FICHE TECHNIQUE

Illumina Connected Analytics

Flux de production informatiques à grande échelle

- Importez, créez et modifiez des flux de travail avec des outils tels que CWL (Common Workflow Language) et Nextflow
- Organisez les données dans un espace de travail sécurisé et partagez-les de manière conforme à l'échelle mondiale
- Interprétez les données dans un environnement informatique flexible qui comprend JupyterLab Notebooks

illumina[MD]

Introduction

Les avancées dans les technologies de séquençage nouvelle génération (SNG) ont modifié de manière conséquente la fréquence à laquelle sont menées les recherches cliniques et en sciences du vivant. À mesure que la vitesse du séquençage augmente et que son coût baisse, la capacité à générer des données dépassera de loin la capacité à extraire des renseignements biologiques et cliniques de ces mêmes données. Relever les défis d'une gestion sécurisée des données, développer l'infrastructure et créer et déployer de nouveaux flux de travail informatiques nécessitent une plateforme flexible et complète. Illumina Connected Analytics (ICA) permet aux utilisateurs de créer, de développer de nouvelles versions et de déployer des pipelines analytiques flexibles tout en maintenant la confidentialité, la sécurité et la conformité des données à grande échelle.

ICA est une plateforme de données génomiques sécurisée qui rend l'informatique opérationnelle et génère des renseignements scientifiques (figure 1, tableau 1).
ICA permet aux utilisateurs de :

- créer et de personnaliser des pipelines d'analyse;
- exécuter des flux de production à grande échelle;
- explorer et de partager des données et des résultats.

Flux de travail rationalisé

ICA est un composant central des laboratoires effectuant des études de SNG avec des systèmes de séquençage Illumina. Tirant profit de l'élasticité des ressources offertes par l'infonuagique, ICA prend en charge des opérations de n'importe quelle ampleur, du criblage occasionnel ou des dizaines de milliers de cellules dans des projets monocellulaires complexes au séquençage d'un génome complet à l'échelle d'une population, le tout avec la même architecture. Les utilisateurs peuvent intégrer facilement leurs instruments à ICA.

Au sein d'ICA, les données peuvent être automatiquement analysées avec les pipelines DRAGEN[MC] prêts à utiliser ou des pipelines personnalisés, en fonction du flux de travail spécifié. Le large éventail d'options d'analyse va du contrôle qualité à l'agrégation des données et aux outils de datalogie avancés pour un traitement rapide et évolutif des données. ICA fournit une plateforme extensible avec un ensemble riche d'interfaces de programmation (API, Application Program Interface) RESTful et un outil d'interface de ligne de commande (CLI, Command-line Interface). Ces API optimisent l'efficacité des flux de travail à mesure que les données sont transférées, consultées et utilisées tout au long de leur cycle de vie, et comportent des API homologuées par l'Alliance mondiale pour la génomique et la santé (GA4GH, Global Alliance for Genomics and Health)[1].

