

Emedgene™

희귀 질환 연구 및
유전 질환 연구에
적용 가능한
효율적인 3차 분석
워크플로우 제공

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

illumina®

소개

차세대 시퀀싱(Next-generation sequencing, NGS) 기술을 활용한 전장 유전체 시퀀싱(whole-genome sequencing, WGS) 및 전장 엑솜 시퀀싱(whole-exome sequencing, WES)은 유전체 전체를 편향 없이 고해상도로 연구할 수 있도록 해 줍니다. 이러한 assay는 유전 질환 관련 유전자 변이의 식별 및 특성 확인을 비롯해 다양한 연구 분야에 적합합니다. 그러나 랩에서 방대한 양의 WGS/WES 데이터를 분석하고 해석하는 단계에서 상당한 병목 현상이 발생합니다. 특히 변이 해석은 많은 시간을 필요로 하는 수작업 단계로, 유전체당 최대 7시간이 소요되며 수많은 반복 단계를 거쳐야 합니다.¹

Illumina는 이 문제를 해결할 수 있도록 랩에 연구 소프트웨어 플랫폼인 Emedgene을 제공하고 있습니다. 이 포괄적인 플랫폼은 랩의 해석 워크플로우를 간소화하고 변이 큐레이션(curation) 부담을 줄여 주기 위한 목적으로 유전학자가 설계하였습니다. 또한 자동화된 설명 가능한 인공지능(explainable artificial intelligence, XAI) 기반의 유전체 분석 플랫폼을 통해 효율성이 2~5배 높고 분석 건당 총 워크플로우 시간은 50~75% 감소된 고처리량 데이터 해석이 가능합니다.

Emedgene은 DRAGEN™ Bio-IT Platform, BaseSpace™ Sequence Hub, Illumina Connected Analytics를 포함하는 Illumina 2차 분석 및 데이터 스토리지 플랫폼과 통합함으로써 모듈식 옵션을 제공하여 랩에서 Illumina의 종합 솔루션을 사용할 수 있도록 합니다. 이 소프트웨어는 단일 염기서열 변이(single nucleotide variant, SNV), 삽입/결실(insertion/deletion, Indel), 유전자 복제수 변이(copy number variation, CNV), 미토콘드리아 DNA(mitochondrial DNA, mtDNA) 변이, 구조적 변이(structural variant, SV), 짧은 연쇄 반복(short tandem repeat, STR)을 모두 아우르는 사용자 정의 변이 분석 및 해석 워크플로우에 WGS, WES 및 표적 패널 데이터의 사용을 지원합니다.

Emedgene은 WES 또는 WGS와 같이 보다 포괄적인 백본(backbone)에서 바이오인포매틱스(bioinformatics, 생명정보학)를 통해 가상 패널(virtual panel)을 정의하는 옵션을 제공하므로 연구 랩이 하나의 실험실 assay에서 표준화를 진행하고 즉시 추가 분석을 수행하는 역량을 갖출 수 있습니다. 또한 백본 assay에서 표준화를 진행하면 새로운 관심 유전자가 발견될 때마다 표적 패널을 업데이트하지 않아도 되는 장점이 있습니다.

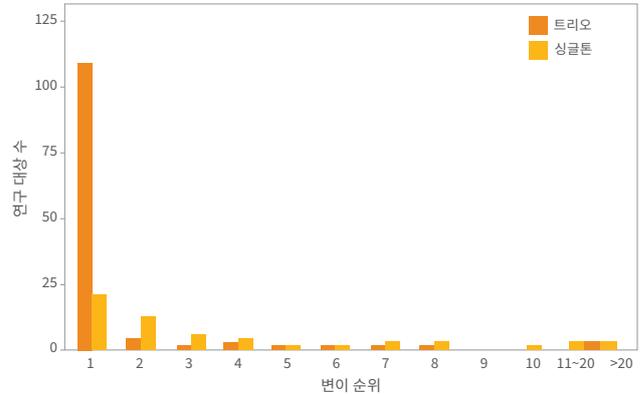


그림 1: XAI를 통한 변이 순위 — Baylor Genetics 연구진이 진행한 검증 연구에서 XAI를 통한 변이 해석은 96.6%의 분석 건에서 트리오(trio, 주황색) 및 싱글톤(singleton, 노란색)을 비롯한 원인 변이를 1~10위로 올바르게 지정함.

