DATA SHEET

Emedgene

Gain efficiency in tertiary analysis workflows for rare and inherited disease research applications

- Streamlines tertiary analysis by as much as 75% using XAI and customizable automation
- Integrates with LIMS, EMR, and other IT systems to simplify and secure the entire assay workflow
- Powers users for growth to meet the evolving demands of science, technology, and scale

illumina®

Introduction

Whole-genome sequencing (WGS) and whole-exome sequencing (WES) using next-generation sequencing (NGS) technologies provide a high-resolution, unbiased view across the entire genome. This makes these assays ideal for various applications, including the identification and characterization of genetic variants associated with disease. However, analyzing and interpreting the vast amounts of data produced by WGS/WES represent a significant bottleneck for laboratories. In particular, variant interpretation is a time-intensive, manual process that requires numerous, repetitive steps and as much as seven hours per genome.[1]

To help labs address this challenge, we offer a research software platform, Emedgene. This comprehensive platform was designed by geneticists to streamline interpretation workflows and reduce the variant curation burden. It features an automated, explainable-AI (XAI)–powered genomic analysis platform that enables high-throughput interpretation with a 2–5× increase in efficiency and 50–75% reduction in total workflow time per subject.

By integrating with Illumina secondary analysis and data storage platforms, including the DRAGEN™ Bio-IT Platform, BaseSpace™ Sequence Hub, and Illumina Connected Analytics, Emedgene offers modular options for a single vendor solution. The software supports data input from WGS, WES, and targeted panels for user-defined workflows for variant analysis and interpretation across single nucleotide variants (SNVs), insertions/deletions (indels), copy number variants (CNVs), mitochondrial DNA variants (mtDNA), structural variants (SVs), and short tandem repeats (STRs).

Emedgene also provides the option to define "virtual panels" bioinformatically from a more comprehensive backbone such as WES or WGS, enabling research labs to standardize on one wet lab assay with immediate reflex analysis capabilities. Also, standardization on a backbone assay removes the need to update targeted panels as new genes of interest come to light.

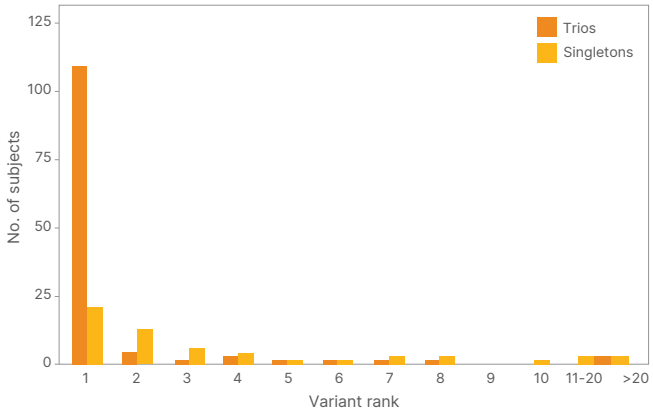

Figure 1: XAI-powered variant ranking—In a validation study by researchers at Baylor Genetics, XAI-powered variant interpretation correctly ranked causal variants in the top 10 in 96.6% of research subjects evaluated, including trios (orange) and singletons (yellow).

XAI-powered tertiary analysis

Proprietary machine learning algorithms deliver evidence-backed insights and help users scale their interpretation workflows. Emedgene highlights prioritized variants for review, based on curated evidence that is automatically linked within the software. This supports the elimination of manual, repetitive tasks around connecting evidence to insights. In a validation study by researchers at Baylor Genetics, Houston, TX, XAI-powered variant interpretation was evaluated with a cohort of 180 WES data sets from research subjects, previously solved by manual review. In 96.6% of samples that were evaluated, the causal variant was identified and ranked in the top 10 for further review (Figure 1).

Automated, explainable variant curation

With a streamlined user experience, Emedgene connects evidence for every variant shortlisted by the XAI. The evidence graph mimics the work performed by a scientist, showing disease-gene associations along with inheritance patterns and phenotypes associated with the subject. Each piece of evidence is linked out to external sources, whether in the scientific literature or a database (Figure 2).

