

Illumina Connected Analytics

Wachsende Datenbestände
für Erkenntnisgewinn nutzen

- Optimierte Reads-zu-Ergebnisse-Lösung zur Verarbeitung von Multiomik-Workflows in großem Maßstab
- Anwenderspezifische Benutzeroberflächen für individuell konfigurierte Workflows und fortschrittliche Datenwissenschaftstools
- Sichere, im Hinblick auf Datenschutz konzipierte Umgebung

illumina®

Einleitung

Fortschritte bei der NGS-Technologie (Next-Generation Sequencing, Sequenzierung der nächsten Generation) haben die Taktfrequenz in den Biowissenschaften und in der klinischen Forschung deutlich erhöht. Mit der steigenden Sequenzierungsgeschwindigkeit und den zugleich sinkenden Kosten übersteigen die Kapazitäten zur Generierung von Daten die Kapazitäten zur Gewinnung biologischer und klinischer Erkenntnisse aus diesen Daten deutlich. Die Herausforderungen in Bezug auf sicheres Datenmanagement, Zusammenarbeit, die Analyse komplexer Daten und die Auswertung von Daten in großem Maßstab lassen sich nur bewältigen, wenn sich der Schritt von der Generierung zur Interpretation der Daten einfach vollziehen lässt. Illumina Connected Analytics (ICA) wurde speziell für das Management, die Analyse und die Interpretation dieser enormen Datenmengen entwickelt.

Bei ICA handelt es sich um eine umfassende cloud-basierte Plattform für Datenmanagement und -analyse, über die Wissenschaftler große Mengen von Multiomik-Daten in einer sicheren, skalierbaren und flexiblen Umgebung sammeln, untersuchen und teilen können (Abbildung 1, Tabelle 1). Vorteile von ICA:

- Direkte Integration in den Datengenerierungs-Workflow, einschließlich Illumina-Sequenziersystemen
- Leistungsstarke Sekundäranalyse auf der DRAGEN™ Bio-IT-Plattform¹
- Skalierbare Aggregation und sichere Speicherung von Daten
- Dynamische, interaktive Datenwissenschaftsumgebung für Machine Learning und künstliche Intelligenz

Optimierter Workflow

Bei ICA handelt es sich um eine zentrale Komponente für Labore, die Illumina-Sequenziersysteme für NGS-Studien nutzen. Dank der Vorteile, die die Flexibilität der Rechenkapazitäten beim Cloud-Computing bietet, eignet sich ICA mit derselben Architektur für Anwendungen jeden Umfangs – vom gelegentlichen Screening Zehntausender Zellen in komplexen Einzelzellprojekten bis hin zur Gesamtgenomsequenzierung bei Populationen. Über BaseSpace™ Sequence Hub² können Anwender ihre Sequenzierungsplattform und entsprechende Daten direkt in die ICA-Umgebung integrieren.

