

NovaSeq™ X 시리즈의 소프트웨어 생태계

기기에 탑재된
2차 분석 파이프라인과
Illumina Connected Software의
통합으로 간소화된 포괄적인
인포매틱스 플랫폼 및 솔루션



소개

NovaSeq X 및 NovaSeq X Plus 시퀀싱 시스템은 생물학 연구 시 가장 복잡한 질문에 대한 답을 찾는 데 요구되는 처리량과 정확도를 제공할 수 있는 획기적인 기술 혁신을 실현하였습니다. NovaSeq X 시리즈는 chemistry 및 광학 기술의 발전뿐만 아니라 기기에 DRAGEN™ 하드웨어를 내장하고 있어 더 빠르고 간소화된 2차 분석 기능을 제공하며 데이터를 80% 정도 무손실 압축할 수 있습니다. 또한 NovaSeq X 시리즈는 Illumina의 차세대 시퀀싱(next-generation sequencing, NGS) 워크플로우 지원을 위해 특별히 개발된 소프트웨어 생태계를 기반으로 운영의 간편성을 제고합니다(그림 1). Illumina Connected Software는 다음과 같이 구성되어 있습니다.

- 랩 운영 — Clarity LIMS™ 소프트웨어
- 런(Run) 계획 및 설정 — Illumina Run Manager 및 BaseSpace™ Sequence Hub
- 2차 분석 — BaseSpace Sequence Hub, DRAGEN 2차 분석, Illumina Connected Analytics
- 3차 분석 — Emedgene™ 소프트웨어, Illumina Connected Insights^a

NovaSeq X 시리즈는 랩 운영, 런 계획 및 데이터 분석을 위한 유연한 로컬 옵션과 클라우드 기반 옵션을 제공하므로 연구자는 바이오인포매틱스(bioinformatics, 생명정보학) 병목 현상 없이 고처리량(high-throughput) 시퀀싱을 수행할 수 있습니다.

이 Technical Note는 NovaSeq X 시리즈에 통합되는 소프트웨어 솔루션 및 플랫폼을 소개하고 NGS 워크플로우의 각 단계별 솔루션을 개략적으로 설명합니다.

NGS 워크플로우의 전 단계를 아우르는 Connected Software

랩 운영 관리

NovaSeq X 시리즈는 클라우드 서비스를 활용하는 Clarity LIMS 소프트웨어와 통합합니다. Clarity LIMS 소프트웨어는 효율적인 샘플 추적 및 워크플로우 관리를 위해 개발된 실험실 정보 관리 시스템(laboratory information management system, LIMS)입니다. 랩에서는 Clarity LIMS 소프트웨어를 사용함으로써 지속적으로 확장되는 선 구성된 프로토콜 메뉴를 지원하는 새로운 워크플로우의 도입 절차를 간소화하고, 워크플로우를 맞춤 설정 및 자동화하며, 감사 추적(audit trail)을 통해 효율적으로 샘플을 실시간 추적할 수 있습니다.

 [Clarity LIMS 소프트웨어에 대해 더 알아보기](#)

