

DRAGEN^{MC} Bio-IT Platform

Analyse secondaire précise,
complète et efficace des
données de séquençage
de nouvelle génération

illumina^{MD}

Introduction

Tirer parti de la puissance du génome via le séquençage de nouvelle génération (SNG) est essentiel pour faire progresser la recherche et la médecine. Pour optimiser les connaissances génétiques issues du SNG, les chercheurs ont besoin d'outils d'analyse de données pouvant traduire de manière précise et efficace des données de séquençage brutes en résultats significatifs. De plus, pour tirer parti des avantages du SNG, les organisations ont besoin de solutions faciles à utiliser qui prennent en charge un certain nombre d'utilisateurs et qui réduisent les obstacles sur le plan financier et technique pour pouvoir l'adopter.

DRAGEN (Dynamic Read Analysis for GENomics, soit Analyse de lecture dynamique pour la génomique en français) Bio-IT Platform d'Illumina a été conçue pour relever d'importants défis associés à l'analyse de données de SNG pour un grand nombre d'applications, notamment les génomes, les exomes, les transcriptomes et les études de méthylome. La plateforme DRAGEN est une suite logicielle d'analyse secondaire qui traite les données de SNG et prend en charge les analyses tertiaires pour approfondir les connaissances. Les outils disponibles en font une solution hautement précise, complète et efficace qui permet aux laboratoires de toutes les tailles et disciplines d'en faire davantage avec leurs données génomiques.

Résultats précis

DRAGEN Bio-IT Platform génère des résultats d'une précision exceptionnelle. 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 référence et les régions difficiles à mapper pour les données de séquençage d'Illumina^{1,2}. Les innovations en matière de génomes présentés sous forme graphique et d'apprentissage automatique (AA) d'Illumina obtenues grâce au logiciel DRAGEN 4.0 démontrent une exactitude des données exceptionnelles dans l'ensemble des technologies de séquençage, avec un score F1 de 99,83 % (mesure combinée de précision et de rappel) dans toutes les régions de référence (figure 1)^{1,2}. DRAGEN 4.0 + Graph (AA activé par défaut) a également obtenu le meilleur score F1 dans la catégorie des appels les plus précis par rapport à l'ensemble des soumissions PrecisionFDA V2 dans les régions du complexe majeur d'histocompatibilité (CMH).

Plateforme complète

DRAGEN Bio-IT Platform répond aux besoins des laboratoires exécutant une grande variété d'applications de SNG, en offrant une couverture complète pour un vaste ensemble de types d'expériences dans une seule plateforme.

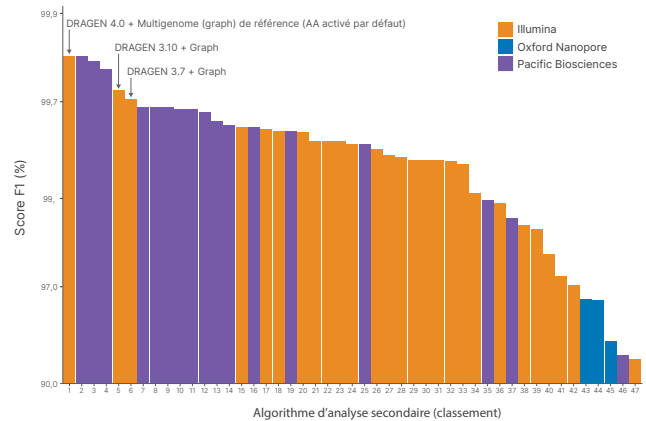


Figure 1 : Précision de DRAGEN 4.0 + Graph (AA activé par défaut) par rapport aux soumissions PrecisionFDA Truth Challenge v2 dans l'ensemble de données de toutes les régions de référence – DRAGEN 4.0 + Graph (AA activé par défaut) montre une précision exceptionnelle, à égalité avec Google DeepVariant sur les données de séquençage 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. Le score F1 (%), indiqué sur l'axe Y, est un calcul des résultats vrais positifs et vrais négatifs par rapport aux résultats totaux^{3,4}.

