



DRAGEN TruSight Oncology 500 ctDNA Analysis Software v1.1

User Guide

ILLUMINA PROPRIETARY

文書番号：1000000143376 v02 JPN

2021年9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を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。

© 2021 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

はじめに	1
インストール要件	1
入力要件	2
DRAGEN TruSight Oncology 500 ctDNA Analysis Softwareの インストール	5
DRAGEN TruSight Oncology 500 ctDNA Analysis Softwareの 実行	6
BCLファイルから開始	7
FASTQファイルから開始	8
シェルスクリプトのコマンドライン引数	9
解析方法	9
入力のバリデーション	10
FASTQファイルの生成	11
FASTQのバリデーション	11
リソースの検証	11
DNAアライメントとリードCollapsing	11
Indel再アライメントとリードステッチング	12
スモールバリエントコール	12
スモールバリエントフィルタリング	13
コピー数バリエントコール	14
フェーズドバリエントコール	14
バリエントのマージ	15
アノテーション	16
腫瘍変異負荷	16
最大体細胞VAF	17
マイクロサテライト不安定性ステータス	17
コンタミネーションの検出	17
結果	18
融合遺伝子コール	18
融合遺伝子フィルタリング	18
品質管理 (QC)	19

解析の出力	19
フォルダー	20
ファイル	21
DRAGEN TruSight Oncology 500 ctDNA Analysis Softwareの アンインストール	30
改訂履歴	31
テクニカルサポート	32

はじめに

Illumina® TruSight™ Oncology 500 ctDNA Local Applは、DRAGEN TruSight Oncology 500 ctDNAアッセイを使用して血漿から分離し調製したDNAライブラリーのシーケンスリードを解析するアプリです。このソフトウェアによってスモールバリエーション用のバリエーションコールファイル (*.vcf) が作成されます。レポートには、腫瘍変異負荷 (TMB) を表すメガベースあたりの変異スコア、マイクロサテライト不安定性 (MSI) ステータスを表す不安定性を示す座位の割合、スモールバリエーション、コピー数バリエーション (CNV)、およびDNA融合が記載されます。

本アプリケーションでは、ベースコールファイル (BCL) を含むイルミナのシーケンスランフォルダーやFASTQフォルダーにステージされたFASTQファイルから二次解析ワークフローを開始できます。

DRAGEN TruSight Oncology 500 ctDNA Analysis Softwareでは、専用のコンピューティングサーバーおよびソフトウェアパッケージであるIllumina DRAGEN Bio-IT Platformを使用して二次解析を実行します。

本書では、DRAGEN TruSight Oncology 500 ctDNA Analysis Softwareのワークフロー、コンピューター要件、インストール、解析ソフトウェア、解析方法、および解析の出力の概要について説明します。本アプリを使用するには、LinuxオペレーティングシステムとDockerソフトウェアコンテナプラットフォームに関する基礎知識が必要です。

DRAGEN TruSight Oncology 500 ctDNA Analysis Softwareは、TruSight Oncology 500 ctDNAアッセイによって生成されたシーケンスの出力を解析するために最適化された唯一の二次解析ソフトウェアです。ソフトウェアの改造は不正確なデータの原因となるおそれがあり、またイルミナのエンドユーザーライセンス契約違反となります。全条項については、エンドユーザーライセンス契約書を参照してください。

インストール要件

以下の項目が、DRAGEN TruSight Oncology 500 ctDNA の解析の最小動作要件です。

ハードウェア

DRAGEN TruSight Oncology 500 ctDNA Analysis Softwareは、Illumina DRAGEN Bio-IT Platform (サーバー) 上でのみ動作します。

ソフトウェア

DRAGEN Bio-ITサーバーにデフォルトでインストールされているソフトウェアは、以下のとおりです。

- Linux CentOS 7.3以降のオペレーティングシステム
- DRAGEN Bio-IT Software v3.6 (v3.6.4が必須)
- 他のLinuxツール

ストレージ

- DRAGEN Bio-ITサーバーには6.4 TBのNVMe SSDが搭載されています。これは、DRAGEN TruSight Oncology 500 ctDNAパイプラインが使用するDRAGEN Bio-ITソフトウェアによる解析用に最適化されています。6.4 TB NVMe SSDは/stagingディレクトリに位置し、解析パイプラインによる1、2回分のランのみを保存するのに適しています。
 - S4フローセルを使用するNovaSeqのシーケンスランでは、最大3 TBの出力を生成することがあります。
 - DRAGEN TruSight Oncology 500 ctDNA Analysis Softwareでは、さらに3 TB分の出力が生じます。
- 解析の出力は、/staging/DRAGEN_TSO500_ctDNA_Analysis_<timestamp>ディレクトリに自動的に書き込まれます。このフォルダーの場所によって、DRAGEN Bio-ITサーバー関連のプロセスで、最適化されたNVMe SSDにデータを読み書きすることが可能になります。
- シーケンスランおよびDRAGEN TruSight Oncology 500 ctDNA Analysis Softwareの出力を長期間保管するためには、ネットワークアタッチストレージが必要です。
- お客様の責任でデータストレージを管理してください。
 - DRAGEN Bio-IT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 Bio-IT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。

入力要件

イルミナのシステムとの互換性

DRAGEN TruSight Oncology 500 ctDNA Analysis Softwareでは、NovaSeq 6000からのシーケンス出力が必要です。

最適なカバレッジを得るためには、『TruSight Oncology 500 ctDNA Reference Guide』（文書番号：1000000092559）に記載のシーケンス設定に従ってください。

サンプルシートの要件

DRAGEN TruSight Oncology 500 ctDNA Analysis Softwareでは、各解析につきサンプルシート1つが必要です。サンプルシートファイルにはサンプルのリスト、インデックスシーケンスおよび解析ワークフローを含めます。

サンプルシートの見本が、DRAGEN TruSight Oncology 500 ctDNA Analysis Softwareのリソースセットに含まれています。このリソースセットは、/staging/illumina/DRAGEN_TSO500_ctDNA/resources/ディレクトリに存在します。リソースセットに含まれるサンプルシートをテンプレートとして利用してください。

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 ctDNA 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

サンプルシートの見本が、DRAGEN TruSight Oncology 500 ctDNA Analysis Softwareのリソースセットに含まれています。このリソースセットは、/staging/illumina/DRAGEN_TS0500_ctDNA/resources/ディレクトリに存在します。リソースセットに含まれるサンプルシートをテンプレートとして利用してください。以下のステップに従って、DRAGEN TruSight Oncology 500 ctDNAのサンプルシートを作成してください。