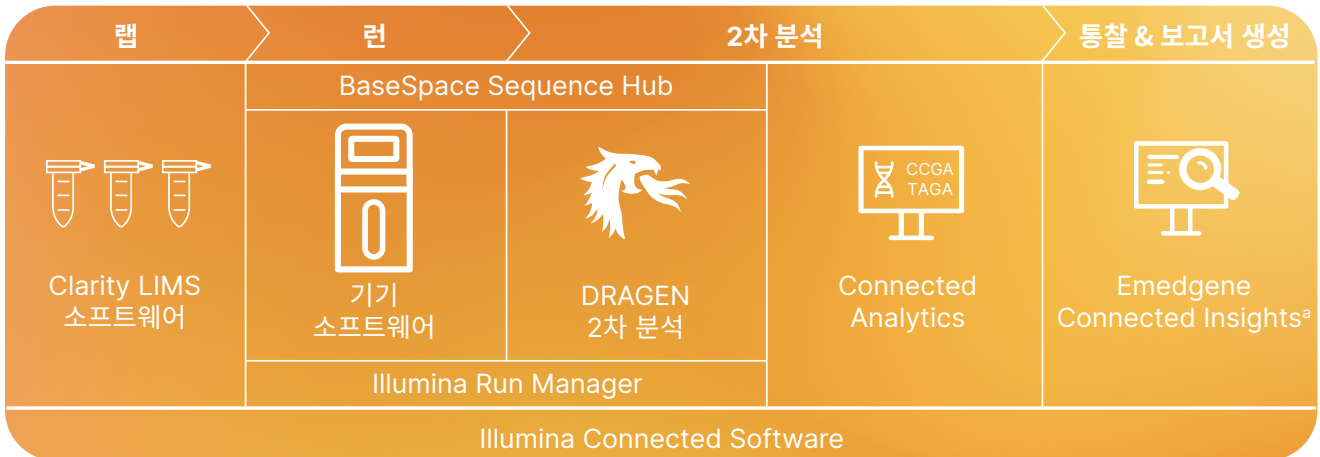


그림 1: 전체 NGS 워크플로우를 지원하는 Illumina Connected Software — NovaSeq X 시리즈는 전체 NGS 워크플로우에 걸쳐 Illumina 소프트웨어 및 인포매틱스 포트폴리오를 완벽하게 통합하여 유연하고 최적화된 맞춤 사용자 경험을 제공합니다.

^a 일부 국가에서는 지원되지 않음. Illumina Connected Insights는 서드 파티 지식 소스(third-party knowledge source)의 API 호출을 통해 사용자 정의 3차 분석 지원.

런 및 2차 분석 설정

Illumina는 다양한 랩의 요구 사항을 충족하는 유연한 런 계획 옵션을 제공합니다(표 1). 런 계획에는 시퀀싱 런 구성 및 DRAGEN 2차 분석 구성이 포함됩니다. 2차 분석은 시퀀싱 시스템 내에서도 클라우드에서 실행할 수 있습니다. 이를 통해 연구자는 간소화된 워크플로우, 더 편리한 데이터 관리, 적어진 터치포인트 수 등의 장점을 누릴 수 있습니다.

Illumina Run Manager

NovaSeq X 시리즈는 Illumina Run Manager를 제공합니다. 연구자는 기기 내에서도 네트워크에 연결된 기기를 통해 Illumina Run Manager에 액세스하여 사용자 및 기기 관리, 앱 구성, 런 설정 등을 수행할 수 있습니다. 직관적인 인터페이스가 오류의 가능성을 줄여 주는 안내 메시지를 제공하므로 연구자는 시퀀싱 런을 계획 및 시작하고, 감사 추적을 통해 라이브러리(library)를 추적하며, 런 진행 상황을 모니터링할 수 있습니다(그림 2).

BaseSpace Sequence Hub

연구자는 간편한 데이터 관리 도구(tool)와 및 분석적인 시퀀싱 도구의 제공을 위해 개발된 유전체학 연구용 클라우드 컴퓨팅 플랫폼인 BaseSpace Sequence Hub에서 NovaSeq X 시리즈를 사용해 시퀀싱 런을 계획할 수 있습니다. BaseSpace Sequence Hub는 직관적인 그래픽 인터페이스를 사용한 원격 런 설정 기능과 실시간 시퀀싱 매트릭스(metrics) 및 기록을 포함하는 런 모니터링 기능을 제공합니다.

