

# Analyse secondaire DRAGEN<sup>MC</sup>

Appel des variants précis,  
complet et efficace pour  
des données de séquençage  
de nouvelle génération

## Introduction

Tirer parti de la puissance du génome via le séquençage de nouvelle génération (SNG) est essentiel pour faire progresser la recherche biomédicale et la médecine de précision. 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.

L'analyse secondaire DRAGEN (Dynamic Read Analysis for GENomics, soit Analyse de lecture dynamique pour la génomique en français) 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 études de méthylome, du génome entier, de l'exome, des transcriptomes et bien d'autres. Le logiciel d'analyse secondaire DRAGEN est une suite d'applications 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

L'analyse secondaire DRAGEN génère des résultats d'une précision exceptionnelle. Lors du 2020 PrecisionFDA Truth Challenge V2 (PrecisionFDA V2), l'analyse secondaire DRAGEN v3.7 a été déclarée la plus précise dans toutes les régions de référence et les régions difficiles à mapper avec les données de séquençage d'Illumina<sup>1,2</sup>. Les versions suivantes continuent d'établir de nouvelles normes en matière de précision, avec des avancées dans des domaines tels que l'apprentissage automatique (AA) et la technologie DRAGEN Multigénome (Graph). La dernière version de l'analyse secondaire DRAGEN v4.3 offre une précision sans précédent de l'appel de petits variants avec un score F1 de 99,89 % (une mesure combinée de précision et de rappel) dans toutes les régions de référence (figure 1). Cela est rendu possible par la référence DRAGEN Multigénome (Graph) de troisième génération, conçue à partir de 128 échantillons avec 256 haplotypes de données de référence pangénomiques conçues en interne, capturant une plus grande diversité génétique. Le nouvel ensemble de paramètres d'appel de mosaïques intégré qui peut être activé pour détecter les variants mosaïques avec des fréquences alléliques aussi faibles que 3 % contribue également à une meilleure précision.

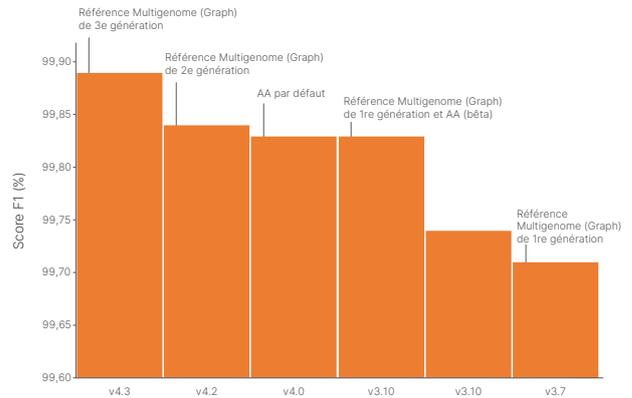


Figure 1 : Précision de l'analyse secondaire DRAGEN : le score F1 (%), indiqué sur l'axe Y, est un calcul des résultats vrais positifs et vrais négatifs par rapport aux résultats totaux<sup>3,4</sup>.

## Analyse complète

Avec une couverture complète du génome et un large éventail d'applications prises en charge, l'analyse secondaire DRAGEN répond aux divers besoins des laboratoires qui effectuent l'analyse du SNG. Les pipelines DRAGEN prennent en charge plusieurs types d'expériences, notamment le séquençage du génome entier (WGS, Whole-Genome Sequencing), le séquençage de l'exome entier, les panels d'enrichissement, le séquençage RNA-Seq unicellulaire, le séquençage ATAC-Seq unicellulaire, le séquençage RNA-Seq en lots et l'analyse de méthylation (tableau 1). Il faudrait plus de 30 outils à code source ouvert pour reproduire partiellement l'étendue des fonctionnalités du logiciel DRAGEN<sup>3,4</sup>.