! | サンプルシートの基準を満たしていない場合、解析に失敗します。

1. シーケンスランフォルダーにSampleSheet.csvという名前を付けてサンプルシートを保存します。
2. [Settings] セクションに、以下のパラメーターを入力します。
 - AdapterRead1 : AGATCGGAAGAGCACACGTCTGAACTCCAGTCA
 - AdapterRead2 : AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
 - AdapterBehavior : trim
 - MinimumTrimmedReadLength : 35
 - MaskShortReads : 35
 - OverrideCycles : U7N1Y143;I8;I8;U7N1Y143
3. [Data] セクションに、以下のパラメーターを入力します。サンプルとサンプルの間に空白の行を残さないでください。
 - Sample_ID : Sample_IDは出力ファイル名の一部になります。サンプルIDでは、大文字と小文字が区別されません。以下の特徴を備えたサンプルIDを入力してください。
 - 各ランに固有のもの。
 - 40文字以下でスペースを入れません。
 - 英数字、アンダースコア、および/またはダッシュのみ。
 - 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ます。例 : Sample1-T5B1_022515。
 - Index_ID
 - Index
 - I7_Index_ID
 - Index2
 - I5_Index_ID
 - Sample_Type : 各サンプルの**DNA**を入力します。
 - Lane : [Lane] 列の入力は、XPフローセルを使用し、かつ1つのインデックスペアを複数のサンプルで用いる場合にのみ必要です。レーンごとのサンプルを指定するためにレーン列を使用します。サンプルごとに1つのレーンを必ず指定してください。1つのサンプルが複数のレーンにまたがる場合は、サンプルごとに複数の行を使用する必要があります (1レーンに対し1行)。

- オプションのパラメーター：「Sample_Name」などの、その他のオプションパラメーターは解析に必須ではありませんが、データセクションに追加することもできます。オプションのパラメーターには、必ず以下の特徴が必要です。
 - 40文字以下でスペースを入れません。
 - 英数字、アンダースコア、および/またはダッシュのみ。

V2サンプルシートのパラメーター

サンプルシートの見本が、DRAGEN TruSight Oncology 500 ctDNA Analysis Softwareのリソースセットに含まれています。このリソースセットは、/staging/illumina/TSO500_ctDNA/resources/ディレクトリに存在します。リソースセットに含まれるサンプルシートをテンプレートとして利用してください。以下のステップに従って、DRAGEN TruSight Oncology 500 ctDNAのサンプルシートを作成してください。

! サンプルシートの基準を満たしていない場合、解析に失敗します。

DRAGEN TruSight Oncology 500 ctDNAでは現在、現行のサンプルシート形式に代わるものとしてv2スタイルのサンプルシート形式をサポートしています。

以下のステップに従って、v2形式のDRAGEN TruSight Oncology 500 ctDNAのサンプルシートを作成してください。

1. [Header] セクションに、以下のパラメーターを入力します。
 - FileFormatVersion : 2
2. 設定セクションとデータセクションには、[3ページの「サンプルシートの作成」](#)に示したものと同じパラメーターを入力します。ただし、TSO500Lの接頭語が付きます（TSO500L_SettingsとTSO500L_Data）。
3. BCLファイルから解析を始める場合は、必ず以下のパラメーターを[BCLConvert_Data] セクションに含めてください。これらの各パラメーターの値が、[TSO500L_Settings] セクションのパラメーターと一致している必要があります。
 - AdapterRead1
 - AdapterRead2
 - OverrideCycles
4. BCLファイルから解析を始める場合は、必ず以下のパラメーターを[BCLConvert_Settings] セクションに含めてください。これらの各パラメーターの値が、[TSO500L_Data] セクションのパラメーターと一致している必要があります。
 - Sample_ID
 - Index
 - Index2
 - Lane (オプション)

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 ctDNA 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
DRAGEN TruSight Oncology 500 ctDNA Analysis Software SampleSheet.csv	解析を実行するためのセットアップテンプレート
『Illumina DRAGEN Computer Server Site Prep & Installation Guide』	DRAGENライセンスのアクティベーション手順を説明
『TruSight Oncology 500 ctDNA Reference Guide』 (文書番号：1000000092559)	TruSight Oncology 500 ctDNA Kitの使用に関する情報を示す
『TruSight Oncology 500 Reference Guide』 (文書番号：1000000067621)	TruSight Oncology 500 Kitの使用に関する情報を示す

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

DRAGEN TruSight Oncology 500 ctDNA Analysis Softwareのインストールには、root権限が必要です。

Dockerの検討事項

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 ctDNA Analysis Softwareをrootとして実行することも可能ですが、この方法は推奨しません。詳細については、Dockerのウェブサイト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、TSO500 DRAGENライセンスがご使用のサーバーで有効化されていることを確認します。イルミナカスタマーサービスから提供されたDRAGENライセンスのインストール手順と『Site Prep Guide』に従います。
2. 以下のいずれかのオプションを選択して、TSO500 DRAGENライセンスをインストールします。
 - ご使用のサーバーがインターネットに接続されている場合は、以下のコマンドを実行します。

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

- ご使用のサーバーがインターネットに接続されていない場合は、イルミナカスタマーサポートがtso500.binファイルを提供します。tso500.binファイルをサーバーの一時的な場所（例えば、/tmp）に保存してから、以下のコマンドを実行します。

```
sudo /opt/edico/bin/dragen_lic -i /tmp/license_TSO.bin -f tso500
```

i | コマンド `dragen_lic` をいずれの時点で使用しても、インストールした DRAGEN ライセンスが表示されます。

3. Dockerバージョン18.09以降をダウンロードしてインストールします。
4. イルミナから届く電子メールに記載されたリンクからインストールスクリプトをダウンロードします。72時間後にリンクの有効期限が切れます。
5. /stagingディレクトリにDRAGEN_TSO500_ctDNA-<version>.runスクリプトを保存します。このスクリプトでは約63 GBの空き領域を必要とするため、イルミナでは、/stagingディレクトリに自己解凍インストールスクリプトを保存し、そこから実行することを推奨します。
6. 以下のコマンドを用いてランスクリプトの許可を更新します。

```
sudo chmod +x /staging/DRAGEN_TSO500_ctDNA-<version>.run
```

7. 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。

```
sudo TMPDIR=/staging /staging/DRAGEN_TSO500_ctDNA-<version>.run
```

セルフテストスクリプトの実行

セルフテストスクリプトでは以下のテストを実施します。約15分かかります。

- 要求されたすべてのサービスが動作しているか確認します。
 - 正しいリソースが所定の位置にあるか確認します。
 - 正しいdockerイメージがインストールされているか確認します。
 - TruSight Oncology 500 ctDNA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確認します。
1. インストールスクリプトが終了したことを確認したら、以下のセルフテストスクリプトを実行します。

```
./usr/local/bin/test_DRAGEN_TSO500_ctDNA-<version>.sh
```