 [BaseSpace Sequence Hub에 대해 더 알아보기](#)

표 1: 런 계획 옵션의 비교

파라미터	Illumina Run Manager	BaseSpace Sequence Hub
사용자 인터페이스	기기 내 또는 온프레미스 (on-premise) 컴퓨터	웹 브라우저
샘플 시트 (Sample Sheet) 생성	v2	v2
DRAGEN 2차 분석 자동 실행	지원	지원

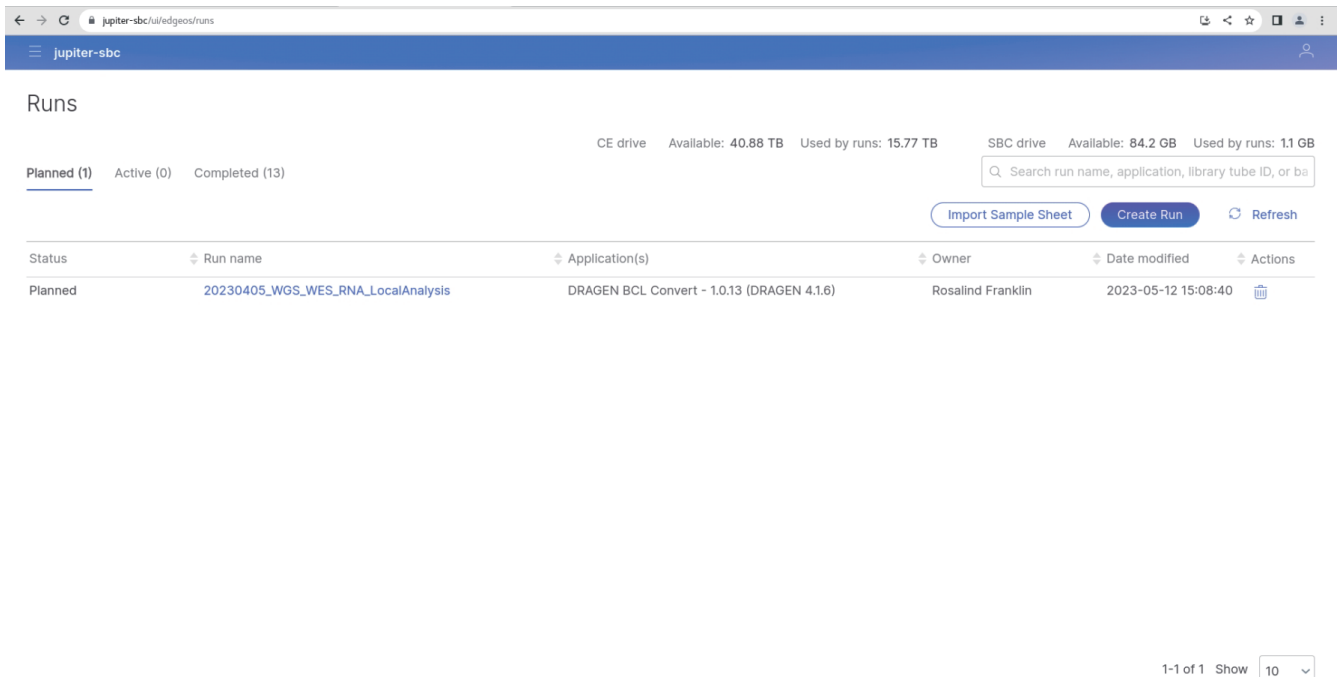


그림 2: Illumina Run Manager 인터페이스 — Illumina Run Manager는 기기 관리, 앱 구성, 런 설정 등을 지원하는 직관적인 사용자 인터페이스를 갖추고 있음.

시퀀싱

NovaSeq X 및 NovaSeq X Plus 시스템은 세심한 인체공학적 디자인을 채택해 운영을 간편하게 하며, 직관적이고 유용한 정보를 제공하는 디스플레이를 갖춘 4K 해상도의 초대형 터치스크린으로 사용자 경험을 최적화합니다. 연구자는 기기 내에서나 BaseSpace Sequence Hub를 통해 런 진행 상황을 모니터링하여 Q30 및 Yield(수율)와 같은 런 QC 메트릭스를 추적할 수 있습니다(그림 3).

