

Emedgene™ 소프트웨어

희귀 질환 및
기타 유전학 연구에
최적화된 사용자 정의
변이 해석 제공

- XAI와 맞춤형 자동화를 통해 생식세포 연구 시 3차 분석 절차 최대 75% 간소화
- LIMS, EMR 및 기타 IT 시스템과의 통합으로 전체 assay 워크플로우의 보안성과 간편성 확보
- NGS와 마이크로어레이 assay의 과학, 기술 및 규모에 대해 높아지는 수요를 충족할 수 있는 성장의 기회 제공

illumina®

소개

차세대 시퀀싱(Next-generation sequencing, NGS)은 희귀 질환 및 기타 유전 질환, 약물유전체학(pharmacogenomics), 세포유전학(cytogenetics), 기타 생식세포 이상(germline condition) 또는 보인자 검사(carrier screening) 연구 시 인간 유전학에 대한 이해를 증진하고 이를 정밀의료(precision medicine)에 적용하는 데 핵심적인 역할을 하고 있습니다. 전장 유전체 시퀀싱(Whole-genome sequencing, WGS), 전장 엑손 시퀀싱(whole-exome sequencing, WES), 표적 패널과 가상 패널은 각각 전체 유전체, 코딩 영역(coding region), 표적 유전자 영역을 편향 없이 고해상도로 연구할 수 있도록 해 줍니다. 그러나 랩에서 이렇게 생성된 방대한 양의 데이터를 분석하고 해석하는 단계에서 상당한 병목 현상이 발생합니다. 특히 변이 해석은 많은 시간을 필요로 하는 수작업 절차로, 유전체당 최대 7시간이 소요될 수 있으며 수많은 반복 단계를 거쳐야 합니다.¹ 따라서 랩에서는 건강한 집단/보인자 집단 스크리닝 프로젝트와 같은 대용량 패널 애플리케이션에 자동화, 변이 큐레이션(curation) 및 효율적인 연구 보고서 생성 도구를 이용해 효율성과 정확도를 크게 높일 방법을 모색하기도 합니다.

Illumina는 변이 해석 워크플로우를 간소화하고 변이 큐레이션 부담을 줄여 주기 위한 목적으로 유전학자들이 설계한 종합적인 연구용 소프트웨어인 Emedgene을 통해 이러한 랩을 지원하고 있습니다. 대용량 사용자 정의 변이 해석 워크플로우를 위한 설명 가능한 인공 지능(explainable artificial intelligence, XAI) 및 자동화를 제공하는 Emedgene 소프트웨어는 수동 해석에 비해 효율성을 2~5배 높이고 분석 건당 총 워크플로우 시간은 50~75% 줄였습니다.

Emedgene 소프트웨어는 DRAGEN™ Secondary Analysis, BaseSpace™ Sequence Hub, Illumina Connected Analytics를 포함하는 Illumina 2차 분석 및 데이터 스토리지 플랫폼과 통합함으로써 모듈식 옵션을 제공하여 랩에서 샘플 처리부터 연구 보고서 생성까지 모든 단계에 Illumina의 종합 솔루션을 활용할 수 있도록 합니다. Emedgene 소프트웨어는 WGS, WES, 표적 패널 및 마이크로어레이(microarray) 데이터의 사용을 지원합니다. 아울러 이 소프트웨어는 단일 염기서열 변이(single nucleotide variant, SNV), 삽입/결실(insertion/deletion, Indel), 유전자 복제수 변이(copy number variation, CNV), 미토콘드리아 DNA 변이(mitochondrial DNA variant, mtDNA), 구조적 변이(structural variant, SV), 짧은 연쇄 반복(short tandem repeat, STR) 등 광범위한 변이 유형을 포괄적으로 지원합니다. Emedgene 소프트웨어는 DRAGEN Secondary Analysis의 포괄적인 변이 검출 기능을 통합함으로써 모자이크 변이, 부분 복제(segmental duplication) 영역 등 분석이 어려운 유전자 및 관심 영역(region of interest, ROI)도 정확하게 검출하여 랩에서 최종 사용자가 후속 해석 시 활용할 수 있도록 하였습니다.

