

illumina®

# TruSight Whole Genome Analysis Application

Documentation relative au produit

PROPRIÉTÉ D'ILLUMINA

Document n° 200049931 v00

Avril 2024

DESTINÉ AU DIAGNOSTIC IN VITRO UNIQUEMENT.

## Historique des modifications

Document	Date	Description de la modification
Document n° 200049931 v00	Avril 2024	Publication initiale.

Ce document et son contenu sont la propriété exclusive d'Illumina, Inc. et ses filiales (« Illumina »), et sont destinés à un usage contractuel de ses clients en lien avec l'utilisation du ou des produits décrits dans la présente et à aucune autre utilisation. Ce document et son contenu ne seront utilisés ou distribués à aucune autre fin et ne seront communiqués, divulgués ou reproduits d'aucune façon sans le consentement écrit préalable d'Illumina. Par le biais de ce document, Illumina ne cède aucune licence en vertu de son brevet, de sa marque de commerce, de son copyright ou de ses droits traditionnels ni des droits similaires d'un tiers quelconque.

Les instructions présentes dans ce document doivent être strictement et explicitement respectées par le personnel qualifié et correctement formé afin d'assurer une utilisation correcte et sécuritaire du ou des produits décrits dans la présente. Tout le contenu de ce document doit être entièrement lu et compris avant d'utiliser le ou les produits.

LE FAIT DE NE PAS LIRE ENTIÈREMENT ET DE NE PAS SUIVRE EXPLICITEMENT TOUTES LES INSTRUCTIONS CONTENUES DANS LA PRÉSENTE PEUT CAUSER DES DOMMAGES AU OU AUX PRODUITS, DES BLESSURES AUX PERSONNES, Y COMPRIS AUX UTILISATEURS OU À D'AUTRES PERSONNES, ET DES DOMMAGES À D'AUTRES BIENS, ET ANNULERA TOUTE GARANTIE APPLICABLE AU OU AUX PRODUITS.

ILLUMINA N'ASSUME AUCUNE RESPONSABILITÉ QUANT AUX DOMMAGES DÉCOULANT D'UNE MAUVAISE UTILISATION DU OU DES PRODUITS DÉCRITS DANS LA PRÉSENTE (Y COMPRIS LES PARTIES DE CELLE-CI OU LE LOGICIEL).

© 2024 Illumina, Inc. Tous droits réservés.

Toutes les marques sont la propriété d'Illumina, Inc. ou de leurs propriétaires respectifs. Pour en savoir plus sur les marques, consultez la page [www.illumina.com/company/legal.html](http://www.illumina.com/company/legal.html).

# Table des matières

Historique des modifications .....	ii
<b>Présentation .....</b>	<b>1</b>
Démarrage .....	1
Paramètres .....	2
<b>Création d'une série .....</b>	<b>4</b>
Révision de la série .....	5
Remise en file d'attente d'une analyse .....	6
<b>Afficher la série et les résultats .....</b>	<b>7</b>
<b>Résumé du fichier de sortie .....</b>	<b>9</b>
Informations concernant le rapport de CQ .....	9
Informations sur le fichier de définition de variant .....	13
Fichiers FASTQ .....	22
Fichiers CRAM .....	23
 <b>Assistance technique .....</b>	 <b>25</b>

# Présentation

TruSight Whole Genome Analysis Application est utilisé pour planifier les séries de séquençage pour TruSight Whole Genome et lancer automatiquement l'analyse une fois la série terminée. L'analyse comprend le démultiplexage, la génération FASTQ, la cartographie de lecture, l'alignement sur le génome de référence humain GrCh38/hg38 avec graphe et la désignation de variant à l'aide de Illumina DRAGEN Server for NovaSeq 6000Dx.

À différentes étapes du flux de travail d'analyse, l'application effectue un contrôle qualité (CQ) selon le séquençage défini, FASTQ et les mesures de la bibliothèque d'échantillons, et génère des rapports avec les résultats. Pour les échantillons qui réussissent toutes les étapes de CQ, l'application génère des fichiers de sortie compatibles pour une utilisation dans les applications germinales en aval.

Le TruSight Whole Genome Analysis Application exécute les définitions de variants DRAGEN, y compris la définition de petit variant, la définition de la variation du nombre de copies (CNV) et la détection d'expansion répétée avec ExpansionHunter.

L'application effectue également une annotation de niveau de confiance faible, intermédiaire ou élevé pour les petits variants et inclut cette annotation dans le fichier de sortie.

## Démarrage

Assurez-vous que le TruSight Whole Genome Analysis Application est installé sur l'instrument NovaSeq 6000Dx qui sera utilisé pour le séquençage dans le cadre de TruSight Whole Genome. Les applications installées se trouvent sur l'écran Applications du NovaSeq 6000Dx Instrument ou dans le Illumina Run Manager à l'aide d'un navigateur sur un ordinateur en réseau. Pour obtenir de l'aide pour planifier l'installation, contactez votre représentant Illumina local.

### Exigences en matière de stockage des données

Reportez-vous à Documentation produit NovaSeq 6000Dx (document n° 200010105) et Documentation produit du DRAGEN Server for NovaSeq 6000Dx (document n° 200014171) pour des informations générales sur la sortie des données et leur stockage.

Les sorties de données TruSight Whole Genome Analysis Application dans un dossier de série et un dossier d'analyse dans un stockage externe. Les exigences minimales de stockage peuvent être estimées à partir de la taille de sortie des données dans chaque dossier pour un seul séquençage présenté ci-dessous.

Configuration	Dossier de série (Go)	Dossier d'analyse (Go)
S2 Flow Cell (6 échantillons)	~430	~350
S4 Flow Cell (16 échantillons)	~1110	~890

## Temps d'analyse approximatif

L'analyse commence automatiquement après la fin d'une série de séquençage et se produit séquentiellement sur les échantillons d'une série. Les fichiers de sortie de données seront disponibles sur le stockage externe une fois l'analyse terminée pour tous les échantillons d'une série et le transfert de copie vers le stockage externe terminé. Au début d'une série de séquençage du côté A et du côté B en même temps, le séquençage sera effectué simultanément. L'analyse de ces séries de séquençage sera effectuée séquentiellement par le TruSight Whole Genome Analysis Application une fois le séquençage terminé. La série qui termine le séquençage et le transfert en premier sera analysée en premier. La deuxième série de séquençage sera transférée et mise en file d'attente pour analyse une fois la première analyse terminée. Reportez-vous à la section [Afficher la série et les résultats à la page 7](#) pour savoir comment déterminer l'état des séries actives ou échouées.

Le temps approximatif jusqu'à ce que les résultats de l'analyse soient disponibles après la fin du séquençage est indiqué ci-dessous pour la situation où les côtés A et B sont chargés simultanément avec la même configuration.

