

Module d'analyse des variants somatiques dans Local Run Manager

Guide du flux de travail pour MiSeqDx

DESTINÉ AU DIAGNOSTIC IN VITRO UNIQUEMENT

Vue d'ensemble	3
Saisie des renseignements sur l'analyse	3
Méthodes d'analyse	5
Affichage des données concernant l'analyse et les échantillons	7
Rapport d'analyse	7
Fichiers de sortie d'analyse	9
Assistance technique	16



Ce document et son contenu sont exclusifs à Illumina, Inc. et ses sociétés affiliées (« Illumina »), et sont exclusivement destinés à l'usage contractuel de son client dans le cadre de l'utilisation du ou des produits décrits dans les présentes et ne peuvent servir à aucune autre fin. 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. Illumina ne cède aucune licence en vertu de son brevet, de sa marque de commerce, de ses droits d'auteur ou de ses droits traditionnels ni des droits similaires d'un tiers quelconque par ce document.

Les instructions contenues dans ce document doivent être suivies strictement et explicitement par un personnel qualifié et adéquatement formé de façon à assurer l'utilisation correcte et sûre du ou des produits décrits dans les présentes. Le contenu intégral de ce document doit être lu et compris avant l'utilisation de ce ou ces produits.

LE MANQUEMENT À LIRE COMPLÈTEMENT ET À SUIVRE EXPLICITEMENT TOUTES LES INSTRUCTIONS CONTENUES DANS LES PRÉSENTES POURRA CAUSER DES DOMMAGES AU(X) PRODUIT(S), DES BLESSURES AUX PERSONNES, UTILISATEURS OU AUTRES, ET DES DOMMAGES AUX AUTRES BIENS.

ILLUMINA DÉCLINE TOUTE RESPONSABILITÉ DÉCOULANT DE L'UTILISATION INAPPROPRIÉE DU OU DES PRODUITS DÉCRITS DANS LES PRÉSENTES (Y COMPRIS LEURS COMPOSANTES ET LE LOGICIEL).

© 2017 Illumina, Inc. Tous droits réservés.

Illumina, MiSeqDx et la conception de bases en flux sont des marques de commerce déposées ou en cours de dépôt d'Illumina, Inc. ou de ses sociétés affiliées aux États-Unis ou dans d'autres pays. Tous les autres noms, logos et marques de commerce sont la propriété de leurs détenteurs respectifs.

Vue d'ensemble

Le module d'analyse des variants somatiques de Local Run Manager est destiné à être utilisé avec le test diagnostique de la TruSeq Custom Amplicon Kit d'Illumina. Lorsqu'il est utilisé avec le module d'analyse des variants somatiques, le test est conçu pour préparer les bibliothèques servant au séquençage de l'ADN de tissus fixés au formol et imprégnés à la paraffine (FFPE). Le test détecte les mutations somatiques à une faible fréquence de variant.

Le module d'analyse évalue de courtes régions d'ADN amplifié, ou amplicons, à la recherche de variants. Un séquençage ciblé des amplicons permet une couverture élevée de régions particulières dans un grand nombre d'échantillons. Le module d'analyse effectue l'analyse secondaire et génère les rapports à la suite des analyses de séquençage ayant recours à la méthode du double brin, qui nécessite des pools d'oligos sens et antisens. Voir la notice d'accompagnement du *test diagnostique de la TruSeq Custom Amplicon Kit Dx (trousse d'amplicons TruSeq personnalisée) (document no 1000000029772)*.

Le module d'analyse des variants somatiques requiert des consommables de séquençage de la MiSeqDx Reagent Kit v3 (trousse de réactifs MiSeqDx v3). Voir la notice d'accompagnement de la *MiSeqDx Reagent Kit v3 (trousse de réactifs MiSeqDx v3) (document no 1000000030849)*.

À propos de ce guide

Le présent guide fournit des instructions sur la configuration des paramètres d'analyse pour le séquençage et l'analyse au moyen du module d'analyse des variants somatiques. Pour obtenir de l'information sur le tableau de bord de Local Run Manager et les paramètres du système, consultez le *Guide de référence du logiciel Local Run Manager (document no 1000000011880)*.

Saisie des renseignements sur l'analyse

Configuration des paramètres

- 1 Ouvrez une session dans Local Run Manager.
- 2 Cliquez sur **Create Run** (Créer l'analyse), et sélectionnez **Somatic Variant** (Variant somatique).
- 3 Entrez un nom qui désignera l'analyse, du séquençage jusqu'à la fin du processus.
Utilisez des caractères alphanumériques, des espaces, des traits de soulignement ou des tirets.
- 4 [Facultatif] Saisissez une description de l'analyse, pour l'identifier.
Utilisez des caractères alphanumériques, des espaces, des traits de soulignement ou des tirets.
- 5 Choisissez le nombre d'échantillons et l'ensemble d'index dans la liste déroulante.