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/tmp/test_DRAGEN_TSO500_ctDNA-<timestamp>.tgzにある出力ファイルを提供してください。

2. **(オプション)** セルフテストスクリプトを実行できるのは、ユーザー1名だけです。複数のユーザーがセルフテストスクリプトを実行するためには、以下のいずれかのオプションを実行します。
 - /staging/tmpディレクトリを削除します。
 - 以下のコマンドを実行して、すべてのユーザーが/staging/tmpディレクトリに書き込むことを許可します。

```
chmod 777 /staging/tmp
```

DRAGEN TruSight Oncology 500 ctDNA Analysis Softwareの実行

DRAGEN TruSight Oncology 500 ctDNA Analysis Softwareのソフトウェアは、/usr/local/binディレクトリにインストールされているTSO500_ctDNA.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かします。引数については、[9ページの「シェルスクリプトのコマンドライン引数」](#)を参照してください。解析の出力は、/staging/DRAGEN_TSO500_ctDNA_Analysis_<timestamp>ディレクトリに自動的に書き込まれます。このフォルダーの場所によって、DRAGEN Bio-ITサーバー関連のプロセスで、最適化されたNVMe SSDにデータを読み書きすることが可能になります。

イルミナのシーケンスランフォルダー内にあるBCLファイルから、またはFASTQフォルダー内にステージされたFASTQファイルから開始できます。いずれの方法にも以下の要件が適用されます。

- シーケンスシステムのランフォルダーへのフルパス<FULL_PATH_TO_RUN_FOLDER>の場所を、`--runFolder`コマンドを使用して選択します。
- サンプルシートがランフォルダー内にあり、`SampleSheet.csv`という名前が付けられています。サンプルシートがランフォルダー内に存在しない、または`SampleSheet.csv`という名前ではない場合は、`--sampleSheet`上書きコマンドを使用してサンプルシートへのフルパスを指定します。

❗ 解析の実行中にファイルを移動しないでください。解析中にファイルを移動すると、解析に失敗する場合や不正確な結果が出力される場合があります。

❗ DRAGENステップ中にCtrl+Cを押すと、現在実行中の解析が停止します。なお、解析の実行中にCtrl+Cを押すとFPGAエラーになるおそれがあります。FPGAエラーを解消するには、サーバーをシャットダウンして再起動します。

- 以下の引数を使用して、BCLファイルからDRAGEN TruSight Oncology 500 ctDNA Analysis Softwareを開始します。

```
TSO500_ctDNA.sh --runFolder <FULL_PATH_TO_RUN_FOLDER>
```

- 以下の引数を使用して、FASTQファイルからDRAGEN TruSight Oncology 500 ctDNA Analysis Softwareを開始します。

```
TSO500_ctDNA.sh --runFolder <FULL_PATH_TO_RUN_FOLDER> --fastqFolder <FULL_PATH_TO_FASTQ_FOLDER>
```

BCLファイルから開始

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ファイルを使用してDRAGEN TruSight Oncology 500 ctDNA Analysis Softwareを実行するためには、ラン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ます。

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	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	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存在しないサンプルシートを使用する場合は、フルパスを指定
CopyComplete.txtファイル	シーケンスランが終了し、データのコピーが完了していることを示す

1. 以下の引数を使用して、BCLファイルからDRAGEN TruSight Oncology 500 ctDNA Analysis Softwareを実行します。

```
DRAGEN_TSO500_ctDNA.sh --runFolder <FULL_PATH_TO_RUN_FOLDER>
```

FASTQファイルから開始

FASTQファイルを使用してDRAGEN TruSight Oncology 500 ctDNA Analysis Softwareを実行するためには、以下の入力が必要です。

- 既存のFASTQフォルダー<FULL_PATH_TO_FASTQ_FOLDER>の場所へのフルパス。
- FASTQファイルは8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。
- 既存のランフォルダーへのフルパス。

1. 以下の引数を使用して、DRAGEN TruSight Oncology 500 ctDNA Analysis Softwareを実行します。

```
DRAGEN_TSO500_ctDNA.sh --runFolder <FULL_PATH_TO_RUN_FOLDER> --fastqFolder <FULL_PATH_TO_FASTQ_FOLDER>
```

FASTQファイルの要件

FASTQファイルから開始する場合は、FASTQファイル編成とサンプルシートの検討事項に関する推奨内容に従います。

FASTQファイル編成

各SAMPLE_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と一緒に入れます。

```
${full_path_to_fastqfolder}/${Sample_ID}/${Sample_ID}_Sample#_Lane#_Read#_001.fastq.gz
```

DRAGEN TruSight Oncology 500 ctDNA 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。

