DATA SHEET

Emedgene™ software

Optimize user-defined, variant interpretation for rare disease and other genetic research applications

- Streamlines tertiary analysis for germline applications 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 for NGS and microarray assays

illumina®

Introduction

Applications of next-generation sequencing (NGS) for researching rare and other genetic diseases, pharmacogenomics, cytogenetics, other germline conditions or carrier screening studies are central to progressing our understanding of human genetics and translating to precision medicine. Whole-genome sequencing (WGS), whole-exome sequencing (WES), targeted and virtual panels provide a high-resolution, unbiased view across the complete genome, coding regions, or targeted gene regions, respectively. However, analyzing and interpreting the vast amounts of data produced by these methods presents significant bottlenecks for laboratories. In particular, variant interpretation is a time-intensive, manual process that requires numerous repetitive steps that can take as much as seven hours per genome.[1] For high-throughput panel applications, such as healthy or carrier population screening projects, labs often look to maximize efficiency and accuracy with tools for automation, variant curation, and efficient research report generation.

To help labs address these challenges, Illumina offers Emedgene software, a comprehensive research software designed by geneticists to streamline interpretation workflows and reduce the variant curation burden. Emedgene software is powered by explainable AI (XAI) and automation for high-throughput, user-defined variant interpretation workflows, enabling a 2–5× increase in efficiency and 50–75% reduction in total workflow time per subject compared to manual interpretation.

By integrating with Illumina secondary analysis and data storage platforms, including DRAGEN™ secondary analysis, BaseSpace™ Sequence Hub, and Illumina Connected Analytics, Emedgene software offers modular options in a single vendor solution from sample processing through research report generation. The software supports data input from WGS, WES, targeted panels, and microarrays. Comprehensive variant types are supported, including single nucleotide variants (SNVs), insertions/deletions (indels), copy number variants (CNVs), mitochondrial DNA variants (mtDNA), structural variants (SVs), and short tandem repeats (STRs). Emedgene software integrates the comprehensive variant calling capabilities of DRAGEN secondary analysis to make sure that even challenging genes and regions of interest, such as mosaic variants, segmental duplication regions, and more, are accurately called as input for downstream interpretation by the lab end user.

Emedgene software also provides the option to define "virtual panels" bioinformatically from a more comprehensive backbone such as WES or WGS, enabling research labs to develop one wet lab assay and select the most relevant gene panel for their application, making assay panel updates more efficient as new gene-disease associations come to light.

Comprehensive XAI tertiary analysis across variant types

Proprietary machine learning algorithms deliver evidence-backed insights and help users scale their interpretation workflows. Emedgene software XAI highlights prioritized variants for review, based on curated evidence that is automatically linked within the software. This focus facilitates the elimination of manual, repetitive tasks that are typically involved in connecting evidence to insights. XAI covers a comprehensive range of variant types, including SNVs, indels, CNVs, SV insertions, STRs, and mtDNA.

In a validation study, researchers at Baylor Genetics, Houston, TX, evaluated XAI-powered variant interpretation 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).[2]

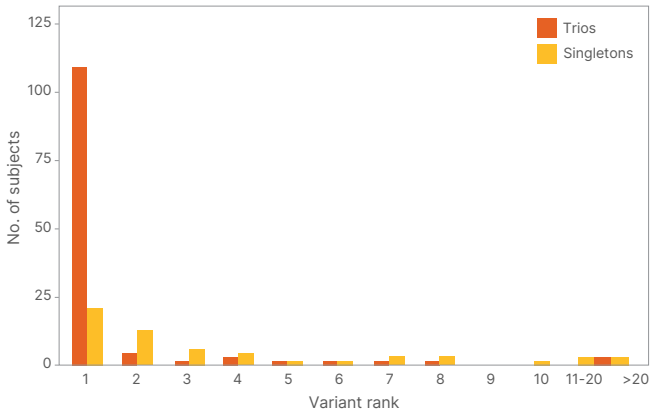

Figure 1: XAI-powered variant ranking—In a validation study by researchers at Baylor Genetics, XAI-powered variant interpretation correctly ranked causal variants in the top 10 for 96.6% of research subjects evaluated, including trios (orange) and singletons (yellow).[2]

Automated, explainable variant curation

As part of the streamlined user experience, Emedgene software 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). The evidence graph enables rapid review and assessment of the output of the XAI platform.