Importation des fichiers de manifeste à analyser

- 1 Assurez-vous que les fichiers de manifeste à importer se trouvent dans un emplacement réseau accessible ou sur une clé USB.
- 2 Cliquez sur **Import Manifests** (Importer les fichiers de manifeste).
- 3 Accédez aux fichiers de manifeste et sélectionnez les fichiers à ajouter.



REMARQUE

Pour rendre les fichiers de manifeste accessibles pour toutes les analyses au moyen du module d'analyse des variants somatiques, ajoutez les fichiers à l'aide de la fonction Module Settings (Paramètres du module). Cette fonction exige un niveau d'accès administrateur. Pour plus de renseignements, consultez le *Guide de référence du logiciel Local Run Manager (document no 1000000011880)*.


Préciser les échantillons à analyser

Préciser les échantillons à analyser en utilisant l'une des options suivantes, conformément aux directives qui suivent.

- ▶ **Saisie des échantillons manuellement** : Utilisez le tableau vide à l'écran Create Run (Créer l'analyse).
- ▶ **Importation des échantillons** : Accédez à un fichier externe aux valeurs séparées par des virgules (*.csv). Un modèle peut être téléchargé à l'écran Create Run (Créer l'analyse).

Après avoir généré le tableau d'échantillons, vous pouvez exporter l'information sur les échantillons dans un fichier externe et utiliser ce fichier comme référence lorsque vous préparez des bibliothèques, ou vous pouvez importer le fichier pour une autre analyse.

Saisie manuelle des échantillons

- 1 Entrez un nom d'échantillon distinct dans le champ Sample Name (Nom de l'échantillon). Utilisez des caractères alphanumériques, des tirets ou des traits de soulignement. La saisie du nom de l'échantillon inscrit automatiquement les données du puits correspondant de l'autre pool.
- 2 [Facultatif] Pour les échantillons de contrôle positif ou négatif, faites un clic droit et sélectionnez le type de contrôle. La saisie du contrôle dans l'un des puits d'échantillon inscrit automatiquement le même contrôle dans le puits correspondant de l'autre pool.
- 3 [Facultatif] Entrez la description de l'échantillon dans le champ Description. Utilisez des caractères alphanumériques, des tirets ou des traits de soulignement. La saisie de la description de l'échantillon inscrit automatiquement les données du puits correspondant de l'autre pool. La description de l'échantillon est associée à l'identifiant de l'échantillon. La description de l'échantillon est écrasée si le même identifiant d'échantillon est utilisé plus tard lors d'une analyse subséquente.
- 4 Sélectionnez l'adaptateur d'index 1 dans la liste déroulante Index 1 (i7).
- 5 Sélectionnez l'adaptateur d'index 2 dans la liste déroulante Index 2 (i5).
- 6 Sélectionnez un fichier de manifeste dans la liste déroulante Manifest (Manifeste). Les échantillons du pool A requièrent un fichier de manifeste différent de celui des échantillons du pool B.
- 7 Choisissez l'une des options pour afficher, imprimer ou enregistrer la présentation de la plaque à titre de référence pour la préparation des bibliothèques :
 - ▶ Cliquez sur l'icône  **Print** (Imprimer) pour afficher la présentation de la plaque. Sélectionnez **Print** (Imprimer) pour imprimer la présentation de la plaque.
 - ▶ Cliquez sur **Export** (Exporter) pour exporter l'information sur les échantillons dans un fichier externe. Assurez-vous que le fichier de manifeste et les renseignements sur l'échantillon sont exacts. Des renseignements erronés peuvent nuire aux résultats.
- 8 Cliquez sur **Save Run** (Enregistrer l'analyse).

Importation des échantillons

- 1 Cliquez sur **Import Samples** (Importer les échantillons) et accédez à l'emplacement du fichier d'information sur les échantillons. Vous pouvez importer deux types de fichiers.
 - ▶ Cliquez sur **Template** (Modèle) à l'écran Create Run (Créer l'analyse) pour établir une nouvelle présentation de plaque. Le fichier modèle contient les bons en-têtes de colonnes en vue de l'importation. Entrez l'information sur les échantillons dans chaque colonne en vue de l'analyse. Supprimez l'information donnée en exemple dans les cellules inutilisées, puis enregistrez le fichier.
 - ▶ Utilisez un fichier d'information sur les échantillons qui a été exporté du module d'analyse des variants somatiques en utilisant la fonction Export (Exporter).
- 2 Cliquez sur l'icône **Print** (Imprimer) pour afficher la présentation de la plaque.
- 3 Sélectionnez **Print** (Imprimer) pour imprimer la présentation de la plaque à titre de référence pour la préparation des bibliothèques.
- 4 [Facultatif] Cliquez sur **Export** (Exporter) pour exporter l'information sur les échantillons dans un fichier externe.
Assurez-vous que le fichier de manifeste et les renseignements sur l'échantillon sont exacts. Des renseignements erronés peuvent nuire aux résultats.
- 5 Cliquez sur **Save Run** (Enregistrer l'analyse).