Les pipelines d'analyse secondaire de DRAGEN prennent en charge plusieurs types d'expériences, notamment le séquençage du génome entier (WGS, Whole-Genome Sequencing), les panels d'enrichissement, le séquençage de l'ARN unicellulaire, le séquençage ATAC-Seq unicellulaire, l'analyse multiomique unicellulaire, le séquençage de l'ARN en vrac et l'analyse de méthylation (tableau 1). Il faudrait plus de 30 outils à code source ouvert pour reproduire partiellement l'étendue des fonctionnalités avec le logiciel DRAGEN^{3,4}.

Grâce à l'ensemble intégré de paramètres d'appel de variants (expansion de répétition, variation structurelle [VS], variation du nombre de copies [VNC], ExpansionHunter) et de paramètres d'appel ciblés tels que *SMN*, *GBA*, *CYP2B6*, *CYP2D6*, et *HLA*, le logiciel DRAGEN offre une large couverture génomique. De plus, DRAGEN Graph. est doté d'une architecture informatique parallèle qui allonge de manière efficace les lectures Illumina et atteint les régions à faible complexité, permettant ainsi d'atteindre les zones du génome difficiles à évaluer en raison de séquences répétées. Cette architecture permet d'améliorer la couverture des gènes potentiellement pertinents sur le plan médical et l'appel des variants mononucléotidiques, des variants du nombre de copies et des variants structurels dans les régions difficiles à mapper.

Analyse efficace

DRAGEN Bio-IT Platform est spécialement conçue pour fournir aux laboratoires la vitesse d'analyse de données et les options de fichier dont ils ont besoin pour tirer le meilleur parti des ensembles de données de SNG. Le logiciel DRAGEN est accéléré par le matériel et utilise une architecture à matrice prédiffusée programmable par l'utilisateur (FPGA, Field Programmable Gate Array) pour obtenir des temps de traitement rapides. L'efficacité des algorithmes d'analyse DRAGEN a permis d'établir deux records mondiaux de vitesse d'analyse de données génomiques^{5,6}. Dans la pratique, DRAGEN Bio-IT Platform peut traiter sur site des données de SNG pour un génome entier équivalent à 34 fois la couverture en 30 minutes sur site environ contre plus de 15 heures environ avec un système traditionnel basé sur une unité centrale de traitement⁷.

Pour faire face aux demandes de stockage de fichiers volumineux de données de SNG, la technologie d'archive des lectures d'origine (ORA, Original Read Archive) de DRAGEN fournit une compression cinq fois sans perte des fichiers FASTQ. La compression sans perte de l'ORA de DRAGEN conserve les détails des fichiers FASTQ et est incroyablement rapide, compressant des fichiers FASTQ de 50 à 70 Go en huit minutes environ. L'ensemble polyvalent de pipelines de DRAGEN peut également accepter des fichiers de données d'entrée et créer des fichiers de sortie à différentes étapes des pipelines (figure 2).

FPGA et accélération par le matériel

La FPGA hautement configurable permet la mise en place ultra efficace par accélération matérielle d'algorithmes d'analyse du génome, comme la conversion de fichiers de définitions des bases (BCL), le mappage, l'alignement, le triage, le marquage des réplicats et l'appel des variants haplotypes. La nature flexible des FPGA permet à Illumina de développer une suite complète de pipelines d'application DRAGEN, avec des mises à jour et des ajouts réguliers pour offrir la meilleure précision, exhaustivité et efficacité possible.

Références personnalisées

DRAGEN Reference Builder permet aux utilisateurs de générer une référence, humaine, non humaine ou non standard; toutes faisant référence aux tableaux de hachage. Les références créées peuvent être utilisées à titre d'entrée pour toutes les applications DRAGEN prenant en charge les fichiers de référence du client. L'application DRAGEN Reference Builder sur BaseSpace^{MC} Sequence Hub requiert un fichier FASTA. La plupart des pipelines DRAGEN comprennent un soutien intégré pour générer hg19, hg38 (avec ou sans HLA*), GRCh37 et hs37d5. DRAGEN Graph Toolkit permet aux utilisateurs d'étendre les fonctions de référence graphique à des références graphiques humaines plus variées également.