또한 Emedgene 소프트웨어는 WES 또는 WGS와 같이 보다 포괄적인 백본(backbone)에서 바이오인포마틱스(bioinformatics, 생명정보학)를 통해 가상 패널을 정의하는 옵션을 제공합니다. 따라서 연구 랩은 단 하나의 웨트 랩(wet lab) assay를 개발해 랩의 애플리케이션과 가장 관련이 있는 유전자 패널을 선택할 수 있어, 새로운 유전자-질병 연관성(gene-disease association)이 발견될 때마다 assay 패널을 더욱 효율적으로 업데이트할 수 있습니다.

다양한 변이 유형에 대한 포괄적인 XAI 기반 3차 분석

사유 기술로 개발된 머신 러닝(machine learning) 알고리즘은 연구자에게 근거를 기반으로 한 정보와 해석 워크플로우를 확장할 기회를 제공합니다. Emedgene 소프트웨어 XAI는 연구자가 소프트웨어 내 자동 연결되어 있는 엄선된 근거를 기반으로 검토를 진행할 수 있도록 우선순위가 설정된 변이를 강조해 줍니다. 이러한 강조 기능은 정보를 뒷받침하는 과학적 근거를 찾기 위해 반복적으로 수작업을 하는 일도 크게 줄여 줍니다. XAI는 SNV, Indel, CNV, SV 삽입, STR, mtDNA를 포함한 포괄적인 변이 유형의 분석을 지원합니다.

미국 텍사스주 휴스턴에 위치한 Baylor Genetics의 연구진은 검증 연구를 실시하여 연구 대상으로부터 수집한 180개의 WES 데이터 세트로 구성된 코호트를 이용해 과거의 수동적인 검토 절차를 대체하는 XAI 기반 변이 해석의 성능을 검증하였습니다. 검증 결과, 96.6%의 샘플에서 원인 변이(causal variant)가 확인되었으며 추가 검토를 요하는 1~10위의 변이로 지정되었습니다(그림 1).²

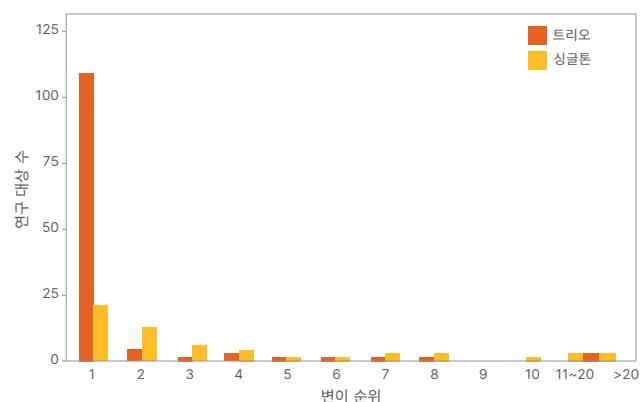


그림 1: XAI를 통한 변이 순위 — Baylor Genetics 연구진이 진행한 한 검증 연구에서 XAI를 통한 변이 해석은 96.6%의 분석 건에서 트리오(trio, 주황색) 및 싱글톤(singleton, 노란색)을 비롯한 원인 변이를 1~10위로 올바르게 지정함²

설명 가능한 자동 변이 큐레이션

간소화된 사용자 경험을 제공하는 Emedgene 소프트웨어는 XAI가 최종 선정한 목록에 포함된 모든 변이에 대한 근거를 제시합니다. 근거 그래프는 과학자가 수행하는 작업을 모방하여 질환-유전자 연관성과 함께 각 연구 대상과 관련이 있는 유전 패턴(inheritance pattern)과 표현형(phenotype)을 보여줍니다. 각각의 근거는 과학 논문, 데이터베이스 등의 외부 지식 소스와 연결되어 있습니다(그림 2). 이러한 근거 그래프는 XAI 플랫폼이 생성하는 데이터의 빠른 검토와 평가를 돕습니다.

Emedgene 소프트웨어는 매월 공공 및 사유 데이터 소스를 업데이트하여 항상 새로운 과학적 발견을 반영하고 있습니다. 이 소프트웨어는 Online Mendelian Inheritance in Man(OMIM) 카탈로그, ClinVar 등 많은 연구자들이 사용하는 외부 데이터베이스와 자연어 처리(natural language processing, NLP) 기술이 적용된 자동 과학 논문 큐레이션 기능을 함께 제공하는 지식 그래프(knowledge graph)의 형태로 데이터를 집산하고 통합하여 XAI를 강화합니다.