Tabelle 1: Übersicht über ICA

	Merkmal	Vorteil
Sicherheit und Datenschutz	Compliance	Einhaltung lokaler, regionaler und globaler gesetzlicher Vorgaben, des HIPPA und der DSGVO sowie der Zertifizierung nach ISO13485 und ISO27001
	Sicherheitsmaßnahmen	Strikte Datentrennung, Verschlüsselung mit TLS 1.2 (Übertragung) und AES 256 (Speicherung)
	Auditpfad	Aktivitätsprotokoll mit Verfolgung, wer wann welche Daten aufgerufen hat
	Single Sign-on (SSO) (optional)	Nutzung einrichtungsspezifischer Anmeldeinformationen für die Zugriffskontrolle
Resourcing	Computing-Ressourcen on demand	Geringere Kosten, indem nur Computing-Ressourcen in der Pipeline-Engine abgerechnet werden
	Skalierung on demand	Skalierung von Cloudspeicherplatz und Computing-Kapazitäten je nach aktuellem Bedarf
	Plattform- und Nutzungs-Dashboard	Anzeigen des Ressourcenbedarfs zur effizienten Übersicht, Verwaltung und Anforderungsprognose
Verwaltung	Projekt- und Benutzerverwaltung	Präzise Datenschutzeinstellungen durch die Verwaltung von Benutzerzugriff und -aktivität
	Gemeinsame Nutzung von Daten	Weltweite Zusammenarbeit in großem Maßstab dank der Überwindung von Datensilos
	Datenarchiv	Kostensenkung durch die Archivierung nicht verwendeter Daten in günstigeren Speicherebenen
Anwenderfreundlichkeit	Direkte Integration von Sequenziersystemen	Unmittelbare Übertragung der Daten von Illumina-Sequenziersystemen
	Grafischer Pipeline-Builder	Erstellung von Pipelines ohne Programmierung
	Tools und Pipelines	Nutzung von vorgefertigten Pipelines und Drittanbieter-Tools
	APIs und CLI	Programmgesteuerte Verknüpfung der Plattform mit vom Anwender bevorzugten Tools
	„Bring your own Cloud“-Konto	Verknüpfung der privaten Cloud
	Datenvisualisierung	Erstellung dynamischer Diagramme und interaktiver Web-Apps zur Darstellung von Daten mit R- und Python-Paketen
Fortschrittliche Tools	Unterstützung für Docker und CWL	Programmierung von Pipelines in einer gemeinsamen Workflow-Sprache und einfacher Start von Analysen in der Cloud
	GA4GH-konforme RESTful-APIs	Programmgesteuerter Zugriff auf Tools und Daten sowie Kompatibilität mit anderen Softwareumgebungen
	JupyterLab-Integration	Ausführung erweiterter Datenanalysen: Erstellen und Trainieren von KI-/ML-Modellen mit R und Python
	Datenaggregation und -abfrage	Datenabfragen auf Populationsebene mithilfe von SQL