Tableau 1 : Aperçu d'ICA

	Caractéristique	Avantage
Sécurité et confidentialité	Conformité	Respectez les normes réglementaires locales, régionales et mondiales, les normes du HIPAA et du RGPD, et les homologations ISO 27001
	Contrôles de sécurité	Maintenez une séparation stricte des données avec un chiffrement des données en transit (TLS 1.2) et entreposées (AES 256)
	Piste de vérification	Maintenez un journal d'activité mentionnant la personne qui a accédé aux données et la date à laquelle elle y a accédé
	Authentification unique (facultatif)	Utilisez les identifiants de l'institution pour contrôler l'accès
Attribution des ressources	Ressources informatiques à la demande	Réduisez les coûts en ne payant que les ressources informatiques dans le moteur du pipeline
	Évolutivité à la demande	Augmentez les besoins en stockage sur nuage et en puissance de calcul en fonction du niveau de demande actuel
	Tableau de bord de plateforme et d'utilisation	Bénéficiez d'un affichage des demandes en ressources pour comprendre, gérer et anticiper les besoins efficacement
Gestion	Gestion des projets et des utilisateurs	Gérez l'accès et l'activité des utilisateurs pour une confidentialité granulaire
	Partage des données	Décompartimentez les données pour une collaboration mondiale à grande échelle
	Archivage des données	Réduisez les coûts en archivant les données non utilisées dans des niveaux de stockage moins coûteux
Exploitabilité	Intégration directe des systèmes de séquençage	Les données proviennent directement des systèmes de séquençage Illumina
	Créateur de pipeline visuel	Créez des pipelines sans écrire de code
	Outils et pipelines	Exploitez des pipelines prêts à l'emploi et importez des outils personnalisés
	API et CLI	Interagissez de façon programmatique avec la plateforme grâce à des outils basés sur les préférences des utilisateurs
	« Apportez votre propre compartiment »	Accédez aux données stockées sur un compte infonuagique privé
	Visualisation des données	Créez des tracés visuels dynamiques et des applications Web interactives pour afficher les données grâce aux modules R et Python
Outils avancés	Prise en charge de Docker, Nextflow et CWL	Écrivez des pipelines en Common Workflow Language (CWL) et lancez des analyses dans le nuage en toute simplicité
	API RESTful homologuées par la GA4GH	Autorisez l'accès programmatique aux outils et aux données et l'interopérabilité avec les autres environnements logiciels
	Intégré à JupyterLab	Effectuez des analyses de données avancées; créez et entraînez des modèles d'IA/apprentissage machine avec R et Python
	Agrégation et interrogation de données	Effectuez des interrogations de données à l'échelle d'une population avec SQL

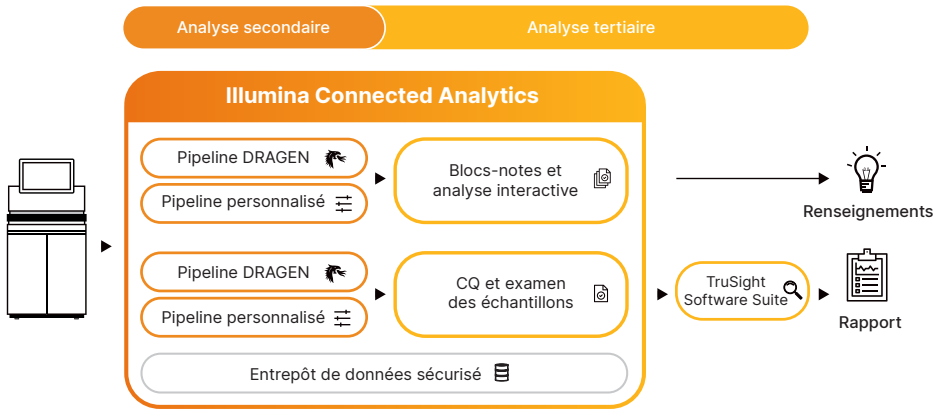

Figure 1 : ICA forme la base de la gestion et de l'analyse des données

Transformation des lectures en données

ICA propose plusieurs options d'analyse secondaire des données, afin de rationaliser le flux de travail des lectures aux résultats. En proposant d'utiliser des pipelines prêts à l'emploi ou de construire et de configurer des pipelines personnalisés, ICA peut prendre en charge pratiquement n'importe quelle application informatique.

Options prêtes à utiliser

ICA propose de puissants outils et pipelines prêts à l'emploi pour le traitement des données, y compris l'accès à la plateforme DRAGEN Bio-IT[2], qui offre une analyse secondaire rapide et précise des données de séquençage (figure 2).

Personnalisation des pipelines

Les bioinformaticiens peuvent importer des outils existants depuis un référentiel d'images Docker, ou construire et éditer de nouveaux pipelines à l'aide de Nextflow, CWL et de l'éditeur de pipeline graphique. Les laborantins et autres scientifiques peuvent lancer des pipelines en toute simplicité grâce à l'interface utilisateur intuitive.

