

Illumina Protein Prep 솔루션으로 정확하고 신뢰할 수 있는 단백질 검출 결과를 얻는 방법

DRAGEN™ Protein Quantification 2차 분석 파이프라인을 이용한 QC 및 데이터 정규화 단계를 기술한 가이드

재현 가능한 결과

여러 샘플 및 플레이트에 걸쳐
일관적인 단백질 검출 결과 확보

분석 자동 실행

DRAGEN Secondary
Analysis로 고품질의
데이터 생성

통합된 통찰력

Illumina Connected
Multiomics로 정규화된 단백질
발현량 파일을 손쉽게 분석

소개

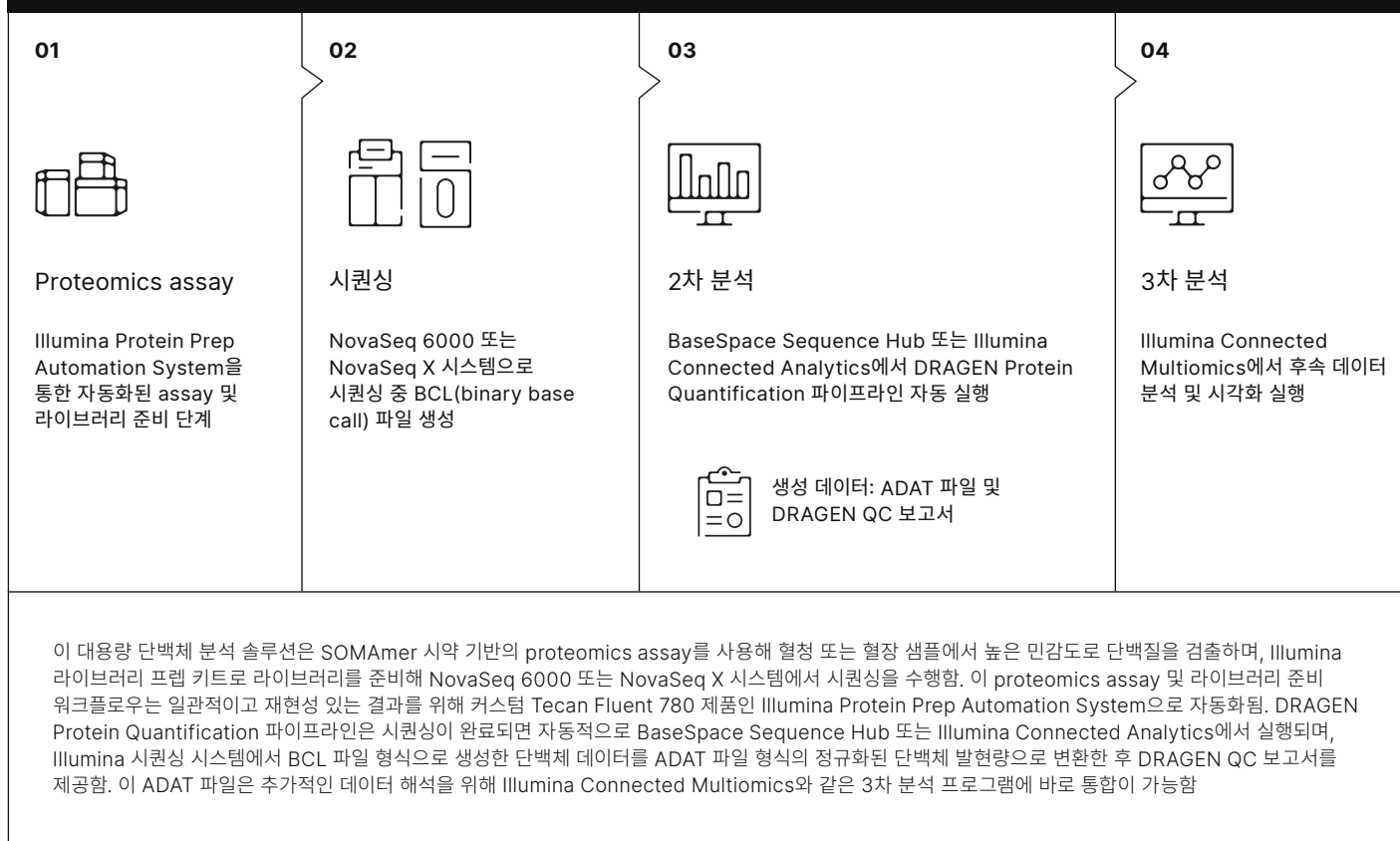
단백질은 인간의 생명 활동에 중요한 기능적인 역할을 수행하며 실시간으로 건강과 질병 상태에 관한 단편적인 정보를 제공합니다. 유의미한 생물학적 통찰력을 얻으려면 정확하고 재현 가능한 단백질 정량이 매우 중요합니다. Illumina Protein Prep은 단백질 캡처에 SOMAmer® (Slow Off-Rate Modified Aptamer, 저속 오프-레이트 변형 압타머) 시약을 사용하여 높은 민감도의 단백질 검출을 지원하는 혁신적인 proteomics assay입니다. 이 assay는 차세대 시퀀싱(next-generation sequencing, NGS) 기반의 리드아웃(readout)과 DRAGEN Secondary Analysis의 바이오인포매틱스(bioinformatics, 생명정보학) 기술을 결합하여 샘플 준비로 시작해 결과 보고로 끝나는 간소화된 대용량 단백질 분석 솔루션을 제공합니다.

