： • IntactExon：エクソン境界に一致 • BrokenExon：エクソン内 • Intronic：イントロン内 • Intergenic：遺伝子重複なし（現在は除外） 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。場合によっては、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： • IntactExon：エクソン境界に一致 • BrokenExon：エクソン内 • Intronic：イントロン内 • Intergenic：遺伝子重複なし（現在は除外） 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、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、遺伝子順序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、偽陽性のバリエント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=excluded_regions` としてラベルされ、バリエント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

解析パフォーマンステスト

イルミナでは、ライブラリー調製、シーケンス、二次解析などのワークフロー全体を対象とする手法を使用して、バリエントコールの解析パフォーマンスをテストします。この手法は、さまざまなバリエントのテストに使用されます。新しいバリエントクラスをコールするようにバリエントコールのパイプラインを拡張する場合、必ずこの手法が使用されます。

in silico特性化法

イルミナでは in silico テストを使用して、拡張した範囲内に含まれる、レアバリエントなどの臨床的に重要なバリエントをコールするソフトウェアの能力をテストします。イルミナでは、EGFR の複合バリエントをコールするソフトウェアの能力については分析的に検証しましたが、他の遺伝子の複合バリエントをコールするソフトウェアの実行可能性については、in silico テスト手法で特性化します。in silico テスト手法では、TruSight Oncology 500 High-Throughput アッセイの BRCA 1 と BRCA 2 における 25 bp を超える挿入と欠失およびエクソンレベル CNV も特性化します。この特性化では、対象バリエントは Cosmic や ClinVar など公開データベースから抽出します。各バリエントは、変異体リードを正常な FFPE バックグラウンドにスパイクインする、または正常な FFPE サンプルのエクソンのカバレッジを増減させてエクソンレベル CNV をシミュレートすることで、さまざまな VAF レベルでシミュレートします。シミュレートしたリードは、フラグメント長、エラー率、ファミリーサイズなど、典型的な FFPE サンプルの予想されるクオリティと一致します。シミュレーション後、ソフトウェアがスパイクインバリエントを使用してサンプルを処理し、結果を判断します。この手法には、対象レアバリエントを含む腫瘍 FFPE サンプルのライブラリー調製とシーケンスは含まれません。ソフトウェアはこれらのバリエントをレポートしますが、分析的検証は実施されません。

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/pipeline_trace.txt</code> を開く。このログファイルには、Nextflow ワークフローマネージャーソフトウェアが実行した各パイプラインステップが表示される。失敗したステップには FAILED とマークされる。各ステップでログファイルが生成され、Logs_Intermediates フォルダのステップ固有のサブフォルダに保存される。そのステップに関連する Logs_Intermediates フォルダのログファイルをレビューして、エラーの原因を特定する
サンプル	統合メトリクスの出力結果ファイル <code>./{AnalysisFolder}/Results/{PairId}/MetricsOutput.tsv</code> を開く。サンプルの解析ステップが失敗した場合、そのサンプルを含むペア ID は、[Analysis Status] セクションの [FAILED_STEPS] の下に表示され、[COMPLETED_ALL_STEPS] に False と表示される。可能な場合は、 <code>./{AnalysisFolder}/Logs_Intermediates</code> の下に表示される失敗したステップの各ログファイルをレビューして、エラーの原因を特定する
マルチノード ギャザー	以下のエラーが表示された場合は、ギャザリング前の、別々のノード解析ラン中に、サンプルまたはペア ID が複数回含まれていなかったかどうかを確認する。エラーが存在する場合は、重複のない状態で解析のいずれかを再実行し、ギャザリングを再試行する ERROR:Gather:Destination file ... already exists - check if the same sample ID is in multiple input folders

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ー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 (%)	%
PCT_READ_ENRICHMENT	ターゲット領域の任意の部分の横断するリードの総リードに対する割合 (%)。このメトリクスは HRD プローブ以外のパネルプローブのみを対象とするため、サンプルに HRD プローブを適用している場合、観察される値が低くなる可能性がある	%
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 (%)	%
PCT_TARGET_50X	50X 超のカバレッジを持つターゲット塩基の割合 (%)	%
(HRD) ALLELE_DOSAGE_RATIO	GIS ステップ中に計算されたノイズのアリル量比測定	カウント
(HRD) MEDIAN_TARGET_HRD_COVERAGE	HRD プローブのターゲットカバレッジの中央値	カウント

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
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
GENE_ABOVE_MEDIAN_CUTOFF	カバレッジカットオフの中央値を超える遺伝子の数	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

電子メール：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
フィリピン	+63 180016510798	
フィンランド	+358 800 918 363	+358 9 7479 0110

地域	フリーダイヤル	国際
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 から入手できます。

製品関連文書：jp.support.illumina.com からダウンロードできます。

リソースと参考文献

イルミナのサポートサイトの DRAGEN TruSight Oncology 500 Analysis Software v2.1 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、トレーニング、適合製品、その他の考慮事項が含まれています。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。

改訂履歴

文書	日付	変更内容
文書番号：200019138 v01	2022 年 8 月	解析パフォーマンステストセクションを更新し、in silico 特性化法セクションを追加
文書番号：200019138 v00	2022 年 8 月	初版リリース



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®

正誤表

ページ番号	章・節	原文	修正内容
p.22	DNA サンプル QC ゲノム DNA のタグメンテーション	DRAGEN TruSight Oncology 500 Analysis Software v2.1 では、QC メトリクスを使用して、コンタミネーションの品質管理をパスする DNA ライブラリーのスモールバリエーションコール、TMB、MSI、および遺伝子増幅の有効性を評価します。ライブラリーが 1 つまたは複数のクオリティメトリクスをパスしなかった場合、対応するバリエーションタイプまたはバイオマーカーはレポートされず、レポートヘッダーに関連する QC カテゴリに FAIL と表示されます。さらに、以下の QC カテゴリの 1 つまたは複数での QC のパスを必要としているコンパニオン診断結果は利用できません。	DRAGEN TruSight Oncology 500 Analysis Software v2.1 では、QC メトリクスを使用して、コンタミネーションの品質管理をパスする DNA ライブラリーのスモールバリエーションコール、TMB、MSI、および遺伝子増幅の有効性を評価します。 以降の「ライブラリーが 1 つまたは複数のクオリティメトリクスをパスしなかった場合、」のくだりは RUO 製品用の記述ではないため、削除。