NovaSeq 6000シーケンスシステムでは、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。

1サンプルに対して出力されるファイルが8個となる例については、20ページの「フォルダー」を参照してください。

FASTQ向けサンプルシートの検討事項

FASTQのファイル名には固有のサンプルIDが必要です。FASTQファイルには必ずUMIを含めます。

シェルスクリプトのコマンドライン引数

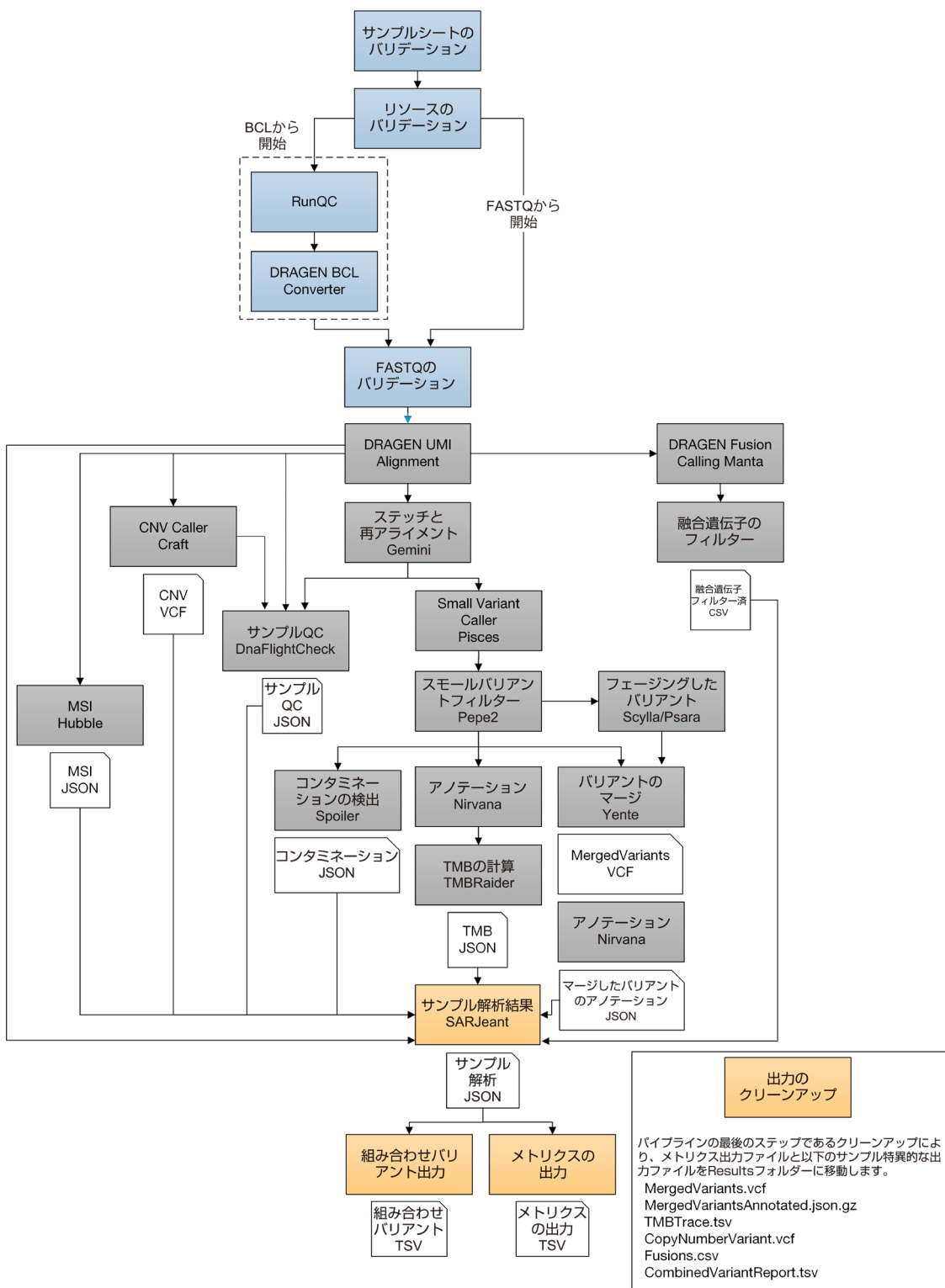
引数	内容説明
--runFolder	(必須) SampleSheet.csv、RunInfo.xml、RunParameters.xml、および Data ディレクトリを含むシーケンスシステムのランフォルダーへのフルパス
--analysisFolder	(オプション) 代替となる解析フォルダーへのフルパス。高い性能を発揮するために、このフォルダーはNVMe SSDパーティションに必須
--fastqFolder	(オプション) 既存のFASTQフォルダーへのフルパス。--runFolder コマンドが、この引数で要求される
--sampleSheet	(オプション) SampleSheet.csvファイルへのフルパス。サンプルシート名がSampleSheet.csvでない場合は、このコマンドが必須
--sampleIDs	(オプション) ランで処理されるカンマ区切りのサンプルIDを指定する。例えば、「Sample_1,Sample_2」となる
--version	(オプション) ソフトウェアのバージョンを表示してから終了する

解析方法

DRAGEN TruSight Oncology 500 ctDNAの解析モジュールでは、DRAGEN TruSight Oncology 500 ctDNAパネル内のがん関連遺伝子のバリエーションをターゲットにしてctDNAライブラリーを評価します。

DRAGEN TruSight Oncology 500 ctDNA Analysis Softwareワークフローでは、以下のセクション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。各解析ステップのソフトウェアコンポーネントとファイルの出力に関する詳細については、[21ページの「ファイル」](#)を参照してください。

図 1 DRAGEN TruSight Oncology 500 ctDNAのワークフロー



入力のバリデーション

要求される解析の入力が存在すること、およびその形式の妥当性を確認します。

FASTQファイルの生成

BCLファイルのデマルチプレックスを実施し、ソフトウェアによりFASTQ形式の中間解析ファイルが生成されます。

FASTQファイルではテキスト形式を用いてシーケンスを表示します。このファイルには、各サンプルのシーケンスリードと関連するクオリティスコアが含まれます。ランに使用したコントロールとフィルターをパスしなかったクラスターは、ファイルから除外されます。各FASTQファイルには、1サンプルのみのリードが含まれ、FASTQファイル名にはそのサンプルの名前が含まれます。

各リードの分子バーコード (UMI) シーケンスは、FASTQの生成ステップ中に取り出され、read1とread2を組み合わせたものがシーケンス識別子の末尾に付け加えられます。

DRAGEN Bio-IT BCL変換ソフトウェアでは、解析モジュールにより生成されたSampleSheet.csvをデマルチプレックスに使用します。

詳細については、[8ページの「FASTQファイルの要件」](#)を参照してください。

FASTQのバリデーション

解析ステップではありませんが、DRAGEN TruSight Oncology 500 ctDNA Analysis Softwareは、FASTQファイルの名前が適切であり固有のものであること、サンプルシートに記載された各サンプルに対応するファイルが存在すること、UMIシーケンスが有効であることを確認します。

リソースの検証

解析ステップではありませんが、このソフトウェアは、リソースフォルダー内に要求される解析の入力が存在すること、およびその形式を確認します。リソースフォルダーには以下のファイルが含まれます。

- マニフェスト
- hg19_decoyリファレンスゲノム
- 複合バリエーションのリファレンスシーケンス
- バリエーションコールおよびフィルタリングに関するベースライン定義
- 解析ステップに必要なデータベース
- リファレンス用のSampleSheet.csvテンプレート
- サニティーチェックテストを実行するためのテストデータ

DNAアライメントとリードCollapsing

アライメントステップでは、UMI Collapsing機能を搭載したDRAGEN Alignerを使用してFASTQファイル内のDNA配列とhg19_decoyゲノムをアライメントします。このステップでは、ゲノム上の位置とUMIタグを基にグループ化したリードのセット (ファミリー) を代表配列にまとめます。この処理によって、非常に頻度の低い (1%未満) 配列バリエーションのシグナルが失われることなく、重複するリードやシーケンスエラーを正確に除去できます。

このアライメントステップでは、アライメントフォルダーに保存される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。