Configuration	Cycle d'analyse 1 (heures)	Cycle d'analyse 2 (heures)
S2 Flow Cell (6 échantillons)	~12	~24
S4 Flow Cell (16 échantillons)	~24	~48

## Paramètres

Sélectionnez le TruSight Whole Genome Analysis Application sur l'écran Applications pour afficher la configuration actuelle et modifier les autorisations.

### Configuration

L'écran de configuration affiche les paramètres d'application suivants :

- **Nom de l'application**
- **Version de l'application**
- **DRAGEN Version**
- **Version de RTA**
- **Date de publication**
- **Organisation**
- **Identifiant du dispositif**
- **Identifiant de production**
- **Trousses de préparation de bibliothèque** : Affiche la trousse de préparation de bibliothèque. Ce paramètre ne peut pas être modifié.

- **Kits d'adaptateurs d'index** : Affiche les jeux d'adaptateurs d'index disponibles pour utilisation.
- **Lectures d'index**
- **Type de lecture**
- **Longueurs d'index**
- **Longueurs de lecture** : Les longueurs de lecture sont définies par défaut lorsque l'ensemble d'index est sélectionné. Ce paramètre ne peut pas être modifié.

## Autorisations

L'administrateur désigné dispose d'un accès Permissions (Autorisations) et peut utiliser les cases à cocher de l'écran Permissions (Autorisations) pour gérer l'accès utilisateur pour le TruSight Whole Genome Analysis Application.

Pour plus d'informations sur les autorisations et la gestion des utilisateurs, reportez-vous à la section System Configuration (Configuration système) du Documentation produit NovaSeq 6000Dx (document n° 200010105).

# Création d'une série

Créez de nouvelles séries dans le mode IVD ou en accédant à Illumina Run Manager (IRM) à l'aide d'un navigateur sur un ordinateur en réseau. Pour accéder à l'instrument à distance, utilisez l'adresse et les informations de compte utilisateur fournies par votre représentant Illumina. Reportez-vous à Documentation produit NovaSeq 6000Dx (document n° 200010105), pour plus d'informations.

Create Run (Créer une série) est la méthode recommandée pour planifier une analyse. Import Sample Sheet (L'importation d'une Feuille d'échantillon) n'est pas recommandée. Les fichiers de feuille d'échantillon sortis dans les dossiers de série et d'analyse ne sont pas adaptés à l'importation pendant la planification de la série.

## Créer des séries

1. Sur l'écran Runs (Séries), sélectionnez **Create Run** (Créer une série).
2. Sélectionnez TruSight Whole Genome Analysis Application, puis sélectionnez **Next** (Suivant).
3. Sur l'écran Run Settings (Paramètres de la série), entrez un nom pour la série. Le nom de la série identifie la série depuis le séquençage jusqu'à l'analyse.
4. [Facultatif] Entrez une description de la série pour identifier davantage la série. Le kit de préparation de bibliothèque est défini par défaut comme TruSight Whole Genome et ne peut pas être modifié.
5. Sélectionnez TruSight Whole Genome index souhaité dans le menu déroulant **Index Adapter Kit** (Kit d'adaptateur d'index).

La longueur de lecture sera définie par défaut et ne peut pas être modifiée. (Les Lectures 1 et 2 utilisent 151 cycles ; les Index 1 et 2 utilisent 10 cycles).

6. Saisissez un ID de tube de bibliothèque (format recommandé DX1234567-LIB), puis sélectionnez **Next** (Suivant).

Si aucun identifiant de tube de bibliothèque n'est spécifié à cette étape, la série planifiée devra être sélectionnée avant le chargement des consommables de séquençage. Si un ID de tube de bibliothèque incorrect est saisi à cette étape, la série planifiée doit être corrigée avant de charger les consommables. Se reporter à [Révision de la série à la page 5](#) pour le protocole afin de corriger la série lorsque les consommables sont prêts à être chargés.

7. Dans les écrans Sample Data (Données d'échantillon) et Sample Settings (Paramètres d'échantillon), les informations relatives à l'échantillon seront saisies. Les données d'échantillon peuvent être saisies manuellement ou en important un fichier de données d'échantillon. L'ID d'échantillon doit être unique pour chaque échantillon et ne peut contenir que des caractères alphanumériques, des traits de soulignement et des tirets. N'incluez pas d'espaces. Well Position (Position du puits) fait référence au puits au format A01 à H04 de la plaque d'indexation. Les informations relatives à la séquence d'index seront automatiquement renseignées lorsque la Well

Position (Position du puits) de la plaque d'index sera saisie. Le sexe doit être saisi en tant que Homme, Femme ou Inconnu. Les champs Library Plate ID (ID de la plaque de bibliothèque) et Library Well ID (ID de puits de bibliothèque) (par ex. format A01) sont obligatoires.

- Pour saisir manuellement les données d'échantillon, ajouter des lignes (6 au total pour la flowcell S2 ou 16 pour la flowcell S4) et saisir les informations requises dans les champs Sample ID et Well Position (Position du puits). Les informations peuvent également être copiées et collées à partir d'Excel. Sélectionnez **Next** (Suivant). Sur l'écran Sample Settings (Paramètres d'échantillon), saisir Library Plate ID (ID de la plaque de bibliothèque), Library Well ID (ID du puits de bibliothèque) et le sexe. Sélectionnez **Next** (Suivant).
  - Pour importer un fichier de données d'échantillon, sélectionnez **Import Samples** (Importer des échantillons) et téléchargez le fichier de données d'échantillon. Les informations seront automatiquement renseignées dans les lignes. Un modèle (\*.csv) peut être téléchargé sur cet écran. Sélectionnez **Next** (Suivant). Sur l'écran Sample Settings (Paramètres d'échantillon), les informations seront automatiquement renseignées dans les lignes à partir du fichier de données d'échantillon importé. Sélectionnez **Next** (Suivant).
8. Sur l'écran Analysis Settings (Paramètres de l'analyse), saisissez le nom du lot enregistré pendant la planification du lot et de la série.
  9. [Facultatif] Sélectionnez le type de flowcell, S2 ou S4.
  10. Confirmez ou désélectionnez la case à cocher Générer des fichiers FASTQ compressés ORA, puis sélectionnez **Next** (Suivant).

**REMARQUE** Le TruSight Whole Genome Analysis Application génère par défaut des fichiers FASTQ compressés ORA. La modification de ce paramètre augmentera la taille de la sortie de données finale.

11. Sur l'écran Run Review (Examen de la série), vérifiez les informations saisies. Si aucune modification n'est nécessaire, sélectionnez **Save** (Enregistrer). Si des modifications sont nécessaires, sélectionnez **Back** (Retour) si nécessaire pour revenir à l'écran approprié.



### ATTENTION

TruSight Whole Genome a été validé pour 6 échantillons lors de l'utilisation de la flowcell NovaSeq 6000Dx S2 et 16 échantillons lors de l'utilisation de la flowcell NovaSeq 6000Dx S4. Assurez-vous que le nombre correct d'échantillons est saisi pour la configuration de flowcell sélectionnée.