Modification d'une analyse

Pour obtenir des instructions sur la modification de l'information dans votre analyse avant le séquençage, consultez le *Guide de référence du logiciel Local Run Manager (document no 1000000011880)*.

Méthodes d'analyse

Le module d'analyse des variants somatiques réalise les étapes d'analyse suivantes, puis enregistre les fichiers de sortie d'analyse dans le dossier Alignment (Alignement).

- ▶ Démultiplexage des lectures d'index
- ▶ Génération des fichiers FASTQ
- ▶ Alignement à une référence
- ▶ Identification des variants

Démultiplexage

Le démultiplexage compare chaque séquence de lecture d'index aux séquences d'indexage définies pour l'analyse. Aucune valeur de qualité n'est prise en compte lors de cette étape.

Les lectures d'index sont identifiées en suivant les étapes ci-dessous :

- ▶ Les échantillons sont numérotés en commençant par 1, selon l'ordre dans lequel ils sont classés pour l'analyse.
- ▶ Le numéro d'échantillon 0 est réservé aux amplifiats qui n'ont pas été assignés à un échantillon.
- ▶ Les amplifiats sont assignés à un échantillon lorsque la séquence d'indexage est identique ou lorsqu'il y a un seul mésappariement par lecture d'index.

Génération des fichiers FASTQ

Après le démultiplexage, le logiciel génère une analyse intermédiaire au format FASTQ, qui est un format texte utilisé pour représenter des séquences. Les fichiers FASTQ contiennent les lectures pour chaque échantillon ainsi que les scores de qualité connexes. Les amplifiats qui n'ont pas passé le filtre sont exclus.

Chaque fichier FASTQ contient des lectures pour un seul échantillon, le nom de cet échantillon est compris dans le nom du fichier FASTQ. Les fichiers FASTQ constituent les principales données d'entrée pour l'alignement. Quatre fichiers FASTQ sont générés pour chaque échantillon, deux pour le pool A, et deux pour le pool B.

Alignement

Au cours de l'étape de l'alignement, l'algorithme de Smith-Waterman par bande aligne les amplifiats provenant de chacun des échantillons avec les séquences d'amplicon définies dans le fichier de manifeste.

L'algorithme de Smith-Waterman par bande effectue des alignements de séquence semi-globale pour déterminer les régions similaires entre deux séquences. Au lieu de comparer l'intégralité de la séquence, l'algorithme de Smith-Waterman compare les segments de toutes les longueurs possibles.

Chaque lecture appariée est évaluée en fonction de son alignement sur les séquences de sonde correspondantes.

- ▶ La lecture 1 est évaluée par rapport au complément inverse des oligos spécifiques aux loci en aval (DLSO).
- ▶ La lecture 2 est évaluée par rapport aux oligos spécifiques aux loci en amont (ULSO).
- ▶ Si le début d'une séquence de lecture correspond à une séquence de sonde ne présentant pas plus de trois différences (mésappariements ou décalages attribuables aux principaux indels), la longueur totale de la lecture est alors alignée sur la séquence d'amplicon cible pour cette séquence.
- ▶ Les indels dans les DLSO et ULSO ne sont pas observés étant donné la chimie du test.

Les alignements sont filtrés en fonction des résultats de l'alignement basés sur les taux de mésappariements par rapport à la région d'intérêt ou à l'amplicon complet, selon la longueur de l'amplicon. Les alignements filtrés sont notés dans les fichiers d'alignement comme étant non alignés et ne sont pas utilisés pour l'appel des variants.

Appel des variants

Développé par Illumina, l'appel des variants Pisces identifie les variants à faible fréquence présents dans un échantillon d'ADN.

L'appel des variants Pisces identifie les SNP en trois étapes :

- ▶ Évaluer séparément chaque position dans le génome de référence.
- ▶ Compter les bases à la position donnée pour les lectures alignées qui chevauchent la position.
- ▶ Établir le score des variants qui mesure la qualité de l'appel en utilisant le modèle Poisson. Les variants ayant un score de qualité inférieur à Q30 sont exclus.

Les variants sont d'abord appelés séparément pour chacun des pools. Ensuite, les variants de chacun des pools sont comparés et combinés dans un fichier de sortie unique. Si un variant satisfait aux critères suivants, la mention PASS (réussite) lui est accordée dans le fichier d'appels de variants (VCF) :



- ▶ Le variant est présent dans les deux pools;
- ▶ Il a une profondeur cumulative de 900x (au moins 450x par pool);
- ▶ Il a une fréquence de variant $\geq 2,6\%$, selon ce qui est indiqué dans le fichier VCF fusionné.

Pisces (Paramètre d'appel des variants somatiques)

Pisces effectue l'appel des variants somatiques pour identifier les variants à faible fréquence présents dans les échantillons d'ADN. L'application peut être utilisée avec tous les échantillons et elle génère des fichiers VCF et gVCF.

