

Illumina Connected Analytics

바이오인포매틱스 운영을
변화시키는 연구 플랫폼

- 간소화된 시퀀싱 시스템과의 통합
- 다양한 샘플 처리 요구 사항에 따라 조정 가능한 처리량
- 워크플로우의 선택, 구축 및 맞춤 설정을 지원하는 직관적인 인터페이스

illumina[®]

소개

차세대 시퀀싱(Next-generation sequencing, NGS) 기술의 발전은 생명 과학 및 임상 연구의 진행 속도를 크게 높였습니다. 시퀀싱 속도가 빨라지고 비용이 절감됨에 따라, 데이터를 생성하는 능력이 데이터에서 생물학적, 임상적 정보를 추출하는 능력을 훨씬 능가하게 될 것입니다. 파이프라인 개발 및 실행 시 어려움을 극복하고, 인포매틱스(informatics, 정보학) 워크플로우의 규모를 조정하며, 지속적으로 안전하게 데이터를 관리하려면 유연하고 종합적인 플랫폼이 필요합니다. Illumina Connected Analytics는 연구자가 데이터 프라이버시, 보안 및 규정 준수 기준을 충족하도록 돕고, 분석 파이프라인의 유연한 빌드, 배포 및 버전 관리를 지원합니다.

Illumina Connected Analytics는 인포매틱스를 운영화하고 과학적 통찰을 얻을 수 있게 해주는 안전한 클라우드 기반의 바이오인포매틱스(bioinformatics, 생명정보학) 데이터 플랫폼입니다(그림 1, 표 1). 연구자는 Illumina Connected Analytics를 활용해 다음과 같은 특징을 가진 워크플로우를 만들 수 있습니다.

- 연결성: Illumina 시스템 및 소프트웨어와의 간소화된 통합
- 확장성: 대용량 데이터 세트의 관리, 분석 및 쿼리 기능 포함
- 유연성: DRAGEN™ 앱과 맞춤형 분석 파이프라인을 활용해 연구자가 워크플로우 구축 및 맞춤 설정 가능

간소화된 워크플로우

Illumina Connected Analytics는 Illumina 시퀀싱 시스템을 사용해 NGS 연구를 수행하는 랩의 워크플로우에 핵심적인 구성 요소입니다. 클라우드 컴퓨팅의 장점 중 하나인 탄력성(elasticity; 수요에 따른 리소스 조정)을 활용하는 Illumina Connected Analytics는 동일한 아키텍처를 기반으로 다양한 규모의 작업을 지원하므로 비정기적 스크리닝부터 수만 개의 세포를 분석하는 복잡한 단일세포 시퀀싱(single-cell sequencing) 프로젝트나 인구집단 규모의 전장 유전체 시퀀싱(whole-genome sequencing, WGS) 프로젝트까지 폭넓게 적용할 수 있습니다. 또한 연구자는 사용 중인 기기를 Illumina Connected Analytics와 원활하게 통합할 수 있습니다.

데이터는 Illumina Connected Analytics 내에서 명시된 워크플로우에 따라 별다른 설정 없이 바로 사용이 가능한 DRAGEN Secondary Analysis 파이프라인이나 커스텀 파이프라인을 통해 자동으로 분석할 수 있습니다. 또한 품질 관리(quality control, QC)부터 신속하고 확장 가능한 데이터 프로세싱을 위한 데이터 집계(data aggregation) 및 고급 데이터 과학(data science) 도구에 이르는 광범위한 분석 옵션도 마련되어 있습니다. Illumina Connected Analytics는 다양한 RESTful API(RESTful application programming interface; REST 아키텍처의 제약 조건을 준수하는 애플리케이션 프로그래밍 인터페이스)와 커맨드라인 인터페이스(command-

line interface, CLI) 도구를 갖춘 확장 가능한 플랫폼을 제공합니다. 이러한 API는 수명 주기 전반에 걸쳐 데이터의 전송, 액세스 및 이용 시 워크플로우의 효율성을 극대화해 줍니다.