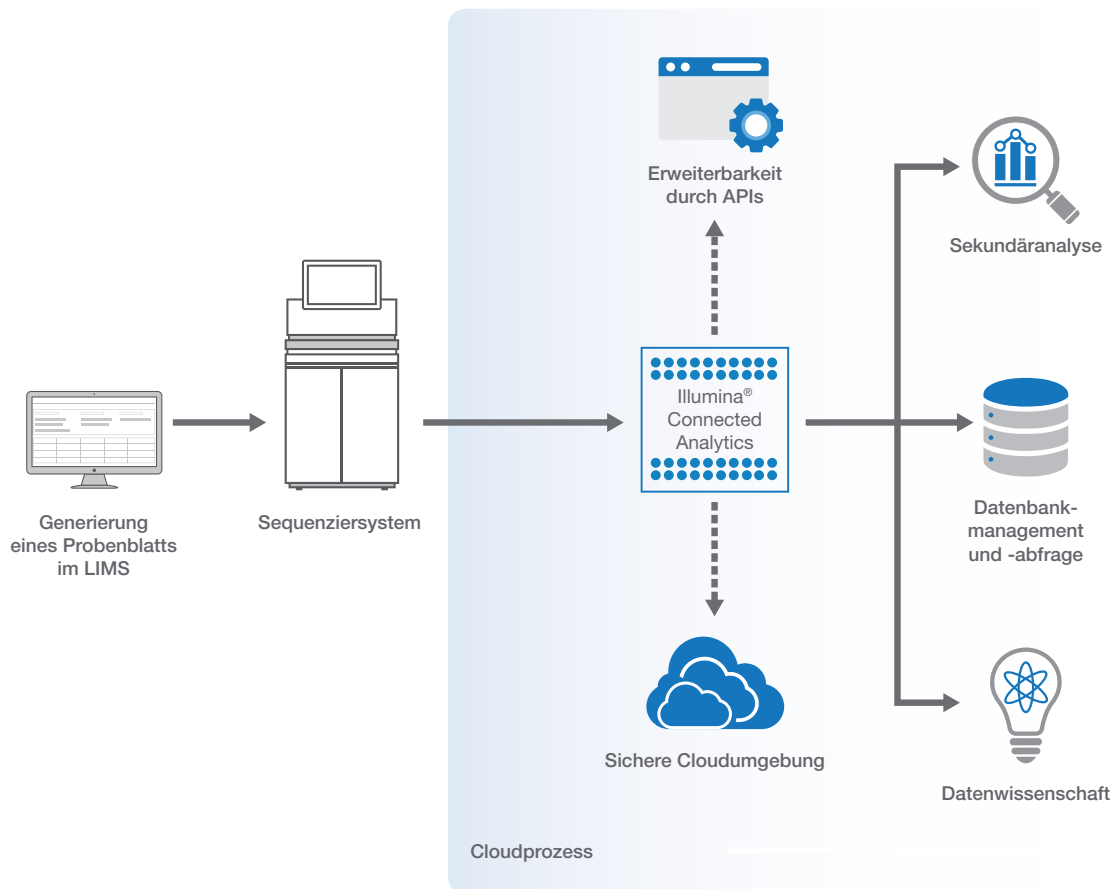


Abbildung 1: ICA bildet die Grundlage für Datenmanagement und -analyse.

Dank der automatisierten Workflows werden Daten während der Generierung in Echtzeit vom Gerät in die Cloud übertragen. Dadurch sind die Reads so schnell wie möglich für die Analyse verfügbar.

Sobald sich die Daten in der ICA-Umgebung befinden, können sie, je nach angegebenem Workflow, automatisch mit gebrauchsfertigen DRAGEN- oder anwenderspezifischen Pipelines analysiert werden. Der breite Umfang an Analyseoptionen reicht von der Qualitätssicherung bis hin zu Datenaggregation und fortschrittlichen Datenwissenschaftstools für die schnelle, skalierbare Datenverarbeitung. ICA bietet eine erweiterbare Plattform mit zahlreichen RESTful-APIs (Application Program Interfaces) sowie einem Befehlszeilentool (Command-Line Interface, CLI). Diese APIs maximieren die Effizienz der Workflows, da Daten während ihres Lebenszyklus übertragen, abgerufen und verwendet werden, und umfassen Global-Alliance-for-Genomics-and-Health (GA4GH)-konforme APIs.³

Datenmanagement und Steuerung

Mit der steigenden Menge generierter Daten steigt auch der Bedarf an Infrastruktur für die Freigabe, Wiederverwendung und Integration von Daten in der Scientific Community, um den Wert einzelner Datensätze zu erhöhen. ICA ermöglicht dies mit zahlreichen Funktionen für die einfache Umsetzung von Best Practices für das Datenmanagement.

Zugriffssteuerung

Dank der differenzierten Zugriffssteuerung kann der Administrator Berechtigungen festlegen und vorhandene einrichtungsinterne Anmeldeinformationen für die Zugriffssteuerung nutzen. Ereignisse und Änderungen werden in einem Auditprotokoll gespeichert.

Für jeden Benutzer wird protokolliert, wann der Zugriff auf die Plattform erfolgt und welche Aktionen auf der Plattform ausgeführt werden. Dies ermöglicht die Durchsetzung von Compliance und Verantwortlichkeit.

Offenes Format

ICA wurde als datentypunabhängige Plattform konzipiert und eignet sich damit für die Anwendung von Multiomik-Konzepten in der Forschung. Die Plattform unterstützt die Analyse zahlreicher Datentypen, darunter Molekül-, Klinik- und Phänotyp-Daten sowie unstrukturierte Daten wie Bilder.