## Révision de la série

Si des modifications sont nécessaires après la création de la série et avant le chargement des consommables pour le séquençage, révisez les séries en mode IVD sur l'instrument ou en accédant à Illumina Run Manager (IRM) à l'aide d'un navigateur sur un ordinateur en réseau.

1. Sélectionnez **Runs** (Séries).

2. Sélectionnez le nom de la série dans l'onglet Planned Runs (Séries planifiées).
3. Sélectionnez **Edit** (Modifier).
4. Mettez à jour les informations de la série ou de l'échantillon si nécessaire. Par exemple, saisissez ou corrigez l'ID du tube de bibliothèque pour qu'il corresponde à celui utilisé lors de la réalisation du flux de travail.
5. Sélectionnez **Next** (Suivant) pour exécuter Run Review (Examen de la série).
6. Sélectionnez **Save** (Enregistrer).
7. Sélectionnez **Exit** (Quitter).

Revenir au séquençage en mode IVD pour répéter le chargement des consommables. La série doit maintenant être automatiquement mise en surbrillance.

En cas de mise à jour de l'ID du tube de bibliothèque pendant le chargement des consommables, revenez à Run Selection (Sélection de la série) dans le logiciel de contrôle et sélectionnez **Refresh** (Actualiser) pour la colonne associée, A ou B. La série doit maintenant être automatiquement mise en surbrillance. Si ce n'est pas le cas, sélectionnez **Back** (Retour) pour répéter Load consumables (Charger les consommables).

## Remise en file d'attente d'une analyse

Reportez-vous à la section Troubleshooting (Dépannage) dans le TruSight Whole Genome Package Insert (document n° 200050132) pour déterminer le type Requeue Analysis (Remettre l'analyse en file d'attente) le plus approprié.

### Remettre l'analyse en file d'attente sans modification

1. Sélectionnez le nom de la Completed Run (Séries complétées) pour afficher les Run Details (Détails de la série).
2. Sélectionnez **Requeue Analysis** (Remettre l'analyse en file d'attente).
3. Sélectionnez **Requeue analysis with no changes** (Remettre l'analyse en file d'attente sans modification).
4. Fournissez des détails dans le champ Reanalysis Reason (Motif de la nouvelle analyse).
5. Sélectionnez **Requeue Analysis** (Remettre l'analyse en file d'attente).
6. Quittez la page et accédez à la page Active Runs (Séries actives) pour confirmer que la file d'attente est en cours.

### Remettre l'analyse en file d'attente avec modification

1. Sélectionnez le nom de la Completed Run (Séries complétées) pour afficher les Run Details (Détails de la série).
2. Sélectionnez **Requeue Analysis** (Remettre l'analyse en file d'attente).

3. Sélectionnez **Edit run settings** et **Requeue analysis** (Modifier les paramètres de la série et Remettre l'analyse en file d'attente)
4. Fournissez des détails dans le champ Reanalysis Reason (Motif de la nouvelle analyse).
5. Sélectionnez **Requeue Analysis** (Remettre l'analyse en file d'attente).
6. Confirmez ou mettez à jour les Run Settings (Paramètres de série), puis sélectionnez **Next** (Suivant).
7. Corrigez les informations d'échantillon si nécessaire en mettant à jour manuellement les champs ou sélectionnez **Download Template** (Télécharger le modèle) pour créer un fichier `sampledata.csv` avec les informations actuelles. Corrigez les informations et supprimez les lignes existantes dans l'onglet Sample Data (Données d'échantillon) avant d'utiliser Import Samples (Importer des échantillons) pour renseigner les données d'échantillon corrigées.
8. Sur l'écran Run Review (Examen de la série), passez en revue les informations sur l'analyse et sélectionnez **Save** (Enregistrer).
9. Sélectionnez **Exit** (Quitter) et accédez à la page Active Runs (Séries actives) pour confirmer que la file d'attente est en cours.

Le dossier de données de série d'origine doit être présent à l'emplacement de stockage externe spécifié dans Run Details (Détails de la série) pour que la nouvelle analyse s'achève avec succès. Si la nouvelle analyse échoue, assurez-vous que la série n'a pas été déplacée ou supprimée.

## Afficher la série et les résultats

1. Dans l'écran Illumina Run Manager principal en mode IVD, sélectionnez **Runs** (Séries).
2. Dans l'onglet Completed Runs (Séries complétées), sélectionnez le nom de la série.  
Cet onglet affiche également les séries terminées en raison d'un échec du séquençage, du transfert de données ou de l'analyse. Les séries actives et leur statut sont affichés dans l'onglet Active Runs (Séries actives). Reportez-vous à Documentation produit NovaSeq 6000Dx (document n° 200010105), pour plus d'informations.
3. Sélectionnez le nom de la série dans l'onglet Completed Runs (Séries complétées) pour afficher les détails de la série et les résultats du chemin d'accès au Analysis Output Folder (Dossier de sortie d'analyse).  
Pour les séries échouées, vérifiez le Status (État) de chaque étape, puis reportez-vous à la section Troubleshooting (Dépannage) du TruSight Whole Genome Package Insert (document n° 200050132).
4. Naviguez jusqu'au dossier d'analyse sur votre lecteur local et ouvrez le rapport consolidé pour examiner le résultat RÉUSSITE/ÉCHEC pour chaque étape du CQ comme suit :
  - Pour le contrôle qualité de séquençage, se reporter au Summary Sequencing QC Result (résumé des résultats du contrôle qualité de séquençage)
  - Pour le CQ FASTQ pour chaque échantillon de la série, se reporter au Summary FASTQ QC Result (Résumé des résultats du CQ FASTQ)

- Pour le CQ de la bibliothèque pour chaque échantillon de la série, se reporter au Summary Sample Library QC Result (Résumé des résultats du CQ de la bibliothèque d'échantillons). Si un résultat ÉCHEC est observé, notez l'étape du CQ et reportez-vous à la section Troubleshooting (Dépannage) du TruSight Whole Genome Package Insert (document n° 200050132).

## Résumé du fichier de sortie

Le TruSight Whole Genome Analysis Application enregistre les fichiers de sortie principaux suivants. Reportez-vous aux sections d'informations sur les fichiers ci-dessous pour connaître l'emplacement des fichiers de sortie principaux.

Les analyses et les échantillons qui ne répondent pas aux critères de validité ne produisent pas de fichiers CRAM, ROH bed ou \*genome.vcf.