Illumina Protein Prep은 하나의 샘플에서 수천 개의 단백질을 동시에 검출하는 바이오마커(biomarker)의 발견을 위한 강력한 플랫폼을 제공합니다. 단, 대용량 단백질 분석 플랫폼은 플레이트 간 편향(plate-to-plate bias)이나 기술적 변동성(technical variability)으로 인해 바이오마커

중에서도 특히 효과 크기(effect size)가 작은 경우 차등 발현(differential expression)을 놓칠 수 있습니다. DRAGEN Protein Quantification 파이프라인은 2차 분석 중 이러한 문제를 해결해 주는 강력한 정규화(normalization) 방법을 적용함으로써 여러 실험에 걸쳐 일관적인 고품질의 데이터를 제공합니다. 클라우드 환경에서는 NovaSeq™ 6000 또는 NovaSeq X 시스템에서 시퀀싱이 완료되면 자동적으로 DRAGEN Protein Quantification 파이프라인이 실행되어 정규화된 단백질 발현량(normalized protein count)을 포함한 DRAGEN QC 보고서와 ADAT 파일을 생성합니다. 연구자는 이후 Illumina Connected Multiomics로 이 파일들의 3차 분석(그림 1)을 실행해 단백질에 관한 보다 심층적인 생물학적 통찰력을 얻을 수 있습니다. 또한 로컬 분석 옵션이 필요한 경우 DRAGEN Protein Quantification 파이프라인을 DRAGEN On-premises Server에 설치하여 사용할 수 있습니다.

이 Technical Note는 고품질의 단백질 분석 데이터를 생성하기 위해 샘플 및 기술적 변동성을 제거하는 DRAGEN Protein Quantification 파이프라인의 QC 및 정규화 단계를 설명합니다.

그림 1: Illumina Protein Prep 워크플로우의 개요



데이터 정규화가 필요한 이유

기술적 변동성은 모든 대용량 proteomics assay의 고유한 특성으로, 샘플 준비 과정, 내재적 샘플 이질성 (intrinsic sample heterogeneity), 하이브리드화 효율성 (hybridization efficiency), PCR 효율성, 시퀀싱 등 다양한 원인으로 인해 발생할 수 있습니다. 샘플 내 기술적 변동성 외에도, 작업자의 차이, 기기 또는 환경적 영향으로 인해 동일한 플레이트의 모든 샘플에 영향을 주는 배치 효과(batch effect)가 발생할 수 있습니다. 따라서 효과적인 정규화 또는 배치 효과 수정 단계를 거치지 않는다면, 이러한 체계적 변동성(systematic variation)으로 인해 실제 생물학적 차이를 확인할 수 없게 되어 부정확한 결과를 얻게 될 수도 있습니다.