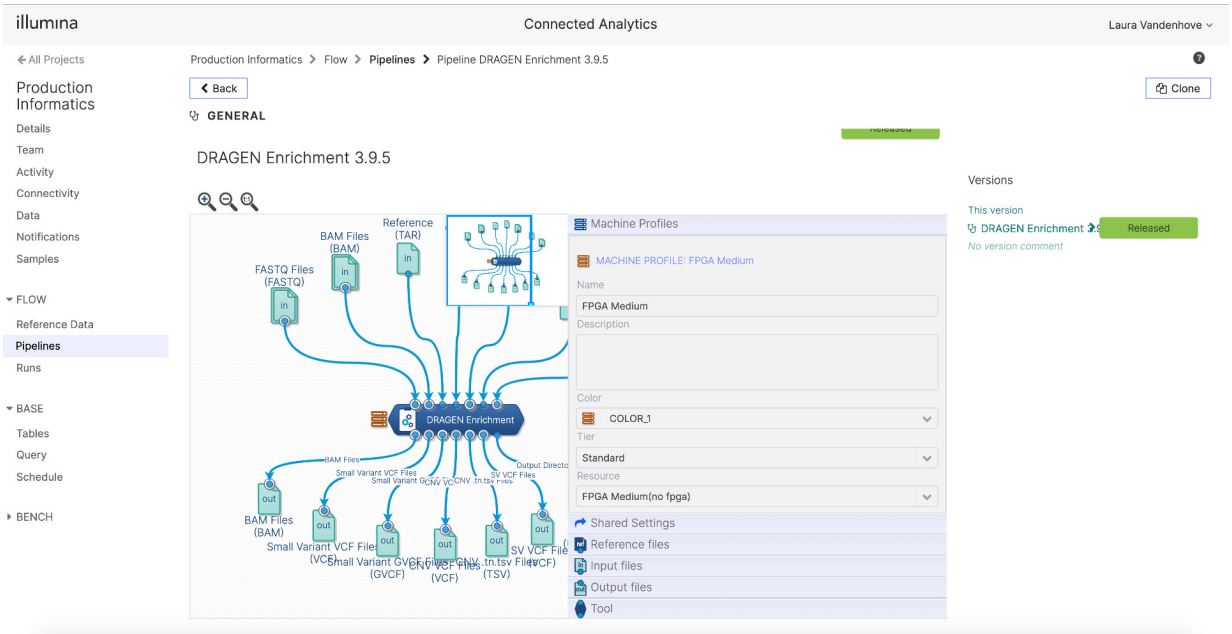

Figure 2 : Pipeline DRAGEN dans ICA : les pipelines DRAGEN prêts à utiliser dans ICA permettent une analyse secondaire des lectures aux rapports rapide et précise.

Gestion et contrôle des données

L'augmentation de la génération de données entraîne un besoin croissant d'une infrastructure permettant le partage, la réutilisation et l'intégration de données au sein de la communauté scientifique pour amplifier la valeur des ensembles de données individuels. Pour répondre à ce besoin, ICA incorpore plusieurs fonctionnalités destinées à permettre l'adoption des pratiques exemplaires en matière de gestion des données.

Contrôle des accès

Un contrôle des accès granulaire permet à un administrateur de définir des autorisations et d'exploiter les identifiants existants de l'institution pour contrôler les accès. Un journal de vérification fait office de registre des événements et des modifications en enregistrant chaque utilisateur lorsqu'il accède à la plateforme ainsi que ses actions lorsqu'il utilise la plateforme, assurant ainsi le respect de la conformité et de la responsabilité.

Format ouvert

ICA est conçue comme une plateforme « agnostique ». Elle peut analyser plusieurs types de données, y compris des données moléculaires, cliniques, phénotypiques, ainsi que des données non structurées comme des images.