Zusammenarbeit

ICA ermöglicht internationale Zusammenarbeit unter Wahrung der Compliance. Daten und Tools lassen sich unmittelbar bereitstellen und mit anderen Anwendern gemeinsam nutzen, wobei die Integrität der Daten sowie der Datenschutz gewährleistet bleiben. Außerdem lassen sich in externen Cloudquellen gespeicherte Daten zur Analyse und Freigabe in die ICA-Umgebung importieren.

Reads in Daten umwandeln

ICA bietet zahlreiche Optionen für die Sekundäranalyse, wodurch der Reads-zu-Ergebnisse-Workflow optimiert wird. Mit der Flexibilität, die die Wahl zwischen dem Einsatz vorgefertigter Pipelines und anwendungsspezifisch erstellter und konfigurierter Pipelines ermöglicht, eignet sich ICA für praktisch jede Informatikanwendung.

Gebrauchsfertige Optionen

ICA umfasst leistungsstarke, sofort einsatzbereite Tools und Pipelines für die Datenverarbeitung wie den Zugriff auf die DRAGEN Bio-IT-Plattform,¹ die eine schnelle, genaue Sekundäranalyse von Sequenzierungsdaten ermöglicht (Abbildung 2).

Anwendungsspezifische Pipelines

Bioinformatiker können vorhandene Tools aus einem Docker-Image-Repository importieren oder neue Pipelines mithilfe von Common Workflow Language (CWL) und dem grafischen Pipeline-Editor erstellen und bearbeiten.

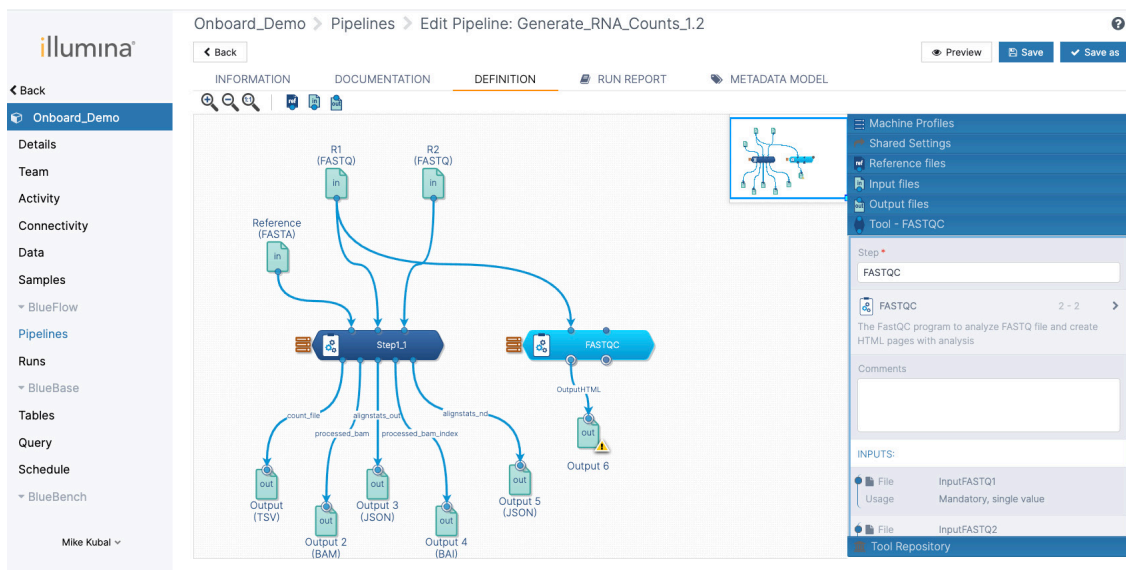


Abbildung 2: DRAGEN-Pipeline in ICA: Benutzer erhalten Zugriff auf gebrauchsfertige Pipelines der DRAGEN Bio-IT-Plattform für die schnelle und genaue Reads-zu-Bericht-Sekundäranalyse.