Fichier de sortie	Description
Rapport consolidé (*.csv)	Contient des mesures de qualité utilisées pour la validité de la série (y compris le rendement total et Q30), des mesures de validité de l'échantillon (y compris le rendement FASTQ), des mesures de CQ de la bibliothèque et des mesures pour information uniquement (FIO) pour tous les échantillons de la série.
Exemple de rapport (*.csv)	Contient les résultats du CQ de séquençage, du FASTQ et du CQ de la bibliothèque d'échantillons. Le rapport contient également la concordance ploïdienne et les mesures FIO pour l'échantillon individuel ainsi que le séquençage associé.
Petit variant et mSNV VCF (*.annotated.hard-filtered-gvcf.gz)	Contient des informations de définition de variants pour les petits variants (SNV, indels) et les SNV mitochondriaux.
CNV VCF (*.cnv.vcf.gz)	Contient des informations de définition de variant pour les gains et les pertes de nombre de copies.
Répétitions VCF (*.repeats.vcf.gz)	Contient des informations de définition de variant sur les extensions STR et SMN1.
ROH BED (*.roh.bed)	Contient des informations sur les régions d'homozygotie.
FASTQ (*.fastq.gz ou *.fastq.ora)	Fichiers intermédiaires contenant les définitions de bases dont la qualité est notée. Les fichiers FASTQ constituent les principales données d'entrée pour l'étape de l'alignement. Si la compression ORA est sélectionnée, le nom du fichier en tient compte.
Alignement CRAM (*.cram)	Contiennent les lectures alignées pour un échantillon donné.

## Informations concernant le rapport de CQ

Le rapport consolidé <<RunID>>\_Consolidated\_Report.csv se trouve dans le répertoire TruSightWholeGenomeAnalysis\_x.x.x\_run-complete et contient des informations sur les mesures de qualité utilisées pour accepter ou refuser des échantillons à différentes étapes de l'analyse. Les

rapports d'échantillons individuels <<Sample\_ID>>\_Sample\_Report.csv se trouvent dans les dossiers <Sample\_ID> du répertoire TruSightWholeGenomeAnalysis\_x.x.x\_run-complete.

Les en-têtes du rapport comprennent les informations suivantes sur l'analyse : la version de l'application, le nom du lot, l'ID du tube du pool de bibliothèques, le nom de l'analyse de séquençage, l'ID de l'analyse de séquençage et le type de flow cell. Les tableaux suivants décrivent les informations incluses dans le rapport consolidé. Le rapport d'échantillon individuel comprend les mêmes informations, à l'exception des mesures Demultiplex.

Tableau 1 Mesures de CQ du séquençage

Indicateurs	Spéc.	Description
Rendement total non indexé (Go)	S.O.	Aucune spécification car des séries à rendement inférieur peuvent entraîner la réussite de bibliothèques d'échantillons. Attendez-vous à $\geq 3\ 000$ Gbp pour S4 et $\geq 1\ 000$ Go pour la flowcell S2.
% total $\geq$ Q30	$\geq 85$	Mesure de la qualité de base au niveau de la série. La spécification minimale est définie car les séries %Q30 trop basses ne permettront pas la réussite des bases Q30 dans le CQ de la bibliothèque d'échantillons.
Résumé des résultats du CQ de séquençage	RÉUSSITE ou ÉCHEC	En cas d'échec du CQ de séquençage, consultez la section Troubleshooting (Dépannage) dans le TruSight Whole Genome Package Insert (document n° 200050132).

Tableau 2 Mesures de démultiplexage

Indicateurs	Spéc.	Description
Pourcentage de lectures identifiées	S.O.	Fraction totale de lectures de filtre réussies dans la série qui ont été attribuées aux échantillons pendant le démultiplexage.
Pourcentage CV	S.O.	Fournit une mesure de l'uniformité des lectures démultiplexées à chaque paire d'index sur la série. Attendez-vous à $< 25\ %$ pour les séries sans échec aux résultats du CQ FASTQ.

Tableau 3 Indicateurs CQ FASTQ

Indicateurs	Spéc.	Description
Rendement par échantillon (bps)	$\geq 90\ 000\ 000\ 000$	Le minimum est défini pour être équivalent à une couverture autosomique moyenne $\sim 26x$ afin de trier les bibliothèques d'échantillons qui ne réussiront pas le CQ pour réduire le temps d'analyse.

Indicateurs	Spéc.	Description
Résumé des résultats du CQ FASTQ	RÉUSSITE ou ÉCHEC	En cas d'échec du CQ FASTQ, consultez la section Troubleshooting (Dépannage) dans le TruSight Whole Genome Package Insert (document n° 200050132).

Tableau 4 Indicateurs de CQ de la bibliothèque d'échantillon

Indicateurs	Spéc.	Description
Couverture autosomique moyenne	$\geq 35$	Couverture moyenne des autosomes. La spécification minimale est définie pour garantir les performances analytiques.
Pourcentage d'autosomes avec une couverture supérieure à 20X	$\geq 93,94$	Mesure de l'uniformité de la couverture qui détecte les problèmes qui ne sont pas nécessairement liés au biais GC. La spécification minimale est définie pour garantir les performances analytiques.
Couverture normalisée entre 60 % et 79 % des bacs GC	$0,82 \leq x \leq 1,13$	Mesure de l'uniformité de la couverture qui détecte le biais GC, en particulier une perte de couverture dans les zones du génome avec un pourcentage de GC plus élevé et une composition de base de % AT plus faible. Les spécifications minimales et maximales sont définies pour garantir les performances analytiques.
Couverture normalisée entre 20 % et 39 % des bacs GC	$0,97 \leq x \leq 1,06$	Mesure de l'uniformité de la couverture qui détecte le biais GC, en particulier une perte de couverture dans les zones du génome avec un pourcentage de GC plus faible et une composition de base de % AT plus élevée. Les spécifications minimales et maximales sont définies pour garantir les performances analytiques.
Couverture mitochondriale moyenne	$\geq 500$	Couverture du chromosome mitochondrial. La spécification minimale est définie pour garantir la limite de détection SNV mitochondriale.
Pourcentage de bases Q30	$\geq 85$	Mesure de la qualité de la base. La spécification minimale est définie pour garantir les performances analytiques.
Contamination estimée de l'échantillon	$\leq 0,005$	Détecte les lectures contaminantes d'autres échantillons. La spécification maximale est définie pour garantir la limite de détection du SNV mitochondrial (le type de variant ayant la plus haute sensibilité à la contamination).

Indicateurs	Spéc.	Description
Résumé du résultat CQ de la bibliothèque d'échantillons	RÉUSSITE ou ÉCHEC	En cas d'échec du CQ de la bibliothèque d'échantillons, consultez la section Dépannage dans le TruSight Whole Genome Package Insert (document n° 200050132).

Tableau 5 Mesures CQ de la ploïdie

Indicateurs	Spéc.	Description
A fourni une ploïdie chromosomique sexuelle	S.O.	Sexe fourni par l'opérateur pendant la création de la série (femme, homme, inconnu).
Estimation de ploïdie	S.O.	La ploïdie sexuelle est estimée à DRAGEN.
Résumé du résultat de la ploïdie	CONCORDANT, DISCORDANT ou ND	CONCORDANT indique un accord entre la ploïdie sexuelle fournie et celle estimée. ND indique le sexe fourni comme Inconnu ou estimation autre que XX ou XY. Pour les résultats DISCORDANTS, soit le mauvais sexe a été saisi lors de la création de la série, soit un échange d'échantillon a pu se produire. Consultez la section Troubleshooting (Dépannage) dans le TruSight Whole Genome Package Insert (document n° 200050132).