XAI를 통한 3차 분석

사유 기술로 개발된 머신 러닝(machine learning) 알고리즘은 연구자에게 근거를 기반으로 한 정보와 해석 워크플로우를 확장할 기회를 제공합니다. Emedgene은 연구자가 소프트웨어 내 자동 연결되어 있는 엄선된 근거를 기반으로 검토를 진행할 수 있도록 우선순위가 설정된 변이를 강조해 줍니다. 따라서 정보를 뒷받침하는 근거를 찾기 위해 반복적으로 수작업을 하는 일도 줄어들습니다. 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 XAI를 통한 변이 해석의 성능을 검증하였습니다. 검증 결과, 96.6%의 샘플에서 원인 변이(casual variant)가 확인되었으며 추가 검토를 요하는 1~10위의 변이로 지정되었습니다(그림 1).

설명 가능한 자동 변이 큐레이션

Emedgene은 간소화된 사용자 경험을 토대로 XAI가 최종 선정된 목록에 포함된 모든 변이에 대한 근거를 제시합니다. 근거 그래프는 과학자가 수행하는 작업을 모방하여 질환-유전자 연관성(disease-gene association)과 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).

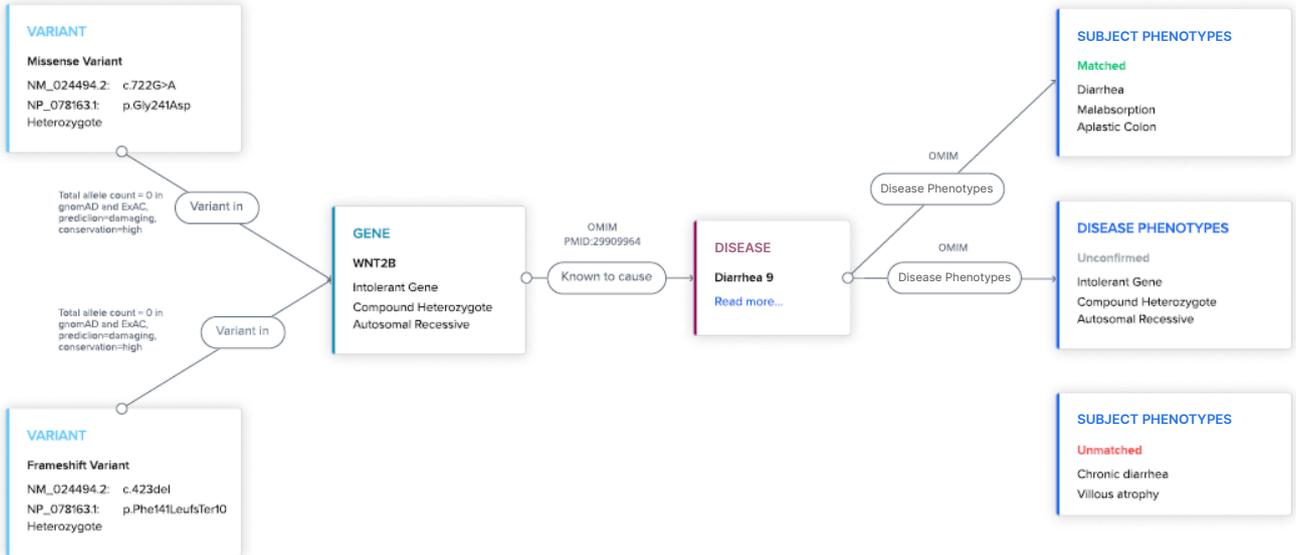


그림 2: 근거 그래프 — 높은 순위의 변이에는 자동으로 엄선된 근거 그래프가 함께 표시되며, 변이 해석을 실시하는 연구자가 이를 빠르고 쉽게 검토할 수 있도록 관련 논문과 데이터베이스의 링크 등 수집된 근거가 제시됨.