Pour obtenir plus de renseignements, rendez-vous à l'adresse github.com/Illumina/Pisces/wiki.

Affichage des données concernant l'analyse et les échantillons

- 1 Dans le tableau de bord de Local Run Manager, cliquez sur le nom de l'analyse.
- 2 À l'onglet Run Overview (Aperçu de l'analyse), vérifiez les indicateurs de l'analyse de séquençage.
- 3 [Facultatif] Cliquez sur l'icône **Copy to Clipboard**  (Copier dans le bloc-notes) pour copier le chemin de sortie du dossier d'analyse.
- 4 Cliquez sur l'onglet Sequencing Information (Renseignements sur le séquençage) pour vérifier les paramètres de l'analyse et les renseignements sur les consommables.
- 5 Cliquez sur l'onglet Samples and Results (Échantillons et résultats) pour voir l'emplacement du rapport d'analyse.
 - ▶ Si l'analyse est répétée, développez le menu déroulant Select Analysis (Sélectionner l'analyse) et choisissez l'analyse appropriée.
- 6 Cliquez sur l'icône **Copy to Clipboard**  (Copier dans le bloc-notes) pour copier le chemin du dossier d'analyse.

Pour plus de renseignements sur les onglets Run Overview (Aperçu de l'analyse) et Sequencing Information (Renseignements sur le séquençage), ainsi que sur la façon de remettre une analyse en file d'attente, consultez le *Guide de référence du logiciel Local Run Manager (document no 1000000011880)*.

Rapport d'analyse

Les résultats d'analyse sont résumés à l'onglet Samples and Results (Échantillons et résultats) et dans un rapport agrégé dans le dossier Alignment (Alignement). Un rapport est aussi offert en format PDF pour chaque échantillon.

Renseignements sur l'onglet Sample and Details (Échantillon et précisions)

- 1 Cliquez sur un échantillon se trouvant dans la liste pour voir le rapport d'échantillon.

Tableau 1 Tableau des renseignements sur l'analyse

Titre de la colonne	Description
Run Status (État de l'analyse)	Indique si l'analyse de séquençage est réussie ou a échoué.
Total Yield (GB) (Rendement total (Go))	Nombre de bases définies dans l'analyse de séquençage. Indique le seuil de réussite, ainsi que l'état (réussite ou échec).
% ≥ Q30	Le pourcentage de lectures dans l'analyse de séquençage ayant un score de qualité de 30(Q30) ou plus. Indique le seuil de réussite, ainsi que l'état (réussite ou échec).
Sample Name (Nom de l'échantillon)	Le nom de l'échantillon indiqué lors de la création de l'analyse.

Titre de la colonne	Description
Total PF Reads (Lectures PF totales)	Le nombre total de lectures passant le filtre.
Read 1% ≥ Q30 (Lecture 1 % ≥ Q30)	Le pourcentage de lectures dans la lecture 1 ayant un score de qualité de 30(Q30) ou plus pour l'échantillon.
Read 2% ≥ Q30 (Lecture 2 % ≥ Q30)	Le pourcentage de lectures dans la lecture 2 ayant un score de qualité de 30(Q30) ou plus pour l'échantillon.
Autosome call rate (Débit d'appel autosome)	Le nombre de positions génomiques dans les autosomes (chromosomes 1 à 22) qui atteignent un seuil de valeur de confiance prédéfini, divisé par le nombre total de positions génomiques autosomiques interrogées. Le débit d'appel est décrit pour chaque échantillon et indiqué par un pourcentage qui est calculé comme suit : 1 moins [nombre de positions autosomiques avec appels incomplets divisé par le nombre total de positions autosomiques séquencées].

Tableau 2 Tableau des rapports d'échantillon

Titre de la colonne	Description
Sample (Échantillon)	Le nom de l'échantillon indiqué lors de la création de l'analyse.
Report Date (Date du rapport)	Date à laquelle le rapport est généré.
Renseignements sur l'échantillon	L'identifiant de l'échantillon fourni lors de la création de l'analyse, le nombre total de lecture ayant passé le filtre dans l'échantillon, le pourcentage de lectures pour l'échantillon ayant un score de qualité de 30(Q30) ou plus et le débit d'appel autosomique.
Amplicon Summary (Résumé de l'amplicon)	Nombre total de régions d'amplicon séquencées et longueur totale dans les paires de bases des amplicons séquencés dans les régions ciblées, pour l'échantillon du pool A et du pool B, et fichier de manifeste utilisé pour chacun des pools. Le fichier de manifeste indique le génome de référence et les régions de référence ciblées utilisés à l'étape de l'alignement.
Read Level Statistics (Statistiques du niveau de lecture)	Nombre et pourcentage des lectures pour l'échantillon couvrant chaque position dans la référence, pour la lecture 1 et la lecture 2, dans le pool A et le pool B.
Variants Summary (Résumé du variant)	Nombre de SNV, d'insertions et de suppressions détectés pour l'échantillon ayant passé les valeurs suggérées, pour déterminer si les résultats de qualité se situent à un niveau acceptable.
Coverage Summary (Résumé de la couverture)	Le nombre total de bases alignées divisé par la taille de la région ciblée, et le pourcentage de régions d'amplicon ayant des valeurs de couverture supérieures au seuil de faible couverture correspondant à 0.2 * couverture moyenne des amplicons, pour l'échantillon, dans le pool A et le pool B.
Coverage Plots (Courbes de couverture)	La couverture par courbe de région d'amplicon affiche la couverture dans les régions d'amplicon, pour l'échantillon. Les régions ayant des valeurs de couverture inférieures au seuil de couverture sont surlignées en rouge. La moyenne de toutes les valeurs est indiquée par une ligne orange. Une courbe est fournie pour la couverture du pool A et du pool B.
Software Versions (Versions du logiciel)	Les versions de logiciels utilisées au moment du séquençage, notamment le logiciel d'exploitation de MiSeq, le logiciel Local Run Manager, le logiciel RTA et la version du module d'analyse des variants somatiques.

