

Illumina Connected Analytics

Transformez vos opérations
bioinformatiques

- Intégration rationalisée avec le système de séquençage
- Débit évolutif pour répondre aux besoins de traitement des échantillons
- Interface intuitive pour sélectionner, créer et personnaliser les flux de travail

Introduction

Les avancées dans les technologies de séquençage de 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 du développement et de la mise en œuvre des pipelines, faire évoluer les flux de travail informatiques et assurer une gestion sécurisée des données nécessitent une plateforme flexible et complète. Illumina Connected Analytics 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.

Illumina Connected Analytics est une plateforme de données bioinformatique sécurisée infonuagique qui rend l'informatique opérationnelle et génère des renseignements scientifiques (figure 1, tableau 1). Illumina Connected Analytics permet aux clients de créer des flux de travail :

- Connectés, via une intégration rationalisée aux systèmes et logiciels d'Illumina
- Évolutifs, avec la possibilité de gérer, d'analyser et de rechercher des ensembles de données massifs
- Flexibles, ce qui permet aux utilisateurs de créer et de personnaliser des flux de travail avec les applications DRAGEN^{MC} et des pipelines d'analyse personnalisés

Flux de travail rationalisé

Illumina Connected Analytics est un composant central pour les laboratoires effectuant des études de séquençage de nouvelle génération (SNG) avec des systèmes de séquençage Illumina. Tirant profit de l'élasticité des ressources offertes par l'infonuagique, Illumina Connected Analytics 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 Illumina Connected Analytics.

Au sein d'Illumina Connected Analytics, les données peuvent être automatiquement analysées avec les pipelines d'analyse secondaire DRAGEN 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. Illumina Connected Analytics 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.

Tableau 1 : Aperçu d'Illumina Connected Analytics

	Caractéristique	Avantage
Sécurité et confidentialité	Conformité	Développée d'après le système Illumina QMS (Quality Management System) conformément aux normes réglementaires locales, régionales et mondiales, notamment les normes du RGPD ² , HIPAA ³ , ISO 27001/27701 ^{4,5} et ISO 13485 ⁶
	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'activités et d'événements pour la surveillance de l'accès
	Authentification multifacteur (AMF)	Appliquez les identifiants institutionnels pour gérer et contrôler l'accès au système
Attribution des ressources	Évolutivité à la demande	Augmentez les besoins en stockage sur nuage et en puissance de calcul en fonction du niveau de demande actuel
	Ressources informatiques à la demande	Réduisez les coûts en ne payant que les ressources informatiques dans le moteur du pipeline
Gestion	Gestion des projets et des utilisateurs	Gérez l'accès et l'activité des utilisateurs pour une confidentialité granulaire
	Partage des données	Décompartmentez les données pour encourager une collaboration mondiale à grande échelle et favoriser la transmission des données
	Archivage des données	Réduisez les coûts de la gestion des données à long terme grâce aux niveaux d'archivage
	Apportez votre propre compartiment infonuagique AWS	Accédez aux données stockées sur un compte infonuagique privé AWS
Facilité d'utilisation et accessibilité	Intégration directe des systèmes de séquençage	Envoyez directement les données depuis les systèmes de séquençage Illumina vers le stockage et l'analyse d'Illumina Connected Analytics
	Outils et pipelines	Exploitez des pipelines DRAGEN prêts à l'emploi ou importez des outils existants
	API et CLI	Interagissez de façon programmatique avec la plateforme, les outils et les données via le CLI ou l'API RESTful
	Interface graphique intuitive	Accédez aux données et aux outils via une interface Web graphique; affichez 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 et lancez des analyses dans le langage en toute simplicité
	Intégré à JupyterLab	Exécutez des scripts personnalisés, créez et entraînez des modèles d'IA/ d'apprentissage machine et interagissez avec les données au sein de la plateforme grâce à des espaces de travail collaboratifs avec des blocs-notes
	Agrégation et interrogation de données	Organisez et interrogez des ensembles de données multiomiques pour alimenter l'analyse tertiaire