표 1: Illumina Connected Analytics의 요약

	항목	장점
보안 및 프라이버시	규정 준수	일반 데이터 보호 규정(General Data Protection Regulation, GDPR), ² 건강 보험 양도 및 책임에 관한 법(Health Insurance Portability and Accountability Act, HIPAA), ³ ISO 27001/27701, ^{4,5} ISO 13485 ⁶ 등 지역적/국제적 규제 기준을 준수하여 Illumina의 품질 경영 시스템(quality management system, QMS)에 따라 개발된 플랫폼
	보안 제어 (Security control)	엄격한 데이터 분리(seggregation), "전송 중 (in-transit)"(TLS 1.2) 데이터와 "유휴 상태 (at rest)"의 데이터(AES-256) 암호화 유지
	감사 추적 (Audit trail)	액세스 모니터링을 위한 지속적인 활동 및 이벤트 로그 추적
리소스	다단계 인증 (Multifactor authentication, MFA)	시스템 액세스 관리 및 제어를 위해 기관 자격 증명(institutional credentials) 적용
	수요에 따른 규모 조정	현재 수요에 맞게 클라우드 스토리지 및 컴퓨터 요구 사항 조정
관리	수요에 따른 컴퓨터 리소스 조정	파이프라인 엔진 내 컴퓨터 리소스에 대한 비용만 지급해 비용 절감
	프로젝트 및 사용자 관리	세밀한 프라이버시를 위한 사용자 액세스 및 활동 관리
	데이터 공유	글로벌 협업을 촉진하고 데이터 전송을 지원하기 위해 데이터 사일로(data silo) 연결
사용성 및 접근성	데이터 아카이브	아카이브 스토리지 계층(Archival storage tier)을 이용해 장기 데이터 관리 비용 절감
	Bring Your Own AWS Cloud Bucket	개인 관리 AWS 클라우드 계정에 저장된 데이터 액세스
고급 도구	시퀀싱 시스템 통합	Illumina 시퀀싱 시스템에서 Illumina Connected Analytics 스토리지 및 분석 도구로 원활한 데이터 스트리밍
	도구 및 파이프라인	별다른 설정 없이 즉시 사용 가능한 DRAGEN 파이프라인 적용 또는 기존 도구 불러오기
	API 및 CLI	CLI 또는 RESTful API를 통해 프로그래밍 방식으로 플랫폼, 도구 및 데이터와 상호 작용
고급 도구	직관적인 그래픽 인터페이스	그래픽 웹 인터페이스를 통해 데이터 및 도구에 액세스하고 R 및 Python 패키지로 데이터 표시
	Docker, Nextflow, CWL 지원	공동 워크플로우 언어(Common workflow language, CWL)로 파이프라인을 쓰고 클라우드에서 손쉽게 분석 실행
	JupyterLab과의 통합	커스텀 스크립트 실행, AI/머신 러닝 모델 빌드 및 교육, 협업 노트북 작업 공간을 통해 플랫폼 내 데이터와 상호 작용
	데이터 집계 및 쿼리	3차 분석 지원을 위해 정형 멀티오믹스 데이터 세트 구성 및 쿼리

Illumina Connected Analytics



Flow
별도의 설정 없이 바로 사용이 가능한 DRAGEN 워크플로우 활용 또는 분석 워크플로우 맞춤 설정

Base
정형 멀티오믹스 데이터 세트 구성 및 쿼리

Bench
R Shiny 앱 실행 및 스크립트 & CLI/API 호출을 위한 협업 노트북

Cohorts
분자/유전체 데이터의 신속한 빌드 및 탐구를 돕는 애드온 연구 설계 도구

그림 1: Illumina Connected Analytics의 모듈식 구성 요소 — 액세스가 제어되는 안전한 데이터 관리 환경에서 데이터의 처리, 집계 및 분석 지원

BaseSpace™ Sequence Hub

모든 Illumina Connected Analytics 구독 제품에 포함되어 있는 BaseSpace Sequence Hub는 Illumina 기기의 연장선상에 있습니다. 연구자는 BaseSpace Sequence Hub와의 통합을 통해 런(run)을 설정하고, 런 품질을 모니터링하며, 시퀀싱 데이터를 클라우드로 바로 스트리밍함으로써 랩 효율성을 높일 수 있습니다. 또한 기기에서 BaseSpace Sequence Hub로 암호화된 데이터가 전송되므로 안전한 환경에서 큐레이션(curation)을 거친 다양한 앱을 실행해 손쉽게 데이터를 관리하고 분석할 수 있습니다.

리드를 데이터로 변환

Illumina Connected Analytics는 다양한 2차 데이터 분석 옵션을 제공하여 리드(read) 생성 단계부터 결과 보고 단계로 구성된 워크플로우를 간소화해 줍니다. 기존 파이프라인을 사용하거나 새로운 파이프라인을 맞춤 제작 및 구성할 수 있는 유연한 옵션을 제공하는 Illumina Connected Analytics는 거의 모든 인포매틱스 애플리케이션을 지원할 수 있습니다.