リードCollapsingでは以下のBAMタグを追加します。

- RX/XU : UMIの組み合わせ。RXを、XUから複製して、BAM/SAM形式に当てはめます。
- XV : 1本のストランド上の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ソフトウェアコンポーネントでは、ローカルIndel再アライメント、ペアリードステッチング、およびリードのフィルターを実施します。ステッチしたリードは、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Indelコールの感度を高めます。Geminiでは、1個のBAMファイルとアライメントに使用したリファレンスゲノム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。

このステップでは以下のBAMタグを追加します。

- XD : 正常にステッチしたリードのみの場合。方向性をサポートするストリングで、順方向、逆方向、ステッチした位置を示します。
- XR : 正常にステッチしたリードのみの場合。リードペアの方向は、順-逆 (FR) または逆-順 (RF) のいずれかです。

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サンプル内のバリエーション候補を同定します。Piscesは、Indel再アライメントとリードステッチング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候補を検出します。

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。最初の6値は変異体サポートを示します。次の6値は合計サポートを示します。

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

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。最終のバリエントコールは、13ページの「smallバリエントフィルタリング」で判定します。

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バリエントコールゲノムの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1,000です。VAFの検出下限は、最小リード深度で0.05%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

表 1 smallバリエントフィルタリングのフィルター情報

ALT	FILTER	注記
.	PASS	WT
., A, C, G, etc ¹	LowDP ²	深度1,000未満のリファレンスまたはフィルターしたバリエント候補
A, C, G, etc ¹	PASS	PASSバリエント
A, C, G, etc ¹	LowSupport、 LowVarSupport	低クオリティスコアでフィルターしたバリエント候補
A, C, G, etc ¹	Blacklist	高いバックグラウンドノイズを伴う位置。バリエント検出には利用不可
A, C, G, etc ¹	VarBias	フィルターしたバリエント候補で、フラグメントの末端でクラスター化した偏りを示す

¹ etclは、表中に記載していない他のバリエントタイプを示します。

² リファレンス位置では、1,000X未満のカバレッジが直接LowDPとなるものの、その位置のバリエントコールがフィルターPASSである場合、LowDPは適用されません。これは、COSMICに登録済みまたは未登録のバリエントのLQ/AQ閾値とアリル頻度により左右されます。

リードCollapsingおよびTMB解析ステップの正味の影響により、標準的なセルフリーDNAサンプルで偽陽性が約1,500/Mbから5/Mb未満に減少します。

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を超えるバリエーションの場合、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に未登録のバリエーションでは、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アレルの位置情報がバリエーション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以上のバリエーションをフィルターし、VarBiasのラベルを付けます。リードCollapsingおよびバリエーション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。

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タイプが原因でカバレッジの偏りが生じます。正常なセルフリーDNAと人工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。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。増幅と欠失は、VCFファイル内にそれぞれDUPとDELとしてアノテーションされます。

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。このワークフローでは、Scyllaを使用して、具体的に、臨床的に重要なEGFRエクソン19の変異体とRETの変異体を検出します。これらは、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。

Scyllaでは、Piscesとは異なり、同一のリード上にあるバリエーションがフェージングされている必要がありません。フェージングが終了すると、新しいVCFが生成されます。

バリエントのマージ

ソフトウェアで、フェージングしたバリエントと、スモールバリエント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ントを統合します。このプロセスでは、染色体、位置、RefおよびAltと一致する完全な重複が除去されます。

フェーズドバリエント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）バリエントのみが追加されます。他のすべてのEGFRバリエントは、バリエントのマージ中に除去されます。

表 2 EGFR バリエント

染色体	位置	リファレンスアリル	代替アリル
chr7	55242482	CATCTCCGAAAGCCAACAAGGAAAT	C
chr7	55242466	GAATTAAGAGAAGCAACAT	G
chr7	55242465	GGAATTAAGAGAAG	AATTC
chr7	55242465	GGAATTAAGAGAAGCAAC	AAT
chr7	55242469	TTAAGAGAAGCAACATCTC	T
chr7	55242467	AATTAAGAGAAGCAACATC	A
chr7	55242469	TTAAGAGAAG	C
chr7	55242467	AATTAAGAGAAGCAACATC	T
chr7	55242465	GGAATTAAGA	G
chr7	55242467	AATTAAGAGAAGCAACATCTC	TCT
chr7	55242467	AATTAAGAGAAGCAAC	T
chr7	55242464	AGGAATTAAGAGAAGC	A
chr7	55242466	GAATTAAGAGAAGCAA	G
chr7	55242464	AGGAATTAAGAGA	A
chr7	55242469	TTAAGAGAAGCAA	T
chr7	55242465	GGAATTAAGAGAAGCAACATC	AAT
chr7	55242469	TTAAGAGAAGCAACATCT	CAA
chr7	55242463	AAGGAATTAAGAGAAG	A
chr7	55242468	ATTAAGAGAAGCAACATCT	A
chr7	55242462	CAAGGAATTAAGAGAA	C
chr7	55242465	GGAATTAAGAGAAGCAA	AATTC
chr7	55242469	TTAAGAGAAGCAA	C
chr7	55242467	AATTAAGAGAAGCAAC	A

染色体	位置	リファレンスアリル	代替アリル
chr7	55242469	TTAAGAGAAGCAACATCTCC	CA
chr7	55242468	ATTAAGAGAAG	GC
chr7	55242465	GGAATTAAGAGAAGCA	G
chr7	55242468	ATTAAGAGAAGCAAC	GCA
chr7	55242465	GGAATTAAGAGAAGCAACA	G
chr7	55249011	AC	CCAGCGTGGAT

エクソン11、13および15内のRET遺伝子の領域は、フェーズドバリエントコールで検出された場合に評価され、追加されます。

アノテーション

Illumina Annotation Engine (Nirvana) でスモールバリエント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。ファイルは、スモールバリエントのフィルター解析ステップ後にアノテーションされ、さらにバリエントのマージステップ後にもアノテーションされます。

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ントのJSONファイルから、さらにスモールバリエントのフィルター解析ステップで生成されたVCFファイルから、TMBメトリクスを生成します。TMBの計算から生殖細胞系列バリエントを除外するために、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ント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。データベースフィルタリングでは、GnomAD exome、genome、および1000 genomesの各データベースを使用します。

以下の式を用いてTMBを算出します。

$$TMB = \text{Eligible Variants} \div \text{Effective Panel Size}$$

TMBの計算に適切なバリエントと有効なパネルサイズを、[表 3](#)で詳細に示します。

表 3 TMB の計算の詳細

計算値	内容説明
適格なバリエント (分子)	<ul style="list-style-type: none"> コーディング領域のバリエント (RefSeq Cds) バリエント頻度が0.2%以上40%未満 カバレッジが1,000X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を超えるバリエントを除外 TET2、TP53、DNMT3AおよびCBLの変異体を除外
有効なパネルサイズ (分母)	<ul style="list-style-type: none"> カバレッジが1,000X以上の総コーディング領域