Fichiers de sortie d'analyse

Les fichiers de sortie d'analyse suivants sont générés pour le module d'analyse des variants somatiques et fournissent des résultats d'analyse pour l'alignement et l'appel des variants. Les fichiers de sortie d'analyse se trouvent dans le dossier Alignment (Alignement).

Nom du fichier	Description
Demultiplexing (*.txt)	Fichiers intermédiaires contenant les résultats sommaires du démultiplexage.
FASTQ (*.fastq.gz)	Fichiers intermédiaires contenant les appels de bases dont la qualité est notée. Les fichiers FASTQ constituent les principales données d'entrée pour l'étape de l'alignement.
Fichiers d'alignement dans le format BAM (*.bam)	Contiennent les lectures alignées pour un échantillon donné.
Fichiers d'appels de variants, par pool, dans le format VCF (*.vcf)	Contiennent les variants appelés à chacune des positions du pool sens ou du pool antisens.
Fichiers d'appels de variants dans le format génome VCF (*.genome.vcf.gz)	Contiennent le génotype pour chacune des positions, qu'elles soient appelées à titre de variant ou de référence.
Fichiers d'appels des variants consensus dans le format VCF (*.vcf.gz)	Contiennent les variants appelés à chacune des positions des deux pools.
AmpliconCoverage_M1.tsv	Contient l'information sur la couverture par amplicon, par échantillon, pour chaque fichier de manifeste fourni. M# représente le numéro de manifeste.

Format de fichier de démultiplexage

Le processus de démultiplexage lit la séquence d'indexage liée à chaque amplifiât pour déterminer de quel échantillon provient l'amplifiât. Le mappage entre les amplifiâts et le numéro de l'échantillon est enregistré dans un fichier de démultiplexage (*.demux) pour chacune des plaques de la Flow Cell.

Les fichiers de démultiplexage utilisent le format de dénomination s_1_X.demux, où X correspond au numéro de la plaque.

Les fichiers de démultiplexage commencent par un en-tête :

- ▶ Version (nombre entier de 4 octets), actuellement 1
- ▶ Nombre d'amplifiâts (nombre entier de 4 octets)

Le reste du nom est composé du numéro des échantillons pour chacun des amplifiâts de la plaque.

Lorsque l'étape du démultiplexage est terminée, le logiciel génère un fichier nommé DemultiplexSummaryF1L1.txt.

- ▶ Dans le nom du fichier, F1 représente le numéro de la Flow Cell.
- ▶ Dans le nom du fichier, L1 représente le numéro de la ligne.
- ▶ Les résultats de démultiplexage sont affichés dans un tableau comportant une ligne par plaque et une colonne par échantillon, y compris l'échantillon 0.
- ▶ Séquences les plus courantes dans les lectures d'index.

Format de fichier FASTQ

FASTQ est un format de fichier texte qui contient les appels de bases et les valeurs de qualité, par lecture. Chaque dossier contient quatre lignes :

- ▶ L'identifiant
- ▶ La séquence
- ▶ Le signe plus (+)
- ▶ Les scores de qualité Phred dans un format codé ASCII + 33

L'identifiant a le format suivant :

@Instrument:IDAnalyse:IDFlowCell:Ligne:Plaque:X:Y NumLecture:MarqueurFiltre:0:NuméroÉchantillon

Exemple :

```
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
+
<>;##=><9=AAAAAAAAAAA9#:<#<;<<<?????#=#
```

Format de fichier BAM

Un fichier BAM (*.bam) est la version binaire compressée d'un fichier SAM utilisée pour représenter des séquences alignées, jusqu'à 128 Mo. Les formats SAM et BAM sont décrits en détail à l'adresse samtools.github.io/hts-specs/SAMv1.pdf.

Les fichiers BAM utilisent le format de dénomination de fichier NomÉchantillon_S#.bam, où le symbole « # » représente le numéro de l'échantillon en fonction de l'ordre dans lequel les échantillons sont classés pour l'analyse.