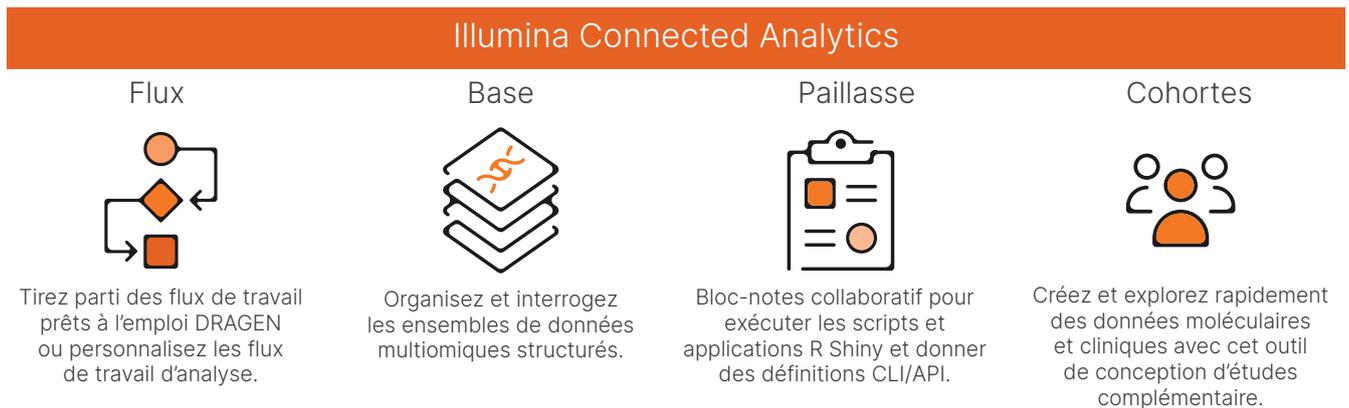


Figure 1 : Composants modulaires Illumina Connected Analytics : traitez, agrégez et analysez les données dans un environnement de gestion des données sécurisé et dont l'accès est contrôlé.

BaseSpace^{MC} Sequence Hub

Compris dans tous les abonnements Illumina Connected Analytics, BaseSpace Sequence Hub est une extension directe de vos instruments Illumina. Améliorez l'efficacité du laboratoire grâce à l'intégration de BaseSpace Sequence Hub en configurant les analyses, en surveillant la qualité des analyses et en diffusant les données de séquençage directement dans le nuage. Le flux de données chiffrées transférées de l'instrument à BaseSpace Sequence Hub permet de gérer et d'analyser facilement les données à l'aide d'un ensemble d'applications soigneusement sélectionnées dans un environnement sécurisé.

Transformation des lectures en données

Illumina Connected Analytics 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, Illumina Connected Analytics peut prendre en charge pratiquement n'importe quelle application informatique.

Options prêtes à utiliser

Illumina Connected Analytics propose de puissants outils et pipelines prêts à l'emploi pour le traitement des données, y compris l'accès à l'analyse secondaire DRAGEN¹ qui fournit une analyse secondaire précise, efficace et complète des données de séquençage (figure 2).

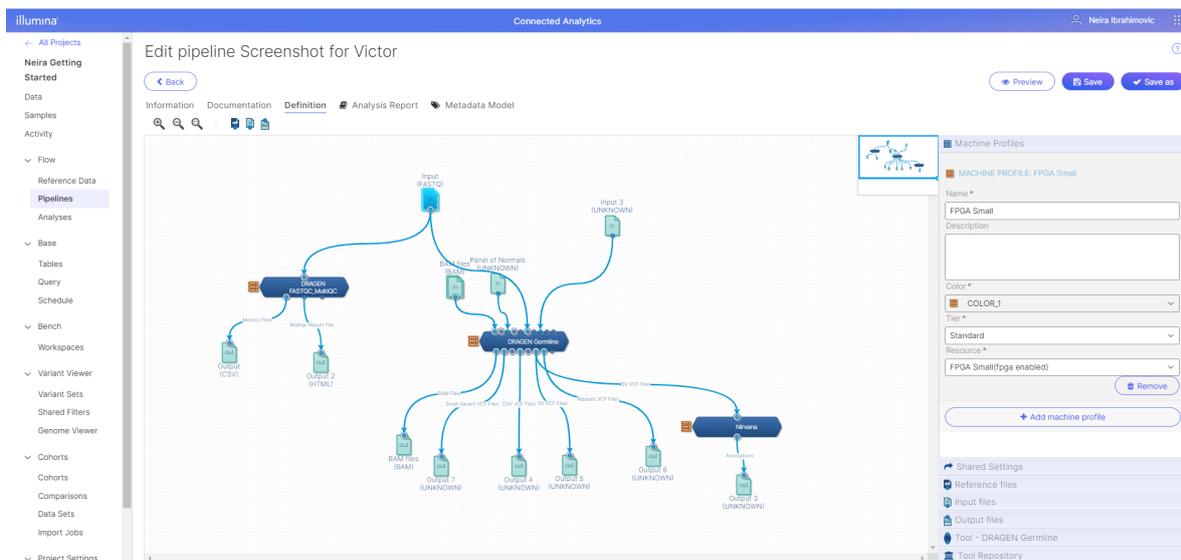


Figure 2 : Flux d'Illumina Connected Analytics : souplesse de la construction, du déploiement et du développement des nouvelles versions de pipelines analytiques.