 [NovaSeq X 시리즈에 대해 더 알아보기](#)

2차 분석

NovaSeq X 시리즈는 가장 일반적으로 사용되는 DRAGEN 2차 분석 파이프라인을 기기에 탑재해 제공합니다. 추가적인 파이프라인은 온프레미스 서버와 클라우드 기반 플랫폼을 통해 사용 가능합니다. DRAGEN 소프트웨어는 한층 더 빨라진 데이터 전송 속도, 더 편리해진 데이터 관리 도구 그리고 데이터 스토리지 비용을 최대 80% 절감시켜 주는 무손실 압축(lossless compression)을 토대로 NGS 데이터의 효율적인 2차 분석을 제공합니다. 또한 DRAGEN Multigenome(그래프) 및 머신러닝(machine learning)을 비롯한 기술의 혁신은 이미 수상을 통해 입증된 바 있는 우수한 정확도를 실현하는 데 기여했습니다.¹

 [DRAGEN 2차 분석에 대해 더 알아보기](#)

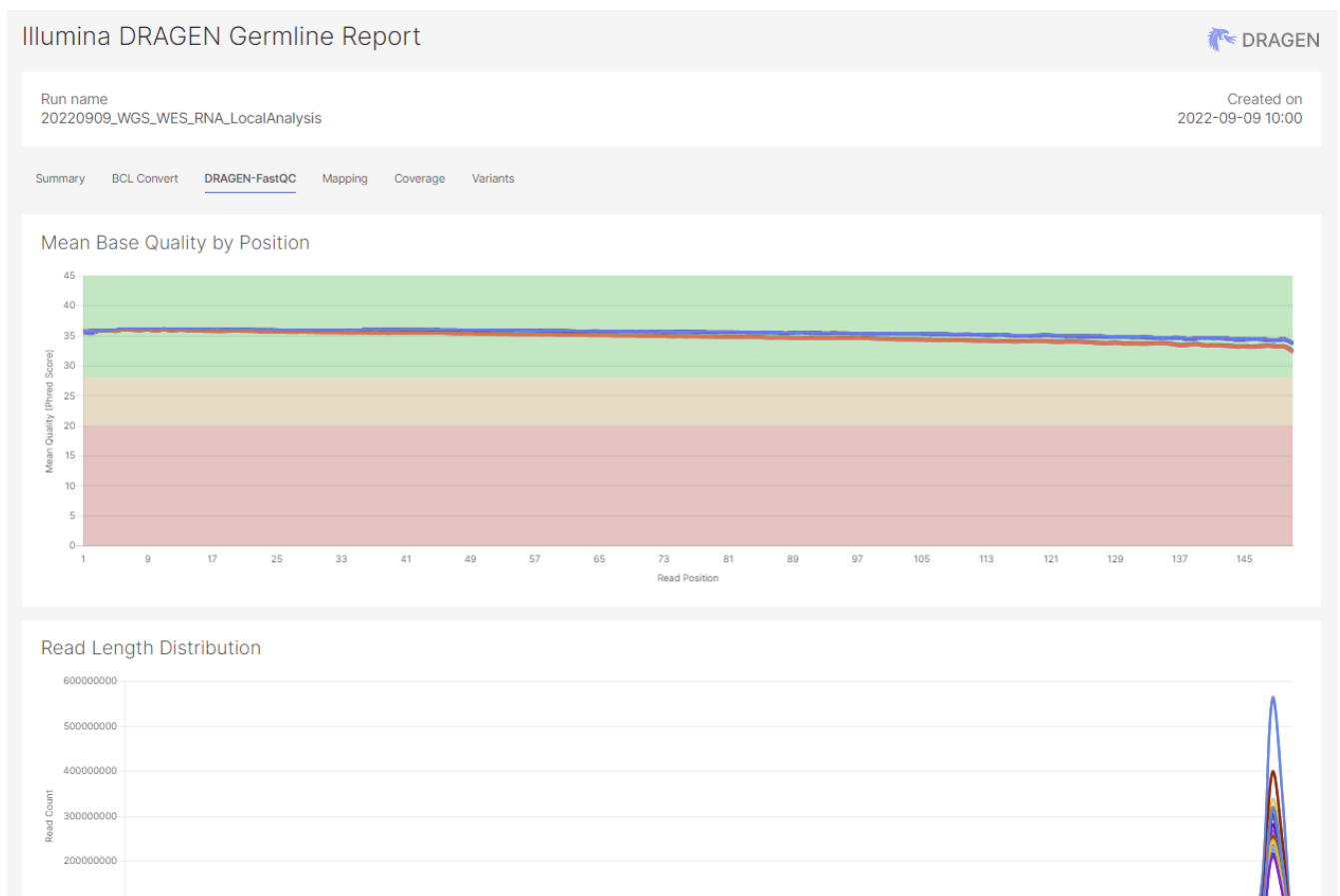


그림 3: 시퀀싱 런 보고서 — NovaSeq X 시리즈는 런 QC 메트릭스 및 FASTQ 보고서 확인 기능을 제공함.

데이터 분석 워크플로우 선택

유연한 런 모드 옵션

연구자는 NovaSeq X 시리즈에서 시퀀싱 런 설정 시 FASTQ 파일 생성, 정렬(alignment), 변이 검출(variant calling) 등 데이터 분석을 자동으로 실행할 수 있는 다양한 옵션을 선택할 수 있습니다(그림 4).

- **Manual** — 연구자가 직접 시퀀싱 런을 설정하며, 선택적으로 후속 분석에 필요한 v2 샘플 시트를 첨부할 수 있음.
