

# Illumina Connected Analytics

Transformez l'afflux de  
données en découverte

- Une solution rationalisée qui permet de traiter les flux de données omiques des lectures aux résultats à grande échelle
- Les interfaces centrées sur l'utilisateur prennent en charge les flux de travail personnalisés et exploitent des outils de datalogie avancés
- Un environnement sécurisé conçu pour assurer la confidentialité des données

**illumina**<sup>®</sup>

## Introduction

Les avancées dans les technologies de séquençage nouvelle génération (NGS) 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. Résoudre les défis d'une gestion sécurisée des données, d'une collaboration, d'une analyse complexe des données et de l'extraction de connaissances de ces données à grande échelle exige de pouvoir passer rapidement de la génération à l'interprétation des données. Illumina Connected Analytics (ICA) est conçu pour gérer, analyser et interpréter cette énorme quantité de données.

ICA est une plateforme infonuagique exhaustive de gestion et d'analyse des données qui permet aux chercheurs d'agrèger, d'explorer et de partager de gros volumes de données multiomiques dans un environnement sécurisé, souple et évolutif (Figure 1, Tableau 1). ICA offre :

- Une intégration directe au flux de travail de génération des données, y compris les systèmes de séquençage Illumina
- Une analyse secondaire puissante grâce à la plateforme DRAGEN<sup>MC</sup> Bio-IT<sup>1</sup>
- Une agrégation des données évolutive et un stockage sécurisé
- Un environnement de datalogie dynamique et interactif pour l'apprentissage machine et l'intelligence artificielle avancés

## Un flux de travail rationalisé

ICA est un composant central des laboratoires effectuant des études de NGS avec des systèmes de séquençage Illumina. Tirant profit de l'élasticité des ressources informatiques offerte 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. Avec BaseSpace<sup>MC</sup> Sequence Hub,<sup>2</sup> l'utilisateur peut intégrer sa plateforme et ses données de séquençage directement à l'environnement ICA.

Tableau 1 : Aperçu d'ICA

	Fonctionnalité	Avantage
Sécurité et confidentialité	Conformité	Respectez les normes réglementaires locales, régionales et mondiales, les normes du HIPAA et du RGPD, et les homologations ISO13485 et ISO27001
	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	Décompartmentez 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 des outils tiers
	API et CLI	Interagissez de façon programmatique avec la plateforme grâce à des outils basés sur les préférences des utilisateurs
	Compte à nuage privé	Connectez-vous à votre propre nuage
	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 et CWL	Écrivez des pipelines en Common Workflow Language (CWL) et lancez des analyses dans le nuage en toute simplicité
	API REST 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