이러한 근거 그래프는 XAI 플랫폼이 생성하는 데이터의 빠른 검토와 평가를 돕습니다.

이 소프트웨어는 Online Mendelian Inheritance in Man(OMIM) 카탈로그, ClinVar 등 많은 연구자들이 사용하는 외부 데이터베이스의 데이터를 XAI 모델의 기반이 되는 지식 그래프(knowledge graph)의 형태로 집산하고 통합합니다. Emedgene이 항상 최신 과학적 발견을 반영할 수 있도록 공공 및 사유 지식 소스를 통해 플랫폼을 매달 업데이트하며, 자연어 처리(natural language processing, NLP) 기술을 채택하여 과학 논문을 자동 엄선하고 있습니다.

자동 ACMG 분류 체계

Emedgene은 SNV, Indel, CNV, SV 및 mtDNA 변이에 대한 미국의학유전학회(American College of Medical Genetics, ACMG)의 분류 체계를 자동화하여 변이 큐레이터(variant curator)가 자동으로 선택된 변이를 검토하는 절차를 간소화해 줍니다(그림 3). 이로써 정량적 검토에 할애되는 해석 시간을 줄이고 큐레이터 간 검토 결과의 일치성을 높일 수 있습니다.

워크플로우 간소화

Emedgene은 사용자 정의 해석 워크플로우를 간소화하고 최적화해 주는 다양한 기능을 제공합니다(그림 4). Emedgene은 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.

SOP 맞춤 설정

연구 랩에서는 이 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

Emedgene은 큐레이션을 거친 랩의 데이터를 저장하는 저장소(repository)를 제공합니다. 이 저장소는 특정 변이와 관련이 있는 변이, 유전자 또는 연구 대상별 큐레이션을 포함하거나, 연구 대상과는 관련이 없는 특정 변이나 유전자에 관한 정보를 포함할 수 있습니다. 랩의 데이터베이스는 변이 해석 및 검토 중에 사용됩니다. 큐레이션을 거친 랩의 데이터를 효과적으로 사용하면 데이터 해석 및 큐레이션 절차를 최적화하여 운영 효율성을 한층 더 높일 수 있습니다.

큐레이션을 거친 데이터 공유

랩은 협업의 힘을 활용하여 안전한 네트워크를 통해 큐레이션을 거친 데이터를 공유함으로써 지식을 확장할 수 있습니다. 네트워크 파트너 및 데이터 공유 권한은 세분화되어 있으며 협업하는 랩 간에 결정됩니다.

