

Plateforme DRAGEN^{MC} Bio-IT intégrée aux systèmes NextSeq^{MC} 1000 et NextSeq 2000

Solution intégrée précise
et efficace pour l'analyse
des données de SNG

- Fournit un flux de travail rationalisé pour une analyse secondaire très précise avec une interface graphique facile à utiliser
- Génère des appels de variants de haute qualité directement sur les systèmes NextSeq 1000 et NextSeq 2000 intégrés.
- Réduit le besoin supplémentaire d'infrastructure de bioinformatique et requiert moins d'interventions manuelles que d'autres offres du logiciel DRAGEN

illumina[®]

Introduction

Libérer la puissance du génome est un objectif partagé par de nombreux laboratoires dans le monde. Les progrès réalisés ces dernières années en matière de technologie améliorent les plateformes de séquençage de nouvelle génération (SNG) et les méthodes qui peuvent permettre de percer les secrets du génome. Les systèmes NextSeq 1000 et NextSeq 2000 prennent en charge les applications à débit faible et moyen, notamment le séquençage des exomes, des transcriptomes et des petits génomes; l'enrichissement de la cible; le séquençage unicellulaire et la métagénomique aléatoires.

L'un des principaux avantages des systèmes NextSeq 1000 et NextSeq 2000 est l'inclusion d'une mise en œuvre intégrée de la plateforme Bio-IT DRAGEN (Dynamic Read Analysis for GENomics). Le logiciel DRAGEN intégré assure aux laboratoires une analyse SNG secondaire exceptionnellement rapide, précise et complète, réduisant ainsi le besoin d'infrastructure informatique et de ressources bioinformatiques supplémentaires. Avec le logiciel DRAGEN intégré, l'analyse secondaire peut être configurée durant la configuration de l'analyse de séquençage, permettant un flux de travail rationalisé avec moins d'interventions manuelles de l'utilisateur et un temps de traitement plus rapide. Pour les laboratoires cherchant un niveau élevé de configurabilité, Illumina offre des options supplémentaires de mise en œuvre du logiciel DRAGEN (Figure 1).

Le logiciel DRAGEN intégré est le complément parfait aux systèmes polyvalents et évolutifs NextSeq 1000 et NextSeq 2000. En incluant plusieurs des meilleurs algorithmes de pipeline, le logiciel DRAGEN intégré permet aux chercheurs de surmonter les congestions dans l'analyse des données et de générer rapidement des résultats précis pour une large gamme d'applications (Tableau 1).

Efficacité

Le logiciel DRAGEN intégrant les systèmes NextSeq 1000 et NextSeq 2000 permet aux laboratoires d'intégrer parfaitement les fonctions de séquençage et d'analyse dans leur configuration d'analyse. Le logiciel DRAGEN intégré permet aux laboratoires d'obtenir des fichiers FASTQ et des fichiers au format d'appel de variants (VCF) pour les applications en aval directement à partir de l'instrument à mesure que l'analyse est effectuée, en enregistrant le transfert de fichiers et la durée d'analyse. Les algorithmes des pipelines de la plateforme DRAGEN intégrée permettent aux utilisateurs novices et experts de réaliser des fonctions d'analyse communes et de réduire les recours à des experts informatiques externes. Les fichiers volumineux de données sont facilement gérés à l'aide de la technologie de stockage des fichiers ORA (Original Read Archive, archive des lectures d'origine) de DRAGEN, qui atteint une compression moyenne quatre fois sans perte des fichiers FASTQ et utilise 80 % d'énergie en moins pour un lieu de travail plus durable.