데이터 정규화 단계는 바로 이러한 기술적 변동성을 수정하고, 데이터를 표준화(standardization)하며, 다양한 샘플, 플레이트 및 assay 런(run) 간 유의미한 비교를 위해 필수적인 요소입니다. DRAGEN Protein Quantification 파이프라인을 활용한 포괄적인 정규화 전략으로 재현 가능하고 신뢰할 수 있는 단백질 정량 데이터를 확보할 수 있습니다.

Illumina Protein Prep 대조물질

Illumina Protein Prep assay는 assay 변동성을 줄이기 위하여 각 샘플에 플레이트 대조물질과 SOMAmer 시약 대조물질을 포함합니다. 두 종류의 대조물질을 모두 사용하여 DRAGEN Protein Quantification 파이프라인으로 정규화된 단백질 발현 카운트를 생성합니다.

플레이트 대조물질

각각의 96-well plate에는 85개의 샘플과 공시료(blank), 보정물질(calibrator), QC 샘플과 같은 플레이트 대조물질을 11개 포함합니다(그림 2). 플레이트 대조물질은 Illumina Protein Prep assay에 포함되어 있으며, 정규화와 QC를 위해 플레이트별로 혈장 또는 혈청 샘플과 함께 분석됩니다. 이러한 대조물질은 기술적 변동성을 감지하고 보정하는 기준점 역할을 하며, 서로 다른 플레이트에서 실행된 샘플들을 정렬하는 기반을 제공합니다.

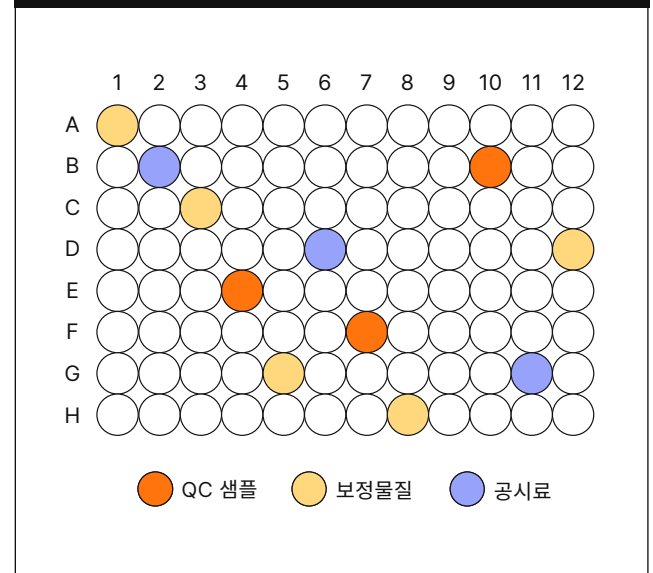
공시료

공시료는 깨끗한 버퍼로, Illumina Protein Prep assay의 비특이적 백그라운드(nonspecific background)에 대한 추정값을 제공합니다. 대조물질로 사용되는 공시료는 혈장 또는 혈청 샘플보다 발현량이 매우 적을 것으로 예상됩니다. 공시료를 분주한 웰에서 높은 발현량이 관찰된다면, 런의 백그라운드가 증가했거나 샘플 오염이 발생했음을 의미할 수 있습니다.

보정물질

보정물질을 사용하면 플레이트 간 기술적 변동성을 수정할 수 있습니다. 이 대조물질로는 분석되는 샘플의 종류에 따라 풀링된 인간의 혈장 또는 혈청이 사용됩니다. 플레이트 보정 정규화 작업은 SOMAmer 시약별 보정물질의 중앙값에 상응하는 레퍼런스 보정물질과 비교하는 방식으로 진행됩니다.

그림 2: 대조물질이 포함된 플레이트의 레이아웃 예시



QC 샘플

QC 샘플은 플레이트가 예상한 결과를 도출하였는지 확인하는 품질 관리 검사를 제공합니다. 이 대조 샘플로는 분석되는 샘플의 종류에 따라 풀링된 인간의 혈장 또는 혈청이 사용됩니다. 플레이트 품질 확인 작업은 SOMAmer 시약별로 QC 샘플의 중앙값에 상응하는 QC 레퍼런스 샘플과 비교하는 방식으로 진행됩니다.