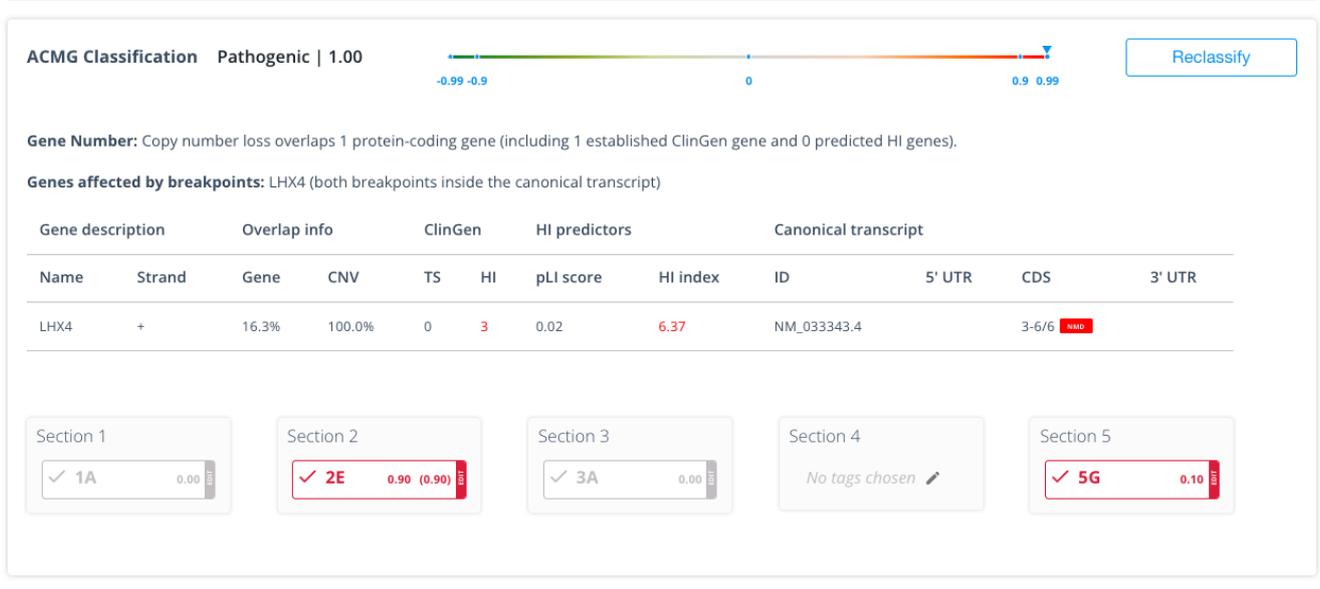


그림 3: 간소화된 자동 ACMG 분류 체계 검토 — SNV, Indel, CNV, SV 및 mtDNA 변이를 포함하는 다양한 변이의 ACMG 분류 체계를 자동으로 생성하여 큐레이터의 검토 작업을 간소화해 주는 소프트웨어

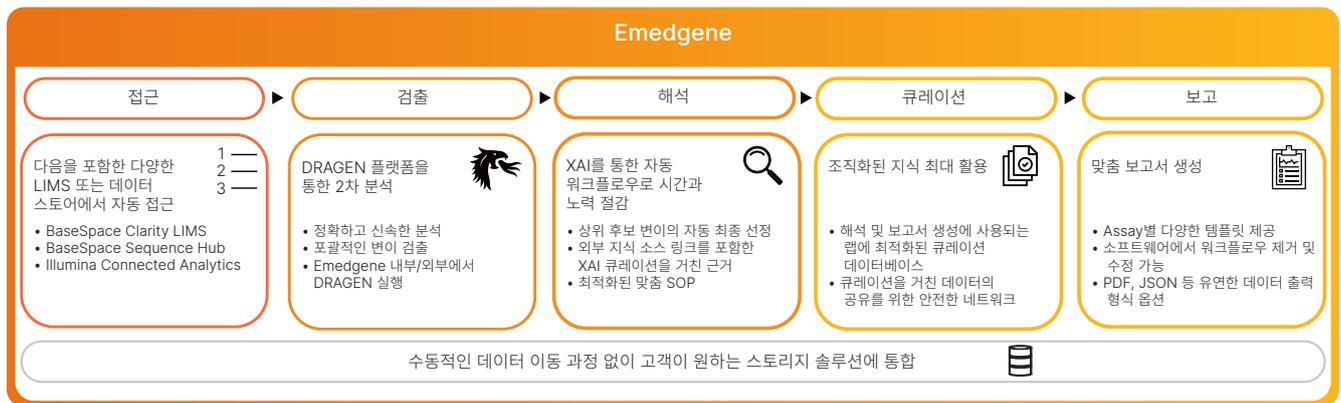


그림 4: Emedgene으로 최적화되는 사용자 정의 변이 해석 워크플로우 — 사용자 정의 해석 워크플로우의 간소화 및 최적화를 지원하는 다양한 소프트웨어의 기능

안전한 네트워크

현재 Emedgene은 일반 데이터 보호 규정(General Data Protection Regulation, GDPR)의 원칙을 준수하고 있습니다. 또한 Emedgene을 Bring Your Own Key 및 기타 보안 설정을 비롯한 랩의 통합 인증(single sign-on, SSO) 정책에 통합하는 옵션도 제공하고 있습니다. 전체 접근 권한 제어, 감사 로깅(audit logging) 및 모니터링을 포함한 데이터 보안 및 보호 메커니즘도 구현되어 있습니다.