Tableau 6 Pour les mesures à titre d'information uniquement

Indicateurs	Description
Insérer la longueur médiane	La cible est de 450 bp, mais on s'attend à une variation par cycle de séquençage et par opérateur. Une plage d'environ 360 à 550 bp est acceptable. Un fonctionnement constant en dehors de cette plage peut entraîner une incidence plus élevée d'échec de l'échantillon.
Pourcentage de lectures cartographiées	Pourcentage de lectures correspondant au génome de référence. Peut être diminué en réponse à une contamination par l'ADNg non humain, à une mauvaise qualité de l'échantillon ou à une longueur d'insert trop petite entraînant des problèmes de cartographie.
Pourcentage de lectures avec alignements supplémentaires	Pourcentage de lectures avec un mappage qui se divise en différents emplacements dans le génome de référence.

Indicateurs	Description
Pourcentage de lectures marquées en double	Attendez-vous à < 20 %. Peut être élevé si le rendement de préparation de la bibliothèque est faible ou si le volume de regroupement est inférieur au volume requis, ou en réponse à des problèmes liés au séquençage.
Pourcentage de bases souples clipsées Lecture 1	Utile pour diagnostiquer la cause première de l'échec de la couverture autosomique moyenne.
Pourcentage de bases souples clipsées Lecture 2	Utile pour diagnostiquer la cause première de l'échec de la couverture autosomique moyenne.
Bases en pourcentage coupées Lecture 1	Utile pour diagnostiquer la cause première de l'échec de la couverture autosomique moyenne.
Bases en pourcentage coupées Lecture 2	Utile pour diagnostiquer la cause première de l'échec de la couverture autosomique moyenne.

## Informations sur le fichier de définition de variant

### Fichiers VCF

Le format de fichier Variant Call Format (\*.vcf) contient des renseignements sur les variants que l'on trouve à des positions spécifiques dans un génome de référence et peut se trouver dans le répertoire <Sample\_ID>/Analysis.

L'en-tête du fichier VCF comprend la version du format de fichier VCF, la version du paramètre de définitions des variants et les annotations utilisées dans le reste du fichier. La dernière ligne de l'en-tête est constituée d'en-têtes de colonnes pour les lignes de données. Chacune des lignes de données d'un fichier VCF contient des renseignements sur une seule position de référence.

Tous les fichiers VCF contiennent un en-tête avec des descriptions de colonnes de sortie et des données de définition de variant dans des colonnes étiquetées CHROM, POS, ID, REF, ALT, QUAL, FILTER, INFO, FORMAT, SAMPLE. Les définitions des valeurs de colonne peuvent varier d'une définition à l'autre.

## Petit variant et mSNV VCF

La sortie est enregistrée sous le fichier `<Sample_ID>.annotated.hard-filtered.gvcf.gz` dans le répertoire `<Sample_ID>/Analysis`.

Un fichier génomique VCF (gVCF) contient des informations sur les variants et les positions déterminées comme étant homozygotes pour le génome de référence. Pour les régions homozygotes, le fichier gVCF comprend des statistiques qui indiquent dans quelle mesure les lectures prennent en charge l'absence de variants ou d'allèles alternatifs. Le fichier gVCF comprend un allèle artificiel `<NON_REF>`. Les lectures qui ne prennent pas en charge la référence ou les variants sont attribuées à l'allèle `<NON_REF>`. DRAGEN utilise ces lectures pour déterminer si la position peut être définie comme référence homozygote, plutôt que de rester sans définition. Le score obtenu représente le niveau de confiance à l'échelle de Phred dans une définition de référence homozygote. En mode germline, le score est FORMAT/GQ.

DRAGEN fournit un filtrage des variants post-VCF basé sur les annotations présentes dans les enregistrements VCF. Le filtrage rigide des variants est décrit ci-dessous. Cependant, en raison de la nature des algorithmes de DRAGEN, qui intègrent l'hypothèse d'erreurs corrélées à partir du cœur de la définition de variants, le pipeline a amélioré ses capacités pour distinguer les véritables variants du bruit, et par conséquent, la dépendance au filtrage post-VCF est considérablement réduite.

Le TruSight Whole Genome Analysis Application fournit une annotation du score de confiance et du niveau de confiance pour les petits variants qui peuvent être utilisés pour améliorer davantage les performances. L'annotation de niveau de confiance n'est pas un filtre de qualité et n'est donc pas directement reflétée dans l'état de qualité des définitions de variant. Par conséquent, il est possible de voir les définitions de variants réussis qui sont néanmoins annotés comme étant de faible confiance.

Tableau 7 En-têtes du fichier VCF

En-tête	Description
CHROM	Chromosome du génome de référence. Les chromosomes apparaissent dans le même ordre que dans le fichier de référence FASTA.
POS	Position de la base unique du variant dans le chromosome de référence. Pour les variants mononucléotidiques (SNV), cette position est la base de référence avec le variant. Pour les indels, cette position est la base de référence immédiatement avant le variant.
ID	Toujours .
REF	Le génotype de référence. Par exemple, une délétion d'un T unique est représentée TT comme référence et T comme alternative. Un variant à simple nucléotide de A à T est représenté A comme référence et T comme alternative.

En-tête	Description
ALT	Allèles qui diffèrent de la lecture de référence. Par exemple, une insertion d'un T unique est représentée A comme référence et AT comme alternative. Un variant à simple nucléotide de A à T est représenté A comme référence et T comme alternative.
QUAL	Score de qualité de l'échelle Phred attribué par le paramètre de définition des variants. Des scores supérieurs indiquent une confiance accrue dans le variant et une moindre probabilité d'erreur. Pour un score de qualité de Q, la probabilité estimée d'une erreur est de $10^{-(Q/10)}$ . Par exemple, l'ensemble de définitions Q30 est associé à un taux d'erreur de 0,1 %. De nombreux paramètres de définition des variants attribuent des scores de qualité en fonction de leurs modèles statistiques, qui sont élevés par rapport au taux d'erreur observé.