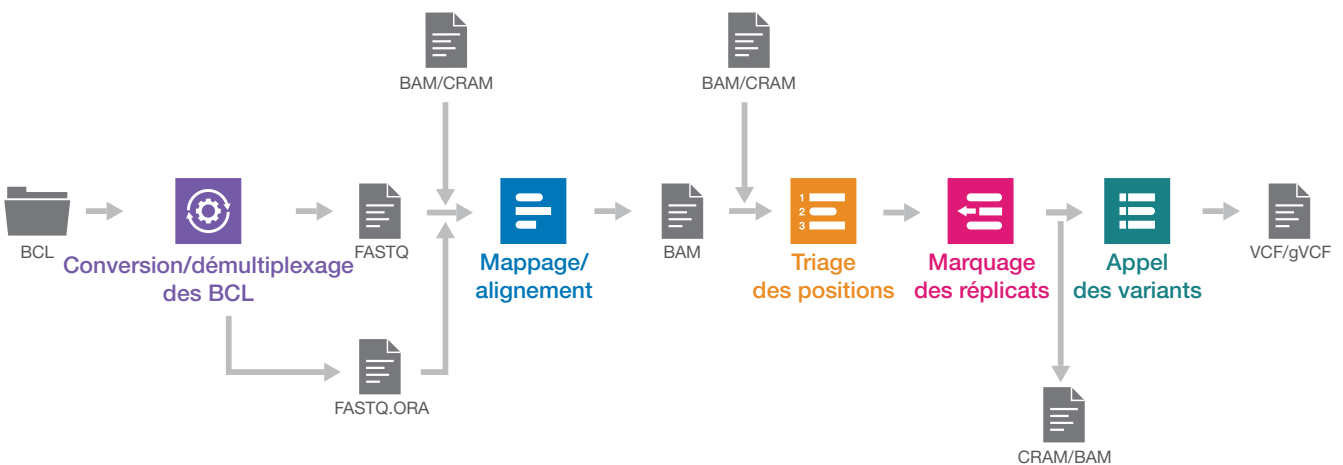


Figure 2 : Souplesse des pipelines DRAGEN – Chaque pipeline DRAGEN comporte un ensemble spécifique d'étapes pour réaliser des analyses précises et efficaces. Le pipeline DRAGEN offre la souplesse nécessaire à la prise en charge d'une variété de fichiers d'entrée et à la production d'un éventail de types de sortie, ce qui permet aux utilisateurs de personnaliser leur expérience et d'obtenir le format de fichiers voulu.

Tableau 1 : DRAGEN Bio-IT Platform prend en charge un large éventail d'applications d'analyses secondaires.

Application	Serveur DRAGEN sur site	Système NextSeq 1000, NextSeq 2000	BaseSpace Sequence Hub	Illumina Connected Analytics		Multi-Cloud BYOL
				Préconfiguré	À concevoir	
Conversion BCL	✓	✓	✓		✓	Non recommandé sur la FGPA
Compression d'ORA DRAGEN	✓	✓			✓	✓
DRAGEN FASTQ + MultiQC	✓	✓	✓	✓	✓	✓
Génome entier	Germinal + somatique	Germinal uniquement	Germinal + somatique	Germinal + somatique	Germinal + somatique	Germinal + somatique
Enrichissement (exome inclus)	Germinal + somatique	Germinal + somatique	Germinal + somatique	Germinal + somatique	Germinal + somatique	Germinal + somatique
Amplicon d'ADN	✓	✓	✓	✓	✓	✓
ARN	✓	✓	✓	✓	✓	✓
ARN unicellulaire	✓	✓	✓	✓	✓	✓
Expression différentielle		✓	✓			
NanoString GeoMx NGS		✓	✓			
Amplicon d'ARN	✓		✓	Prochainement	✓	✓
Méthylation	✓		✓	✓	✓	✓
Métagénomique			✓			
Détection de pathogènes ARN			✓			
Lignée COVID	✓		✓	Prochainement		
TruSight Oncology 500 ctDNA	✓			✓		
ScATAC-Seq	✓		✓	✓	✓	✓
Imputation	✓		✓	✓	✓	✓
Définition d'allèles PGx Star	✓		✓	✓	✓	✓