Monthly updates of public and proprietary data sources makes sure that Emedgene software stays current with scientific discoveries. 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 and includes automatic curation of scientific literature using natural language processing (NLP).

Automatic ACMG classifications

Using automated American College of Medical Genetics (ACMG) classifications for SNV, indel, CNV, SV, and mtDNA variants allows Emedgene software to streamline review of automated classification tag selections by a variant curator (Figure 3). This decreases the interpretation time of quantitative tags and can increase concordance between curators.

Streamlining workflows

Emedgene software includes multiple features that help streamline and optimize user-defined interpretation workflows (Figure 4). Emedgene software 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.

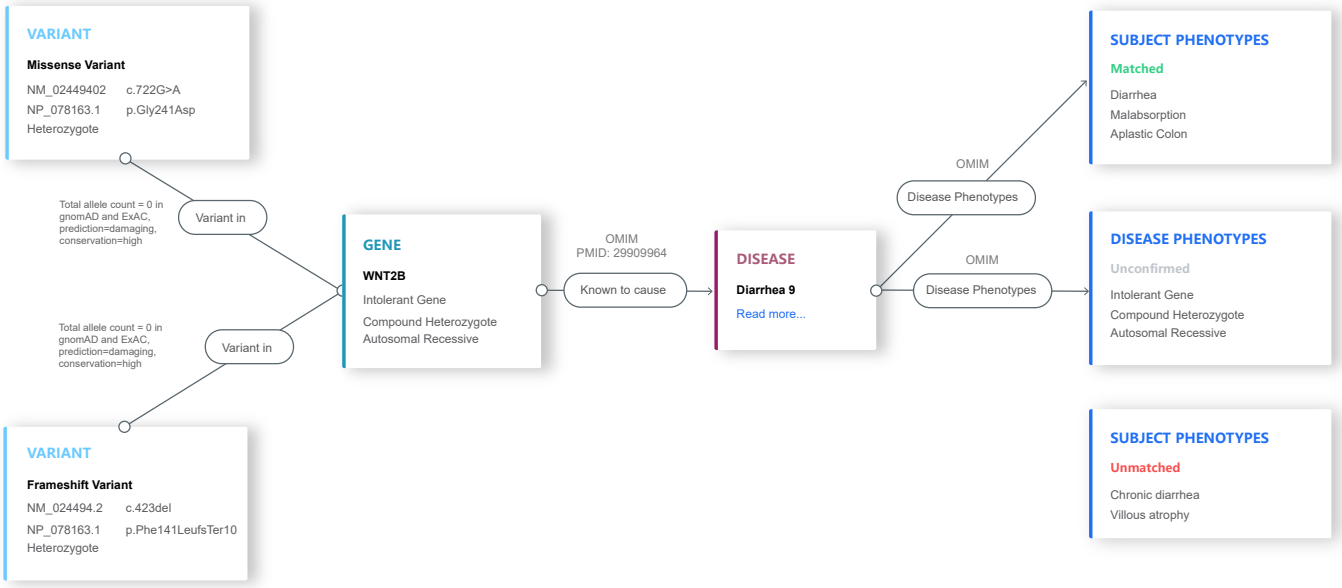

Figure 2: Emedgene software evidence graph—Highly ranked variants are displayed with an automatically curated evidence graph compiling evidence for quick and easy review by a variant interpretation research scientist, including links to relevant literature and databases.