- **Local** — 연구자가 Illumina Run Manager로 시퀀싱 런을 로컬에서 설정하면, 기기에 내장된 DRAGEN을 통해 또는 동일한 네트워크에 연결된 온프레미스 컴퓨터에서 데이터 분석이 자동으로 실행됨.
- **Cloud** — 연구자가 BaseSpace Sequence Hub에서 시퀀싱 런을 설정하면, 클라우드의 BaseSpace Sequence Hub 또는 Connected Analytics에서 데이터 분석이 자동으로 실행됨.
- **Hybrid** — 연구자가 BaseSpace Sequence Hub에서 시퀀싱 런을 설정하면, 기기에 내장된 DRAGEN을 통해 데이터 분석이 기기에서 자동으로 실행됨.[†]

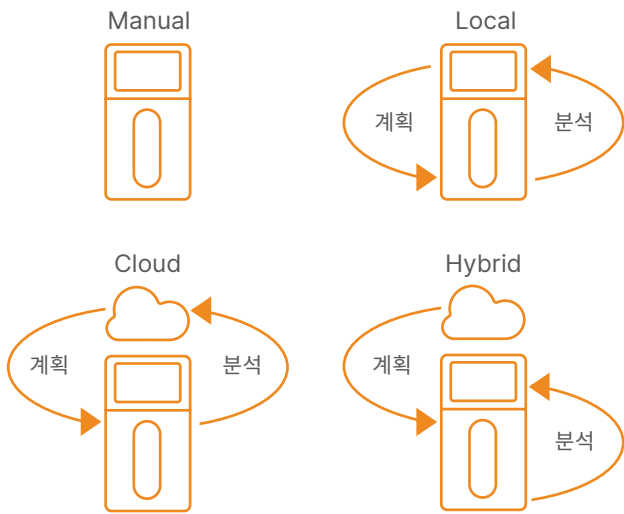


그림 4: 런 모드 옵션 — 연구자는 NovaSeq X 시리즈에서 런 설정 시 다양한 런 모드 옵션 중 하나를 선택할 수 있음.

[†] Hybrid 런 모드는 향후 소프트웨어 업데이트를 통해 제공될 예정.

NovaSeq X 시리즈에 내장된 DRAGEN

NovaSeq X 시리즈에 내장된 DRAGEN은 정확한 자동 2차 분석을 수행하며, 컴퓨팅 속도를 높여 주고, 보다 빠른 데이터 전송과 편리한 데이터 관리를 위해 무손실 데이터 압축을 지원합니다. DRAGEN 소프트웨어는 기기 내에서도 클라우드에서 여러 가지 2차 분석 파이프라인을 동시에 실행할 수 있습니다.