Extensibilité

DRAGEN Bio-IT Platform permet aux laboratoires d'étendre leurs opérations selon leurs besoins tout en maintenant de faibles coûts et un temps de traitement court. Le logiciel DRAGEN permet d'accroître les capacités de recherche de plusieurs façons :

- Maintenir le rythme du NovaSeq^{MC} 6000 System :**
Un serveur DRAGEN unique peut démultiplexer en moins de deux heures une analyse effectuée par le système NovaSeq 6000 avec Flow Cell S4.
- Capacité de transmission par rafales :** Pendant les périodes d'augmentation des volumes d'échantillons, les laboratoires peuvent bénéficier d'une capacité supplémentaire via les options d'accès à la plateforme DRAGEN parallèles disponibles (figure 3).
- Opérations accrues :** Une plateforme DRAGEN unique peut être utilisée pour traiter tous les pipelines DRAGEN et prendre en charge des types d'échantillons. La précision, l'exhaustivité et l'efficacité des outils DRAGEN permettent aux utilisateurs d'accroître leurs opérations sans compromettre le temps de traitement et la qualité des résultats.
- Des exomes aux génomes :** Le passage du séquençage de l'exome entier (WES) au séquençage du génome entier (WGS) entraîne une augmentation importante des données générées. Le logiciel DRAGEN permet aux clients d'effectuer des analyses en passant du séquençage d'exomes à celui de génomes sans investissements considérables en infrastructure informatique additionnelle ou en solutions infonuagiques.

- Ensembles de données très volumineux :** La plateforme DRAGEN fournit un flux de travail simplifié pour les analyses de cohorte à grande échelle et dispose de nombreux pipelines qui sont utilisés conjointement pour les appels de petits et de grands variants avec une précision élevée à partir d'échantillons de cohorte. Le logiciel DRAGEN permet l'agrégation et le génotypage de milliers, voire de millions de fichiers gVCF (genome Variant Call Format) et regroupe les nouveaux lots sans retraiter les lots existants. Le pipeline de génotypage combiné DRAGEN réalise les appels de variants conjointement dans de multiples génomes et peut atteindre de grandes cohortes grâce à une analyse rapide sans pour autant compromettre la précision⁸. Le déploiement de la plateforme DRAGEN platform pour les données du [1000 Genomes Project](#) a permis, par exemple, un appel de variants précis et à grande échelle de plusieurs échantillons et l'identification des régions dans lesquelles les données de couverture ne sont pas uniformes ou s'écartent des hypothèses.

Accessibilité multiplateforme

La suite de pipelines DRAGEN est accessible par le biais de solutions sur site, sur instrument ou en nuage, offrant ainsi la possibilité aux laboratoires de choisir la solution qui convient le mieux à leurs besoins.