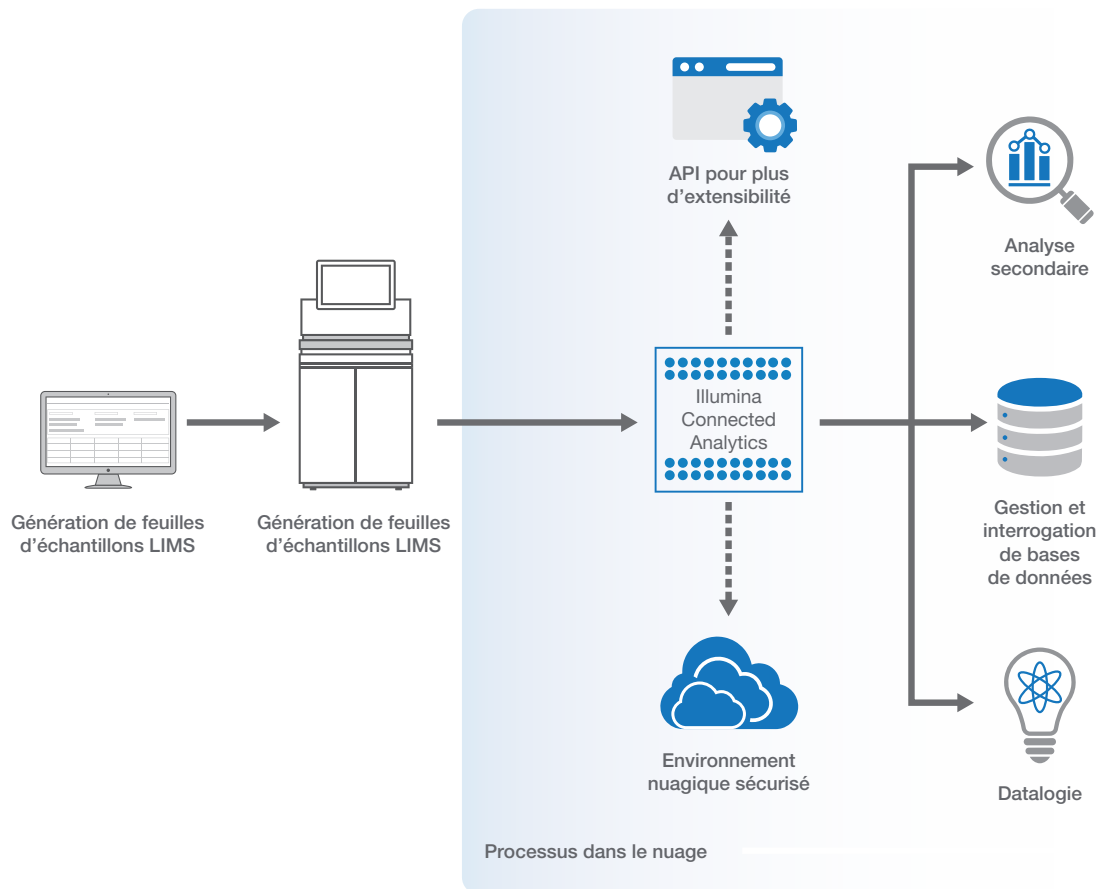


Figure 1 : ICA forme la base de la gestion et de l'analyse des données

Les flux de travail automatisés diffusent les données de l'instrument au nuage à mesure qu'elles sont générées en temps réel et garantissent la disponibilité des lectures dans ICA dans les plus brefs délais.

Une fois dans l'environnement ICA, les données peuvent être automatiquement analysées avec les pipelines 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. ICA fournit une plateforme extensible avec un ensemble riche d'interfaces de programmation (API) REST et un outil d'interface de ligne de commande (CLI). 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).<sup>3</sup>

## 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

Afin de soutenir une approche multiomique de la recherche, ICA a été 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 l'environnement ICA pour être analysés et partagés.

## Transformez les 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,<sup>1</sup> 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 du Common Workflow Language (CWL) et de l'éditeur de pipeline graphique.

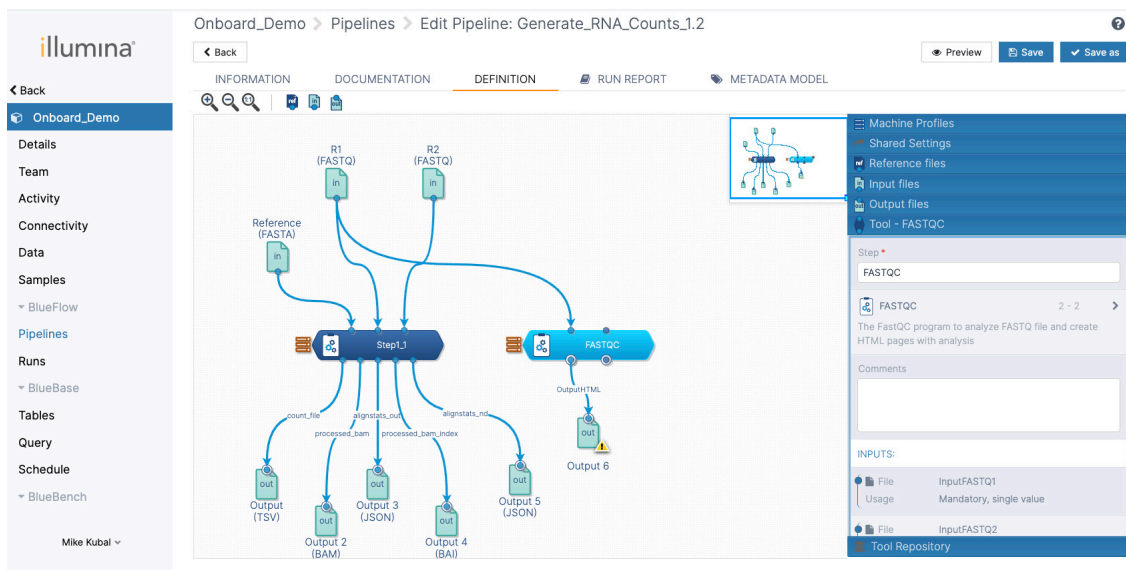


Figure 2 : Pipeline DRAGEN dans ICA — Les utilisateurs peuvent accéder à des pipelines prêts à utiliser depuis la plateforme DRAGEN Bio-IT pour une analyse secondaire des lectures aux rapports rapide et précise.

Les laborantins et autres scientifiques peuvent aussi lancer des pipelines en toute simplicité grâce à l'interface utilisateur intuitive. Pour accélérer le développement des pipelines, les utilisateurs peuvent aussi accéder à ICA Reference Solutions, une collection de pipelines d'analyse pouvant être optimisés pour s'adapter à des besoins spécifiques.

## Apprentissage continu

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 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.