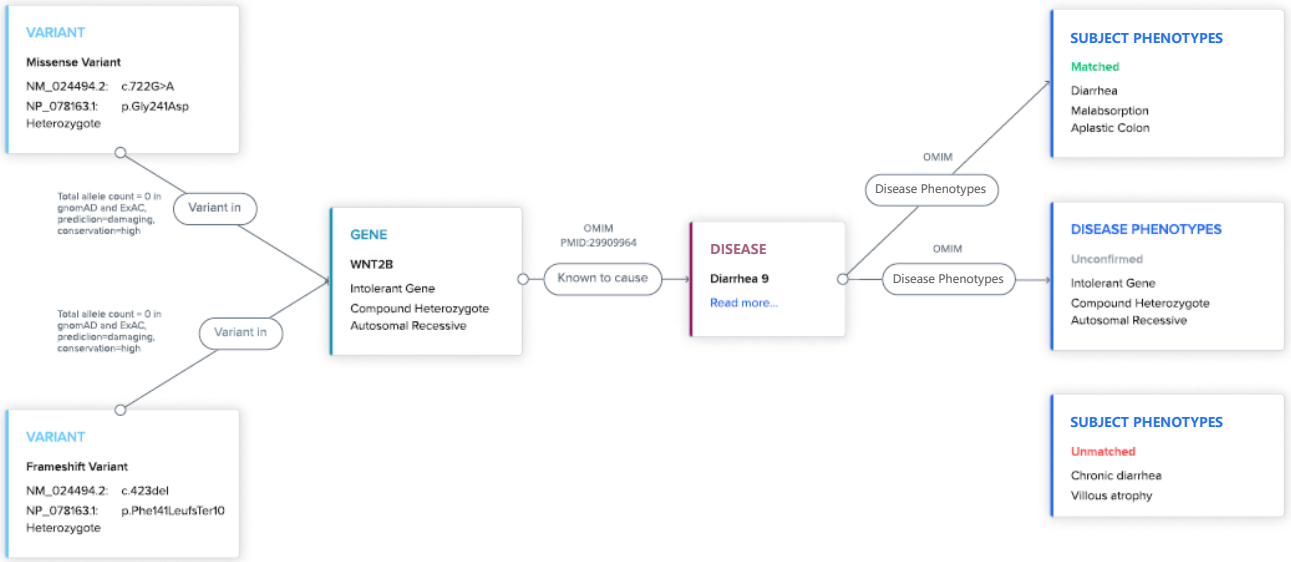

Figure 2: Evidence graph—Highly ranked variants are displayed with an automatically curated evidence graph compiling evidence for a quick and easy review by a variant interpretation research scientist, including links out to relevant literature and databases.

The evidence graph enables rapid review and assessment of the output of the XAI platform.

The software aggregates and integrates data from preferred external databases, such as the Online Mendelian Inheritance in Man (OMIM) catalog, ClinVar, and others into a knowledge graph that powers the XAI models. To make sure that Emedgene stays current with scientific discoveries, the platform is updated monthly from public and proprietary sources with automatic curation of scientific literature using natural language processing (NLP).

Automatic ACMG classifications

By automating American College of Medical Genetics (ACMG) classifications for SNV, indel, CNV, SV, and mtDNA variants, Emedgene streamlines review of automated selections by a variant curator (Figure 3). This decreases the interpretation time of quantitative sections, and can increase concordance between curators.

Streamlining workflows

Multiple features streamline and optimize user-defined interpretation workflows (Figure 4). Emedgene is assay agnostic and compatible with a range of data inputs, including FASTQ and VCF file formats. The software integrates with the user's storage solution to achieve a "no data movement" experience.

Customizable SOPs

This flexible and customizable platform enables research laboratories to maximize efficiency and scale their specific interpretation workflows. Configurable standard operating procedures (SOPs) can be implemented by creating sets of predefined custom filters, workflows, reporting, and the inclusion of validation and production environments.