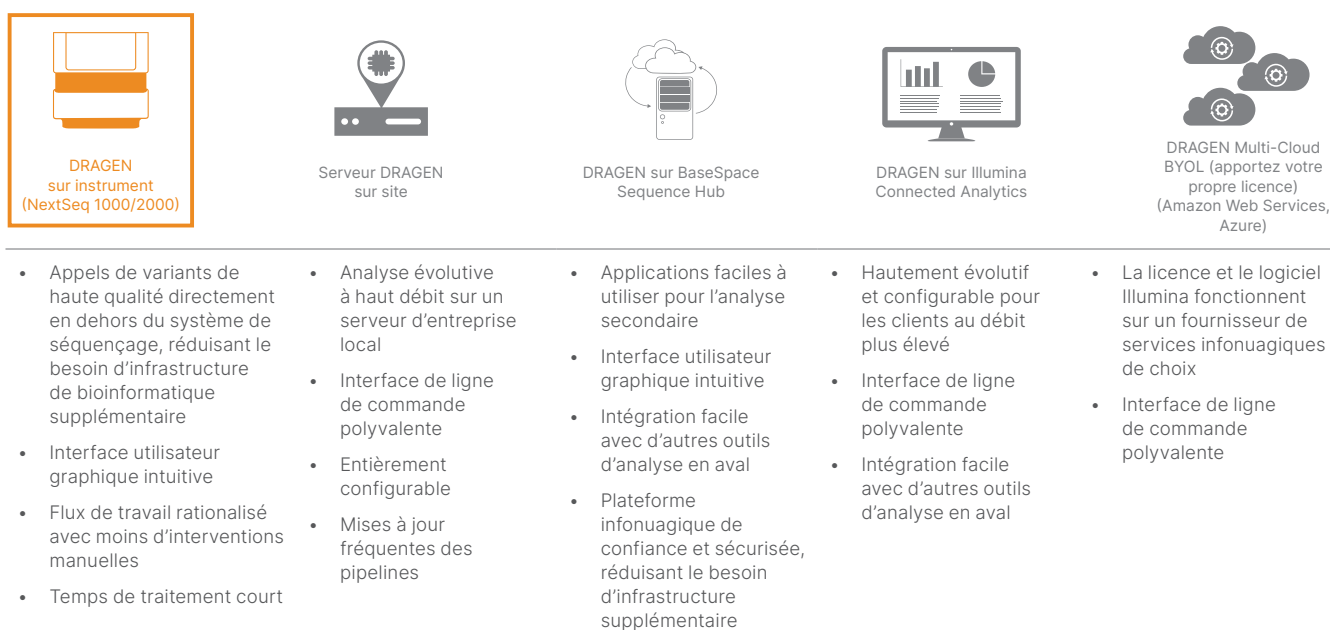


Figure 1 : Options pour la mise en place des pipelines de DRAGEN Bio-IT pour répondre aux besoins en analyse de SNG de tous les laboratoires.

Tableau 1 : Applications du logiciel DRAGEN intégrant les systèmes NextSeq 1000 et NextSeq 2000

Application ^a	Description
Conversion BCL	Convertit les fichiers BCL produits par les systèmes de séquençage Illumina en fichiers FASTQ
Compression d'ORA DRAGEN	Produit une compression sans perte et basée sur la référence de fichiers FASTQ
DRAGEN FASTQ + MultiQC	Mesure des indicateurs FastQC accélérés par le matériel sans durée d'analyse supplémentaire
Génome entier	Effectue un mappage du génome humain, un alignement et un appel de petits variants. Germinal uniquement
Enrichissement (exome inclus)	Effectue un appel de petits variants dans les échantillons de variants germinaux, ou un appel de variants de basse fréquence dans les échantillons somatiques
Amplicon d'ADN	Analyse la variation génétique dans des régions génomiques spécifiques. Utilise le pipeline d'ADN de la plateforme DRAGEN avec une étape supplémentaire pour écarter les primeurs et remanier les alignements, s'assurant ainsi que les séquences des primers ne contribuent pas aux appels de variants
ARN	Offre un aligneur (d'épissage) de séquençage de l'ARN avec un filtrage d'ARNr facultatif durant l'alignement, réduisant ainsi la durée de l'analyse et la taille du fichier
ARN unicellulaire	Traite un large éventail d'ensembles de données de séquençage de l'ARN à partir de cellules uniques à partir de lectures en matrices d'expression cellulaire par gène
Expression différentielle	Exécute l'algorithme DESeq2 sur les données de quantification d'ARN produites par le pipeline d'ARN de la plateforme DRAGEN. Gènes de sortie et transcriptions qui sont exprimés différemment entre deux groupes d'échantillons
SNG NanoString GeoMx	Simplifie l'analyse pour les clients qui utilisent à la fois les instruments SNG GeoMx et Illumina pour les flux de travail spatiaux génomiques
Méthylation	Traite des données de méthylation du séquençage de bisulfite et de pyridine assistée de la TET (TAPS, tet-assisted pyridine sequencing)

