

Software Emedgene™

Ottimizzazione
dell'interpretazione delle
varianti definita dall'utente
per le malattie rare e altre
applicazioni di ricerca genetica

- Semplificazione dell'analisi terziaria per le applicazioni per la linea germinale fino al 75% grazie all'uso di XAI e all'automazione personalizzabile
- Integrazione con LIMS, EMR e altri sistemi IT per semplificare e rendere sicuro l'intero flusso di lavoro dei saggi
- Grande potenzialità di crescita per soddisfare le esigenze in continua evoluzione delle applicazioni scientifiche, tecnologiche e di scala per saggi NGS e microarray



Introduzione

Le applicazioni di sequenziamento di nuova generazione (NGS, next-generation sequencing) per la ricerca di malattie genetiche rare e di altro tipo, la farmacogenomica, la citogenetica, altre condizioni della linea germinale o studi di screening dei portatori sono fondamentali per lo sviluppo della nostra comprensione della genetica umana e la trasformazione in medicina di precisione. Il sequenziamento dell'intero genoma (WGS, whole-genome sequencing), il sequenziamento dell'intero esoma (WES, whole-exome sequencing), i pannelli mirati e quelli virtuali forniscono una visione imparziale e ad alta risoluzione rispettivamente sull'intero genoma, sulle regioni codificanti o sulle regioni geniche mirate. Tuttavia, l'analisi e l'interpretazione delle grandi quantità di dati prodotti da questi metodi rappresentano difficoltà significative per i laboratori. In particolare, l'interpretazione delle varianti è un processo manuale dispendioso in termini di tempo e che prevede numerose fasi ripetitive che possono richiedere fino a sette ore per genoma.¹ Per le applicazioni di pannelli a elevata processività, come i progetti di screening della popolazione sana o portatrice, i laboratori spesso cercano di massimizzare l'efficienza e l'accuratezza attraverso strumenti per l'automazione, la gestione delle varianti e la generazione efficiente di report di ricerca.

Per aiutare i laboratori ad affrontare queste sfide, Illumina offre Emedgene, un software di ricerca completo progettato da genetisti per semplificare i flussi di lavoro di interpretazione e ridurre il carico della gestione delle varianti. Il software Emedgene si basa sull'intelligenza artificiale spiegabile (XAI, explainable AI) e sull'automazione dei flussi di lavoro di interpretazione delle varianti definiti dall'utente a elevata processività, con un aumento dell'efficienza da due a cinque volte e una riduzione del 50-75% del tempo totale necessario per il flusso di lavoro di ogni soggetto rispetto all'interpretazione manuale.

Grazie all'integrazione con le piattaforme di analisi secondaria e archiviazione dei dati di Illumina, tra cui l'analisi secondaria DRAGEN™, BaseSpace™ Sequence Hub e Illumina Connected Analytics, il software Emedgene offre opzioni modulari in una soluzione di unico fornitore, dall'elaborazione dei campioni alla generazione di report di ricerca. Il software supporta l'input di dati da WGS, WES, pannelli mirati e microarray. Sono supportati tipi di varianti completi, tra cui varianti a singolo nucleotide (SNV, single nucleotide variant), inserzioni/delezioni (indel), varianti del numero di copie (CNV, copy number variant), varianti del DNA mitocondriale (mtDNA, mitochondrial DNA), varianti strutturali (SV, structural variant) e brevi ripetizioni in tandem (STR, short tandem repeat). Il software Emedgene integra le funzionalità complete di identificazione di varianti dell'analisi secondaria

DRAGEN per garantire che anche i geni e le regioni di interesse difficili, come le varianti a mosaico, le regioni di duplicazione segmentale e di altro tipo, siano accuratamente identificati come input per l'interpretazione a valle da parte dell'utente finale del laboratorio.

Il software Emedgene offre anche la possibilità di definire bioinformaticamente i "pannelli virtuali" da una struttura portante più completa come WES o WGS, consentendo ai laboratori di ricerca di sviluppare un saggio di laboratorio umido e di selezionare il pannello di geni più rilevante per la loro applicazione, rendendo più efficienti gli aggiornamenti del pannello dei saggi man mano che emergono nuove associazioni gene-malattia.