## Prend en charge pratiquement n'importe quelle application génomique

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 Jupyter 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 ses propres pipelines ou modifier les pipelines existants 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<sup>4</sup> ou scikit-learn,<sup>5</sup> 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.

The screenshot displays the 'Edit LoadFeatures' window with the following configuration:

- Search:** .final.count.tsv
- Generated by Pipelines:** <any pipeline>
- Target BlueBase Table:** FeatureCounts
- Write preference:** Append to table
- Data format:** TSV
- Delimiter:** Tab
- Custom delimiter:** (empty)
- Header rows to skip:** 0
- Annotation Fields:** (empty)
- References:**
  - Include references
  - Data Reference
  - Sample Reference
  - Account Reference
  - Pipeline Reference
  - Pipeline Execution Reference
- Advanced Options:**
  - Encoding:** UTF-8
  - Number of errors allowed:** 0
  - Null marker:** (empty)
  - Quote:** (empty)
  - Ignore unknown values
  - Allow quoted newlines
  - Allow jagged rows

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.

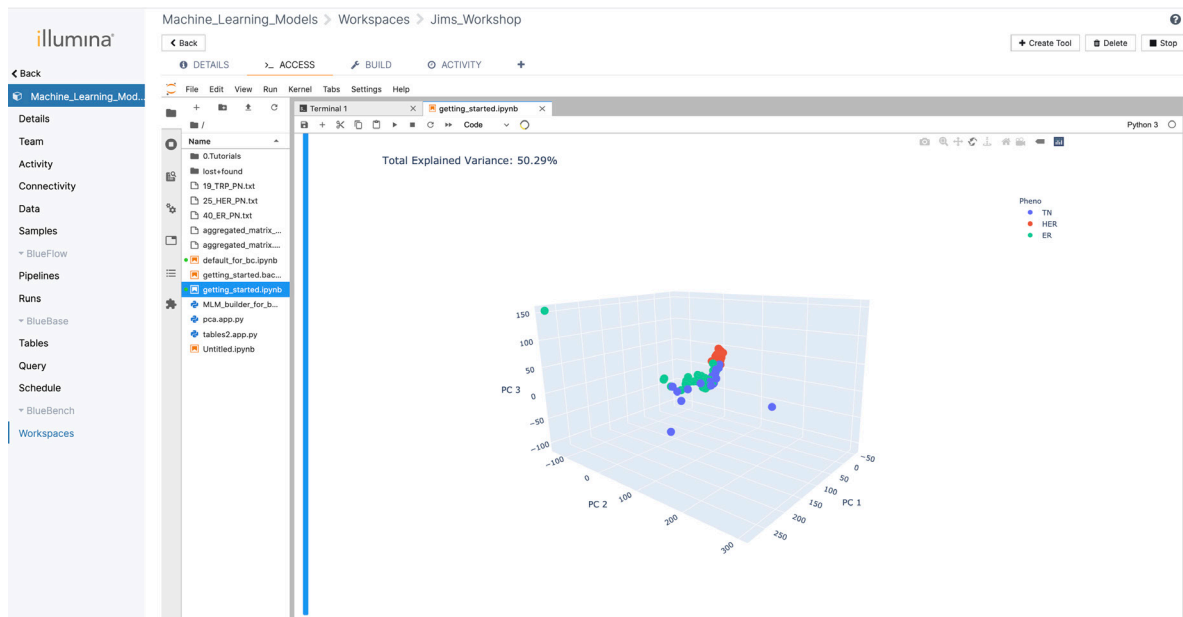


Figure 4 : Analyse interactive et visualisation — ICA prend en charge les Blocs-notes Jupyter pour une exploration visuelle de données multidimensionnelles.

## Un environnement donnant priorité à la sécurité avec prise en charge de la conformité

La sécurité est une préoccupation capitale lorsque l'on travaille avec des données dans un environnement infonuagique. ICA emploie différentes mesures physiques, électroniques et administratives pour respecter toutes les exigences en matière de sécurité des données, même les plus strictes :

- Les données téléchargées 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) et bénéficient des pratiques exemplaires AWS Well-Architected, qui sont conformes à un grand nombre de normes de sécurité acceptées par le secteur<sup>6</sup>

- 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)
- Rapports de vérification pour la traçabilité de la provenance des données

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

- Les lois en vigueur sur la protection des données comme le Règlement général sur la protection des données (RGPD)<sup>7</sup> et le Health Insurance Portability and Accountability Act (HIPAA)<sup>8</sup>
- Les normes de l'Organisation internationale de normalisation (ISO) 13485 – Système de gestion de la qualité<sup>9</sup> et ISO 27001 – Système de gestion de la sécurité de l'information<sup>10</sup>
- La garantie de résidence des données afin d'assurer le respect des exigences réglementaires et de conformité locales



## Options flexibles

Pour plus de flexibilité d'achat, ICA est offerte sous forme d'abonnement annuel. La facturation utilise des iCredits en fonction de l'utilisation des outils et du stockage.<sup>11</sup> Les iCredits peuvent être achetés au préalable, ou facturés mensuellement.