a. Des pipelines d'application supplémentaires sont disponibles sur BaseSpace Sequence Hub, Illumina Connected Analytics, DRAGEN Multi-Cloud et le serveur DRAGEN sur site

Précision

Les pipelines d'analyse DRAGEN génèrent des résultats exceptionnellement précis. Lors du 2020 Precision FDA Truth Challenge V2 (PrecisionFDA V2), la plateforme DRAGEN v3.7 a été déclarée la plus précise dans toutes les régions de comparaison et les régions difficiles à mapper pour les données de séquençage d'Illumina. D'autres innovations en matière de génomes présentés sous forme graphique et d'apprentissage automatique Illumina obtenues grâce à la plateforme DRAGEN v3.10 ont permis d'établir une nouvelle norme quant à l'exactitude des données dans l'ensemble des technologies de séquençage, obtenant ainsi un score F1 de 99,83 % dans tous les ensembles de données des régions de comparaison, (un calcul des résultats vrais positifs et vrais négatifs par rapport aux résultats totaux) (Figure 2). DRAGEN 3.10, Graph. et les outils d'apprentissage automatique Illumina ont également obtenu le meilleur score F1 dans la catégorie « appels les plus précis » par rapport à l'ensemble des soumissions PrecisionFDA V2 dans les régions du complexe majeur d'histocompatibilité (CMH).^{1,2}

La plateforme Bio-IT DRAGEN exceptionnellement efficace permet d'obtenir des durées d'analyse rapides par l'utilisation de la technologie de réseau prédéfini programmable par l'utilisation (FPGA, field-programmable gate array). La carte FGPA intégrée, propre aux systèmes NextSeq 1000 et NextSeq 2000, permet l'accélération matérielle des pipelines informatiques d'analyse secondaire pour de nombreuses applications. Ces pipelines DRAGEN sont améliorés en continu et des pipelines supplémentaires sont libérés pour assurer la meilleure fonctionnalité, précision et vitesse possible. La plateforme a été conçue autour des besoins d'analyse des chercheurs, pour leur permettre de passer moins de temps et d'efforts à l'exécution des pipelines au niveau de la production et se concentrer ainsi davantage sur les résultats.

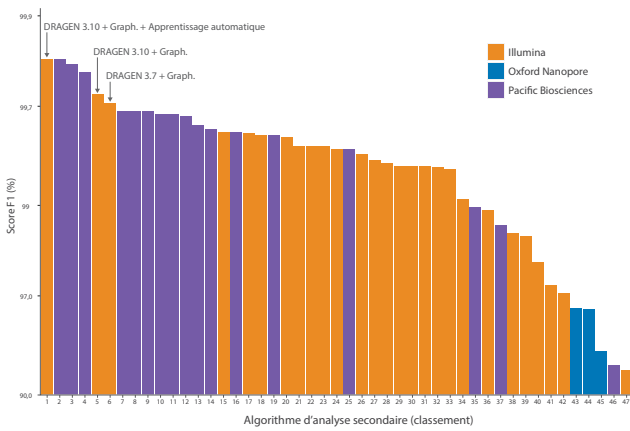


Figure 2 : Précision de l'appel de petits variants avec DRAGEN 3.10, Graph. et des outils d'apprentissage automatique par rapport aux soumissions PrecisionFDA Truth Challenge v2 dans l'ensemble de données Toutes les régions de comparaison — La méthode d'analyse DRAGEN 3.10 + Graph. + Apprentissage automatique est arrivé à la première place, ex aequo avec la soumission de lecture la plus performante de Pacific Biosciences. DRAGEN 3.10 + Graph. a démontré sa supériorité par rapport à DRAGEN 3.7 + Graph. grâce à des améliorations en matière de graphique et de références/contigs ALT.