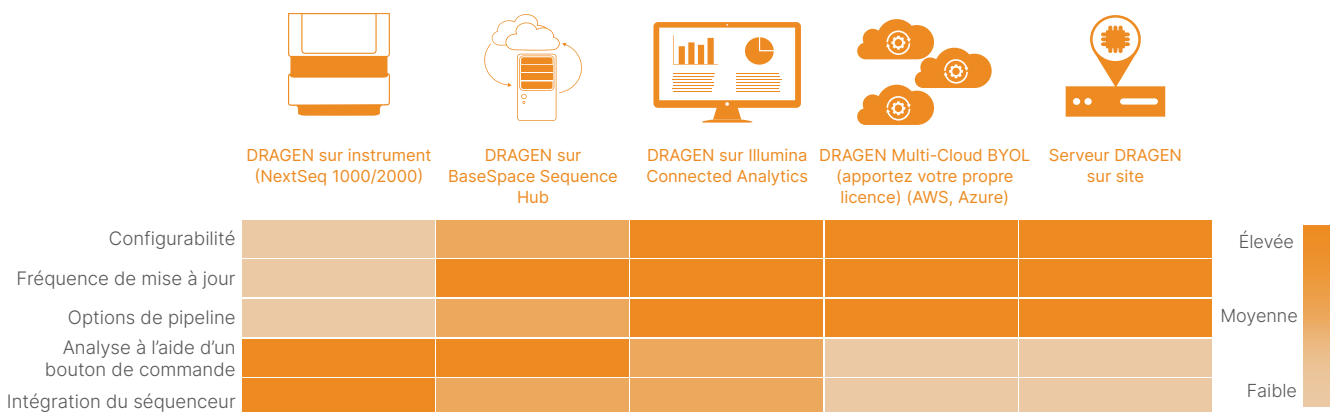


Figure 3 : Options d'accès aux pipelines de DRAGEN et fonctionnalités conçues pour répondre aux besoins en analyse de SNG de tous les laboratoires.

Serveur DRAGEN sur site

Le serveur DRAGEN sur site s'appuie sur une solution de stockage locale pour recueillir et stocker les données de SNG. Une fois les données de séquençage brutes transférées de l'instrument de séquençage au lieu de stockage local à l'aide de la connexion réseau locale, elles sont transférées au serveur DRAGEN pour exécuter le flux de travail choisi. Après analyse, le logiciel verse les fichiers de sortie générés dans la solution de stockage locale. Le serveur DRAGEN sur site :

- prend en charge un éventail de niveaux de l'interface de ligne de commande;
- remplace jusqu'à 30 ordinateurs traditionnels;
- traite des données de SNG pour un génome humain entier à 34 fois la couverture en moins de 30 minutes environ.

Intégration de DRAGEN Bio-IT Platform aux systèmes NextSeq^{MC} 1000 et NextSeq 2000

Le logiciel DRAGEN est intégré aux systèmes NextSeq 1000 et NextSeq 2000 pour offrir une solution d'analyse secondaire précise et rapide. La suite logicielle DRAGEN intégrée fournit un certain ensemble de pipelines pour prendre en charge plusieurs applications de SNG courantes ([tableau 1](#)) et dispose d'une interface conviviale qui permet aux utilisateurs experts et non experts de réaliser rapidement des analyses et de donner des résultats. DRAGEN Bio-IT Platform intégrée :

- donne accès à certains pipelines informatiques DRAGEN;
- permet aux utilisateurs de générer des résultats en seulement deux heures;
- utilise des algorithmes de pipeline intuitifs pour réduire les recours à des experts informatiques externes.

BaseSpace Sequence Hub

La suite DRAGEN infonuagique disponible sur BaseSpace Sequence Hub allie l'analyse précise et efficace à l'écosystème sécuritaire et aux fonctions d'extensibilité polyvalentes. Le logiciel DRAGEN disponible sur BaseSpace Sequence Hub permet aux laboratoires de toutes les tailles et de toutes les disciplines d'effectuer des analyses secondaires à l'aide d'un bouton de commande. BaseSpace Sequence Hub est une extension directe de vos instruments Illumina. Le flux de données chiffrées transférées de l'instrument à BaseSpace Sequence Hub vous permet de gérer et d'analyser facilement vos données à l'aide d'un ensemble d'applications soigneusement sélectionnées. BaseSpace Sequence Hub, optimisé par Amazon Web Services (AWS) :

- fournit une solution facile à utiliser à l'aide d'un bouton de commande pour effectuer des analyses DRAGEN;
- utilise une interface utilisateur graphique intuitive permettant aux utilisateurs experts et non experts de réaliser des opérations efficaces;
- donne accès à des ressources informatiques importantes sans devoir effectuer des dépenses pour une infrastructure supplémentaire.