## Études multiomiques évolutives

À mesure que la génération de données NGS gagne en rapidité et baisse en coût, il deviendra essentiel de disposer de plateformes de données avancées qui permettent aux chercheurs de passer des lectures aux rapports facilement et à grande échelle. Avec des solutions puissantes qui permettent une collaboration mondiale en centralisant l'accès aux données distribuées, des pipelines prêts à utiliser et personnalisables, l'accès à des outils de datalogie, et un environnement sécurisé conformément aux réglementations internationales, ICA permet aux utilisateurs d'exploiter pleinement le potentiel des données multiomiques.

## En savoir plus

Consultez [www.illumina.com/ConnectedAnalytics](http://www.illumina.com/ConnectedAnalytics)

## Références

1. Plateforme Illumina DRAGEN Bio-IT | Identification de variants et analyse génomique secondaire. Site d'Illumina. [www.illumina.com/products/by-type/informatics-products/dragen-bio-it-platform.html](http://www.illumina.com/products/by-type/informatics-products/dragen-bio-it-platform.html). Consulté le 22 octobre 2020.
2. BaseSpace Sequence Hub | Calculs génomiques infonuagiques. Site d'Illumina. [www.illumina.com/basespace](http://www.illumina.com/basespace). Consulté le 11 janvier 2021.
3. Favoriser un partage responsable des données génomiques au bénéfice de la santé humaine. Site de l'Alliance mondiale pour la génomique et la santé. [www.ga4gh.org](http://www.ga4gh.org). Consulté le 22 octobre 2020.
4. TensorFlow. Site de TensorFlow. [tensorflow.org](http://tensorflow.org). Consulté le 11 janvier 2021.
5. scikit-learn : apprentissage machine en Python. Site de scikit-learn. [scikit-learn.org/stable/](http://scikit-learn.org/stable/). Consulté le 11 janvier 2021.
6. Sécurité infonuagique — Amazon Web Services (AWS). Site d'Amazon. [aws.amazon.com/security](http://aws.amazon.com/security). Consulté le 22 octobre 2020.
7. Consignes de conformité au Règlement général sur la protection des données (RGPD). Site du RGPD. [gdpr.eu](http://gdpr.eu). Consulté le 11 janvier 2021.
8. Département américain de la Santé et des Services sociaux (HHS). Confidentialité des renseignements médicaux. Site du HHS. [hhs.gov/hipaa/index.html](http://hhs.gov/hipaa/index.html). Consulté le 11 janvier 2021.
9. Organisation internationale de normalisation (ISO). ISO-ISO 13485:2016-Dispositifs médicaux — Systèmes de gestion de la qualité — Exigences réglementaires. Site de l'ISO. [iso.org/standard/59752.html](http://iso.org/standard/59752.html). Consulté le 11 janvier 2021.
10. Organisation internationale de normalisation (ISO). ISO-ISO/CEI 27001 — Gestion de la sécurité de l'information. Site de l'ISO. [iso.org/isoiec-27001-information-security.html](http://iso.org/isoiec-27001-information-security.html). Consulté le 11 janvier 2021.
11. iCredits pour le stockage et l'analyse des données | Illumina Analytics. Site d'Illumina. [www.illumina.com/products/by-type/informatics-products/icredits.html](http://www.illumina.com/products/by-type/informatics-products/icredits.html). Consulté le 22 octobre 2020.

## Renseignements relatifs à la commande

Produit	N° de référence
ICA Enterprise (Entreprise)	20038994
ICA Data Science (Datalogie)	20044877
Illumina Analytics – 1 iCredit	20042038
Illumina Analytics – 1000 iCredits	20042039
Illumina Analytics – 5000 iCredits	20042040
Illumina Analytics – 50 000 iCredits	20042041
Illumina Analytics – 100 000 iCredits	20042042
Facturation de la consommation	20012931

a. N'inclut pas les fonctionnalités de datalogie.  
b. Donne accès aux Blocs-notes (Jupyter, R) et à l'infrastructure d'IA/apprentissage machine.  
c. Le montant à ne pas dépasser est indiqué par le montant sur le devis. Les clients seront facturés mensuellement sur la base de la consommation de puissance de calcul, de stockage, et d'applications tierces jusqu'au montant associé à la référence N° 20012931.

illumina®

+ (1) 800 809 4566 numéro sans frais (États-Unis) |

+ (1) 858 202 4566 tél.

techsupport@illumina.com | [www.illumina.com](http://www.illumina.com)

© 2020 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](http://www.illumina.com/company/legal.html).  
N° de publication 986-2020-009-A FRA. QB11606.