Über die intuitive Benutzeroberfläche können Laboranwender und andere Wissenschaftler Pipelines einfach ausführen. Der Zugang zu ICA Reference Solutions, einer Sammlung für spezifische Anforderungen optimierbarer Pipelines, ermöglicht Anwendern die schnellere Entwicklung von Pipelines.

Kontinuierliches Lernen

ICA automatisiert komplexe Aggregations- sowie Integrationsschritte und schafft dadurch ein funktionales Wissensmanagementsystem mit Daten aus Millionen von Proben (Abbildung 3). Die Plattform erfasst sämtliche verfügbaren Datentypen, darunter Genomik-, Phänotyp- und Metadaten, Annotationen sowie weitere zugehörige Informationen. Anwender können bedarfsabhängig eigene Datenmodelle definieren, eigene Abfragen erstellen und Beziehungen zwischen Datensätzen untersuchen. Die auf der ICA-Plattform aggregierten Daten bilden einen Informationsschatz, in dem sich u. a. neue Biomarker identifizieren, Patientenpopulationen stratifizieren und die Assay-Performance im Zeitverlauf verfolgen lassen.

Für praktisch alle Genomik- anwendungen geeignet

Vor dem Hintergrund des Umfangs der laufenden Datenexploration ist die Möglichkeit zur Entwicklung und Anpassung von Algorithmen von höchster Bedeutung. Ein interaktives Programmiermodul auf Basis verbreiteter Jupyter Notebooks (Python und R) ermöglicht Datenwissenschaftlern die Analyse aggregierter Daten in einer nahtlosen und sicheren Umgebung (Abbildung 4).

Während der Entwicklung von Methoden und Algorithmen können Anwender in einer Sandbox-Umgebung eigene Pipelines entwickeln oder vorhandene anpassen. Hier besteht die Möglichkeit, Machine-Learning-Modelle schnell zu erstellen, zu testen und fortlaufend zu optimieren. Benutzer haben Zugriff auf zahlreiche Standardbibliotheken wie TensorFlow⁴ oder scikit-learn⁵ und können einfach eigene anwendungsspezifische Bibliotheken integrieren. Beim Übergang in die Produktionsphase können Anwender die Notebooks mit ICA in Tools konvertieren. Diese Tools stehen anschließend im ICA-Tools-Repository zur Verfügung und werden in Produktionspipelines integriert.

Abbildung 3: ICA ermöglicht die Aggregation von Daten, Data Mining und kontinuierliches Lernen: Anwender können Zusammenhänge zwischen Datensätzen untersuchen, um anwendungsspezifische Fragen zu beantworten.