자동 ACMG 분류 체계

Emedgene 소프트웨어는 SNV, Indel, CNV, SV 및 mtDNA 변이에 대한 미국의학유전학회(American College of Medical Genetics, ACMG)의 자동화된 분류 체계를 사용해 변이 큐레이터(variant curator)의 자동 분류 태그 선택(automated classification tag selection)을 검토하는 절차를 간소화해 줍니다(그림 3). 이로써 정량적 태그 검토에 할애되는 해석 시간을 줄이고 큐레이터 간 검토 결과의 일치성을 높일 수 있습니다.

워크플로우 간소화

Emedgene 소프트웨어는 사용자 정의 해석 워크플로우의 간소화 및 최적화를 지원하는 다양한 기능을 포함하고 있습니다(그림 4). Emedgene 소프트웨어는 assay에 대한 지식이 없이도 사용이 가능하며, FASTQ 및 VCF(variant call format)를 포함한 다양한 데이터 형식을 지원합니다. 또한 이 소프트웨어는 사용자의 데이터 스토리지 솔루션에 통합되므로 마치 데이터 이동이 없는 듯한 경험을 제공할 수 있습니다.

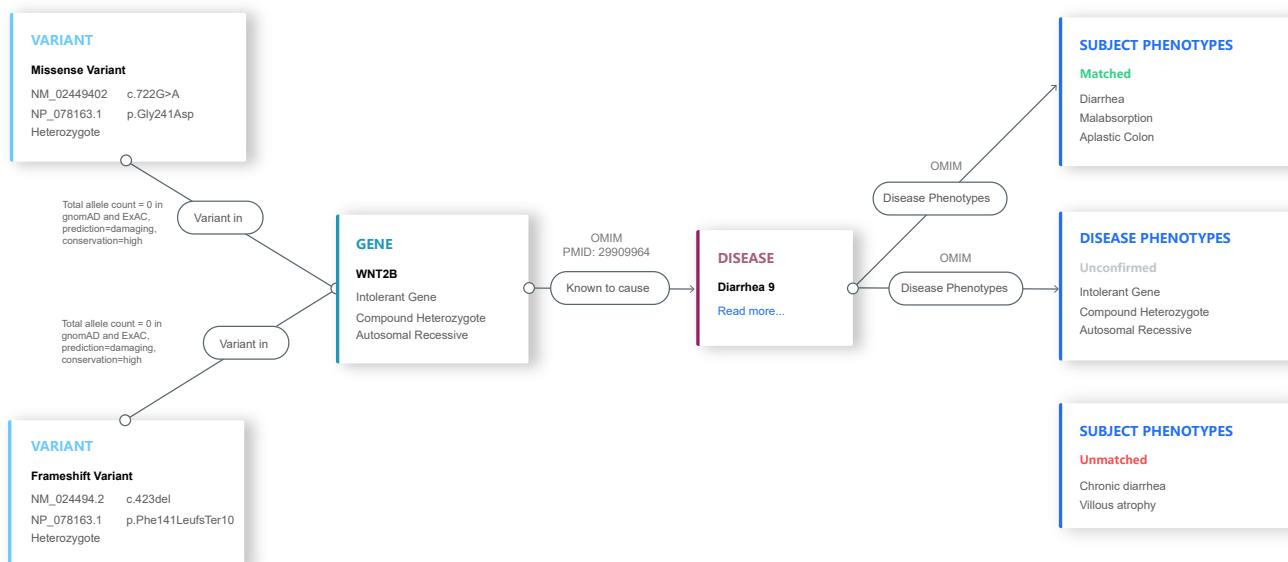


그림 2: Emedgene 소프트웨어 근거 그래프 — 높은 순위의 변이에는 자동으로 엄선된 근거 그래프가 함께 표시되며, 변이 해석을 실시하는 연구자가 이를 빠르고 쉽게 검토할 수 있도록 관련 논문과 데이터베이스의 링크 등 수집된 근거가 제시됨