Les fichiers BAM contiennent une section d'en-tête et une section pour les alignements.

- ▶ **Header** (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.
- ▶ **Alignments** (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
- ▶ **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

Les fichiers d'index BAM (*.bam.bai) fournissent un index du fichier BAM correspondant.

Format de fichier VCF

Le Variant Call Format (VCF) est un format de fichier courant, développé par la communauté scientifique travaillant sur la génomique. Il contient des renseignements sur les variants que l'on trouve à des positions spécifiques dans un génome de référence. Les fichiers VCF se terminent par le suffixe .vcf.

L'en-tête du fichier VCF comprend la version du format de fichier VCF, la version du paramètre d'appel des variants et les annotations utilisées dans le reste du fichier. L'en-tête du fichier comprend aussi le fichier du génome de référence et le fichier BAM. 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 un seul variant.

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 FASTQ.
POS	Position de la base unique du variant dans le chromosome de référence. Pour les SNP, cette position est la base de référence avec le variant pour les indels ou les délétions. Il s'agit de la base de référence immédiatement avant le variant.
ID	Numéro rs pour le variant obtenu à partir du fichier dbSNP.txt le cas échéant. S'il existe plusieurs numéros rs à cet emplacement, la liste est délimitée par des points-virgules. Si aucune entrée dbSNP n'existe à cette position, un marqueur de valeur manquante ('.') est utilisé.
RÉF	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.
ALT	Allèles qui diffèrent de la lecture de référence. Par exemple, une délétion d'un T unique est représentée TT 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 d'appel 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 d'appels Q30 est associé à un taux d'erreur de 0,1 %. De nombreux paramètres d'appel des variants attribuent des scores de qualité en fonction de leurs modèles statistiques, qui sont élevés par rapport au taux d'erreur observé.

Annotations du fichier VCF

En-tête	Description
FILTER (Filtre)	<p>Si tous les filtres sont passés, PASS (Réussite) s'inscrit dans la colonne Filter (Filtre).</p> <ul style="list-style-type: none"> • LowDP : appliqué aux sites dont la profondeur de couverture est inférieure à 450x dans l'un des pools. Pour les positions d'amplicon couvertes par les lectures avant et arrière, la couverture de lecture unique est équivalente à 900x. • LowGQ : la qualité du génotypage (GQ) est inférieure à la limite. • q30 : score de qualité < 30. • LowVariantFreq : la fréquence du variant est inférieure au seuil donné. • PB : biais des pools sondes. Le variant n'a pas été trouvé ou a été trouvé à une faible fréquence dans un ou deux pools sondes. • R3x6 : nombre de répétitions adjacentes (longueur de 1 à 3 pb) aux appels de variants ≥ 6. • SB : le biais lié au brin est supérieur au seuil donné.
INFO	<p>Entrées possibles dans la colonne INFO :</p> <ul style="list-style-type: none"> • AC : comptage des allèles dans les génotypes pour chaque allèle ALT dans le même ordre que celui indiqué. • AF : fréquence des allèles pour chaque allèle ALT dans le même ordre que celui indiqué. • AN : nombre total d'allèles dans les génotypes appelés. • CD : indicateur précisant que le PNS est survenu dans la région de codage d'au moins une entrée du gène de référence. • DP : profondeur (nombre d'appels de bases alignés à une position et utilisés dans l'appel de variant). • Exon : liste séparée par des virgules d'une lecture de régions exoniques à partir du gène de référence. • FC : conséquence fonctionnelle. • GI : liste séparée par des virgules d'une lecture d'identifiants génétiques à partir du gène de référence. • QD : fiabilité et qualité des variants basées sur la profondeur. • TI : liste séparée par des virgules d'une lecture d'identifiants de transcrits à partir du gène de référence.
FORMAT	<p>La colonne Format dresse la liste des champs séparés par deux-points. Par exemple, GT:GQ. La liste des champs fournie dépend du paramètre d'appel des variants utilisé. Champs offerts :</p> <ul style="list-style-type: none"> • AD : entrée du format X, Y, où X est le nombre d'appels de référence et Y le nombre d'appels autres. • DP : profondeur de lecture approximative; les lectures avec MQ=255 ou de mauvais appariements sont filtrées. • GQ : qualité du génotype. • GQX : qualité du génotype. GQX est le minimum de la valeur GQ et de la colonne QUAL. En général, ces valeurs sont identiques; en prenant en compte la valeur minimale, GQX est la mesure la plus conservatrice pour mesurer la qualité du génotype. • GT : génotype. 0 correspond à la base de référence, 1 correspond à la première entrée dans la colonne ALT, etc. La barre oblique (/) indique qu'aucun renseignement relatif à la mise en phase n'est disponible. • NC : fraction des bases non appelées, ou qualité des appels de bases inférieure au seuil minimum. • NL : degré de bruit; une estimation du bruit des appels de bases à cette position. • PB : biais des pools sondes. Les valeurs près de 0 indiquent davantage de biais relativement au pool sonde et moins de fiabilité de l'appel de variant. • SB : biais lié au brin à cette position. Des valeurs négatives importantes indiquent un biais moins important; des valeurs près de 0 indiquent un biais plus important. • VF : fréquence du variant; pourcentage de lectures prenant en charge l'allèle alternatif.
SAMPLE (Échantillon)	La colonne relative aux échantillons indique les valeurs précisées dans la colonne FORMAT.