Lab-specific curation

Emedgene features a repository for a lab's curated data, which can include variant, gene, and subject specific curations pertaining to a specific variant or subject independent information about a variant or gene. A lab's database is used during variant interpretation and review. Effectively using a lab's curated data can help optimize data interpretation and curation, resulting in additional efficiency gains.

Share curated data

Labs can leverage the power of collaboration to scale knowledge by sharing curated data across a private network of labs. Network partners and data sharing permissions are granular and determined by participating laboratories.

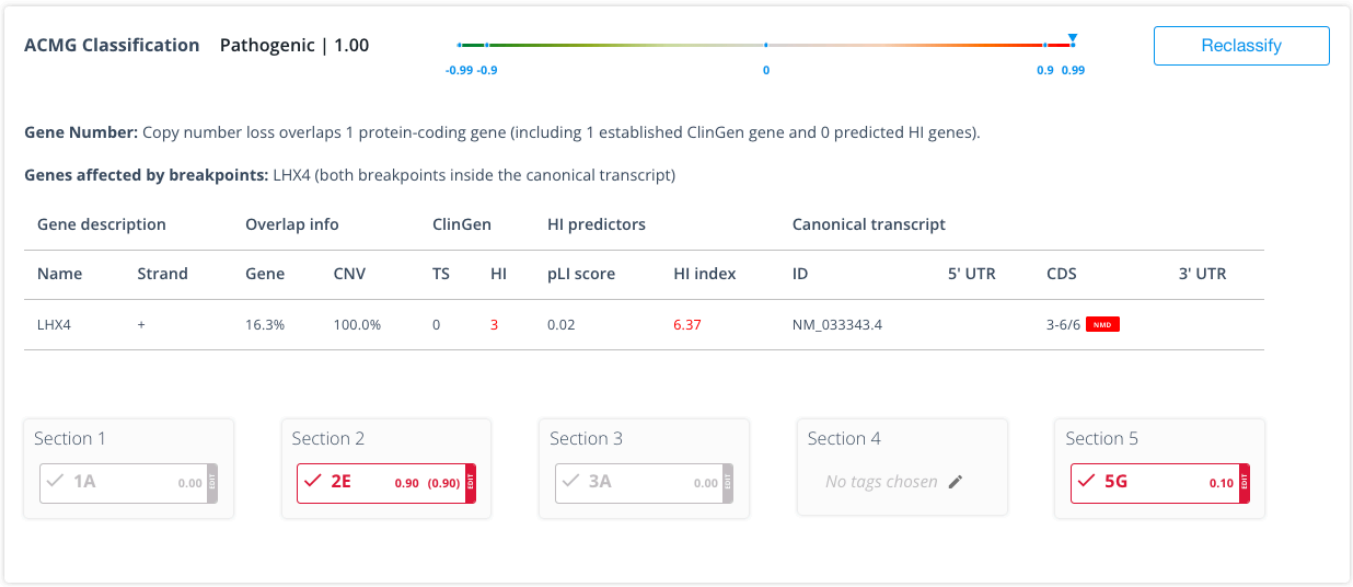

Figure 3: Streamlined review of automated ACMG classifications—The software automatically generates ACMG classifications for variants, including SNVs, indels, CNVs, SVs, and mtDNA variants, enabling streamlined review by a curator.