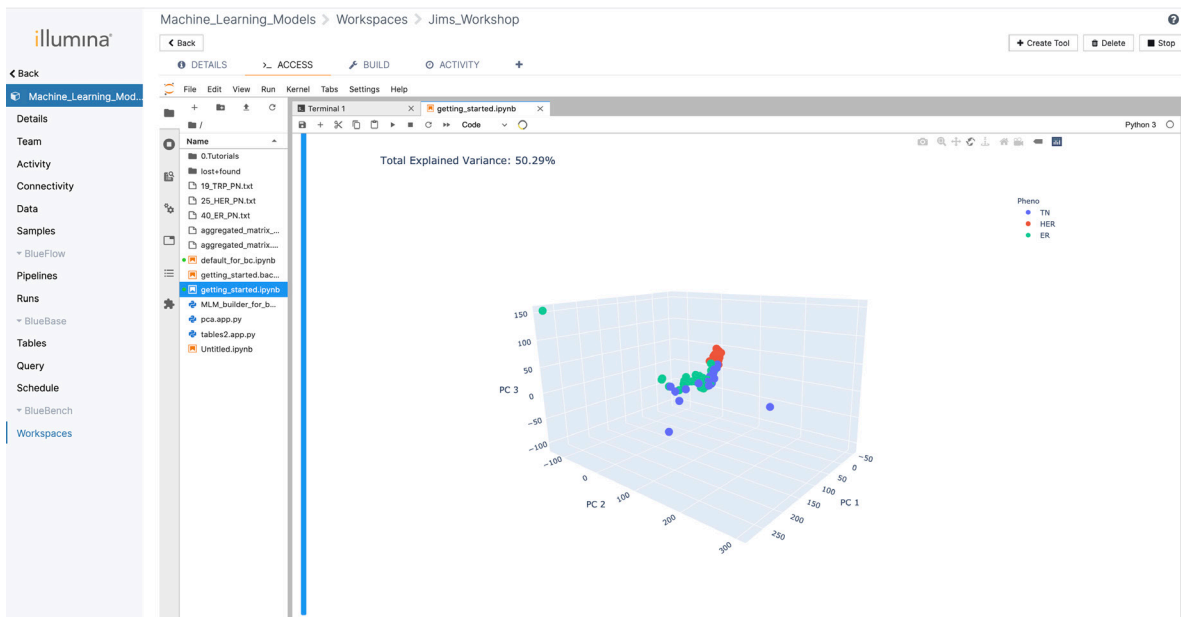


Abbildung 4: Interaktive Analyse und Visualisierung: ICA unterstützt den Einsatz von Jupyter Notebooks zur visuellen Auswertung mehrdimensionaler Daten.

Sichere Umgebung zur Gewährleistung der Compliance

Sicherheit ist bei der Verarbeitung von Daten in einer Cloudumgebung von höchster Bedeutung. ICA gewährleistet mit unterschiedlichen physischen, elektronischen und administrativen Sicherheitsmaßnahmen, dass selbst strengste Datenschutzbestimmungen erfüllt werden:

- Von Sequenzierungsgeräten hochgeladene Daten werden mit Standard AES 256 verschlüsselt und durch TLS (Transfer Layer Security) geschützt.
- Die in ICA enthaltenen Daten werden bei Amazon Web Services (AWS) gehostet. Hierbei kommen AWS Well-Architected-Best-Practices zum Einsatz, die mit zahlreichen branchenüblichen Sicherheitsstandards konform sind.⁶
- Der Authentifizierungsdienst ermöglicht die optionale Verwaltung von Benutzern und Kennwörtern der Einrichtung per SAML 2.0.
- Auditberichte zur Nachverfolgbarkeit des Ursprungs von Daten.

ICA eignet sich auch für den Einsatz bei Kunden, die in regulierten Bereichen tätig sind, und erfüllt strenge Bestimmungen:

- Aktuelle Datenschutzbestimmungen wie die Datenschutz-Grundverordnung (DSGVO)⁷ und der Health Insurance Portability and Accountability Act (HIPAA)⁸
- Qualitätsmanagementsystem⁹ gemäß ISO (International Organization for Standardization) 13485 und Informationssicherheitsmanagementsystem gemäß ISO 27001¹⁰
- Garantierter Datenspeicherort zur Gewährleistung der Einhaltung lokaler regulatorischer und Compliance-Bestimmungen

Flexible Optionen

ICA ist in einem flexiblen Jahresabonnement erhältlich. Die Abrechnung erfolgt in iCredits anhand der Nutzung von Speicherplatz und Tools.¹¹ iCredits können vorab erworben oder monatlich in Rechnung gestellt werden.

Skalierbare Multiomik-Studien

Die Generierung von NGS-Daten wird immer schneller und billiger. Daher benötigen Wissenschaftler unbedingt fortschrittliche Datenplattformen, um schnell und einfach von Reads zu Berichten zu gelangen. Mit leistungsstarken Lösungen für die globale internationale Zusammenarbeit durch zentralen Zugriff auf verteilte Daten, gebrauchsfertige und anpassbare Pipelines, Zugriff auf Datenwissenschaftstools und eine sichere Umgebung, die weltweit mit gesetzlichen Vorgaben konform ist, sorgt ICA dafür, dass Anwender das Potenzial ihrer Multiomik-Daten voll ausschöpfen können.