Un grand choix d'applications

Le logiciel DRAGEN intégrant les systèmes NextSeq 1000 et NextSeq 2000 présente des pipelines d'analyse polyvalents et peut créer plus fichiers de sortie à différentes étapes du flux de travail (Figure 3).

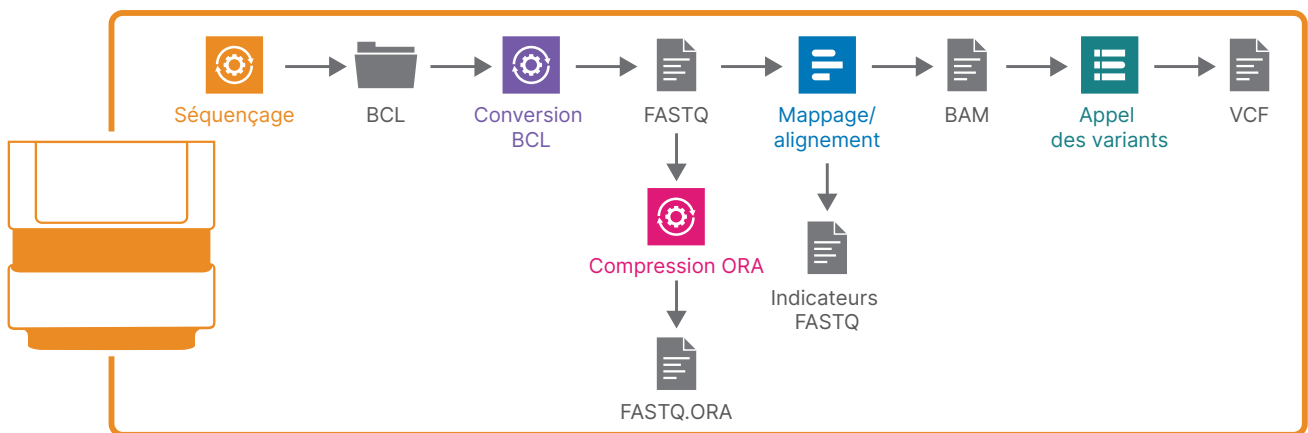


Figure 3 : Le logiciel DRAGEN intégré sur les systèmes NextSeq 1000 et NextSeq 2000 assure un flux de travail de traitement des données et une analyse secondaire — La solution de logiciel DRAGEN intégré fournit un outil très efficace pour produire les résultats dans les formats requis par les laboratoires pour les applications en aval.

La sélection de pipelines d'analyse inclure prend en charge une grande variété de types d'analyse, notamment l'analyse de cellules uniques, d'exome et d'ARN (Tableau 1).

Pipelines d'ARN unicellulaire de la plateforme DRAGEN intégrée

Les pipelines d'ARN unicellulaire de la plateforme DRAGEN sont des solutions rapides et évolutives qui peuvent traiter plusieurs ensembles de données de séquençage de l'ARN à partir de cellules uniques avec des formats de débit faciles à utiliser, notamment des diagrammes de génération d'amplifiats de type cellulaire basés sur les activités d'expression de l'ARN (Figure 4). De plus, il présente des extensions utiles pour traiter les ensembles de données multiplexées comprenant plusieurs échantillons (par ex. utilisation du démultiplexage de génotype ou hachage cellulaire) et comptage de l'expression des protéines de surface cellulaire. Il prend également en charge plusieurs types de préparation de bibliothèques pour être compatible avec les outils d'analyse en aval.

Le CQ intégré des bibliothèques de l'expression des cellules uniques et le pipeline d'analyse unicellulaire sont exécutés en moins de temps par rapport aux solutions infonuagiques et sans que la précision ne soit compromise. Le pipeline unicellulaire réalise aussi bien la configuration de l'analyse que l'expression quantifiée par cellule sur les systèmes NextSeq 1000 et NextSeq 2000 à partir d'une seule intervention manuelle, réduisant ainsi le besoin de ressources informatiques supplémentaires.