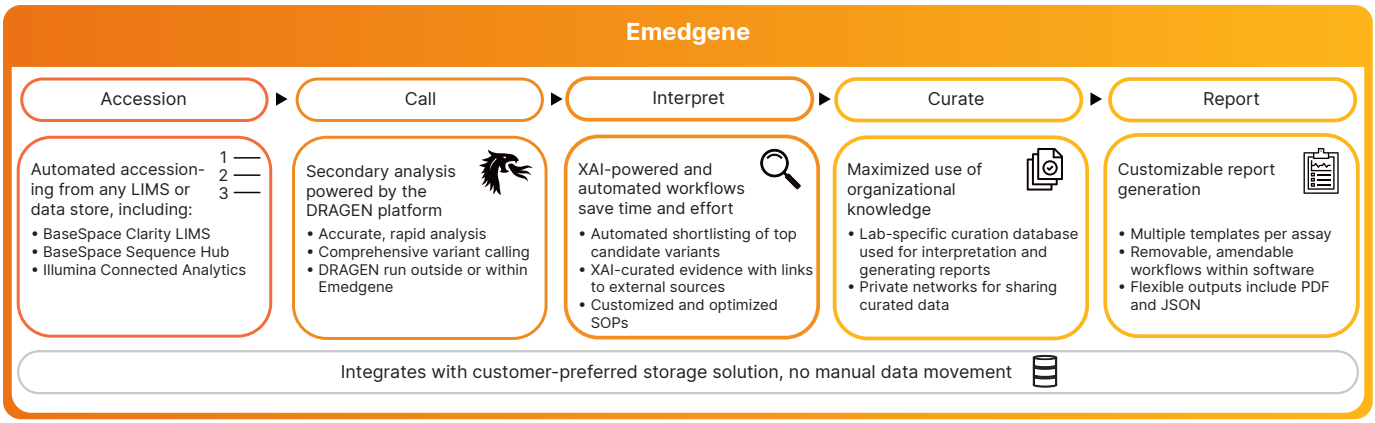

Figure 4: Optimizing user-defined variant interpretation workflows with Emedgene—The software has numerous features to streamline and optimize user-defined interpretation workflows.

Private, secure network

Emedgene currently follows the principles of the General Data Protection Regulation (GDPR). It offers options to integrate with a lab's single sign-on policy, including Bring Your Own Key and other security settings. Data security and protection mechanisms have been implemented, including full access control and audit logging and monitoring.

Customizable report generation

Users can customize, edit, and automatically populate reports following a user-defined workflow within a convenient editing tool similar to Microsoft Word. With support for amended and revised workflows, reports can be sent for additional review within the software. For ease of data sharing and increased flexibility, reports can be downloaded in PDF or JSON format.

Secondary analysis powered by the DRAGEN platform

Emedgene is compatible with any secondary variant caller and can be integrated with the DRAGEN (Dynamic Read Analysis for GENomics) Bio-IT Platform. As part of the PrecisionFDA Truth Challenge V2, the DRAGEN platform was reported as the most accurate (F1 score) and most comprehensive solution across variants assessed in all benchmark regions as compared against other participating tools.[2] The proven performance of the DRAGEN platform upstream of Emedgene further addresses common challenges in genomic analysis, such as optimal variant calling accuracy, compatibility with a breadth of assay and variant types, and how to accelerate the processing of massive volumes of genomic data. Secondary analysis upstream of Emedgene can be performed using a DRAGEN platform (standalone instance) or by directly including the DRAGEN analysis in Emedgene interface. Sequencing data can be stored in any customer-preferred solution without the need for manual data movement.

Integration with other platforms and systems

Many labs struggle to keep pace with integrating new genomics technology, instruments, and methodologies. Emedgene simplifies the process, integrating with the NovaSeq™ X Series, NovaSeq 6000, NextSeq™ 2000, or other Illumina sequencing systems for streamlining WGS and WES analysis. Also, Emedgene represents the final piece in the rare disease research workflow of an integrated, DNA-to-report WGS solution that includes Illumina library preparation and sequencing. Compatibility with Application Programming Interfaces (APIs) enables integration with other institutional laboratory information management systems (LIMS).

Summary

Emedgene offers an XAI-powered and automation-enabled solution for rare and other hereditary disease tertiary analysis research applications. It increases data interpretation efficiency and enables labs to scale their operations. With integration across Illumina sequencing and software systems, including the DRAGEN platform for accurate, comprehensive, and efficient secondary analysis variant calling, Emedgene powers research laboratories for continued growth.

Learn more

Emedgene, illumina.com/products/by-type/informatics-products/connected-insights-germline.html

Ordering information

For qualified inquiries, Illumina offers a supported evaluation experience, allowing customers to work with example research data sets available in Emedgene or upload and evaluate their own data sets within the software. Contact an Illumina sales representative for more information.

References

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®

1.800.809.4566 toll-free (US) | +1.858.202.4566 tel
techsupport@illumina.com | www.illumina.com

M-GL-01057 v1.0