기기에서 실행할 수 있는 자동 DRAGEN 분석 파이프라인은 다음과 같습니다.

- DRAGEN BCL Convert
- 전장 유전체 시퀀싱(Whole-genome sequencing, WGS) 연구를 위한 DRAGEN Germline
- 전장 엑솜 시퀀싱(Whole-exome sequencing, WES) 연구를 위한 DRAGEN Enrichment
- 전장 전사체 시퀀싱(Whole-transcriptome sequencing) 연구를 위한 DRAGEN RNA

NovaSeq X에 내장된 DRAGEN의 분석 런 타임은 연구자가 선택한 워크플로우, 활성화된 기능, 샘플의 수, 샘플의 품질 및 수율에 따라 다를 수 있습니다(표 2).

연이어 시퀀싱 런을 진행할 때는 런 사이에 6.5시간에서 7시간 정도의 “대기 시간(latency window)”이 발생한다는 점을 고려해야 합니다. 대기 시간 동안에는 워시(wash), 수동 런 설정 및 다음 런을 위한 클러스터 생성(cluster generation) 단계가 진행됩니다. DRAGEN 분석이 대기 시간 내에 완료될 경우, 첫 번째 시퀀싱 런의 분석에 미치는 영향 없이 워시 단계 직후 두 번째 런을 시작할 수 있습니다. 만약 첫 번째 시퀀싱 런의 DRAGEN 분석이 6.5시간 안에 완료되지 않을 경우, 연구자는 두 번째 런을 시작하지 말고 기다렸다가 분석이 완료되면 런을 시작해야 합니다(그림 5).

기타 DRAGEN 2차 데이터 분석 액세스 옵션

DRAGEN 2차 분석 옵션을 검토할 때는 랩과 연구자의 요구 사항을 고려하는 것이 중요합니다. 사용 용이성과 가능한 한 적은 수의 터치포인트를 우선시하는 랩에서는 NovaSeq X 시리즈에 내장된 DRAGEN을 선택할 수 있으며, 확장성과 유연성에 더 중점을 두는 랩이라면 Connected Analytics 내에서도 DRAGEN Server를 통해 DRAGEN 소프트웨어를 맞춤 설정해 사용해 볼 수 있습니다. NovaSeq X 시리즈 사용 시 현재 제공 중인 클라우드 기반 솔루션이나 온프레미스 솔루션을 통해 확장된 DRAGEN 파이프라인 메뉴에 액세스할 수 있으므로 랩에서 요구 사항에 가장 적합한 옵션을 선택할 수 있습니다(표 3).

표 2: DRAGEN 런 타임^a

DRAGEN 구성	샘플 수 ^b	런 타임 ^c
BCL Convert(ORA 압축 포함)	24,576개	약 1시간 30분
RNA(Differential expression 포함)	전사체(mRNA) 1,536개	약 4시간 30분
Enrichment(Germline)	엑솜 512개	약 4시간 30분 ~ 5시간 30분
Germline(Map align 지원; variant calling 미지원)	전장 유전체 48개(30×)	약 3시간 30분
Germline(Small variant calling 지원)	전장 유전체 48개(30×)	약 5시간 30분 ~ 7시간 ^d
Germline(모든 variant caller 지원)	전장 유전체 48개(30×)	약 7시간 30분 ^d

- a. 런 타임은 샘플의 품질, 분석 설정값 및 샘플 구성에 따라 차이가 있음. 상기 데이터는 Illumina 생성 데이터의 대표적인 런 타임을 나타냄. 연구자의 실제 런 타임은 상이할 수 있음.
- b. 10B 플로우 셀 두 개 사용 시 샘플 수는 Illumina 내부 데이터에 근거함.
- c. 동일한 구성의 10B 플로우 셀 두 개를 동시에 시작하거나 10B 플로우 셀을 한 개 사용한 경우
- d. 기기 내에서 진행되는 분석 시간으로 인해 다음 시퀀싱 런이 곧바로 시작되지 않음. 6.5시간의 대기 시간이 있고 연이은 시퀀싱 수행을 위해 DRAGEN 분석이 그 시간 안에 반드시 완료되는 것으로 가정함.