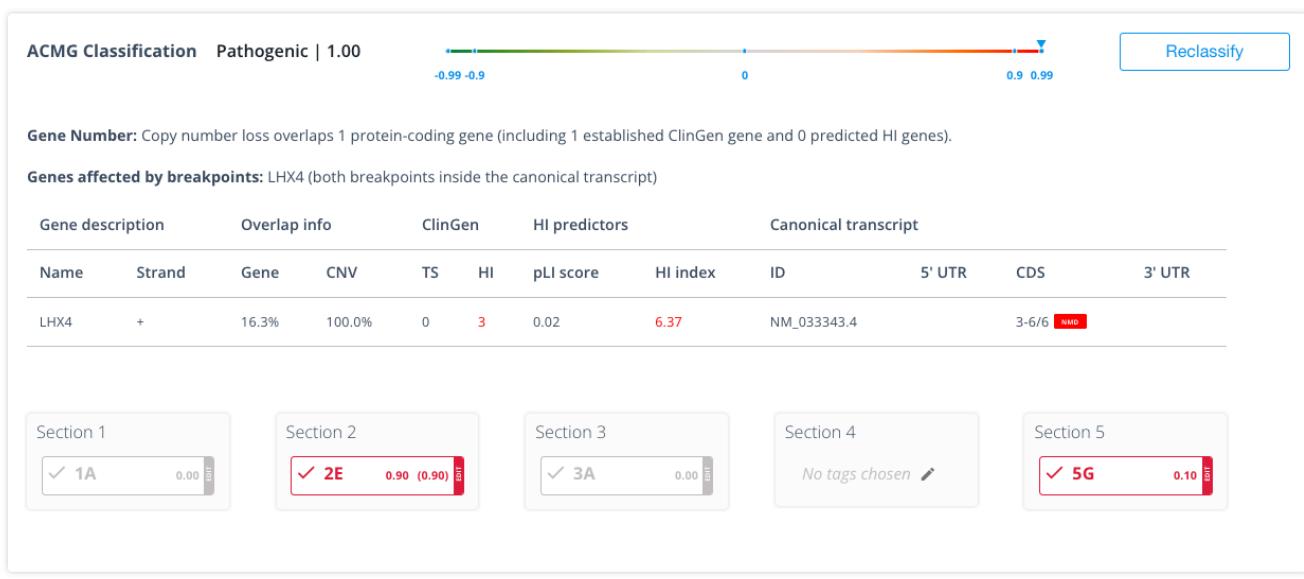


그림 3: 자동 ACMG 분류 체계 — SNV, Indel, CNV, SV 및 mtDNA 변이를 포함하는 다양한 변이의 ACMG 분류 체계를 자동으로 생성하여 변이 해석 연구 과학자의 검토 작업을 간소화해 주는 소프트웨어



그림 4: 사용자 정의 해석 워크플로우의 간소화 및 최적화를 지원하는 다양한 기능이 포함되어 있는 Emedgene 소프트웨어

SOP 맞춤 설정

연구 랩에서는 Emedgene 소프트웨어가 제공하는 유연하고 맞춤화가 가능한 플랫폼을 통해 운영 효율성을 크게 높이고 특정 해석 워크플로우를 확장할 수 있습니다. 다양한 사전 정의된 맞춤형 필터, 워크플로우, 보고서를 생성하고 검증 및 프로덕션 환경을 포함함으로써 자체적인 표준 작업 지침서(standard operating procedure, SOP)를 구현할 수 있습니다.

랩에 최적화된 큐레이션

Emedgene 소프트웨어는 랩의 큐레이션을 거친 변이와 유전자를 포함한 데이터를 저장하는 저장소(repository)를 제공함으로써 자동으로 해석 템플릿을 작성하고 케이스에 편리하게 주석(annotation)을 달 수 있게 해 줍니다. 큐레이션된 데이터를 유지하고 추가하면 운영 효율성을 크게 향상할 수 있습니다. 큐레이션된 데이터베이스는 변이별, 유전자별로 검색이 가능하므로 이전에 분석했던 모든 케이스에 쉽게 접근할 수 있습니다.

큐레이션을 거친 데이터 공유

랩은 협업의 힘을 활용하여 안전한 네트워크를 통해 큐레이션된 데이터를 공유함으로써 지식을 확장할 수 있습니다. 네트워크 파트너 및 데이터 공유 권한은 세분화되어 있으며 협업하는 랩 간에 결정됩니다. 이 옵트인(opt-in) 공유 데이터 기능은 신뢰할 수 있는 파트너와 보다 심층적인 지식을 공유할 수 있게 해 줍니다.