바로 사용 가능한 옵션

Illumina Connected Analytics는 시퀀싱 데이터의 정확하고 효율적이며 포괄적인 2차 분석을 제공하는 DRAGEN Secondary Analysis¹에 대한 액세스를 포함해 강력하고 별다른 설정 없이 바로 사용이 가능한 데이터 프로세싱 도구 및 파이프라인을 제공합니다(그림 2).

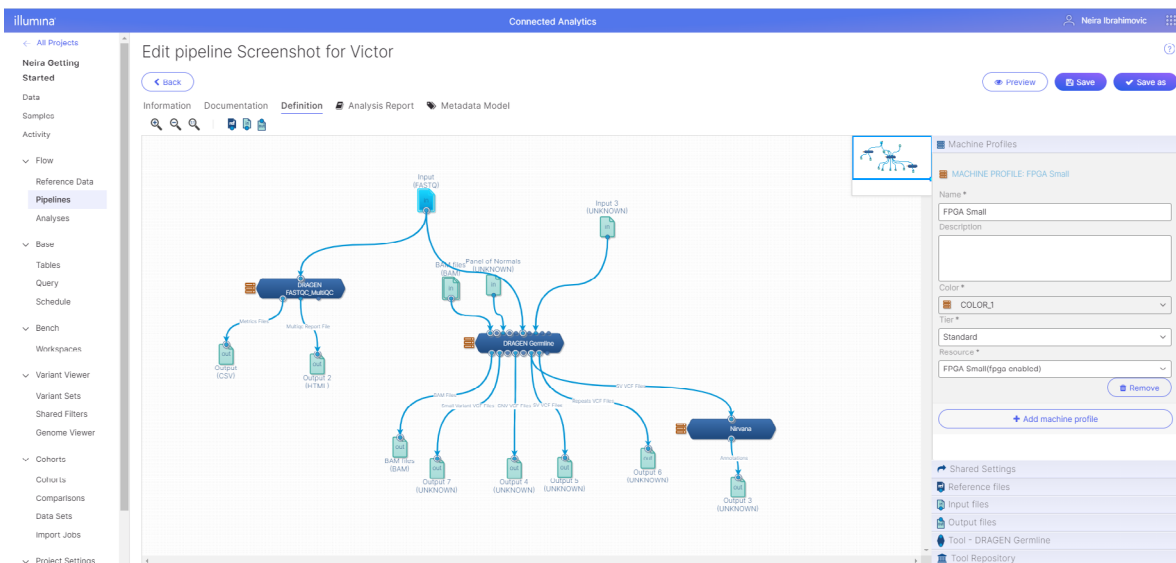


그림 2: Illumina Connected Analytics Flow — 분석 파이프라인의 유연한 빌드, 배포 및 버전 관리 지원

파이프라인 맞춤화

바이오인포매틱스 전문가는 Docker Hub 이미지 저장소에서 기존의 분석 도구를 불러오거나 Nextflow,⁷ CWL 및 그래픽 파이프라인 데이터를 사용해 새로운 파이프라인을 제작하고 편집할 수 있습니다. 랩 작업자와 다른 과학자는 직관적으로 디자인된 사용자 인터페이스(user interface, UI)를 통해 손쉽게 파이프라인을 실행할 수 있습니다.

데이터 관리 및 제어

데이터 생성량이 증가함에 따라 과학계에서는 개별 데이터 세트의 가치를 더 높여 줄 데이터 공유, 재사용 및 통합을 지원하는 인프라에 대한 필요성이 대두되었습니다. Illumina Connected Analytics는 이러한 요구를 충족하고자 데이터 관리 모범 사례를 채택하도록 설계된 여러 가지 기능을 포함하고 있습니다.

액세스 제어

관리자(Administrator)는 세밀한 액세스 제어(fine-grained access control)를 통해 권한을 설정하고 기존의 기관 자격 증명을 활용해 액세스를 제어할 수 있습니다. 감사 로그(Audit log)는 이벤트 및 변경 사항의 기록으로, 각 사용자의 플랫폼 액세스 시간과 플랫폼에서 진행한 작업을 로깅함으로써 규정 준수 및 책임을 요구합니다.

오픈 포맷

Illumina Connected Analytics는 특정 데이터 포맷에 국한되지 않는 플랫폼(data-agnostic platform)으로 설계되어 분자 데이터, 유전체 데이터, 표현형(phenotype) 데이터, 이미지 등의 비정형 데이터를 비롯한 다양한 데이터 타입의 분석을 지원합니다.