Illumina Connected Analytics

DRAGEN Bio-IT Platform disponible sur Illumina Connected Analytics est une plateforme infonuagique exhaustive de gestion et d'analyse des données qui permet aux chercheurs de gérer, d'analyser et d'interpréter de gros volumes de données multiomiques dans un environnement sécurisé, souple et évolutif. Illumina Connected Analytics :

- donne accès à la suite DRAGEN Bio-IT complète, disponible sous la forme de pipelines prêts à l'emploi ou d'outils individuels pour les pipelines personnalisés;
- prend en charge des flux de travail hautement automatisés et des solutions personnalisées pour les études optimisées à débit élevé;
- fournit un environnement hautement sécurisé avec une garantie de résidence des données, une authentification unique, des journaux de vérification et un contrôle des accès conforme à la loi américaine sur l'assurance maladie (HIPAA, Health Insurance Portability and Accountability Act) et aux principes du Règlement général sur la protection des données (RGPD) de l'Union européenne.

DRAGEN Multi-Cloud BYOL (apportez votre propre licence)

Les laboratoires disposant de services infonuagiques existants peuvent exécuter des analyses DRAGEN sur AWS ou Microsoft Azure en achetant DRAGEN Bio-IT Platform en tant qu'image virtuelle. DRAGEN Multi-Cloud BYOL (apportez votre propre licence) :

- utilise les mêmes options de logiciel et de ligne de commande que le serveur DRAGEN sur site (à l'exception des pipelines TruSight^{MC} Oncology 500);
- est idéal pour les utilisateurs avancés cherchant à intégrer les pipelines DRAGEN à leur environnement infonuagique et leurs flux de travail existants.

Résumé

DRAGEN Bio-IT Platform est une suite puissante d'outils logiciels permettant de réaliser des analyses secondaires précises, complètes et efficaces des données de SNG. Les nombreuses options de DRAGEN Bio-IT Platform offrent la possibilité aux laboratoires de choisir la solution qui convient le mieux au type et à la taille de leurs projets. Au fur et à mesure que des progrès seront apportés à la technologie de SNG, des mises à jour régulières de DRAGEN Bio-IT Platform garantiront la meilleure performance possible des pipelines existants, tandis que de nouveaux pipelines continueront d'être ajoutés à mesure que les applications deviennent disponibles.

En savoir plus

Illumina DRAGEN Bio-IT Platform, illumina.com/dragen

Page d'assistance de la plateforme DRAGEN, support.illumina.com/sequencing/sequencing_software/dragen-bio-it-platform

Communiquez avec nous, illumina.com/company/contact-us.html#/united-states/customer-care

Références

1. Food and 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. Illumina. DRAGEN Wins at PrecisionFDA Truth Challenge V2 Showcase Accuracy Gains from Alt-aware Mapping and Graph Reference Genomes. illumina.com/science/genomics-research/articles/dragen-wins-precisionfda-challenge-accuracy-gains.html. Accessed March 14, 2022.
4. Internal data on file. Illumina, Inc. 2022.
5. BioIT World. Children's Hospital Of Philadelphia, Edico Set World Record For Secondary Analysis Speed. bio-itworld.com/news/2017/10/23/children-s-hospital-of-philadelphia-edico-set-world-record-for-secondary-analysis-speed. Accessed March 14, 2022.
6. San Diego Union Tribune. Rady Children's Institute sets Guinness world record. <https://www.sandiegouniontribune.com/95899028-132.html>. Published February 12, 2018. Accessed March 14, 2022.
7. Miller NA, Farrow EG, Gibson M, et al. A 26-hour system of highly sensitive whole genome sequencing for emergency management of genetic diseases. *Genome Med.* 2015;7:100. doi: 10.1186/s13073-015-0221-8.
8. Illumina. Accurate and Efficient Calling of Small and Large Variants from PopGen Datasets Using the DRAGEN Bio-IT Platform. www.illumina.com/science/genomics-research/articles/popgen-variant-calling-with-dragen.html. Accessed March 14, 2022.

illumina^{MD}

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 la page www.illumina.com/company/legal.html.
M-GL-00680 FRA v4.0