시각화 중심의 CNV 해석 워크플로우

강화된 전장 유전체 및 염색체 시각화 도구가 한 번의 클릭으로 편리하게 변이 레벨 데이터에 대한 액세스를 제공하여 CNV 해석을 가속화합니다. 종합적인 집단 및 지식 베이스(knowledge base) 시각화 트랙(track)과 간편한 조직화된 데이터, 과거 데이터 및 큐레이션된 데이터 시각화 기능을 결합함으로써 NGS 또는 마이크로어레이 데이터의 간소화된 세포유전학적(cytogenetic) 해석 솔루션을 제공합니다.

DRAGEN Secondary Analysis 해석

DRAGEN Secondary Analysis 소프트웨어는 NGS 데이터 처리 앱과 정확하고 포괄적이며 효율적인 변이 검출(variant calling) 앱을 제공합니다. DRAGEN Secondary Analysis는 precisionFDA Truth Challenge V2에서 챌린지에 참가한 다른 도구들과 비교했을 때 전체 벤치마크 영역(all benchmark regions)에서 다양한 변이에 걸쳐 가장 정확(F1 점수)하며 가장 포괄적인 솔루션인 것으로 확인되었습니다.³ 이 솔루션의 입증된 성능은 최적의 변이 검출 정확도, 다양한 assay 및 변이 유형과의 호환성, 방대한 유전체 데이터세트의 효율적인 처리 등 유전체 분석 시 발생하는 일반적인 문제에 대한 해결책을 제시합니다.

또한 Multigenome(그래프) 레퍼런스로 유전적 다양성(genetic diversity)을 포착하고 인종 편향(ethnicity bias)을 줄이며 다른 앱보다 SNV 정확도를 높여줍니다. DRAGEN Secondary Analysis는 상동성이 높은 영역(high-homology region)에서도 포괄적으로 변이를 검출합니다.

Emedgene 소프트웨어는 DRAGEN Secondary Analysis와 가장 포괄적인 통합을 통해 최적화된 2차 분석 및 3차 분석 솔루션을 제공합니다. DRAGEN 파이프라인은 로컬 인스턴스, BaseSpace Sequence Hub, Illumina Connected Analytics(ICA)를 통해 액세스하거나, Emedgene 소프트웨어 내에서 바로 구성하고 액세스하여 손쉽게 통합된 플러그 앤 플레이(plug-and-play) 솔루션으로 사용할 수 있습니다. 뛰어난 유연성을 자랑하는 Emedgene 소프트웨어는 어떤 2차 variant caller로 생성한 VCF 파일과도 호환이 가능하므로 랩은 기존의 NGS 워크플로우에 한 업체가 제공하는 솔루션만을 활용할 수 있게 됩니다.

맞춤 보고서 생성

연구자는 Microsoft Word와 유사한 인터페이스를 제공하는 편리한 편집 도구로 사용자 정의 워크플로우에 따라 보고서를 맞춤화, 편집 및 자동 작성할 수 있습니다. 수정 및 변경된 워크플로우를 지원하므로 추가 검토를 위해 보고서를 소프트웨어 내에서 전송할 수도 있습니다. 향상된 유연성을 제공하기 위해 연구자가 데이터를 간편하게 공유할 수 있는 PDF 또는 JSON 형식으로 보고서를 다운로드할 수 있도록 했습니다.

기타 플랫폼 및 시스템과의 통합

새로운 유전체 분석 기술, 기기 및 방법을 빨리 통합하는데에는 많은 어려움이 따릅니다. Emedgene 소프트웨어는 NovaSeq™ X 시리즈, NovaSeq 6000, NextSeq™ 2000 또는 Illumina의 다른 시퀀싱 시스템과 통합하여 NGS 분석 절차를 간소화해 줍니다. 또한 Emedgene 소프트웨어는 Illumina 라이브러리 준비 및 시퀀싱 기술이 통합된 DNA 준비부터 보고서 생성까지 모든 단계를 아우르는 솔루션의 유전 질환 연구 워크플로우를 구성하는 하나의 핵심 요소입니다.

Emedgene 소프트웨어는 탄탄한 애플리케이션 프로그래밍 인터페이스(application programming interface, API)를 통해 다른 기관의 실험실 정보 관리 시스템(laboratory information management system, LIMS)과의 통합을 지원합니다. 또 기존의 IT 인프라와의 유연한 통합이 가능하도록 BaseSpace Sequence Hub, Illumina Connected Analytics, Amazon Web Services의 Amazon S3, Azure Blob 등을 위한 데이터 이동 없는 스토리지 통합도 지원합니다.