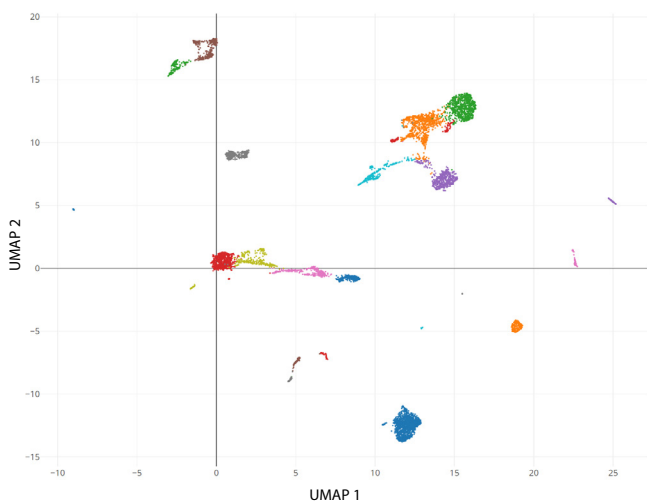


Figure 4 : Exemple de diagramme de génération d'amplifiats de type cellulaire à partir du pipeline d'ARN unicellulaire de la plateforme DRAGEN — L'analyse d'approximation et de projection uniforme de variétés (Uniform Manifold Approximation and Projection, UMAP) permet une visualisation des cellules individuelles par type ou états fonctionnels en fonction de profils d'expression complexes multidimensionnels.⁴

Pipeline d'enrichissement de la plateforme DRAGEN intégrée

Le pipeline d'enrichissement de la plateforme DRAGEN intégrée fournit une analyse rapide, comportant l'étalonnage avancé des modèles d'erreur pour une précision accrue à partir de l'enrichissement et des panels d'exome de type hybridation et des indicateurs basés sur la cible d'entrée. Le logiciel peut être exécuté en mode germinale ou somatique, en effectuant un appel de petits variants germinaux ou un appel de variants somatiques (à faible fréquence). L'appel de variation du nombre de copies (VNC) et l'appel de variants structurels peuvent être activés en mode germinale. Le pipeline d'enrichissement de la plateforme DRAGEN est plus rapide et plus précis que l'analyse avec le paramètre de définition de variants de l'Aligneur Burrows-Wheeler (BWA) et de l'outil Genome Analysis Toolkit (GATK), notamment dans l'appel d'indels (Tableau 2, Figure 5). Une Flow Cell sur les systèmes NextSeq 1000 et NextSeq 2000 peut généralement être traitée en moins de deux heures.

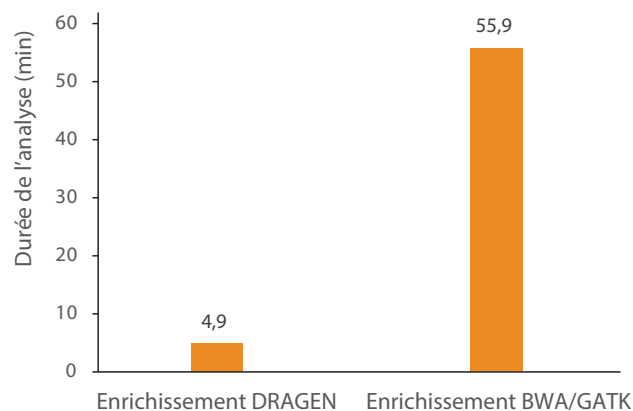


Figure 5 : Durée d'analyse moyenne par échantillon pour le pipeline d'enrichissement DRAGEN et l'application d'enrichissement BWA sur BaseSpace^{MC} Sequence Hub — L'analyse germinale d'enrichissement de 24 SNG montre des durées d'analyse par échantillon plus rapide pour le pipeline d'enrichissement DRAGEN que la même analyse utilisant l'enrichissement BWA suivi d'appels de variants avec le paramètre de définition des variants GATK.