SOMAmer 시약 대조물질

SOMAmer 시약 대조물질은 하이브리드화 전 알려진 농도를 사용해 각각의 샘플에 추가되며, 샘플을 정규화하고 품질을 확인하기 위한 샘플 특이적 배율(scale factor)의 계산에 활용됩니다.

데이터 정규화의 단계

원시 단백질 발현량 데이터의 정규화는 Illumina Protein Prep 워크플로우 중 샘플과 플레이트에서 발생할 수 있는 체계적 편향(systematic bias)을 제거하기 위한 여러 단계를 포함합니다. 이 조정 작업은 하이브리드화 정규화(hybridization normalization), 외부 레퍼런스 중앙값 정규화(external reference median normalization), 보정(calibration) 그리고 플레이트 스케일링(plate scaling)의 4단계로 구성되며, 비연속적으로 진행됩니다(그림 3).

하이브리드화 정규화

이 단계에서는 assay의 하이브리드화 단계와 라이브러리 준비 단계 중 발생할 수 있는 샘플 간 변동성을 수정합니다. 샘플 특이적 배율은 각 샘플 내 하이브리드화 정규화 대조물질의 발현량을 플레이트 내 모든 샘플의 중앙값에 비교해 계산합니다.

외부 레퍼런스 중앙값 정규화

이 단계에서는 대조물질 외 샘플 간 총 단백질 풍부도(total protein abundance) 측정값의 차이를 수정합니다. 배율은 관찰된 단백질 측정값을 각 단백질에 대한 레퍼런스 기댓값에 비교해 계산합니다.

보정

보정은 앞서 플레이트 대조물질 섹션에 기술된 5개의 보정물질 샘플을 사용해 개별 SOMAmer 시약에 영향을 주는 배치 효과를 수정하는 작업입니다. 이 단계는 여러 플레이트와 런을 통해 생성된 데이터의 정확한 비교를 가능하게 해 줍니다. 보정 단계에서는 SOMAmer 시약별로 보정물질과 보정물질에 대한 레퍼런스 기댓값 간 편차가 수정됩니다.

그림 3: DRAGEN Protein Quantification 파이프라인에서 실행되는 정규화의 단계

01	02	03	04	05	06	07
샘플 준비	결합	SOMAscan assay	하이브리드화	캡처 및 워시	Post-PCR 클린업	시퀀싱
1.5시간	3.5시간	2.3시간	19시간	3.3시간	0.5시간	시퀀싱 런 타임은 사용하는 샘플 개수, 시스템 및 플로우 셀에 따라 상이
외부 레퍼런스 중앙값 정규화			하이브리드화 정규화			
플레이트 스케일링 및 보정						

보정은 두 단계로 진행됩니다.

- 1단계: 보정물질 중앙값을 시퀀싱 시스템에서 얻은 레퍼런스와 비교한 후 각 SOMAmer 시약을 적절히 조정합니다.
- 2단계: 변경된 보정물질 중앙값을 시퀀싱 시스템에서 얻은 레퍼런스와 비교함으로써 여러 시스템에서 생성한 데이터를 비교할 수 있습니다. 보정물질 중앙값을 레퍼런스 값에 맞출 때 사용되는 배율이 플레이트 내 모든 샘플에 적용됩니다.

플레이트 스케일링

플레이트 스케일링은 여러 보정물질에 걸친 각 SOMAmer 시약 측정값의 중앙값을 이용해 플레이트 간 총 단백질 측정값의 변동성을 외부 보정 레퍼런스에 맞춰 수정하는 작업입니다.

스케일링은 두 단계로 진행됩니다.