안전한 네트워크

Emedgene 소프트웨어는 건강 보험 양도 및 책임에 관한 법(Health Insurance Portability and Accountability Act, HIPAA)의 요건을 지지하며 유럽 연합의 일반 데이터 보호 규정(General Data Protection Regulation, GDPR) 원칙을 준수합니다. Illumina는 기관에서 데이터 보안을 위해 기술적 조치와 조직적 조치를 모두 취하고, 개인의 권리를 보호할 수 있는 데이터 처리 활동에 대한 보안 대책을 포함하도록 요구하는 설계와 기본 설정에 의한 GDPR 개인정보 보호(GDPR data protection by design and by default)를 시행합니다. 또한 Illumina는 관리적, 물리적 그리고 기술적 보안 대책을 비롯해 HIPAA의 요구 사항을 지지하는 여러 기능을 제공합니다.

Emedgene 소프트웨어는 ISO 27001 등 국제 및 국내 데이터 프라이버시 요구 사항을 따르며, SOC(Service Organization Control) II Type 1 및 Type 2를 준수합니다. 기관의 통합 인증(single sign-on, SSO) 정책 통합과 같은 고급 맞춤형 보안 옵션도 제공합니다. 또 직접 암호화 키를 생성하고 제어(Bring Your Own Key)하며 데이터를 이중 암호화하는 옵션도 사용 가능합니다.

요약

Emedgene 소프트웨어는 희귀 질환 및 유전 질환 연구 데이터의 3차 분석에 활용할 수 있는 XAI를 통한 자동화 지원 솔루션을 제공합니다. 랩은 Emedgene 소프트웨어를 이용하여 변이 해석 효율성을 높이고 운영을 확대할 수 있습니다. Emedgene 소프트웨어는 정확하고 포괄적이며 효율적인 변이 검출에 사용하는 DRAGEN Secondary Analysis 등 Illumina의 다양한 시퀀싱, 마이크로어레이 및 소프트웨어 솔루션과 통합함으로써 연구 랩에 지속적인 성장의 기회를 제시하는 강력한 도구를 제공합니다.

상세 정보

Emedgene 소프트웨어

제품 목록

제품	카탈로그 번호
Emedgene software—Genome Equivalent Sample (FASTQ)	20073772
Emedgene software—Genome Equivalent Sample (VCF)	20073773
Emedgene software—Annual Support and Maintenance (Basic)	20073776
Emedgene software—Annual Support and Maintenance (Professional)	20073777
Emedgene software—Annual Support and Maintenance (Enterprise)	20073778
Emedgene software—Training at Customer Site	20073774
Custom report (Annual)	20073778

추가적으로 랩에 필요한 제품은 제공되는 견적서에 포함됨

Illumina는 고객분들의 문의 사항을 검토해 한정적으로 지원을 통한 제품 평가 기회를 제공하고 있습니다. 이를 통해 Emedgene 소프트웨어에서 예시 연구 데이터 세트를 선택해 작업해 보시거나, 소프트웨어에서 자체적으로 보유 중인 데이터 세트를 업로드 후 평가해 보실 수 있습니다. 자세한 정보는 Illumina 영업 담당자에게 문의하시기 바랍니다.

참고 문헌

1. Austin-Tse CA, Jobanputra V, Perry DL, et al. [Best practices for the interpretation and reporting of clinical whole genome sequencing](#). *NPJ Genom Med.* 2022;7:27. doi.org/10.1038/s41525-022-00295-z.
2. Meng L, Attali R, Tammy T, et al. [Evaluation of an automated genome interpretation model for rare disease routinely used in a clinical genetic laboratory](#). *Genet Med.* 2023;25(6):100830. doi:10.1016/j.gim.2023.100830
3. Olson ND, Wagner J, McDaniel J, et al. [PrecisionFDA Truth Challenge V2: Calling variants from short and long reads in difficult-to-map regions](#). *Cell Genomics.* 2022;2(5):100129. doi.org/10.1016/j.xgen.2022.100129.



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-00273 KOR

연구 전용입니다. 진단 절차에는 사용할 수 없습니다.

문서 번호: M-KR-00273 KOR
영어 원본: M-GL-01057 v2.0 | 7