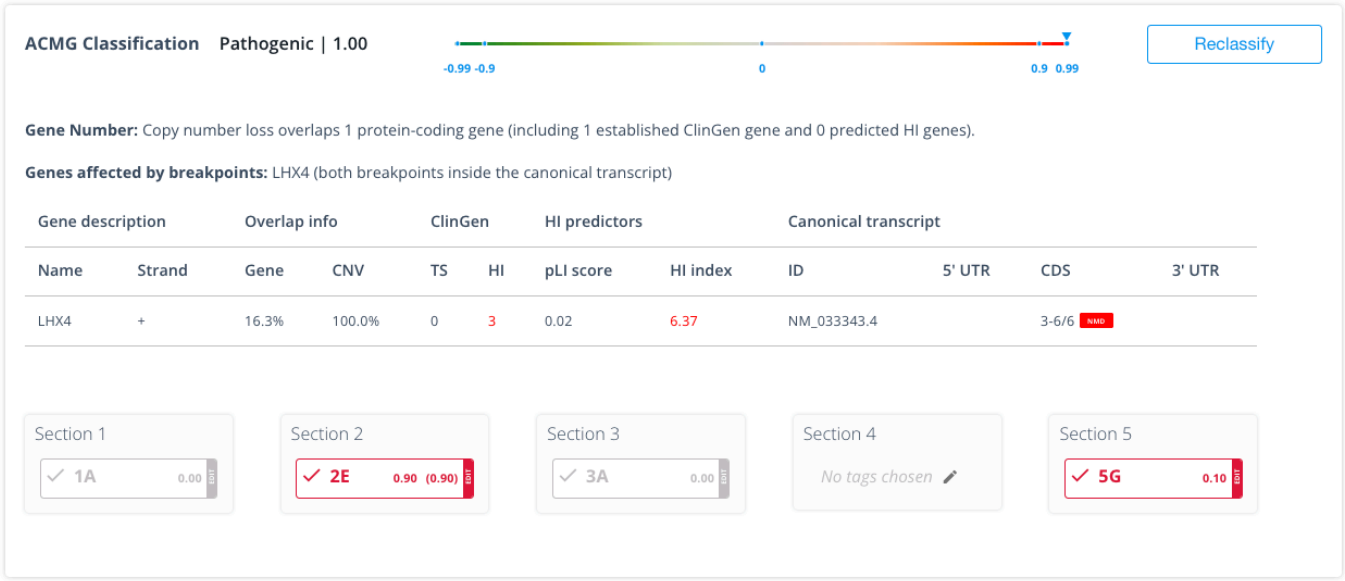

Figure 3: Automated ACMG classifications—The software automatically generates ACMG classifications for variants, including SNVs, indels, CNVs, SVs, and mtDNA variants, enabling streamlined review by a variant interpretation research scientist.

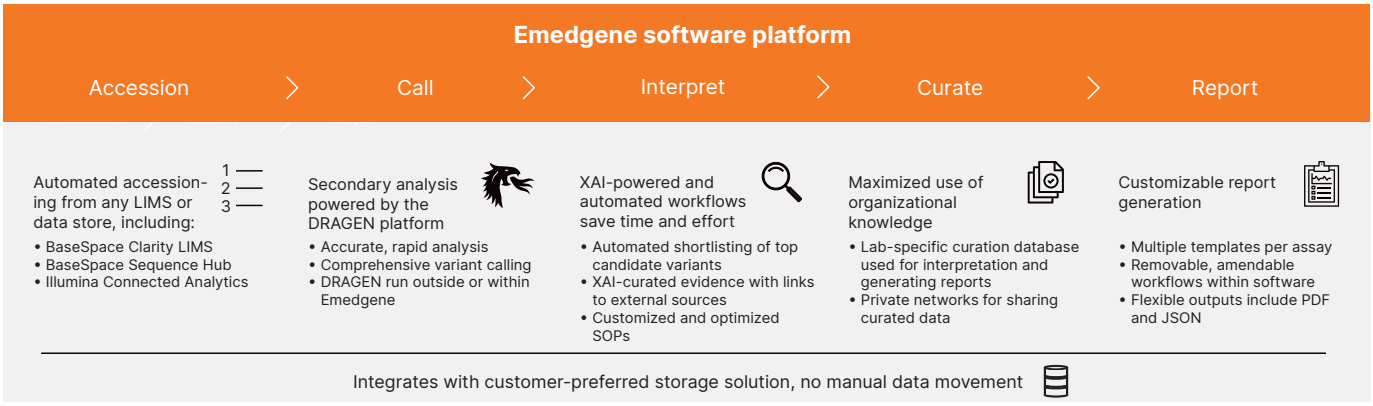

Figure 4: Emedgene software has numerous features to streamline and optimize user-defined interpretation workflows.