시나리오 A: DRAGEN 분석이 대기 시간 내 완료됨



시나리오 B: DRAGEN 분석이 대기 시간 내 완료되지 않음



그림 5: 연이은 시퀀싱 런 계획 — NovaSeq X 시리즈의 한쪽 칸에서만 연이어 시퀀싱 런을 수행하려는 경우, 연구자는 DRAGEN 분석이 런 사이 요구되는 대기 시간 내에 완료될 것인지를 고려해야 함. 시나리오 A에서와 같이 DRAGEN 분석이 대기 시간 안에 완료된다면 두 번째 런을 곧바로 시작할 수 있으나, 시나리오 B에서와 같이 DRAGEN 분석이 대기 시간 안에 완료되지 않는다면 분석이 완료되기를 기다렸다가 두 번째 런을 시작해야 함.

표 3: DRAGEN 배포 옵션

배포 옵션	NovaSeq X 시리즈에 결합된 DRAGEN	BaseSpace Sequence Hub 내 DRAGEN	Connected Analytics 내 DRAGEN	DRAGEN Server
설명	기기에 내장된 효율적인 자동 워크플로우	직관적인 그래픽 인터페이스와 간편히 버튼만 누르면 되는 분석 기능	규정을 준수하는 안전한 클라우드에서 사용하는 맞춤형 파이프라인	최신 파이프라인과 맞춤형 로컬 서버
터치포인트 수 ^a	1개(시퀀싱 설정)	2개(시퀀싱 설정, 2차 분석)	1개(시퀀싱 설정)	3개(시퀀싱 설정, 수동 파일 전송, 2차 분석)
분석을 위한 파일 전송 횟수	0회(기기 내에서 생성된 VCF 파일)	1회(BaseSpace Sequence Hub로 스트리밍된 BCL 파일)	1회(Connected Analytics로 스트리밍된 BCL 파일 - 액세스 가능한 클라우드 버킷)	2회(네트워크 스토리지로 전송된 후 DRAGEN Server로 전송된 BCL, FASTQ 또는 BAM 파일)
지원되는 앱	BCL Convert, WGS(Germline), Enrichment(Germline + Somatic), RNA	모든 DRAGEN 파이프라인: RNA, DNA amplicon, Single-cell RNA, Differential expression, RNA amplicon, Methylation, Metagenomics, RNA pathogen detection, COVID lineage, scATAC-Seq, Imputation		

a. 시퀀싱 설정 단계에서 VCF 생성 단계까지

BaseSpace Sequence Hub 내 DRAGEN

BaseSpace Sequence Hub에서 현재 제공되는 클라우드 기반의 DRAGEN 제품은 정확하고 효율적인 분석과 함께 안전한 소프트웨어 생태계와 유연한 확장성을 제공합니다. 암호화된 데이터가 기기에서 BaseSpace Sequence Hub로 전송되면 DRAGEN 소프트웨어로 여러 가지 염색된 앱을 활용해 간단히 버튼만 눌러 2차 분석을 실행할 수 있습니다.

Connected Analytics 내 DRAGEN

Connected Analytics는 실시간 데이터 전송을 위해 시퀀싱 시스템에 통합되는 안전하고 유연한 바이오인포매틱스 플랫폼입니다. 또한 최적화된 고처리량 연구를 위해 고도로 자동화된 워크플로우와 맞춤 솔루션을 지원합니다. 국제 표준을 준수하기 위해 데이터 레지던시(data residency) 요구 사항을 충족하고, 통합 인증(single sign-on, SSO) 기능을 지원하며, 감사 로그(audit log)를 제공하고, 액세스가 제어되는 매우 안전한 연구 환경을 제공합니다. Connected Analytics Cohorts 모듈은 연구자가 대규모 샘플 코호트를 활용해 연구를 만들고 최적화할 수 있도록 도와 탄탄한 연구 설계에 기여합니다.