Pour l'analyse germinale, l'analyse secondaire DRAGEN comprend un ensemble de paramètres d'appel de variants comme ExpansionHunter et de paramètres d'appel ciblés comme *SMN*, *GBA*, *CYP2B6*, *CYP2D6* et *HLA*. DRAGEN v4.3 intègre également MRJD, une nouvelle fonction d'appel qui permet de couvrir les gènes difficiles dans les régions de duplication segmentaire, tels que *PMS2*, *SMN1*, *SMN2*, *STRC*, *NEB*, *TTN* et *IKBK*. Ces outils permettent l'analyse d'un large éventail de variations génétiques, y compris les variations mononucléotidiques, les insertions et les suppressions (indels), les expansions de répétitions et les variations structurelles dans les régions génomiques étendues. De plus, la référence DRAGEN Multigénome (Graph) améliore la qualité du mappage, ce qui permet d'améliorer la précision de l'appel des variants et d'atteindre les zones du génome difficiles à évaluer en raison de la complexité des séquences. Cette architecture permet d'améliorer la couverture des gènes potentiellement pertinents sur le plan médical et permet l'appel des variants mononucléotidiques, des petits variants indels, des variants du nombre de copies et des variants structurels dans les régions difficiles à mapper.

Tableau 1 : L'analyse secondaire DRAGEN prend en charge un large éventail d'applications d'analyses secondaires<sup>a</sup>

Application	Serveur sur site	Systèmes de séquençage intégrés d'Illumina		Plateformes infonuagiques d'Illumina	
	Serveur DRAGEN	Série NovaSeq X	Système NextSeq 1000, NextSeq 2000	BaseSpace Sequence Hub	Illumina Connected Analytics
Conversion BCL	✓	✓	✓	✓	Personnalisé uniquement
DRAGEN ORA Compression	✓	✓	✓		Personnalisé uniquement
DRAGEN FASTQ + MultiQC	✓	✓	✓	✓	✓
Génome entier	Germinal + somatique	Germinal + somatique	Germinal + somatique	Germinal + somatique	Germinal + somatique
Enrichissement (exome inclus)	Germinal + somatique	Germinal + somatique	Germinal + somatique	Germinal + somatique	Germinal + somatique
DRAGEN Amplicon	✓		ADN uniquement	✓	✓
ARN	✓	✓	✓	✓	✓
ARN unicellulaire	✓		✓	✓	✓
NanoString GeoMx NGS			✓	✓	
Méthylation	✓	✓		✓	✓
Métagénomique	✓ <sup>b</sup>			✓	
Détection de pathogènes ARN				✓	
COVID	COVIDSeq. Lignée COVID		COVIDSeq (nuage uniquement)	COVIDSeq. Lignée COVID	
Gamme TruSight Oncology 500	✓			✓ <sup>c</sup>	✓
scATAC-seq	✓			✓	✓
Imputation	✓			✓	✓
Définition d'allèles PGx Star	✓	✓	✓	✓	✓
Illumina Complete Long Read				✓	✓
Analyse secondaire DRAGEN pour RPIP et UPIP	✓			✓	Bêta

a. La version logicielle de base DRAGEN varie selon les plateformes. Communiquez avec un représentant local pour obtenir de plus amples renseignements.

b. Applications de métagénomique via le classificateur k-mer, plus d'outils prochainement.

c. Abonnement à Illumina Connected Analytics requis.

## Analyse efficace

Le logiciel DRAGEN est conçu pour fournir aux laboratoires la vitesse d'analyse de données dont ils ont besoin pour optimiser l'efficacité du traitement de leurs ensembles de données de SNG. L'analyse secondaire DRAGEN est accélérée 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<sup>5,6</sup>. Dans la pratique, l'analyse secondaire DRAGEN peut traiter sur site des données de SNG pour un génome entier équivalent à 40 fois la couverture en 35 minutes avec tous les paramètres d'appel<sup>7</sup> contre plus de 8 heures avec les méthodes à code source ouvert couramment utilisées appelant un nombre limité de types de variants<sup>7</sup>.

Pour faciliter le stockage, la gestion et le partage des fichiers volumineux de données de séquençage de nouvelle génération (SNG), la technologie d'archive des lectures d'origine (ORA, Original Read Archive) de DRAGEN permet une compression sans perte des fichiers FASTQ jusqu'à

cinq fois au format fastq.gz traditionnel. 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 et prenant en charge un large éventail d'espèces couramment étudiées. L'analyse secondaire DRAGEN propose un ensemble polyvalent de pipelines qui 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 fréquents pour offrir les meilleures précision, exhaustivité et efficacité possibles.

\* Basé sur les données internes d'Illumina selon les normes HG001-HG007 sur le DRAGEN Server v4, sans nouveaux paramètres d'appel spécialisés comme MRJD et VNTR disponibles dans DRAGEN v4.3.