出力は、TMBの計算に使用したバリエントを含む*_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。

最大体細胞VAF

最大体細胞VAFは、ctDNA中の腫瘍割合として代用されます。TMBステップでは、データベースとVAF情報を用いて、潜在的な体細胞ステータス別にバリエントをフラグ付けします。残りのバリエントについては、TP53、DNMT3A、CBL、TET2の潜在的なクローン性造血変異体を除去します。残りのバリエントにおける4番目のバリエントアリル頻度が、サンプルごとのMaxSomaticVAFとしてレポートされます。

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、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マイクロサテライト不安定性のエビデンスがあるか評価します。これらはJensen-Shannon (JS) 距離 (情報エントロピーメトリクス) に基づきます。パネルには、6、7個の1塩基リピートを持つ、選択したMSI Siteが合計で2,408個あります。500個以上のスパンニング重複除去 (spanning duplex collapsed) リードを持つMSI Siteの場合、テストサンプルとベースライン正常サンプルの比較を用い、さらに任意のベースライン正常サンプル2個を用いてJS距離を計算します。テストサンプルとベースライン正常サンプルの比較において、JS距離が有意に大きい場合 (p 値 ≤ 0.01 、JS距離の差 ≥ 0.02)、そのSiteは不安定とみなされます。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。最終的なMSIスコアは、全不安定性Siteにわたる全JS距離を集計したものです。入力はDNAアライメントとリードCollapsingステップからのBAMファイルで、出力はMSIメトリクスファイルです。

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のフィルターステップ中に生成されたVCFファイル内にある外来のヒトDNAのコンタミネーションを検出します。ソフトウェアで、コンタミネーションの p 値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。コンタミネーションのp値は、コピー数イベントにより生じた予想外に低いアリル頻度を持つサンプルを識別することにより、ゲノム全体での低頻度SNPの均一性を評価します。

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%の期待値から外れています。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、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
- 変異サポートのカウント
- 合計深度のカウント

結果

MetricsOutput.tsvファイルには、全サンプルの品質管理メトリクスであるRunQc解析ステータス、コンタミネーション、スモールバリエーションコール用のQCメトリクス、TMB、MSI、CNV、融合遺伝子が含まれます。また、MetricsOutput.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。

CombinedVariantOutput.tsvファイルは、サンプルごとに作成されます。このファイルには、解析の詳細、シーケンスランの詳細、TMB、MSI、スモールバリエーション、遺伝子コピー数バリエーション、および融合遺伝子が含まれます。

融合遺伝子コール

DRAGEN Mantaで融合遺伝子コールを実施します。入力はBAMファイルであり、出力はVCFファイルです。

候補となる融合遺伝子コールを、イルミナが開発した構造多型検出法であるMantaを使用して生成します。Mantaでは、最初にゲノムをスキャンして、スプリットリードとスパニングリードを基に構造多型（SV）や大きなIndelと考えられるエビデンスを検出します。このエビデンスを、ブレイクエンドに関連する可能性があるゲノムの全領域を結び付けたエッジを持つグラフに示します。次に、Mantaで個々のグラフのエッジを解析し、エッジに関連するSVを検出してスコア付けします。このプロセスのサブステップには、以下の項目が含まれます。

- エッジに関連するSV候補の推定
- SVブレイクエンドのアセンブリの試み
- SVのスコアリング、ジェノタイプングおよびフィルタリング
- VCFへの出力

融合遺伝子フィルタリング

DRAGEN TruSight Oncology 500 ctDNAの解析モジュールでは、Mantaによって数百～数千の融合遺伝子候補が1サンプル内でコールされます。大多数の融合遺伝子候補（約99%）は偽陽性です。

融合遺伝子フィルタリングツールのDNA Fusion Filter (DNAFF) が、真の融合遺伝子コールを偽陽性と区別します。DNAFFフィルターは、以下の機能を果たします。

- サポートリードペアが1個のみの融合遺伝子やシーケンスエラーが起これやすいリピート領域と重複する融合遺伝子などの偽の融合遺伝子を除去します。
- 以下の基準に基づき、不確かなサポートリードをすべての融合遺伝子候補に対しフィルターします。
 - 融合遺伝子コンティグとの配列同一性が低いリードをフィルターします。
 - 融合遺伝子をサポートしないリードである誤ったリードをフィルターします。例えば、疑わしい補足的アライメントがあるリードです。
 - UMI情報を基に、リードの重複を除去します。
- 最終の融合遺伝子出力には以下の規則が適用されます。
 - 少なくとも1つの融合遺伝子ブレイクポイントが23個のターゲット遺伝子の範囲内に入ること。
 - 融合遺伝子ペアがCOSMICでレポートされている場合、2個以上のサポートリードがあること。
 - 融合遺伝子ペアがCOSMICでレポートされていない場合、3個以上のサポートリードがあること。

品質管理 (QC)

DRAGEN TruSight Oncology 500 ctDNA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。Read 1とRead 2に関し、レポートでは、クオリティスコア (Qスコア) 測定値であるQ30以上の塩基の平均割合を示します。Qスコアは不適切なベースコールの確率の予測値です。

DNA QCメトリクス

このステップでは、DNAサンプルのメトリクスを算出しレポートを作成します。入力は、DNAアライメントとリードCollapsingステップからのBAMファイル、Indel再アライメントとリードステッチングステップからのBAMファイル、およびCRAFTでノーマライズしたBinCount.tsvファイルです。出力はJSONファイルです。

解析の出力

解析が完了すると、DRAGEN TruSight Oncology 500 ctDNA Analysis Softwareでは、TSO500_ctDNA_Analysis_YYYYMMDD-HHMMSSのフォルダー名で、ユーザーが指定した場所に解析の出力が生成されます。

各解析ステップで、解析フォルダー内のLogs_Intermediatesフォルダー内にサブフォルダーが作成されます。

解析の出力を表示するには、以下の手順に従います。

1. 解析の出力フォルダーが含まれるディレクトリに移動します。
2. フォルダーを開き、表示したいファイルを選択します。

フォルダー

このセクションでは、解析中に生成された各出力フォルダーを図で示します。

 **Results** : パイプラインによる最終の結果ファイルが含まれます。


 MetricsOutput.tsv: 全サンプルのサマリーメトリクスが含まれます。

 **Sample1**


 Sample1_CombinedVariantOutput.tsv: 検出したすべてのスモールバリエーションとバイオマーカーを1つのファイルにしたTSVが含まれます。