Customizable SOPs

Emedgene software offers a flexible and customizable platform to enable 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 software features a repository for a lab's curated data, which can include variant and gene curations that are automatically used to conveniently annotate cases and autopopulate interpretation templates. Maintaining and growing curated data results in significant efficiency gains. The curated database is searchable by variant and gene, for easy access to all previously assessed cases.

Share curated data

Through the power of collaboration, labs can 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. This opt-in shared data feature enables partners to share deeper knowledge with trusted partners.

Visualization-driven workflow for CNV interpretation

Enhanced genome- and chromosome-wide visualizations accelerate the interpretation of CNVs with convenient one-click access to variant-level data. Comprehensive population and knowledge base visualization tracks are coupled with easy visualization of organizational, historical, and curated data to provide a streamlined solution for cytogenetic interpretation from NGS or microarray data.

DRAGEN secondary analysis integration

DRAGEN secondary analysis software is a suite of applications that processes NGS data and performs accurate, comprehensive, and efficient variant calling. As part of the PrecisionFDA Truth Challenge V2, DRAGEN secondary analysis was reported as the most accurate (F1 score) and the most comprehensive solution across variants assessed in all benchmark regions as compared against other participating tools.[3] The proven performance of the solution addresses common challenges in genomic analysis, such as optimal variant calling accuracy, compatibility with a range of assay and variant types, and efficient processing of massive genomic data sets. The multigenome (graph) reference captures genetic diversity, reduces ethnicity bias, and improves SNV accuracy in comparison with other applications. DRAGEN secondary analysis also comprehensively calls variants in high-homology regions.

Emedgene software provides the most comprehensive integration possible with DRAGEN secondary analysis to generate an optimized secondary through tertiary analysis solution. DRAGEN pipelines can be accessed via a local instance, BaseSpace Sequence Hub, Illumina Connected Analytics (ICA), or pipelines can be configured and accessed directly within Emedgene software for an easily-integrated "plug-and-play" solution. For maximum flexibility, Emedgene software is also compatible with the variant call format (VCF) output of virtually any secondary variant caller, allowing labs to take advantage of a single-vendor solution for their NGS workflow.

Customizable report generation

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

Integration with other platforms and systems

Keeping pace with and integrating new genomics technology, instruments, and methodologies are challenging. Emedgene software simplifies the process by integrating with the NovaSeq™ X Series, NovaSeq 6000, NextSeq™ 2000, or other Illumina Sequencing Systems for streamlining NGS analysis. Emedgene software also represents a critical piece in the genetic disease research workflow by integrating a DNA-to–report solution with Illumina library preparation and sequencing technologies.

Emedgene software offers robust Application Programming Interfaces (APIs) to enable integration with other institutional laboratory information management systems (LIMS). No-data-movement storage integrations are available for BaseSpace Sequence Hub, Illumina

Connected Analytics, Amazon S3 for Amazon Web Services, Azure Blob, and more for flexible integration into existing information technology infrastructure.

Private, secure network

Emedgene software supports Health Insurance Portability and Accountability Act (HIPAA) requirements and follows the principles of the European Union, General Data Protection Regulation (GDPR). Illumina implements GDPR data protection, which requires organizations to implement technical and organizational measures for data security and to integrate safeguards into data processing activities to protect individual rights, by design and by default. Also, Illumina offers several features to support the requirements of HIPAA, including administrative, physical, and technical safeguards.

Emedgene software also conforms to global and local data privacy requirements, including ISO 27001, and is compliant with SOC 1 and 2 type II (Service Organization Control). Custom advanced security options include integration with organizational single sign-on policy. An option to bring your own key and doubly encrypt data is also available.

Summary

Emedgene software offers an XAI-powered and automation-enabled solution for tertiary analysis of rare disease and genetic research data. It increases variant interpretation efficiency and enables labs to scale their operations. With integration across Illumina sequencing, microarray, and software solutions, including DRAGEN secondary analysis for accurate, comprehensive, and efficient variant calling, Emedgene software provides research labs with a powerful tool for continued growth.

Learn more

Emedgene software

Ordering information

Product	Catalog no.
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
Any additional items required for the lab will be included in the prepared quotation.	

For qualified inquiries, Illumina offers a supported evaluation experience, allowing customers to work with example research data sets available in the Emedgene software or to 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. Meng L, Attali R, Talmy 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.

illumina®

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

M-GL-01057 v2.0