Personnalisation des pipelines

Les bioinformaticiens peuvent importer des outils d'analyse existants depuis un référentiel d'images Docker Hub⁷, ou construire et éditer de nouveaux pipelines à l'aide de Nextflow, Common Workflow Language (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.

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, Illumina Connected Analytics 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

Illumina Connected Analytics 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

Illumina Connected Analytics 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 Illumina Connected Analytics pour être analysés et partagés.

Agrégation et interrogation de données

Illumina Connected Analytics 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. 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 Illumina Connected Analytics 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.

Utilisation des cohortes pour explorer les données clinico-génomiques à grande échelle

Illumina Connected Analytics Cohorts est un outil de conception d'études permettant de créer et d'explorer rapidement des cohortes (figure 3). Illumina Connected Analytics Cohorts accélère la découverte génomique pour ouvrir la voie à des études significatives et permet une meilleure prise de décision sur la conception des études en agrégeant les données multiomiques et phénotypiques en un seul endroit.

Augmentez les cohortes avec les ensembles de données publiques harmonisées, notamment les données du The Cancer Genome Atlas (TCGA), du Broad Rare Genomes Project, du 1000 Genomes Project et du Gene Expression Omnibus. Créez des hypothèses en quelques minutes plutôt qu'en plusieurs semaines en tirant parti d'un navigateur de cohorte intégré. Accédez à des ensembles de données publiques hautement organisées. Utilisez des outils visuels simples pour sélectionner et visualiser les cohortes, évitant ainsi la nécessité d'effectuer des explorations complexes.

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

Le développement et la personnalisation des algorithmes sont des composants essentiels d'Illumina Connected Analytics pour prendre en charge l'exploration approfondie des données. Un module de programmation interactif, qui exploite les blocs-notes populaires Jupyter, Python et R, permet aux scientifiques de données d'analyser les données agrégées dans un environnement fluide et sécurisé (figure 4).

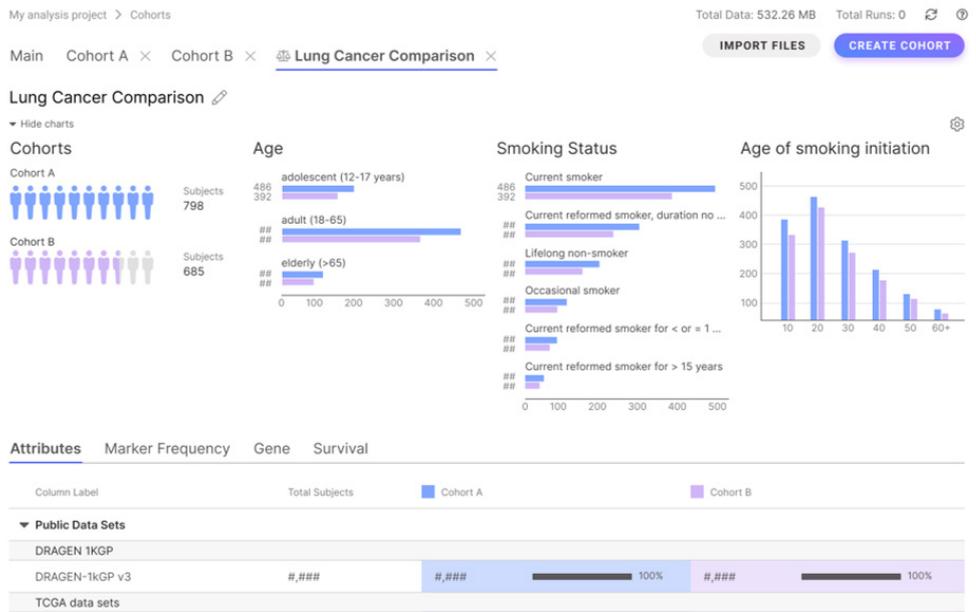


Figure 3 : Illumina Connected Analytics Cohorts : créez et explorez rapidement des données moléculaires et cliniques à l'aide de ce module complémentaire dans Illumina Connected Analytics.



Figure 4 : Illumina Connected Analytics Bench : des outils informatiques flexibles et interactifs sont intégrés à Illumina Connected Analytics Bench, notamment les blocs-notes Jupyter, Python et R.