Collaboration

ICA permet une collaboration sans frontières géographiques tout en préservant la conformité. Les données et les outils peuvent être transmis à d'autres utilisateurs et partagés avec eux de manière instantanée tout en préservant l'intégrité et la confidentialité des données. De plus, les données et les outils analytiques hébergés dans une source infonuagique externe peuvent être importés dans ICA pour être analysés et partagés.

Agrégation et interrogation de données

ICA automatise les étapes d'agrégation et d'intégration complexes pour créer un système fonctionnel de gestion des connaissances qui englobe des données provenant de millions d'échantillons (figure 3). Elle acquiert presque tous les types de données disponibles, que ce soit des données génotypiques, phénotypiques, des métadonnées, des annotations, et d'autres informations connexes. Les utilisateurs peuvent définir leurs propres modèles de données, écrire leurs propres interrogations, et explorer les liens entre les ensembles de données au besoin. Les données agrégées sur ICA représentent une mine d'informations pouvant être utilisées pour découvrir de nouveaux biomarqueurs, stratifier les populations de patients, surveiller la performance d'un test dans le temps, et plus encore.

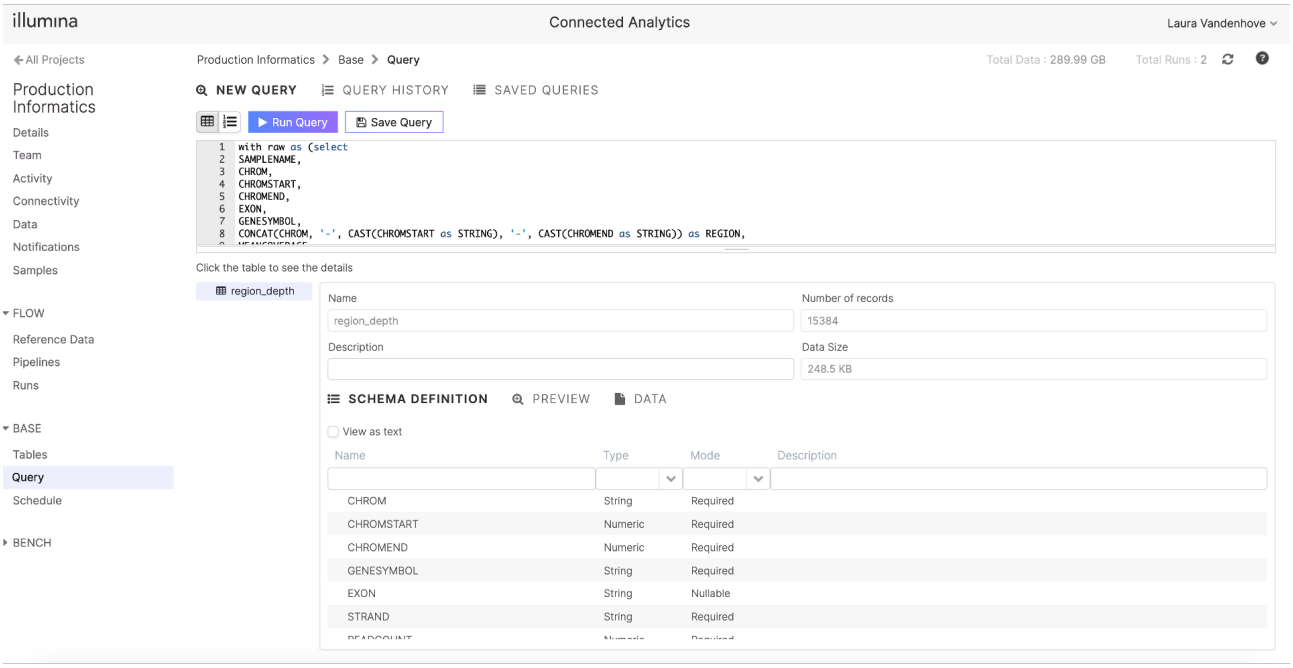

Figure 3 : ICA permet l'agrégation et l'exploration de données, ainsi que l'apprentissage continu : l'utilisateur peut explorer des liens entre les ensembles de données pour répondre à ses questions.