Tableau 8 Annotations du fichier VCF

En-tête	Description
FILTER	<ul style="list-style-type: none"> <li>• <b>RÉUSSITE</b> : Tous les filtres ont été passés.</li> <li>• <b>DRAGENSnpHardQUAL</b> : Définir si vrai :QUAL &lt; 10,41</li> <li>• <b>DRAGENIndelHardQUAL</b> : Définir si vrai :QUAL &lt; 7,83</li> <li>• <b>LowDepth</b> : Définir si vrai :DP &lt;= 1</li> <li>• <b>LowGQ</b> : Définir si vrai :GQ = 0</li> <li>• <b>PloidyConflict</b> : la définition de génotype du paramètre de définition de variant n'est pas compatible avec la ploïdie chromosomique</li> <li>• <b>base_quality</b> : site filtré, car la qualité de base médiane des lectures alt à ce locus n'atteint pas le seuil</li> <li>• <b>filtered_reads</b> : site filtré, car une trop grande fraction de lectures a été filtrée</li> <li>• <b>fragment_length</b> : site filtré, car la différence absolue entre la longueur médiane des fragments des lectures alt et la longueur médiane des fragments des lectures ref à ce locus dépasse le seuil</li> <li>• <b>low_af</b> : la fréquence de l'allèle n'atteint pas le seuil</li> <li>• <b>low_depth</b> : site filtré, car la profondeur de lecture est trop faible</li> <li>• <b>low_frac_info_reads</b> : site filtré, car la fraction de lectures informatives est inférieure au seuil</li> <li>• <b>low_normal_depth</b> : site filtré, car la profondeur de lecture normale de l'échantillon est trop faible</li> <li>• <b>long_indel</b> : site filtré, car la longueur de l'indel est trop longue</li> <li>• <b>mapping_quality</b> : site filtré, car la qualité de cartographie médiane des lectures alt à ce locus n'atteint pas le seuil</li> <li>• <b>multiallelic</b> : site filtré, car plus de deux allèles alt passent la LOD de la tumeur</li> <li>• <b>non_homref_normal</b> : site filtré, car le génotype normal de l'échantillon n'est pas une référence homozygote</li> <li>• <b>no_reliable_supporting_read</b> : site filtré, car il n'existe aucune lecture somatique fiable</li> <li>• <b>panel_of_normals</b> : vu dans au moins un échantillon dans le panel des échantillons normaux vcf</li> <li>• <b>read_position</b> : site filtré, car la médiane des distances entre le début/la fin de la lecture et ce lieu est inférieure au seuil</li> <li>• <b>RMxNRepeatRegion</b> : site filtré, car tout ou partie de l'allèle variant est une répétition de la référence</li> <li>• <b>strand_artifact</b> : site filtré en raison d'une distorsion sévère du brin</li> <li>• <b>str_contraction</b> : site filtré en raison d'une erreur de PCR suspectée où l'allèle alt est une unité répétée de moins que la référence</li> <li>• <b>too_few_supporting_reads</b> : site filtré, car il y a trop peu de lectures de soutien dans l'échantillon de tumeur</li> <li>• <b>weak_evidence</b> : le score de variant somatique n'atteint pas le seuil</li> </ul>

En-tête	Description
INFO	<ul style="list-style-type: none"><li>• <b>DB</b> : adhésion dbSNP.</li><li>• <b>FS</b> : valeur p à l'échelle de Phred utilisant le test exact de Fisher pour détecter la distorsion du brin.</li><li>• <b>QD</b> : confiance/qualité par profondeur pour le variant.</li><li>• <b>R2_5P_bias</b> : score basé sur la distorsion intersexuelle et la distance à partir de 5.</li><li>• <b>SOR</b> : rapport de cotes symétrique du tableau de contingence 2x2 pour détecter la distorsion du brin.</li><li>• <b>DP</b> : profondeur de lecture approximative (informative et non informative) ; certaines lectures peuvent avoir été filtrées en fonction de mapq, etc.</li><li>• <b>END</b> : position d'arrêt de l'intervalle.</li><li>• <b>FractionInformativeReads</b> : fraction de lectures informatives sur le nombre total de lectures.</li><li>• <b>MQ</b> : qualité de mappage RMS.</li><li>• <b>MQRankSum</b> : score Z du test de la somme des rangs de Wilcoxon des qualités de mappage de lecture Alt vs. Ref.</li><li>• <b>ReadPosRankSum</b> : Z-score du test de somme des rangs de Wilcoxon du biais de position de lecture Alt vs. Ref.</li><li>• <b>SOMATIC</b> : au moins un variant à cette position est un variant somatique.</li><li>• <b>ILENS</b> : longueurs d'index pour chaque variant d'ALAT.</li><li>• <b>SCORE</b> : score de confiance pour chaque type de variant présent au centre sous la forme (type de variant) : (score de confiance).</li><li>• <b>TIER</b> : niveau de confiance pour chaque type de variant présent au centre sous la forme (type de variant) : (niveau de confiance).</li></ul>

En-tête	Description
FORMAT	<p>La colonne FORMAT dresse la liste des champs séparés par deux-points, par exemple GT:GQ.</p> <ul style="list-style-type: none"> <li>• <b>AD</b> : profondeurs alléliques (en ne comptant que les lectures informatives sur le total des lectures) pour les allèles ref et alt dans l'ordre indiqué.</li> <li>• <b>AF</b> : fractions d'allèles pour les allèles alt dans l'ordre indiqué.</li> <li>• <b>DP</b> : profondeur de lecture approximative (les lectures avec MQ=255 ou de mauvais appariements sont filtrées).</li> <li>• <b>F1R2</b> : nombre de lectures dans l'orientation de la paire F1R2 prenant en charge chaque allèle.</li> <li>• <b>F2R1</b> : nombre de lectures dans l'orientation de la paire F2R1 prenant en charge chaque allèle.</li> <li>• <b>GP</b> : probabilités postérieures à l'échelle de Phred pour les génotypes tels que définis dans la spécification VCF.</li> <li>• <b>GQ</b> : qualité du génotype.</li> <li>• <b>GT</b> : génotype.</li> <li>• <b>ICNT</b> : nombre de lectures informatives INDEL basées sur le modèle de confiance de référence.</li> <li>• <b>MB</b> : statistiques de composants par échantillon pour détecter le biais intersexuel.</li> <li>• <b>MIN_DP</b> : DP minimum observé dans le bloc GVCF.</li> <li>• <b>PL</b> : probabilités normalisées à l'échelle Phred pour les génotypes tels que définis dans la spécification VCF.</li> <li>• <b>PRI</b> : probabilités antérieures à l'échelle de Phred pour les génotypes.</li> <li>• <b>PS</b> : informations d'identification de phase physique, où chaque ID unique au sein d'un échantillon donné (mais pas entre les échantillons) relie les enregistrements au sein d'un groupe de phase.</li> <li>• <b>SB</b> : statistiques de composants par échantillon qui comprennent le test exact de Fisher pour détecter la distorsion du brin.</li> <li>• <b>SPL</b> : probabilités normalisées, à l'échelle de Phred pour les SNP basées sur le modèle de confiance de référence.</li> <li>• <b>SQ</b> : qualité pour le variant somatique.</li> </ul>
SAMPLE	La colonne relative aux échantillons indique les valeurs précisées dans la colonne FORMAT.