1. 1단계: 시퀀싱 시스템(즉, NovaSeq 6000 또는 NovaSeq X 시스템)에서 얻은 레퍼런스를 기반으로 측정값을 조정합니다.
2. 2단계: 다양한 시퀀싱 시스템에서 일관되게 유지되는 레퍼런스를 활용하고 보정 과정을 정밀하게 다듬어, 서로 다른 기기에서 생성된 데이터 간 비교를 가능하게 합니다. DRAGEN Protein Quantification 보고서에는 1단계 스케일링을 통해 얻은 QC 매트릭스만이 표시됩니다.

마지막 QC 검사 퍼센트

QC 단계는 정규화의 가장 마지막 순서로, 각 플레이트 내 QC 대조물질을 이용해 진행됩니다. 이 단계에서는 QC 샘플을 사용한 3번의 반복 실험에서 얻은 각 SOMAmer 측정치의 중앙값을 외부 QC 레퍼런스와 비교함으로써 SOMAmer 시약 특이적 배율 및 QC 지표(QCCheckTailPercent)를 할당합니다.

DRAGEN Protein Quantification 파이프라인은 Illumina Protein Prep 워크플로우의 특유한 단계에서 발생한 관련이 없는 변동성을 최소화해 주는 다양한 정규화 절차를 실행합니다(그림 4).

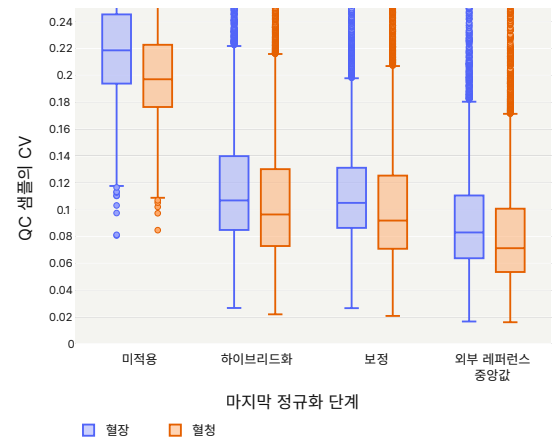
DRAGEN Protein Quantification 데이터

DRAGEN Protein Quantification 파이프라인은 Illumina Protein Prep 데이터 분석 시 매우 중요한 두 가지 파일을 생성합니다. 하나는 각 정규화 단계에서 생성되는 일련의 ADAT 파일이며, 다른 하나는 런의 QC 결과를 요약한 DRAGEN QC 보고서입니다.

ADAT 파일

DRAGEN Protein Quantification 파이프라인은 원시 발현량 데이터를 처리한 후 실행된 정규화 단계별로 ADAT 확장자를 가진 탭(tab)으로 구분된 ASCII 텍스트 파일을 생성합니다. ADAT 파일은 일련의 샘플(행)에 대한 일련의 분석물질(열)의 측정값을 포함하며, 분석물질의 설명과 샘플의 설명을 제공합니다.

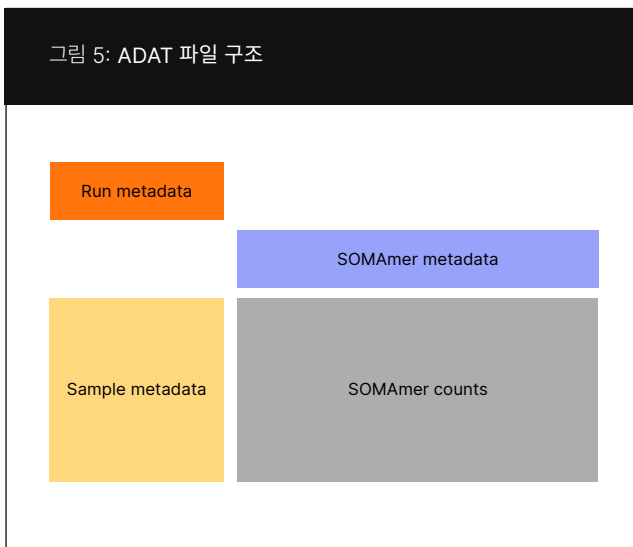
그림 4: 데이터 변동성에 대한 정규화의 효과



원시 assay 데이터 결과의 정규화로 감소된 플레이트 내 변동 계수 (coefficient of variation, CV). 상기 데이터는 혈장 및 혈청 QC 샘플과 Illumina Protein Prep 6K assay를 사용해 생성됨. 약 5.5%의 CV 중앙값^a을 도출하는 Illumina Protein Prep 9.5K assay도 현재 제공 중

a. 건강한 공여자의 샘플을 사용해 계산한 예상 CV 중앙값

그림 5: ADAT 파일 구조



ADAT 파일은 다음과 같은 구조를 가진 탭으로 구분된 파일입니다(그림 5).