 Sample1_MergedSmallVariantsAnnotated.json.gz: フェージングしたバリエーションを含む、すべてのスモールバリエーションに関する情報が含まれます。

 Sample1_CopyNumberVariants.vcf: コピー数バリエーションを記載したVCFファイルが含まれます。

 Sample1_MergedSmallVariants.genome.vcf: フェージングしたバリエーションを含む、すべてのスモールバリエーションに関する情報を記載したVCFファイルが含まれます。

 Sample1_Fusions.csv: すべてのDNA融合遺伝子が含まれます。

 Sample1_TMB_Trace.tsv: TMB値の計算方法に関する情報を記載したトレースファイルが含まれます。

 **Logs_Intermediates** : パイプラインの各ステップの全中間ファイルが含まれます。各出力フォルダーには、実行した全ステップ、出力ログおよび中間ファイルを記載した dsdm.json ファイルが含まれます。

 **AlignCollapseFusionCaller**

 **Annotation**

 **Cleanup**

 **CnvCaller**

 **CombinedVariantOutput**

 **Contamination**

 **DnaFusionFiltering**

 **DnaQCMetrics**

 **FastqGeneration (アプリを BCL から開始する場合のみ)**

 **FastqValidation**

 **MergedAnnotation**

 **MetricsOutput**

 **Msi**

- 📁 PhasedVariants
- 📁 ResourceVerification
- 📁 RunQc (アプリを BCL から開始する場合のみ)
- 📁 SampleAnalysisResults
- 📁 SamplesheetValidation
- 📁 SmallVariantFilter
- 📁 StitchedRealigned
- 📁 Tmb
- 📁 VariantCaller
- 📁 VariantMatching

Logs_Intermediates内のすべてのログは、実行中のDockerコンテナから生成されます。実行中のDockerコンテナ（例えば、ランフォルダー、サンプルシート、FASTQフォルダー）への入力は、サーバー上にある本来の場所からコンテナ内の以下の場所にまで及びます。

入力	実行中の Docker コンテナの場所
ランフォルダー	/opt/illumina/run-folder
サンプルシート	/opt/illumina/SampleSheet.csv
FASTQフォルダー	/opt/illumina/fastq-folder
リソース	/opt/illumina/resources
解析の出力フォルダー	/opt/illumina/analysis-folder

ログメッセージ内のパスは、サーバーのパスではなく、実行中のコンテナ内のパスを示します。

ファイル

このセクションでは、解析中に生成されたサマリー出力ファイルを示します。

メトリクスの出力

ファイル名：MetricsOutput.tsv

メトリクス出力ファイルは、*.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。

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2つの場合に、[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

サンプルQCメトリクス

サンプル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	バリエントクラス
CONTAMINATION_SCORE	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スコア ≤ 1672	すべて
および		または	
CONTAMINATION_P_VALUE	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の評価に用いられる。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p値は、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	コンタミネーションスコア > 1672 および コンタミネーションp値 ≤ 0.049	

メトリクス	内容説明	推奨閾値	バリエントクラス
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	≥ 1300	スモールバリエント、TMB、融合遺伝子、MSI
PCT_EXON_1000X	1,000Xフラグメントカバレッジを持つエクソン塩基の割合	≥ 80.0	スモールバリエント、TMB
GENE_SCALED_MAD	遺伝子の倍率変化でノーマライズした絶対偏差の中央値	≤ 0.059*	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	≥ 6.0	CNV

* GENE_SCALED_MADでの推奨閾値0.059は、実際のセルフリーDNAにのみ適用されます。

マージしたバリエントゲノムVCF

ファイル名：{SAMPLE_ID}_MergedSmallVariants.genome.vcf

マージしたバリエントゲノムVCFは、スモールバリエントゲノムVCFとEGFR複合バリエントVCFを統合したものです。

スモールバリエントゲノムVCFのファイル名は、SmallVariantFilter/{sample_ID}/{sample_ID}_SmallVariants.genome.VarPosFilter.vcfです。Pepeが生成したスモールバリエントゲノムVCFには、すべてのターゲット化した区間（左詰で25 bpまで）のバリエントコールステータスが含まれます。

EGFR複合バリエントVCFのファイル名は、PhasedVariants /{sample_ID}/{sample_ID}.Complex.vcfです。Scyllaが生成したEGFR複合バリエントVCFには、フェージングしたEGFRバリエントが含まれます。ゲノムVCFの [FILTER] 列から、バリエントステータスを判定します。詳細については、表 4を参照してください。

表 4 ゲノム VCF の [FILTER] 情報

ALT	FILTER	注記
.	PASS	WT
., A, C, G, etc	LowDP	コールなし (DP < 1000)
A, C, G, etc	PASS	PASSバリエント

ALT	FILTER	注記
A, C, G, etc	LowSupport	以下の条件でフィルターしたバリエント候補 <ul style="list-style-type: none"> • AQ、LQまたはBFQの閾値に未達 または • 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ッチしたサポートがゼロ
A, C, G, etc	LowVarSupport	フィルターしたバリエント候補（変異体サポート < 閾値）
A, C, G, etc	Blacklist	高いバックグラウンドノイズを伴う位置。バリエント検出には利用不可
A, C, G, etc	VarBias	フィルターしたバリエント候補で、フラグメントの末端でクラスタ化された偏りを示す

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

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。

バリエントの結果の定義は、Sequence Ontologyのウェブサイトに掲載されています。

TMB JSON

ファイル名：{Sample_ID}.tmb.json

TMB jsonファイルには、DNAサンプルごとの腫瘍変異負荷メトリクスが含まれます。イルミナでは、TMBの計算において分子の一部に同義変異と非同義変異の両方を表すTmbPerMbの使用を推奨します。

メトリクス	内容説明
TmbPerMb	結果に関係なく、適格な全バリエントを用いて判定されるTMB値
AdjustedTmbPerMb	バリエントコールを信頼するために十分なカバレッジを持つパネル内の位置の総数に対して、偏りで補正したTMB値
NonsynonymousTmbPerMb	バリエントコールを信頼するために十分なカバレッジを持つ、パネル内の非同義変異位置のTMB値
AdjustedNonsynonymousTmbPerMb	バリエントコールを信頼するために十分なカバレッジを持つ、パネル内の非同義変異位置に対して、偏りで補正したTMB値
TotalRegionSizeMb	パネルの合計サイズ（Mb）
CodingRegionSizeMb	コーディング領域のパネルサイズ（Mb）