협업

Illumina Connected Analytics는 규정 준수를 유지하는 방식으로 지리적 경계를 뛰어넘는 협업을 가능케 합니다. 연구자는 데이터 무결성(integrity) 및 프라이버시를 유지하면서 데이터와 도구를 즉시 전송하고 다른 사용자와 공유할 수 있습니다. 또한 외부 클라우드 소스에서 호스팅되는 데이터 및 분석 도구를 Illumina Connected Analytics로 불러와 분석하고 공유하는 것도 가능합니다.

데이터 집계 및 쿼리

Illumina Connected Analytics는 수백만 개의 샘플에서 얻은 데이터를 아우르는 기능적 지식 관리 시스템을 만들기 위해 복잡한 데이터 집계 및 통합 단계를 자동화합니다. Illumina Connected Analytics는 유전형(genotype) 데이터, 표현형 데이터, 메타데이터, 주석(annotation), 기타 관련 정보 등 거의 모든 타입의 데이터를 캡처합니다. 연구자는 필요시 직접 데이터 모델을 정의하거나 직접 쿼리를 쓰거나 데이터 세트 간의 관계를 파악할 수 있습니다. Illumina Connected Analytics에 집계된 데이터는 새로운 바이오마커의 발견, 환자 집단의 계층화, 시간 경과에 따른 assay 성능의 모니터링 등에 활용할 수 있는 풍부한 정보를 제공합니다.

Cohorts를 통한 연구 규모에 맞는 임상유전체 데이터 탐구

Illumina Connected Analytics Cohorts는 코호트를 신속하게 구축하고 탐구하는 데 사용되는 연구 설계 도구입니다(그림 3). Illumina Connected Analytics Cohorts는 유전체학적 발견이 유의미한 연구로 이어지는 속도를 높여주고, 멀티오믹스(multiomics) 및 표현형 데이터를 한 곳에 취합하여 연구 설계 시 연구자가 더 나은 의사 결정을 내릴 수 있도록 합니다.

또 The Cancer Genome Atlas(TCGA), Broad Rare Genomes Project, 1000 Genomes Project, Gene Expression Omnibus와 같은 표준화된 공공 데이터 세트(harmonized public data set)를 포함하고 있어, 연구자는 이를 이용해 코호트를 확장할 수 있습니다. 가설은 통합된 코호트 브라우저를 활용해 몇 주가 아닌 단 몇 분 안에 세울 수 있습니다. 또한 연구자는 엄격한 큐레이션을 거친 공공 데이터 세트에 액세스할 수 있습니다. 간단한 시각적 도구를 사용해 코호트를 선택하고 시각화할 수 있어 복잡한 쿼리를 실행하지 않아도 됩니다.

통찰력을 이끌어 내는 안전한 노트북 환경

알고리즘 개발 및 맞춤 설정은 심층적인 데이터 탐색을 지원하는 Illumina Connected Analytics의 핵심 구성 요소입니다. 널리 사용되는 Jupyter Notebook, Python 및 R을 활용하는 대화형 프로그래밍 모듈이 데이터 과학자가 안전한 환경에서 집계된 데이터를 원활하게 분석할 수 있도록 해 줍니다(그림 4).



그림 3: Illumina Connected Analytics Cohorts — Illumina Connected Analytics에서 애드온 모듈을 통한 분자/유전체 데이터의 신속한 빌드 및 탐구 지원

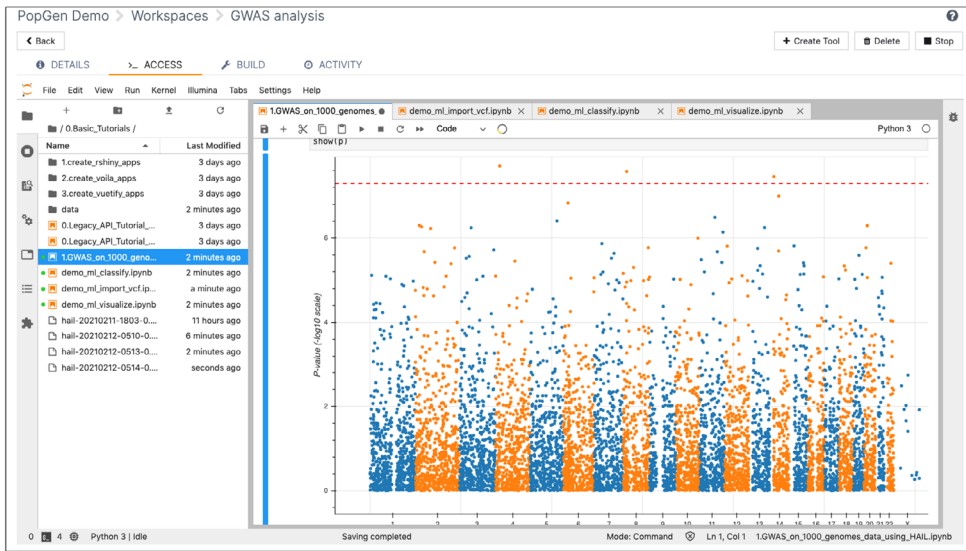


그림 4: Illumina Connected Analytics Bench — Jupyter 노트북, Python, R 등 유연한 대화형 컴퓨팅 도구가 내장되어 있는 Illumina Connected Analytics Bench