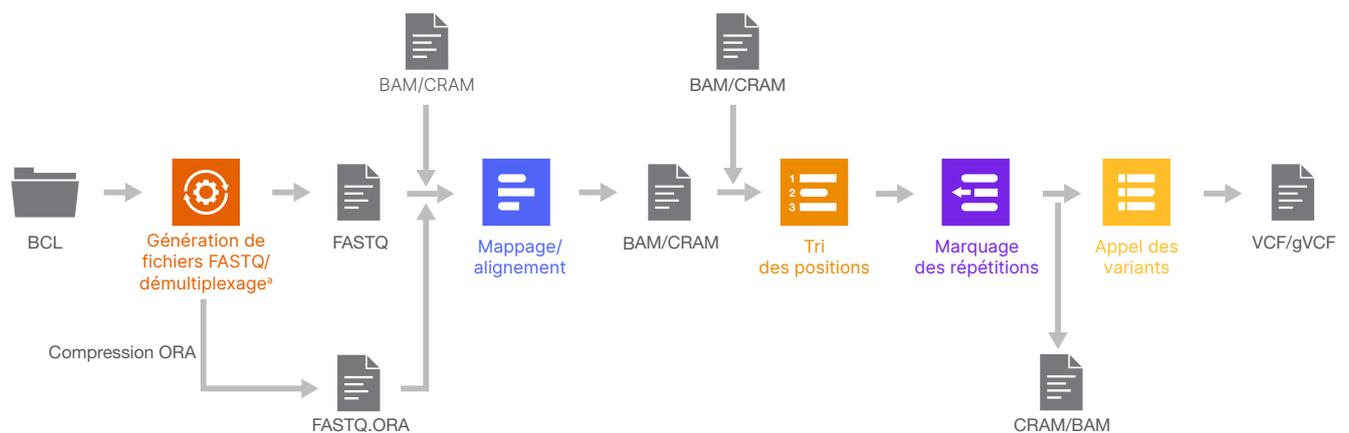


Figure 2 : Souplesse des pipelines de l'analyse secondaire DRAGEN : chaque pipeline DRAGEN comporte un ensemble spécifique d'étapes pour réaliser des analyses précises et efficaces. Le pipeline de variants germinaux du génome entier DRAGEN fourni en exemple offre la souplesse nécessaire à la prise en 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.

a. La conversion BCL est également disponible en tant qu'outil autonome.

## Références personnalisées

L'analyse secondaire de DRAGEN permet aux utilisateurs de générer une référence personnalisée humaine, non humaine ou non standard. Les références créées peuvent être utilisées à titre d'entrée pour toutes les applications DRAGEN prenant en charge les fichiers de références personnalisées. La plupart des pipelines DRAGEN comprennent un soutien intégré pour générer hg19, hg38 (avec ou sans HLA), GRCh37, CHM13v2 et hs37d5. Le logiciel DRAGEN permet aux utilisateurs d'étendre les capacités de référence multigénomique standard graphique à des populations diverses et spécifiques.

## Extensibilité

L'analyse secondaire DRAGEN 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 de la série NovaSeq<sup>MC</sup> X et des systèmes NextSeq<sup>MC</sup> 1000 et NextSeq 2000 :** la plateforme DRAGEN intégrée peut exécuter plusieurs applications simultanées (quatre applications simultanées avec un maximum d'une conversion BCL et trois autres pipelines de votre choix) par Flow Cell lors d'une même analyse.
- Capacité de transmission par rafales :** pendant les périodes d'augmentation des volumes d'échantillons, les laboratoires peuvent tirer parti d'une capacité infonuagique supplémentaire avec l'analyse secondaire

DRAGEN sur Illumina Connected Analytics ou les applications DRAGEN sur BaseSpace Sequence Hub (figure 3).