- **Run metadata:** 전체 런 또는 개별 플레이트와 관련이 있는 메트릭스(예: RunID, 마지막 QC 검사 퍼센트와 관련된 메트릭스) 포함
- **SOMAmer metadata:** 각 SOMAmer 시약 또는 단백질 표적과 관련이 있는 메트릭스(예: SOMAmer ID, 특정 유전자 및 표적 ID) 포함
- **Sample metadata:** 각 샘플과 관련이 있는 메트릭스(예: Sample ID, Plate ID, QC) 포함
- **SOMAmer counts:** 샘플별 각 SOMAmer의 상대적인 풍부도를 반영한 NGS 리드 수의 행렬

이 형식의 파일은 샘플 수뿐만 아니라 분석물질의 수와 종류 그리고 샘플 서술자(descriptor)에 관한 유연한 옵션을 제공하며, 생물학적 통찰력을 얻기 위한 3차 분석에도 활용할 수 있습니다.

DRAGEN QC 보고서

DRAGEN Protein Quantification 파이프라인의 QC 보고서는 ADAT 파일과 함께 제공되며, 각각의 샘플과 특정 플레이트 전체의 QC 결과를 요약합니다. 개별 샘플의 QC 결과는 하이브리드화 정규화 단계와 외부 레퍼런스 중앙값 정규화 단계를 통해 얻습니다. 플레이트의 QC 결과는 보정 정규화 단계와 QC 단계를 통해 얻습니다.

ADAT 파일을 활용한 3차 분석

정규화된 ADAT 파일은 즉시 Illumina Connected Analytics에서 Illumina Connected Multiomics로 추가적인 데이터 분석을 실행할 때 활용할 수 있습니다. Illumina Connected Multiomics는 데이터의 관리 및 선별을 위한 그래픽 사용자 인터페이스(graphical user interface, GUI)를 지원하며, 차등 단백질 발현(differential protein expression), 경로 분석(pathway analysis), 유전자 세트 인리치먼트 분석(gene set enrichment analysis, GSEA) 등 고급 데이터 해석에도 사용해 볼 수 있습니다. 또한 ADAT 파일의 콘텐츠도 R(SomadatalO)과 Python(Canopy)에서 제공하는 오픈소스 패키지를 이용해 분석할 수 있습니다.

요약

원시 단백질 데이터를 정규화하는 작업은 Illumina Protein Prep 워크플로우 중 발생할 수 있는 assay와 샘플의 편향을 최소화하는 데 매우 중요합니다. DRAGEN Protein Quantification 파이프라인은 몇 가지 정규화 및 보정 단계를 통해 체계적 변동성을 줄이는 동시에 실제 생물학적 변동성의 제거 위험을 최소화함으로써 보다 정확한 결과를 제공합니다. 이 파이프라인은 클라우드 또는 On-premises Server에서 지원됩니다. 클라우드 환경에서 사용 시 NovaSeq 6000 시스템이나 NovaSeq X 시스템에서 시퀀싱이 완료되면 DRAGEN Protein Quantification 파이프라인이 자동적으로 실행되어 정규화된 단백질 발현량을 포함하는 DRAGEN QC 보고서 및 ADAT 파일을 생성합니다. 이렇게 생성된 ADAT 파일을 Illumina Connected Multiomics를 통한 3차 분석에 활용하면, 단백질에 대한 더욱더 심층적인 생물학적 통찰력을 얻을 수 있습니다.

상세 정보

[Illumina Protein Prep](#)

[DRAGEN Secondary Analysis](#)

[Illumina Connected Analytics](#)



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

© 2025 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-03205 v1.0 KOR