Fichiers VCF du génome

Les fichiers VCF du génome (gVCF) sont des fichiers VCF v4.1 qui respectent un ensemble de conventions pour la représentation de tous les sites du génome dans un format raisonnablement compact. Les fichiers gVCF (*.genome.vcf.gz) comprennent tous les sites de la région d'intérêt dans un fichier unique, pour chacun des échantillons.

Le fichier gVCF affiche « aucun appel » aux positions qui ne passent pas tous les filtres. L'indicateur de génotype (GT) ./. indique qu'il n'y a aucun appel.

Pour plus de renseignements, rendez-vous à l'adresse Web sites.google.com/site/gvcftools/home/about-gvcf.

Fichiers VCF consensus et fichiers VCF par pool

Le flux de travail des variants somatiques génèrent deux ensembles de fichiers d'appels de variants.

- ▶ **Per-pool VCF files** (Fichiers VCF par pool) : contiennent les variants appelés dans le pool sens ou le pool antisens. Les fichiers par pool sont enregistrés dans le dossier VariantCallingLogs (Journaux d'appels des variants).
- ▶ **Consensus VCF files** (Fichiers VCF consensus) : contiennent les variants appelés dans les deux pools. Les fichiers consensus sont enregistrés dans le dossier Alignment (Alignement).

Les fichiers VCF consensus et fichiers VCF par pool comprennent à la fois des fichiers VCF (*.vcf) et gVCF (*.genome.vcf) et suivent la convention de nomenclature suivante, dans laquelle S# représente l'ordre dans lequel se trouve l'échantillon pour l'analyse :

- ▶ **Reports for all sites** (Rapports pour tous les sites) : Nomdel'échantillon_S#.genome.vcf
- ▶ **Reports variants only** (Rapports sur les variants seulement) : Nomdel'échantillon_S#.vcf

Le logiciel compare les fichiers VCF par pool et combine les données de chacune des positions pour créer un fichier VCF consensus pour l'échantillon.

Les appels de variants provenant de chacun des pools sont fusionnés dans les fichiers VCF consensus selon les critères suivants.

Critères	Résultats
Un appel de référence dans chaque pool.	Reference call (Appel de référence)
Un appel de référence dans un pool et un appel de variant dans l'autre pool.	Filtered variant call (Appel de variant filtré)
Les appels de variants concordent et ont des fréquences similaires dans chaque pool.	Variant call (Appel de variants)
Les appels de variants concordent, mais les fréquences sont très différentes entre les deux pools.	Filtered variant call (Appel de variants filtré)
Les appels de variants ne concordent pas dans les deux pools.	Filtered variant call (Appel de variants filtré)

Les indicateurs de chaque pool sont fusionnés en fonction des valeurs suivantes.

Indicateurs	Valeur
Profondeur	Addition des profondeurs de chacun des deux pools
Fréquence des variants	Nombre total de variants divisé par la profondeur totale de la couverture
Score de qualité	Valeur minimale des deux pools

Fichier de couverture des amplicons

Un fichier de couverture des amplicons est généré pour chaque fichier de manifeste. Dans le nom du fichier, M# désigne le numéro du manifeste tel qu'il figure dans le tableau d'échantillons de l'analyse.

Chaque fichier comprend une ligne d'en-tête contenant les identifiants d'échantillon associés au manifeste. Sous la ligne d'en-tête, il y a trois colonnes comportant les renseignements suivants :

- ▶ L'identifiant de la cible, tel qu'il figure dans le manifeste.
- ▶ La profondeur de couverture des lectures passant le filtre.
- ▶ La profondeur de couverture totale.

Fichiers de sortie supplémentaires

Les fichiers de sortie supplémentaires fournissent des renseignements complémentaires ou résument les résultats de l'analyse ainsi que les erreurs associées. Bien que ces fichiers ne soient pas requis pour l'évaluation des résultats de l'analyse, ils peuvent être utilisés à des fins de dépannage. Tous les fichiers se trouvent dans le dossier Alignment (Alignement), à moins d'indication contraire.