TMBトレース

TMBの計算のみに限定されます。このファイルをバリエントを解釈する目的で使用しないでください。

ファイル名：{Sample_ID}_TMB_Trace.tsv

TMBトレースファイルは、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あり、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	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）。領域が複数の遺伝子と重複する場合、遺伝子名を文字「:」で連結する
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、これらを文字「;」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	exome、genome、genomes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
IncludedInTMBNumerator	バリエントがTMBの計算で使用される場合、TRUE

Sample Analysis Results Json

ファイル名：{SampleID}_SampleAnalysisResults.json

Sample Analysis Results Json (SARJ) ファイルは、サンプルごとに作成された結果ファイルを集計したものです。SARJファイルは、ダウンストリーム出力の生成に用いられます。このファイルには、フィルターをパスしたバリエーションとパスしなかったバリエーションのアノテーションが含まれています。

コピー数バリエーションVCF

ファイル名：{Sample_ID}_CopyNumberVariants.vcf

このファイルには、DRAGEN TruSight Oncology 500 ctDNAパネルでターゲット化したCNV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。

コピー数バリエーションVCFの [QUAL] 列にある値は、以下の式で表されるp値のPhred変換値です。

$$Q = -10x \log_{10}(p\text{-value})$$

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、(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。

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。

$$n = \frac{[(200 \times Y) - 2 \times (100 - X)]}{X}$$

例えば、腫瘍純度30%のテストサンプルにおいて、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。

マージしたスモールバリエントVCF

ファイル名：{Sample1}_MergedSmallVariants.genome.vcf

このファイルには、フェージングしたバリエントと他のすべてのスモールバリエントの両方が含まれます。このマージしたVCFには、フェージングしたバリエント（複合）VCFとスモールバリエントVCFの両方からのヘッダーセクションが含まれます。フェージングしたバリエントとスモールバリエントの両方に認められるバリエントは、フェージングしたバリエントとしてのみ表示されます。

すべてのフィルターをパスしたバリエントコールは、イルミナのAnnotation Engineを使用し、以下の情報によりアノテーションされます（RefSeq transcriptを使用）。

- HGNC遺伝子
- 転写産物
- エクソン
- バリエントの結果
- HGVSコーディングシーケンス名
- HGVSタンパク質シーケンス名
- COSMIC ID

融合遺伝子CSV

ファイル名：{Sample_ID}_Fusions.csv

このファイルには、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。

融合遺伝子の列を以下の表に記載します。Microsoft Excelを使用してこのファイルを表示する場合、日付に変換できる遺伝子（例えば、MARCH1）が、自動的にdd-mm形式（1-Mar）に変換されます。

融合遺伝子のオブジェクトフィールド	内容説明
Sample	入力サンプルID
Name	融合遺伝子名
Chr1	最初のブレイクポイントの染色体
Pos1	最初のブレイクポイントから1塩基上流の位置
Chr 2	2番目のブレイクポイントの染色体
Pos2	2番目のブレイクポイントから1塩基上流の位置
Direction	融合遺伝子片の方向性
Alt	融合遺伝子コールをサポートするリード数

融合遺伝子のオブジェクトフィールド	内容説明
Depth	融合遺伝子ブレイクポイントでのリード総数
VAF	バリエーションアリル頻度
Gene1	融合遺伝子の開始点に関連する遺伝子（セミコロンで区切られる）
Gene2	融合遺伝子の終結点に関連する遺伝子（セミコロンで区切られる）
Contig	融合遺伝子の配列
Filter	融合遺伝子がすべての融合遺伝子フィルターをパスしたかどうかを示す
Is_Cosmic_GenePair	遺伝子ペアがCOSMICでレポートされているかどうかを示す（True/False）

以下の表に、方向列の値の意味を示します。値は、Samtoolsで用いられる形式です。

方向	VCF形式	内容説明
L1R2	t p	breakend1の左側とbreakend2の右側が結合
L1rL2	t p	breakend1の左側とbreakend2の左側の逆相補鎖が結合
L2R1]p t	breakend2の左側とbreakend1の右側が結合
rR2R1]p t	breakend2の右側の逆相補鎖とbreakend1の右側が結合

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、ペアサンプルを基に（PairIDを使用する場合）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。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）
- コピー数バリエーション
- TMB
- MSI
- 融合遺伝子

解析の詳細

以下の情報は、組み合わせバリエーション出力解析の詳細セクションにあります。

- サンプルID
- 出力日
- 出力時間
- 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）

シーケンスランの詳細

解析をランフォルダーから開始する場合、以下の情報は、組み合わせバリエーション出力にあります。

- ラン名
- ラン日
- サンプルインデックスID
- 装置ID
- 装置制御ソフトウェアのバージョン
- 装置タイプ
- RTAバージョン
- SBS試薬カートリッジのロット番号
- クラスター試薬カートリッジのロット番号

バリエーションのフィルターの規則

- 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 Assayでターゲット化されていない重複遺伝子において、バリエーションが正規の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ファイルでiSNP、Indel、またはFlankingに指定された領域に位置する場合。
- **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、正規のRefSeq transcript（MergedSmallVariantsAnnotated.jsonに記録）がある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 Whitelist-スモールバリエーション内にある正規の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 Whitelist-スモールバリエーション内にある正規の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>または。

- **融合遺伝子バリエント**：融合遺伝子バリエント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ントコール（KeepFusionフィールドがtrue）。
 - 融合遺伝子のホワイト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。
- **スプライスバリエント**：遺伝子EGFR、METおよびARIに含まれる、フィルターをパスしたスプライスバリエント。

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

DRAGEN TruSight Oncology 500 ctDNA Analysis Softwareのインストールには、`uninstall_TSO500_ctDNA-<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（`TSO500_ctDNA.sh`、`test_TSO500_ctDNA-<VERSION>.sh`、および`uninstall_TSO500_ctDNA-<VERSION>.sh`）
 - `/staging/illumina/TSO500_ctDNA`に存在するリソース
 - `tso500-ctdna:<VERSION>`のdockerイメージ
1. 以下のコマンドをrootとして実行し、DRAGEN TruSight Oncology 500 ctDNA Analysis Softwareをアンインストールします。

```
uninstall_TSO500_ctDNA-<version>.sh
```

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

改訂履歴

文書	日付	変更内容
文書番号： 1000000143376 v02	2021年9月	リファレンスコールの最小リード深度と、VAFの検出下限を修正。 EGFRバリエーションの表を追加。
文書番号： 1000000143376 v01	2021年1月	要求されるDRAGENバージョンを明記。
文書番号： 1000000143376 v00	2020年11月	初版リリース。

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

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 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。



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。

© 2021 Illumina, Inc. All rights reserved.

illumina®