- Opérations accrues :** une instance DRAGEN unique peut exécuter un large éventail de pipelines DRAGEN et de types d'échantillons pris en charge. L'exhaustivité et l'efficacité du logiciel DRAGEN permettent aux utilisateurs d'accroître leurs opérations sans compromettre le temps de traitement et la qualité des résultats.
- Transition vers les génomes :** les pipelines prédéfinis DRAGEN permettent une transition facile des panels ciblés aux exomes, puis aux génomes.
- Initiatives génomiques à grande échelle sur les populations :** l'analyse secondaire 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 appeler les variations génétiques avec une précision élevée. Le DRAGEN gVCF Genotyper permet l'agrégation de milliers, voire de millions de fichiers gVCF (genome Variant Call Format) et intègre les nouveaux lots sans retraiter les lots existants. La compression ORA permet de réduire les coûts de stockage.
- Applications de séquençage à grande profondeur :** l'analyse secondaire DRAGEN prend en charge l'analyse du séquençage à des profondeurs élevées avec une grande efficacité pour une couverture moyenne de plus de 300 fois pour les génomes et de 1 000 fois pour les exomes. Les capacités du séquençage à grande profondeur sont précieuses pour des applications telles que la recherche en oncologie et les études sur les maladies génétiques rares.

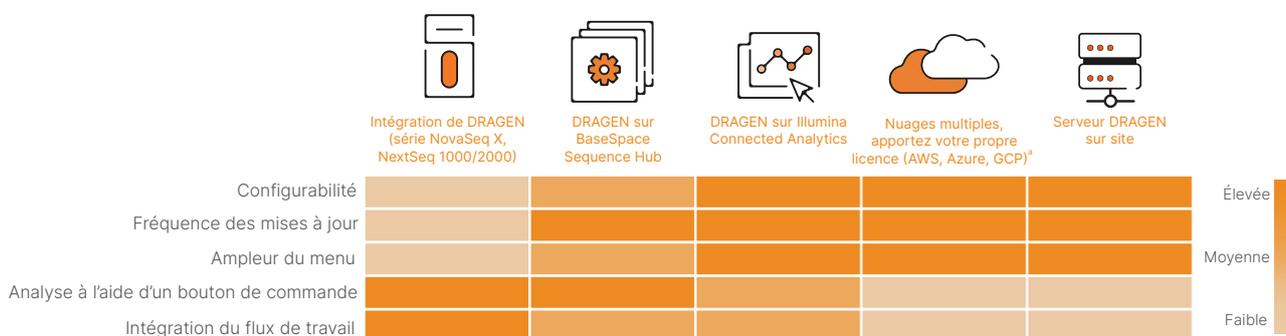


Figure 3 : Options d'accès aux pipelines DRAGEN et fonctionnalités conçues pour répondre aux besoins en analyse de SNG de tous les laboratoires.

a. Communiquez avec votre représentant Illumina pour obtenir des renseignements sur l'accès à Amazon Web Service (AWS), Azure ou Google Cloud Platform (GCP, accès anticipé).

## 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).

### 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 :

- permet une configuration souple des fonctionnalités DRAGEN par l'entremise d'une interface de ligne de commande;
- remplace jusqu'à 30 ordinateurs traditionnels;
- traite des données de SNG pour un génome humain entier à 40 fois la couverture en 35 minutes environ.

### Intégration de DRAGEN à la série NovaSeq X

La série NovaSeq X comprend l'analyse secondaire DRAGEN intégrée, permettant de réaliser des analyses précises, automatisées et rationalisées, conçue pour prendre en charge un volume extraordinaire de données générées par la série NovaSeq X. La suite logicielle intégrée DRAGEN permet l'analyse secondaire et la compression ORA avec les applications de séquençage de nouvelle génération (SNG) (tableau 1) couvrant la conversion BCL, les variants germinaux, les variants somatiques, l'enrichissement, l'ARNet la méthylation. L'intégration de DRAGEN :

- permet d'analyser plusieurs pipelines d'analyse secondaire en parallèle;
- permet d'exécuter jusqu'à quatre applications simultanées par Flow Cell lors d'une analyse unique;
- permet une compression de données sans perte jusqu'à cinq fois et de faire des économies sur les coûts de stockage;
- permet de faire des économies d'analyse, qui sur cinq ans peuvent dépasser le prix d'achat du système NovaSeq X.