Lors de la phase de développement des méthodes et des algorithmes, l'utilisateur peut développer ou modifier des pipelines dans un environnement de bac à sable. Dans cet environnement, l'utilisateur a accès à des bibliothèques standard et peut facilement importer ses bibliothèques personnalisées, comme TensorFlow⁸ ou scikit-learn⁹ pour créer des scripts personnalisés et complexes afin de traiter les données au sein d'Illumina Connected Analytics de manière collaborative. Lorsque l'utilisateur est prêt à passer en phase de production, Illumina Connected Analytics permet de convertir les blocs-notes en outils. Ces outils sont ensuite disponibles dans le référentiel d'Illumina Connected Analytics 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. Illumina Connected Analytics 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 AES256¹⁰ et protégées par le protocole TLS (Transport Layer Security).
- Les données contenues dans Illumina Connected Analytics 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-Architected¹¹.
- Illumina Connected Analytics est hébergée dans plus de 10 régions AWS au niveau mondial pour garantir la conformité aux réglementations locales afin de maintenir les données génomiques dans des zones géographiques particulières.
- Le service d'authentification est pris en charge par Security Assertion Markup Language (SAML) 2.0 qui gère les utilisateurs et les mots de passe institutionnels (facultatif).
- Les rapports de vérification prennent en charge la traçabilité de la provenance des données.

Illumina Connected Analytics répond également aux besoins des clients travaillant dans des environnements réglementés, qui doivent se conformer à des exigences strictes. Illumina Connected Analytics a été développée conformément au processus du cycle de vie du développement du logiciel Illumina d'après Illumina QMS (Illumina Quality Management System).

De plus, les processus Illumina QMS ont adopté les meilleures pratiques du secteur et les normes pertinentes, notamment :

- Système de gestion de la sécurité des renseignements ISO 27001:2013⁴ et techniques de sécurité ISO 27701:2019⁵ de la norme de l'Organisation internationale de normalisation (ISO)
- Meilleures pratiques du système de gestion de la qualité (SGQ) organisationnelle ISO 13485⁶
- Règlement général sur la protection des données (RGPD)²
- Health Insurance Portability and Accountability Act (HIPAA)³
- 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 Enterprise Annual Subscription	20038994
ICA Professional Annual Subscription	20044876
ICA Training and Onboarding	20049422
ICA Cohorts Annual Subscription	20065842
ICA Cohorts Annual Subscription	20065842
Illumina Analytics – 1 iCredit	20042038
ICA Enterprise Svc and Compliance Add-on	20066830

Abréviations : ICA : Illumina Connected Analytics

En savoir plus

[Illumina Connected Analytics](#)

Références

1. Illumina. Illumina DRAGEN secondary analysis. www.illumina.com/products/by-type/informatics-products/dragen-secondary-analysis.html. Consulté le 1er mars 2024.
2. General Data Protection Regulation (GDPR) Compliance Guidelines. Site du RGPD. gdpr.eu. Consulté le 1er mars 2024.
3. US Department of Health and Human Services. Health Information Privacy. Site du HHS. hhs.gov/hipaa/index.html. Consulté le 1er mars 2024.
4. International Organization for Standardization. ISO-ISO/IEC 27001—Information security management. Site de l'ISO. iso.org/isoiec-27001-information-security.html. Consulté le 1er mars 2024.
5. International Organization for Standardization. ISO/IEC 27701:2019—Security techniques. iso.org/standard/71670.html. Consulté le 1er mars 2024.
6. International Organization for Standardization. ISO 13485—Medical devices. Site de l'ISO. iso.org/iso-13485-medical-devices.html. Consulté le 26 mars 2024.
7. Docker. Docker Hub Image Container Library. <https://hub.docker.com>. Consulté le 26 mars 2024.
8. TensorFlow. Site de TensorFlow. tensorflow.org. Consulté le 1er mars 2024.
9. scikit-learn: machine learning in Python. Site de scikit-learn. scikit-learn.org/stable/. Consulté le 1er mars 2024.
10. National Institute of Standards and Technology. Advanced Encryption Standard (AES). nist.gov/publications/advanced-encryption-standard-aes-0. Consulté le 26 mars 2024.
11. Cloud Security—Amazon Web Services (AWS). Site d'Amazon. aws.amazon.com/security. Consulté le 1er mars 2024.
12. Illumina. iCredits for Data Storage & Analysis. illumina.com/products/by-type/informatics-products/icredits.html. Consulté le 1er mars 2024.



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.
M-GL-00684 FRA v3.0.