연구자는 방법 및 알고리즘 개발 단계에 샌드박스(sandbox) 환경에서 파이프라인을 개발하거나 수정할 수 있습니다. 연구자는 샌드박스 환경에서 표준 라이브러리(standard library)에 대한 액세스 권한이 있어 TensorFlow⁸나 scikit-learn⁹과 같이 개인에 맞춤형 라이브러리를 손쉽게 불러와 Illumina Connected Analytics 내에서 협업을 통해 데이터를 처리하기 위한 맞춤형 스크립트나 복잡한 스크립트를 만들 수 있습니다. 생산 단계로 넘어 갈 준비가 되면, Illumina Connected Analytics는 연구자가 노트북을 도구로 변환할 수 있도록 해 줍니다. 이렇게 변환된 도구는 Illumina Connected Analytics 도구 저장소에서 사용이 가능하며 생산 파이프라인에 통합됩니다.

보안 및 규정 준수를 핵심으로 하는 플랫폼

연구에서 유전체 데이터를 활용할 때 가장 중요한 것은 보안입니다. Illumina Connected Analytics는 가장 엄격한 데이터 보안 요구 사항도 충족할 수 있도록 다음과 같은 다양한 디지털 조치와 관리적 조치를 취하고 있습니다.

- 시퀀싱 기기에서 업로드된 데이터는 AES-256¹⁰을 사용해 암호화되며, 전송 보안 계층(transfer layer security, TLS)을 사용해 보호됩니다.
- Illumina Connected Analytics 내의 데이터는 업계에서 인정받는 다양한 보안 표준을 준수하기 위해 Amazon Web Services(AWS) Well-Architected 모범 사례¹¹를 사용해 AWS에서 호스팅됩니다.
- Illumina Connected Analytics는 유전체 데이터를 특정 지리적 영역 내 유지하도록 하는 지역 규정을 준수하기 위해 전 세계 10곳이 넘는 AWS 리전에서 호스팅됩니다.
- 인증 서비스는 기관 사용자와 비밀번호(선택 사항)를 관리하기 위해 보안 검증 마크업 언어(security assertion markup language)인 SAML 2.0의 지원을 받습니다.
- 감사 보고서를 통해 데이터 유래(provenance)를 추적할 수 있습니다.

Illumina Connected Analytics는 반드시 엄격한 요구 사항을 준수해 규제된 환경에서 운영해야 하는 고객도 지원합니다.

Illumina Connected Analytics는 Illumina QMS의 소프트웨어 수명 주기 프로세스에 따라 개발되었습니다. 또한 Illumina QMS 내 프로세스에는 다음을 포함한 업계 모범 사례와 관련 표준이 적용되어 있습니다.

- 국제 표준화 기구(International Organization for Standardization, ISO)의 ISO 27001:2013 Information Security Management System⁴ 및 ISO 27701:2019 Security Techniques⁵
- ISO 13485 조직 QMS 모범 사례⁶
- GDPR²
- HIPAA³
- 지역별 규정 및 규제 준수 요건에 대처하기 위해 데이터 레지던시(data residency) 확보

제품 목록

제품명	카탈로그 번호
ICA Enterprise Annual Subscription	20038994
ICA Professional Annual Subscription	20044876
ICA Training and Onboarding	20049422
ICA Cohorts Annual Subscription	20065842
Illumina Analytics - 1 iCredit	20042038
ICA Enterprise Svc and Compliance Add-on	20066830
ICA = Illumina Connected Analytics	

상세 정보

[Illumina Connected Analytics](#)

참고 문헌

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



무료 전화(한국) 080-234-5300
techsupport@illumina.com | www.illumina.com

© 2024 Illumina, Inc. All rights reserved.
모든 상표는 Illumina, Inc. 또는 각 소유주의 자산입니다.
특정 상표 정보는 www.illumina.com/company/legal.html을 참조하십시오.
M-KR-00248 KOR