### Intégration de DRAGEN aux systèmes NextSeq 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. Le logiciel est accessible par le biais d'une interface graphique conviviale qui permet aux utilisateurs experts et non experts de réaliser les analyses nécessaires et de donner rapidement des résultats. Le logiciel intégré DRAGEN fournit un certain ensemble de pipelines pour prendre en charge plusieurs applications de séquençage de nouvelle génération (SNG) courantes (tableau 1) et comprend l'analyse de référence Multigénome (Graph) et l'apprentissage automatique pour un appel des variants de haute qualité. L'intégration de DRAGEN :

- offre la plus grande précision de tout système de séquençage de pailasse avec l'analyse secondaire intégrée DRAGEN;
- 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 puissantes sans devoir effectuer des dépenses pour une infrastructure supplémentaire.

## Illumina Connected Analytics

Illumina Connected Analytics est une plateforme bioinformatique exhaustive qui permet aux chercheurs de gérer, d'analyser et d'interpréter de gros volumes de données multiomiques dans un environnement sécurisé, souple et évolutif. Accédez à la suite d'analyse secondaire DRAGEN sur Illumina Connected Analytics, disponible sous la forme de pipelines prêts à l'emploi ou d'outils individuels à intégrer dans les pipelines personnalisés.

## Résumé

L'analyse secondaire DRAGEN est une suite puissante d'outils logiciels permettant de réaliser des analyses précises, complètes et efficaces des données de SNG. Les nombreuses options de déploiement du logiciel DRAGEN offrent aux laboratoires la possibilité de choisir la solution qui convient le mieux au type et à la taille de leurs projets. De plus, les utilisateurs peuvent combiner diverses options de déploiement pour mieux répondre à leurs besoins de performance et de flux de travail. Au fur et à mesure que des progrès seront apportés à la technologie de SNG, des mises à jour régulières de l'analyse secondaire DRAGEN 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

[Analyse secondaire DRAGEN](#)

[Page d'assistance de l'analyse secondaire DRAGEN](#)

[Communiquez avec nous](#)



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

© 2024 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](http://www.illumina.com/company/legal.html).  
M-GL-00680 FRA v10.0

## Références

1. The Food and Drug Administration. Truth Challenge V2: Calling Variants from Short and Long Reads in Difficult-to-Map Regions. <https://precision.fda.gov/challenges/10/results>. Consulté le 14 mars 2022.
2. Catreux S, Jain V, Murray L, et al. DRAGEN sets new standard for data accuracy in PrecisionFDA benchmark data. Optimizing variant calling performance with Illumina machine learning and DRAGEN graph. Site Web d'Illumina. <https://www.illumina.com/science/genomics-research/articles/dragen-shines-again-precisionfda-truth-challenge-v2.html>. Publié le 12 janvier 2022. Consulté le 11 juillet 2024.
3. Mehio R, Ruehle M, Catreux S, et al. DRAGEN wins at PrecisionFDA Truth Challenge V2 showcase accuracy gains from alt-aware mapping and graph reference genomes. Site Web d'Illumina. [illumina.com/science/genomics-research/articles/dragen-wins-precisionfda-challenge-accuracy-gains.html](https://www.illumina.com/science/genomics-research/articles/dragen-wins-precisionfda-challenge-accuracy-gains.html). Publié le 9 novembre 2020. Consulté le 11 juillet 2024.
4. Données internes. 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](https://www.bio-itworld.com/news/2017/10/23/children-s-hospital-of-philadelphia-edico-set-world-record-for-secondary-analysis-speed). Publié le 23 octobre 2017. Consulté le 11 juillet 2024.
6. San Diego Union Tribune. Rady Children's Institute sets Guinness world record. <https://www.sandiegouniontribune.com/2018/02/12/rady-childrens-institute-sets-guinness-world-record/>. Publié le 12 février 2018. Consulté le 11 juillet 2024.
7. Betschart RO, Thiéry A, Aguilera-Garcia D, et al. Comparison of calling pipelines for whole genome sequencing: an empirical study demonstrating the importance of mapping and alignment. *Sci Rep.* 2022;12(1):21502. Publié le 13 décembre 2022. doi:10.1038/s41598-022-26181-3
8. Gross A, Maciucia S, Cox A, et al. Accurate and Efficient Calling of Small and Large Variants from PopGen data sets Using the DRAGEN Bio-IT Platform. Site Web d'Illumina. [www.illumina.com/science/genomics-research/articles/popgen-variant-calling-with-dragen.html](https://www.illumina.com/science/genomics-research/articles/popgen-variant-calling-with-dragen.html). Publié le 24 mai 2021. Consulté le 14 mars 2022.