맞춤 보고서 생성

연구자는 Microsoft Word와 유사한 인터페이스를 제공하는 편리한 편집 도구 내에서 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. 간편한 데이터 공유와 향상된 유연성을 위해 연구자가 PDF 또는 JSON 형식으로 보고서를 다운로드할 수 있도록 했습니다.

DRAGEN 플랫폼을 통한 2차 분석

Emedgene은 어떤 2차 variant caller와도 호환이 가능하며, DRAGEN(Dynamic Read Analysis for GENomics) Bio-IT Platform에도 통합할 수 있습니다. PrecisionFDA Truth Challenge V2에서 DRAGEN 플랫폼은 챌린지에 참가한 다른 도구와 비교했을 때 모든 벤치마크 영역에서 다양한 변이에 걸쳐 가장 정확(F1 점수)하며 가장 포괄적인 솔루션인 것으로 확인되었습니다.² Emedgene의 전 단계에 사용 시 입증된 DRAGEN 플랫폼의 성능은 최적의 변이 검출 정확도, 다양한 assay 및 변이 유형과의 호환성, 방대한 양의 유전체 데이터 처리 속도를 가속화하는 방법 등 유전체 분석 시 발생하는 일반적인 문제에 대한 해결책을 제시해 줍니다. Emedgene의 전 단계에 실행하는 2차 분석에는 DRAGEN 플랫폼(독립형 인스턴스)을 사용하거나 Emedgene 인터페이스에 포함된 DRAGEN 분석 기능을 사용할 수 있습니다. 시퀀싱 데이터는 수동적인 데이터 이동 없이 연구자가 원하는 솔루션에 저장할 수 있습니다.

기타 플랫폼 및 시스템과의 통합

많은 랩에서 발 빠르게 새로운 유전체 분석 기술, 기기 및 방법을 통합하는 데 어려움을 겪습니다. Emedgene은 NovaSeq™ X 시리즈, NovaSeq 6000, NextSeq™ 2000 또는 Illumina의 다른 시퀀싱 시스템과 통합하여 WGS 및 WES 분석 절차를 간소화해 줍니다. 또한 Emedgene은 Illumina 라이브러리 준비 및 시퀀싱 단계를 포함하는 DNA 준비부터 보고서 생성까지 모든 단계를 아우르는 통합형 WGS 솔루션의 희귀 질환 연구 워크플로우에서 마지막 단계에 사용됩니다. Emedgene은 애플리케이션 프로그래밍 인터페이스(application programming interface, API)와 호환되므로 다른 기관의 실험실 정보 관리 시스템(laboratory information management system, LIMS)과의 통합을 지원합니다.

요약

Emedgene은 희귀 질환과 기타 유전 질환 연구 시 3차 분석에 활용할 수 있는 XAI를 통한 자동화 지원 솔루션을 제공합니다. 랩은 Emedgene을 이용하여 데이터 해석 효율성을 높이고 운영을 확대할 수 있습니다. Emedgene은 정확하고 포괄적이며 효율적인 2차 분석 변이 검출에 사용하는 DRAGEN 플랫폼 등 Illumina의 다양한 시퀀싱 시스템 및 소프트웨어와 통합함으로써 연구 랩에 지속적인 성장의 기회를 제공합니다.

상세 정보

Emedgene: illumina.com/products/by-type/informatics-products/emedgene.html

제품 목록

Illumina는 고객분들의 문의 사항을 검토해 한정적으로 지원을 통한 제품 평가 기회를 제공하고 있습니다. 이를 통해 Emedgene에서 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. 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.

illumina®

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-KR-00168 KOR