8. US Department of Health & Human Services. Health Information Privacy. HHS-Website. [hhs.gov/hipaa/index.html](https://www.hhs.gov/hipaa/index.html). Aufgerufen am 11. Januar 2021.
9. International Organization for Standardization. ISO-ISO 13485:2016-Medical devices—Quality management systems—Requirements for regulatory purposes. ISO-Website. [iso.org/standard/59752.html](https://www.iso.org/standard/59752.html). Aufgerufen am 11. Januar 2021.
10. International Organization for Standardization. ISO-ISO/IEC 27001—Information security management. ISO-Website. [iso.org/isoiec-27001-information-security.html](https://www.iso.org/isoiec-27001-information-security.html). Aufgerufen am 11. Januar 2021.
11. iCredits for Data Storage and Analysis | Illumina Analytics. Illumina-Website. www.illumina.com/products/by-type/informatics-products/icredits.html. Aufgerufen am 22. Oktober 2020.

Weitere Informationen

Besuchen Sie www.illumina.com/ConnectedAnalytics

Quellen

1. Illumina DRAGEN Bio-IT Platform | Variant calling & secondary genomic analysis. Illumina-Website. www.illumina.com/products/by-type/informatics-products/dragen-bio-it-platform.html. Aufgerufen am 22. Oktober 2020.
2. BaseSpace Sequence Hub | Cloud-based genomics computing. Illumina-Website. www.illumina.com/basespace. Aufgerufen am 11. Januar 2021.
3. Enabling responsible genomic data sharing for the benefit of human health. Global Alliance for Genomics & Health-Website. www.ga4gh.org. Aufgerufen am 22. Oktober 2020.
4. TensorFlow. TensorFlow-Website. [tensorflow.org](https://www.tensorflow.org). Aufgerufen am 11. Januar 2021.
5. scikit-learn: machine learning in Python. scikit-learn-Website. scikit-learn.org/stable/. Aufgerufen am 11. Januar 2021.
6. Cloud Security—Amazon Web Services (AWS). Amazon-Website. aws.amazon.com/security. Aufgerufen am 22. Oktober 2020.
7. General Data Protection Regulation (GDPR) Compliance Guidelines. GDPR-Website. gdpr.eu. Aufgerufen am 11. Januar 2021.

Bestellinformationen

Produkt	Katalog-Nr.
ICA Enterprise	20038994
ICA Data Science	20044877
Illumina Analytics – 1 iCredit	20042038
Illumina Analytics – 1.000 iCredits	20042039
Illumina Analytics – 5.000 iCredits	20042040
Illumina Analytics – 50.000 iCredits	20042041
Illumina Analytics – 100.000 iCredits	20042042
Verbrauchsabhängige Abrechnung	20012931

a. Datenwissenschaftsfunktionen nicht enthalten.
 b. Bietet Zugriff auf Notebooks (Jupyter, R) und KI-/ML-Framework.
 c. Der Maximalbetrag entspricht dem Betrag im Angebot. Kunden erhalten monatliche Rechnungen für Rechenkapazitäten, Speicherplatz und Drittanbieter-Apps bis zum für Katalog-Nr. 20012931 gültigen Maximalbetrag.

illumina[®]

1.800.809.4566 (USA, gebührenfrei) | +1.858.202.4566

(Tel. außerhalb der USA)

techsupport@illumina.com | www.illumina.com

© 2020 Illumina, Inc. Alle Rechte vorbehalten. Alle Marken sind Eigentum von Illumina, Inc. bzw. der jeweiligen Eigentümer.

Spezifische Informationen zu Marken finden Sie unter www.illumina.com/company/legal.html.

Pub.-Nr. 986-2020-009-A DEU. QB11606.