Sécurisation de l'environnement des blocs-notes pour approfondir les connaissances

Avec la myriade d'explorations de données en cours, il est essentiel de pouvoir développer et personnaliser des algorithmes. Un module de programmation interactif, qui exploite l'application populaire JupyterLab Notebooks (blocs-notes en Python et R), permet aux scientifiques de données d'analyser les données agrégées dans un environnement fluide et sécurisé (figure 4).

Lors de la phase de développement des méthodes et des algorithmes, l'utilisateur peut développer ou modifier des pipelines dans un environnement de test. Dans cet environnement, l'utilisateur peut rapidement construire, tester et itérer sur des modèles d'apprentissage machine au besoin. L'utilisateur a accès à une large gamme de bibliothèques standard, comme TensorFlow[3] ou scikit-learn[4], et peut facilement importer ses bibliothèques personnalisées. Lorsque l'utilisateur est prêt à passer en phase de production, ICA permet de convertir les blocs-notes en outils. Ces outils sont ensuite disponibles dans le référentiel d'ICA et incorporés aux pipelines de production.

La sécurité et la conformité sont la priorité

La sécurité est une préoccupation capitale lorsque l'on travaille avec des données génomiques appliquées à la recherche, à la thérapeutique et au diagnostic chez les humains. ICA emploie différentes mesures numériques et administratives pour respecter toutes les exigences en matière de sécurité des données, même les plus strictes :

- Les données téléversées à partir des instruments de séquençage sont chiffrées conformément à la norme AES 256 et protégées par le protocole TLS (Transport Layer Security).
- Les données contenues dans ICA sont hébergées sur la plateforme Amazon Web Services (AWS) pour respecter la conformité avec un grand nombre de normes de sécurité acceptées par le secteur en utilisant les pratiques exemplaires AWS Well-Architected[5].
- Le service d'authentification est pris en charge par SAML 2.0 qui gère les utilisateurs et les mots de passe de l'institution (facultatif).
- Les rapports de vérification prennent en charge la traçabilité de la provenance des données.