Pipeline d'ARN de la plateforme DRAGEN intégrée

Le pipeline d'ARN de la plateforme DRAGEN intégrée effectue l'analyse secondaire des transcriptions d'ARN. Le pipeline d'ARN offre plusieurs modes de fonctionnement, notamment l'alignement uniquement à titre de référence et l'alignement à l'aide de l'annotation avec la détection des fusions de gènes. Un filtre facultatif d'ARN ribosomal (ARNr) peut être activé pour réduire la durée de l'analyse et la taille du fichier. Le module de fusions de gènes tire parti de l'aligneur d'ARN DRAGEN à épissure pour effectuer une analyse des lectures fractionnées sur les alignements (chimériques) supplémentaires pour détecter les points de rupture potentiels, tout en ajoutant une durée de traitement minimale à l'analyse globale. La détection des fusions de gènes, la quantification des transcriptions et l'appel des variants peuvent être activés à l'aide des fonctionnalités du logiciel DRAGEN intégré.

Tableau 2 : Résultats sur la précision du pipeline d'enrichissement de la plateforme DRAGEN et d'enrichissement BWA/GATK

	SNV			Indel		
	Précision (%)	Rappel (%)	F1	Précision (%)	Rappel (%)	F1
Enrichissement DRAGEN	99,7	94,64	97,11	96,54	83,83	89,74
Enrichissement BWA/GATK	99,82	92,77	96,17	96,71	71,32	82,1

Récapitulatif

Le logiciel DRAGEN Bio-IT intégrant les systèmes NextSeq 1000 et NextSeq 2000 offrent une bioinformatique facile à configurer et une expérience utilisateur simplifiée dans une mise en œuvre locale. Il offre une analyse secondaire des données SNG générées sur l'instrument qui est plus rapide que les pipelines DRAGEN sur BaseSpace^{MC} Sequence Hub. Les pipelines intégrés, tels que les pipelines unicellulaires, d'enrichissement et d'ARN offrent de nombreux avantages en termes de vitesse et de précision par rapport aux offres actuelles.

Renseignements relatifs à la commande

Produit	N° de référence
Système de séquençage NextSeq 1000	20038898
Système de séquençage NextSeq 2000	20038897
Mise à niveau de NextSeq 1000 vers NextSeq 2000	20047256

En savoir plus

Systèmes NextSeq 1000 et NextSeq 2000, illumina.com/systems/sequencing-platforms/nextseq-1000-2000.html

Plateforme Bio-IT DRAGEN, illumina.com/products/by-type/informatics-products/dragen-bio-it-platform.html

Assistance DRAGEN, support.illumina.com/sequencing/sequencing_software/dragen-bio-it-platform.html



Numéro sans frais aux États-Unis : + (1) 800 809 4566 | Téléphone : + (1) 858 202 4566
techsupport@illumina.com | www.illumina.com

© 2022 Illumina, Inc. Tous droits réservés. Toutes les marques de commerce sont la propriété d'Illumina, Inc. ou de leurs détenteurs respectifs. Pour obtenir des renseignements sur les marques de commerce, consultez le lien www.illumina.com/company/legal.html.

M-GL-00729 v1.0 FRA

Références

1. Food & Drug Administration. Truth Challenge V2: Calling Variants from Short and Long Reads in Difficult-to-Map Regions. precision.fda.gov/challenges/10. Accessed March 14, 2022.
2. Illumina. DRAGEN Sets New Standard for Data Accuracy in PrecisionFDA Benchmark Data. Optimizing Variant Calling Performance with Illumina Machine Learning and DRAGEN Graph. illumina.com/science/genomics-research/articles/dragen-shines-again-precisionfda-truth-challenge-v2.html. Accessed March 14, 2022.
3. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011;43(5):491-498. doi:10.1038/ng.806
4. McInnes, L, Healy J, and Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. [arXiv preprint arXiv:1802.03426](https://arxiv.org/abs/1802.03426). Revised Sep 18, 2020. Accessed May 5, 2022. doi.org/10.48550/arXiv.1802.03426