DRAGEN On-premise Server

DRAGEN Server는 로컬 저장소에 NGS 데이터를 수집하고 보관합니다. 데이터는 로컬 네트워크 연결을 통해 기기에서 스토리지 솔루션으로 전송된 후 선택한 분석 파이프라인의 실행을 위해 다시 DRAGEN Server로 전송됩니다.

3차 분석 및 해석

자동화된 설명 가능한 인공지능(explainable artificial intelligence, XAI) 기반의 변이 해석 기능을 제공하는 Emedgene은 연구 시 3차 분석 워크플로우를 간소화하기 위해 개발된 포괄적인 소프트웨어 플랫폼입니다. Emedgene은 희귀 유전 질환 및 유전 질환 연구 시 단일 업체 솔루션 지원을 위해 모듈식 옵션을 제공하고 있습니다.

 [Emedgene에 대해 더 알아보기](#)

Illumina Connected Insights는 강력한 애플리케이션 프로그래밍 인터페이스(application programming interface, API)의 통합을 통해 광범위한 지식 소스를 활용하여 체세포 종양 연구, 희귀 질환 연구(곧 지원 예정) 등과 관련이 있는 임상 시험, 의약품 라벨 및 가이드라인의 우선순위 설정이 적용된 보고서를 생성함으로써 간소화된 변이 해석을 제공하고 있습니다.

 [Connected Insights에 대해 더 알아보기](#)

Correlation Engine은 엄격한 큐레이션을 거친 공공 데이터와 함께 민간 오믹스(omics) 데이터를 생물학적 맥락에서 살펴볼 수 있도록 해 주는 대화형 오믹스 지식베이스(knowledgebase)입니다. 연구자는 이 소프트웨어를 이용해 더 짧은 시간 안에 과학적 발견을 이룰 수 있을 것입니다.

 [Correlation Engine에 대해 더 알아보기](#)

요약

NovaSeq X 및 NovaSeq X Plus 시퀀싱 시스템은 우수한 정확도를 바탕으로 데이터 집약적 연구와 대규모 유전체학 연구 시 뛰어난 시퀀싱 성능을 발휘합니다. 수많은 기술 혁신과 발전을 통해 탄생한 NovaSeq X 시리즈는 Illumina의 다양한 바이오인포매틱스 소프트웨어 도구와 통합하여 전체 NGS 워크플로우에 걸쳐 운영의 간편성, 유연성 및 설정 가능성을 크게 높여 줍니다. 연구자는 간편하게 버튼만 누르면 되는 앱부터 맞춤 설정이 가능한 커맨드 라인(command line) 인터페이스에 이르기까지 워크플로우의 단계별로 로컬 옵션과 클라우드 기반 옵션을 모두 제공하는 NovaSeq X 시리즈를 이용해 연구의 요구 사항에 맞춰 기기에서 NGS 워크플로우를 설정하고 바이오인포매틱스 병목 현상 없이 고처리량 시퀀싱을 수행할 수 있습니다.

상세 정보

[NovaSeq X 및 NovaSeq X Plus 시퀀싱 시스템](#)
[Illumina Connected Software](#)

참고 문헌

1. Mehio R, Ruehle M, Catreux S, et al. DRAGEN Wins at Precision- FDA Truth Challenge V2 Showcase Accuracy Gains from Alt-aware Mapping and Graph Reference Genomes. [illumina.com/science/genomics-research/dragen-wins-precisionfda-challenge-showcase-accuracy-gains.html](https://www.illumina.com/science/genomics-research/dragen-wins-precisionfda-challenge-showcase-accuracy-gains.html). Accessed January 12, 2023.



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

© 2023 Illumina, Inc. All rights reserved.
모든 상표는 Illumina, Inc. 또는 각 소유주의 자산입니다.
특정 상표 정보는 www.illumina.com/company/legal.html을 참조하십시오.
M-GL-01587 v1.0 KOR