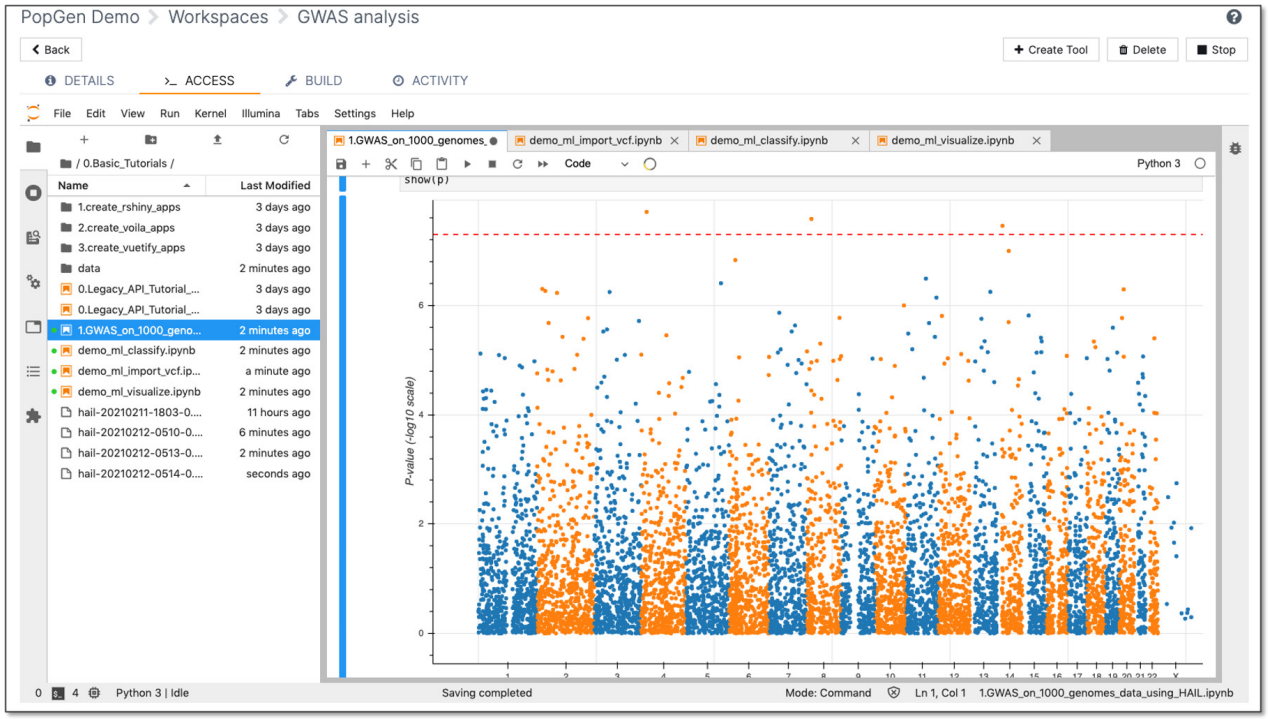

Figure 4 : Analyse interactive et visualisation : ICA prend en charge les Blocs-notes Jupyter pour une exploration visuelle de données multidimensionnelles.

ICA répond également aux besoins des clients qui travaillent dans des environnements réglementés, qui doivent se conformer à des exigences strictes :

- Lois en vigueur sur la protection des données comme le Règlement général sur la protection des données (RGPD)[6] et la Health Insurance Portability and Accountability Act (HIPAA)[7]
- Système de gestion de la sécurité des renseignements de la norme de l'Organisation internationale de normalisation (ISO) 27001[8]
- Garantie de résidence des données afin d'assurer le respect des exigences réglementaires et de conformité locales

Renseignements relatifs à la commande

Produit	N° de référence
ICA Professional Annual Subscription	20044876
ICA Enterprise Annual Subscription	20038994
ICA Enterprise Compliance Add-on	20066830
ICA Training and Onboarding	20049422

En savoir plus

Consultez illumina.com/ConnectedAnalytics

Références

1. Enabling responsible genomic data sharing for the benefit of human health. Site de l'Alliance mondiale pour la génomique et la santé. www.ga4gh.org. Consulté le 22 octobre 2020.
2. Illumina DRAGEN Bio-IT Platform | Variant calling & secondary genomic analysis. Site d'Illumina. www.illumina.com/products/by-type/informatics-products/dragen-bio-it-platform.html. Consulté le 22 octobre 2020.
3. TensorFlow. Site de TensorFlow. tensorflow.org. Consulté le 11 janvier 2021.
4. scikit-learn: machine learning in Python. Site de scikit-learn. scikit-learn.org/stable/. Consulté le 11 janvier 2021.
5. Cloud Security—Amazon Web Services (AWS). Site d'Amazon. aws.amazon.com/security. Consulté le 22 octobre 2020.
6. General Data Protection Regulation (GDPR) Compliance Guidelines. Site du RGPD. gdpr.eu. Consulté le 11 janvier 2021.
7. US Department of Health & Human Services. Health Information Privacy. Site du HHS. hhs.gov/hipaa/index.html. Consulté le 11 janvier 2021.
8. International Organization for Standardization. ISO-ISO/IEC 27001—Information security management. Site de l'ISO. iso.org/isoiec-27001-information-security.html. Consulté le 11 janvier 2021.
9. iCredits for Data Storage and Analysis | Illumina Analytics. Site d'Illumina. www.illumina.com/products/by-type/informatics-products/icredits.html. Consulté le 22 octobre 2020.

illumina[MD]

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

M-GL-00684 FRA v2.0.