Nom du fichier	Description
AnalysisLog.txt	Journal de traitement qui décrit chaque étape survenue au cours de l'analyse du dossier d'analyse actuel. Ce fichier ne contient pas de messages d'erreur. Il est situé dans le dossier Alignment (Alignement).
AnalysisError.txt	Journal de traitement qui répertorie toutes les erreurs qui se sont produites au cours de l'analyse. Le fichier sera vide si aucune erreur ne se produit. Il est situé dans le dossier Alignment (Alignement).
DemultiplexSummaryF1L1#.txt	Résultats de démultiplexité des rapports dans un tableau avec une rangée par plaque et une colonne par échantillon. Le symbole « # » représente la ligne 1, 2, 3 ou 4 de la Flow Cell. Il est situé dans le dossier Alignment (Alignement).
AmpliconRunStatistics.xml	Contient des statistiques récapitulatives propres à l'analyse. Il est situé dans le dossier Alignment (Alignement).

Dossier d'analyse

Le dossier d'analyse contient les fichiers générés par le logiciel Local Run Manager.


La relation entre le dossier de sortie et le dossier d'analyse se résume comme suit :

- ▶ Durant le séquençage, le logiciel Real-Time Analysis (RTA) remplit le fichier de sortie avec des fichiers générés lors de l'analyse des images, des appels de bases et de la notation de la qualité.
- ▶ RTA copie les fichiers dans le dossier d'analyse en temps réel. Après que RTA ait affecté un score de qualité à chacune des bases de chaque cycle, le logiciel enregistre le fichier RTAComplete.txt dans les deux dossiers.
- ▶ Lorsque le fichier RTAComplete.txt est présent, l'analyse commence.
- ▶ Pendant que l'analyse se poursuit, Local Run Manager enregistre les fichiers de sortie dans le dossier d'analyse, puis recopie les fichiers dans le dossier de sortie.










Dossiers Alignment (Alignement)

Chaque fois que l'analyse est remise en file d'attente, Local Run Manager crée un dossier Alignment (Alignement) nommé **Alignment_N**, où N est un numéro séquentiel.

Structure de dossier

 **Alignment** (Alignement) : Contient des fichiers *.bam, *.vcf et FASTQ, ainsi que des fichiers propres au module d'analyse.



 **Date and Time Stamp** (Horodatage) : horodatage de l'analyse (AAAAMMJJ_HHMMSS)

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

 **FASTQ**

 **Sample1** (Échantillon 1)

 **Stats** (Statistiques)

-  DemuxSummaryF1L1.txt
-  FastqSummaryF1L1.txt

 **Data** (Données)


 **Intensities** (Intensités)

 **BaseCalls** (Appels de bases)

 **L001** : contient un sous-dossier par cycle, chacun contenant des fichiers *.bcl.

 **L001** : contient des fichiers *.locs, un pour chaque plaque.

 **RTA Logs** (Journaux RTA) : contient les fichiers journaux de l'analyse du logiciel RTA.

 **InterOp** : contient les fichiers binaires utilisés pour déclarer les indicateurs de l'analyse de séquençage.

 **Logs** (Journaux) : contient les fichiers journaux décrivant les étapes réalisées au cours du séquençage.

 RTAComplete.txt

 RunInfo.xml

 runParameters.xml

Assistance technique

Pour obtenir de l'assistance technique, communiquez avec l'assistance technique d'Illumina.

Site Web : www.illumina.com
 Courriel : techsupport@illumina.com

Numéros de téléphone de l'assistance clientèle d'Illumina

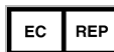
Région	Sans frais	Numéro régional
Amérique du Nord	+1 800 809-4566	
Allemagne	+49 8001014940	+49 8938035677
Australie	+1.800.775.688	
Autriche	+43 800006249	+43 19286540
Belgique	+32 80077160	+32 34002973
Chine	400 635 9898	
Danemark	+45 80820183	+45 89871156
Espagne	+34 911899417	+34 800300143
Finlande	+358 800918363	+358 974790110
France	+33 805102193	+33 170770446
Hong Kong	800960230	
Irlande	+353 1800936608	+353 016950506
Italie	+39 800985513	+39 236003759
Japon	0800 111 5011	
Norvège	+47 800 16836	+47 21939693
Nouvelle-Zélande	0 800 451 650	
Pays-Bas	+31 8000222493	+31 207132960
Royaume-Uni	+44 8000126019	+44 2073057197
Singapour	+1 800 579 2745	
Suisse	+41 565800000	+41 800200442
Suède	+46 850619671	+46 200883979
Taiwan	00806 65 1752	
Autres pays	+44,1799 534000	

Fiches signalétiques (SDS) : disponibles sur le site Web d'Illumina à l'adresse support.illumina.com/sds.html.

Documentation produit : disponible en téléchargement au format PDF sur le site Web d'Illumina. Rendez-vous sur support.illumina.com, sélectionnez un produit, puis cliquez sur **Documentation & Literature** (Documentation).



Illumina
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 Cambridge Limited
Chesterford Research Park, Little Chesterford
Saffron Walden, CB10 1XL
ROYAUME-UNI

Commanditaire australien :
Illumina Australia
1 International Court
Scoresby, Victoria, 3179
Australie

DESTINÉ AU DIAGNOSTIC IN VITRO UNIQUEMENT

© 2017 Illumina, Inc. Tous droits réservés.

illumina®