## Variations de nombre de copies VCF

L'étape de comptage cible est la première étape de traitement pour le pipeline DRAGEN CNV, produisant `<Sample_ID>.target.counts.gz`, puis la correction des biais GC est effectuée, générant un fichier `*.target.counts.gc-corrected.gz`. L'étape de normalisation produit le fichier `*.tn.tsv.gz`. Le Logiciel hôte DRAGEN génère de nombreux fichiers intermédiaires.

\*.seg.called.merged est le fichier de définition final qui contient les événements d'amplification et de suppression.

En plus du fichier de segment, DRAGEN émet les définitions au format VCF standard. La sortie est enregistrée dans <Sample\_ID>.cnv.vcf.gz dans le répertoire <Sample\_ID>/Analysis.

Définitions des colonnes spécifiques à la définition CNV :

La colonne POS est la position de départ du variant. Selon la spécification VCF, si l'un des allèles ALT est un allèle symbolique, tel que <DEL>, alors la base de rembourrage est requise et POS indique la coordonnée de la base précédant le polymorphisme. Toutes les coordonnées du VCF sont basées sur 1.

La colonne ID est utilisée pour représenter l'événement. Le champ ID code le type d'événement et les coordonnées de l'événement.

La colonne REF contient un N pour tous les événements CNV.

La colonne ALT spécifie le type d'événement CNV. Étant donné que le VCF ne contient que des événements CNV, seule l'entrée DEL ou DUP est utilisée.

La colonne QUAL contient un score de qualité estimé pour l'événement CNV, qui est utilisé dans le filtrage matériel.

La colonne FILTER (Filtre) contient PASS (RÉUSSITE) si l'événement CNV passe tous les filtres, sinon la colonne contient le nom du filtre défaillant.

La colonne INFO contient des informations représentant l'événement. L'entrée REFLen indique la durée de l'événement. L'entrée SVTYPE est toujours CNV. L'entrée END indique la position de fin de l'événement.

Les champs FORMAT sont décrits dans l'en-tête.

- GT : génotype.
- SM : rapport de copie linéaire de la moyenne du segment
- CN : nombre de copies estimé
- BC : nombre de bacs dans la région
- PE : nombre de lectures d'extrémité mal appariées aux points de rupture de début et de fin

## Répète VCF

ExpansionHunter effectue un réalignement basé sur une séquence graphique des lectures qui proviennent de l'intérieur et autour de chaque répétition cible. ExpansionHunter génotype ensuite la longueur de la répétition dans chaque allèle en fonction de ces alignements graphiques.

De plus amples informations et analyses sont disponibles dans les documents ExpansionHunter suivants :

- Dolzhenko et al., *Detection of long repeat expansions from PCR-free whole-genome sequence data* 2017

- Dolzhenko et al., *ExpansionHunter : A sequence-graph based tool to analyze variation in short tandem repeat regions* 2019

Le catalogue de variants TruSight Whole Genome Analysis Application STR contient des spécifications sur les répétitions causant la maladie situées dans les gènes AFF2, AR, ATN1, ATXN1, ATXN10, ATXN2, ATXN3, ATXN7, ATXN8OS, C9ORF72, CACNA1A, CBL, CNBP, CSTB, DIP2B, DMPK, FMR1, FXN, GLS, HTT, JPH3, NIPA1, NOP56, NOTCH2NL, PABPN1, PHOX2B, PPP2R2B et TBP.

Les résultats du génotypage répété sont générés sous forme de fichier VCF distinct, qui fournit la longueur de chaque allèle à chaque répétition identifiable définie dans le fichier de catalogue de spécification des répétitions. Le nom du fichier est <Sample\_ID>.repeats.gz et se trouve dans le répertoire <Sample\_ID>/Analysis.

Certaines colonnes sont spécifiques à la définition de l'extension répétée :

Tableau 9 Champs VCF de base

Champ	Description
CHROM	Identifiant du chromosome
POS	Position de la première base avant la région de répétition dans la référence
ID	Toujours .
REF	La base de référence en position POS
ALT	Liste des allèles répétés au format <STR <sub>n</sub> > . N est le nombre d'unités de répétition.
QUAL	Toujours .
FILTER	Le filtre LowDepth est appliqué lorsque la profondeur totale du locus est inférieure à 10x ou que le nombre de lectures qui couvrent une ou les deux extrémités de rupture est inférieur à 5.

Tableau 10 Champs INFO supplémentaires

Champ	Description
END	Position de la première base avant la région de répétition dans la référence
REF	Numéro de copie de référence
REPID	ID de variant du catalogue de variants
RL	Longueur de référence en bp
RU	Répéter l'unité dans l'orientation de référence
VARID	ID de variant du catalogue de variants

Tableau 11 Champs GENOTYPE (par échantillon)

Champ	Description
AD	Profondeurs alléliques pour les allèles réf et alt dans l'ordre indiqué
ADFL	Nombre de lectures de flanc cohérentes avec l'allèle
ADIR	Nombre de lectures de flanc cohérentes avec l'allèle
ADSP	Nombre de lectures de flanc cohérentes avec l'allèle
DST	Résultats (+ détectés, - non détectés, ? indéterminés) du test représenté par le variant
GT	génotype
LC	Couverture du locus
REPCI	Intervalle de confiance pour le REPCN
REPCN	Nombre d'unités répétées dans l'allèle
RPL	Log-Rapport de vraisemblance pour la présence de l'allèle de référence
SO	Type de lectures qui soutiennent l'allèle. Les valeurs peuvent être <i>SPANNING</i> , <i>FLANKING</i> , ou <i>INREPEAT</i> . Ces valeurs indiquent si les lectures sont étendues, flancs ou entièrement contenues dans la répétition.

Le fichier <Sample\_ID>.repeats.vcf.gz inclut la sortie SMN ainsi que toutes les répétitions ciblées. La sortie SMN est représentée comme une seule définition SNV à la position affectant l'épissage dans l'état SMN1 (NM\_000344.3:c.840C/T) avec atrophie musculaire rachidienne (SMA) dans les champs personnalisés suivants.

Tableau 12 Résultats SMA dans le fichier de sortie repeats.vcf

Champ	Description
VARID	SMN marque la définition SMN.
GT	Définition de génotype à cette position en utilisant un modèle de génotype normal (diploïde).
DST	Définition de statut SMA : + indique détecté - indique non détecté ? indique indéterminé
AD	Nombre total de lectures qui soutiennent l'allèle C et T.
RPL	Rapport de probabilité Log10 entre les modèles non affectés et affectés. Les scores positifs indiquent que le modèle non affecté est plus probable.

## ROH BED

Les régions d'homozygotie (ROH) sont détectées dans le cadre de la définition du petit variant. La définition détecte et produit les séries d'homozygotie à partir des définitions de génome entier sur les chromosomes humains autosomiques. Les chromosomes sexuels sont ignorés sauf si le caryotype sexuel de l'échantillon est XX, tel que déterminé par l'estimateur de Ploidy. La sortie ROH permet aux outils en aval de dépister et de prédire la consanguinité entre les parents du sujet proband.

Une région est définie comme des définitions de variants consécutifs sur le chromosome sans grand écart entre ces variants. En d'autres termes, les régions sont rompues par chromosome ou par de grands espaces sans définition SNV. La taille de l'écart est définie sur 3 Mbases.

La définition ROH produit un fichier de sortie ROH nommé `<Sample_ID>.roh.bed` dans le répertoire `<Sample_ID>/Analysis`. Chaque ligne représente une région d'homozygotie. Le fichier lit contient les colonnes suivantes :

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

Où

- Le score est une fonction du nombre de variants homozygotes et hétérozygotes, où chaque variant homozygote augmente le score de 0,025, et chaque variant hétérozygote réduit le score de 0,975.
- Les positions de début et de fin sont un intervalle basé sur 0, en demi-ouverture.
- `#homozygote` est le nombre de variants homozygotes dans la région.
- `#heterozygote` est le nombre de variants hétérozygotes dans la région.

La définition produit également un fichier de mesures nommé `<Sample_ID>.roh_metrics.csv` qui répertorie le nombre de ROH importantes et le pourcentage de SNP dans les ROH importantes (> 3 Mo).

## Indicateurs d'estimation de la ploïdie

L'estimateur de ploïdie s'exécute par défaut. L'estimateur de ploïdie utilise les lectures du mappeur/aligneur pour calculer la profondeur de couverture du séquençage pour chaque autosome et allosome dans le génome humain. Le caryotype sexuel de l'échantillon est ensuite estimé à l'aide des rapports entre la couverture médiane des chromosomes sexuels et la couverture autosomique médiane. XX ou XY, et CONCORDANT, DISCORDANT ou ND (non déterminé) par rapport aux données d'échantillon fournies sont rapportés dans le rapport consolidé. Les résultats détaillés, y compris chaque couverture médiane normalisée par conflit, sont rapportés dans le fichier `<Sample_ID>.ploidy_estimation_metrics.csv`.

## Fichiers FASTQ

FASTQ (\*.fastq.gz, \*.fastq.ora) est un format de fichier texte qui contient les définitions de bases et les valeurs de qualité, par lecture. Chaque fichier contient les informations suivantes :

- L'identifiant de l'échantillon
- La séquence

- Le signe plus (+)
- Les scores de qualité Phred sont identifiés dans un format codé ASCII + 33

Le logiciel génère un fichier FASTQ pour chaque échantillon, lecture et voie. Par exemple, pour chaque échantillon d'une série à deux extrémités, le logiciel génère deux fichiers FASTQ : un pour la Lecture 1 et un pour la Lecture 2. En plus de ces fichiers FASTQ d'échantillon, le logiciel génère deux fichiers FASTQ par voie contenant tous les échantillons inconnus. Les fichiers FASTQ pour Lecture d'index 1 et Lecture d'index 2 ne sont pas générés car la séquence est incluse dans l'en-tête de chaque entrée FASTQ. Le format du nom de fichier est construit à partir des champs spécifiés dans la feuille d'échantillon et utilise le format de nommage de fichier `<Sample_ID>_S#_L00#_R#_001.fastq.gz`

Les fichiers FASTQ sont enregistrés dans le répertoire `<Sample_ID>/Conversion` directory. Dans le répertoire FASTQ du dossier d'analyse, vous pouvez trouver le répertoire Logs avec les journaux de conversion BCL-à-FASTQ, et le répertoire Reports qui contient divers fichiers de mesures lus, et `SampleSheet.csv` utilisé pour la conversion FASTQ. Les fichiers FASTQ provenant de lectures indéterminées se trouvent dans le répertoire `Undetermined/Conversion` du dossier Analysis.

L'identifiant de l'échantillon est formaté comme suit :

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

Exemple :

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

```
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
```

```
+
```

```
<>;##=><9=AAAAAAAAAA9#:<#<;<<<????#=#
```

## Fichiers CRAM

Les fichiers Compressed Reference-oriented Alignment Map ou CRAM (\*.cram) sont stockés dans le répertoire `<Sample_ID>/Analysis` et contiennent des en-têtes et des enregistrements d'alignement relatifs au fichier de référence génomique utilisé lors de l'alignement. Le chemin d'accès au fichier de référence est répertorié dans le fichier `<Sample_ID>/Analysis/<Sample_ID>-replay.json`, en tant que paramètre `--ht-reference`, par défaut défini sur `hg38.fa`.

Les fichiers CRAM contiennent une section d'en-tête et une section pour les alignements :

- **En-tête**—Contient les renseignements sur la totalité du fichier, tels que le nom de l'échantillon, sa longueur et la méthode d'alignement. Les alignements dans la section des alignements sont associés aux renseignements spécifiques figurant dans la section d'en-tête.
- **Alignements**—Contient le nom de la lecture, la séquence de lecture, la qualité de la lecture, les renseignements sur l'alignement et des balises personnalisées. Le nom de la lecture comprend le chromosome, les coordonnées de départ, la qualité de l'alignement et la chaîne du descripteur de correspondance.

La section des alignements comprend les renseignements suivants sur chacune des lectures ou des paires de lectures :

- AS : qualité de l'alignement des paires de bases appariées.
- RG : groupe de lecture, qui indique le nombre de lectures pour un échantillon spécifique.
- BC : marqueur du code à barres qui indique l'identifiant de l'échantillon démultiplexé associé à la lecture.
- SM : qualité de l'alignement à paire de bases unique.
- XC : chaîne du descripteur de correspondance.
- XN : marqueur du nom de l'amplicon, qui enregistre l'identifiant de l'amplicon associé à la lecture

Pour afficher les enregistrements d'alignement, les `samtools` peuvent être utilisés sous `samtools view --reference <path_to_reference_folder>/hg38.fa <Sample_ID>.cram`.

Un fichier d'index et un fichier de somme de contrôle sont également générés.

# Assistance technique

Pour une assistance technique, contactez le support technique Illumina.

**Site Web :** [www.illumina.com](http://www.illumina.com)

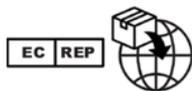
**E-mail :** [techsupport@illumina.com](mailto:techsupport@illumina.com)

**Fiches de données de sécurité (SDS)** : disponibles sur le site Web d'Illumina à l'adresse [support.illumina.com/sds.html](http://support.illumina.com/sds.html).

**Documentation sur les produits** : disponible en téléchargement sur [support.illumina.com](http://support.illumina.com).



Illumina, Inc.  
5200 Illumina Way  
San Diego, Californie 92122 États-Unis  
+(1) 800 809 ILMN (4566)  
+(1) 858 202 4566 (en dehors de l'Amérique du Nord)  
techsupport@illumina.com  
www.illumina.com



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

**Promoteur australien**

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

DESTINÉ AU DIAGNOSTIC IN VITRO UNIQUEMENT.

© 2024 Illumina, Inc. Tous droits réservés.

**illumina**<sup>®</sup>