Analisi terziaria XAI completa su tutti i tipi di varianti

Gli algoritmi proprietari di apprendimento automatico forniscono approfondimenti basati sulle evidenze e consentono agli utenti di scalare i flussi di lavoro di interpretazione. La XAI del software Emedgene evidenzia le varianti prioritarie da esaminare, basandosi su evidenze valutate e collegate automaticamente all'interno del software. Questa focalizzazione facilita l'eliminazione delle attività manuali e ripetitive tipicamente coinvolte nel collegamento delle evidenze alle informazioni. La XAI copre una gamma completa di tipi di varianti, tra cui SNV, indel, CNV, inserzioni SV, STR e mtDNA.

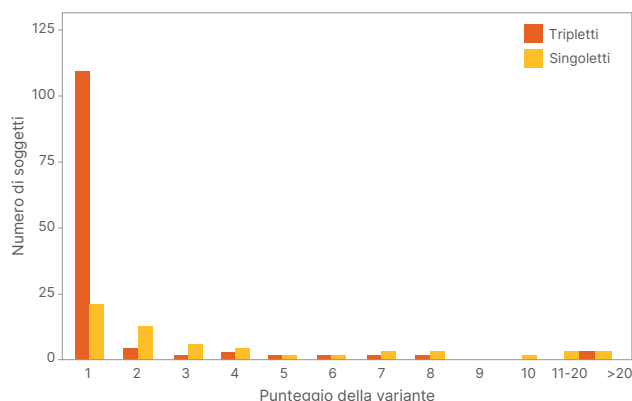


Figura 1: classifica delle varianti basata su XAI. In uno studio di convalida condotto dai ricercatori di Baylor Genetics, l'interpretazione delle varianti basata su XAI ha classificato correttamente le varianti causali tra le prime 10 nel 96,6% dei soggetti della ricerca valutati, compresi i tripletts (in arancione) e i singoletti (in giallo).²

In uno studio di convalida, i ricercatori di Baylor Genetica di Houston (Texas, Stati Uniti) hanno valutato l'interpretazione delle varianti basata su XAI con una coorte di 180 set di dati WES provenienti da soggetti di ricerca, precedentemente risolti mediante revisione manuale. Nel 96,6% dei campioni valutati, la variante causale è stata identificata e classificata tra le prime 10 per un'ulteriore revisione (Figura 1).²

Descrizione e valutazione delle varianti automatizzata

Come parte dell'esperienza utente semplificata, il software Emedgene collega le evidenze per ogni variante selezionata dalla XAI. Il grafico delle evidenze riproduce il lavoro svolto da un ricercatore, mostrando le associazioni tra malattia e gene insieme agli schemi ereditari e ai fenotipi associati al soggetto. Ogni elemento di evidenza è collegato a fonti esterne, che possono essere nella letteratura scientifica o in un database (Figura 2). Il grafico delle evidenze consente un esame e una valutazione rapidi dei risultati della piattaforma XAI.

Gli aggiornamenti mensili delle fonti di dati pubbliche e proprietarie assicurano il continuo allineamento del software Emedgene alle scoperte scientifiche. Il software aggrega e integra i dati provenienti dai database esterni preferiti, come il catalogo Online

Mendelian Inheritance in Man (OMIM), il ClinVar e altri, in una rete di conoscenze alla base dei modelli XAI e include la gestione automatica della letteratura scientifica tramite un sistema di elaborazione del linguaggio naturale (NLP, natural language processing).

Classificazioni automatiche ACMG

L'uso delle classificazioni automatizzate dell'American College of Medical Genetics (ACMG) per le varianti SNV, indel, CNV, SV e mtDNA consente al software Emedgene di ottimizzare la revisione delle selezioni automatiche dei tag di classificazione da parte di un descrittore di varianti (Figura 3). Questo riduce il tempo di interpretazione dei tag quantitativi e può aumentare la concordanza tra i descrittori.

Ottimizzazione dei flussi di lavoro

Il software Emedgene include molteplici funzioni che semplificano e ottimizzano i flussi di lavoro interpretativi definiti dall'utente (Figura 4). Il software Emedgene è imparziale rispetto ai saggi e compatibile con una serie di input di dati, compresi i formati di file FASTQ e VCF. Il software si integra con la soluzione di archiviazione dell'utente per fornire un'esperienza che non preveda lo spostamento dei dati.

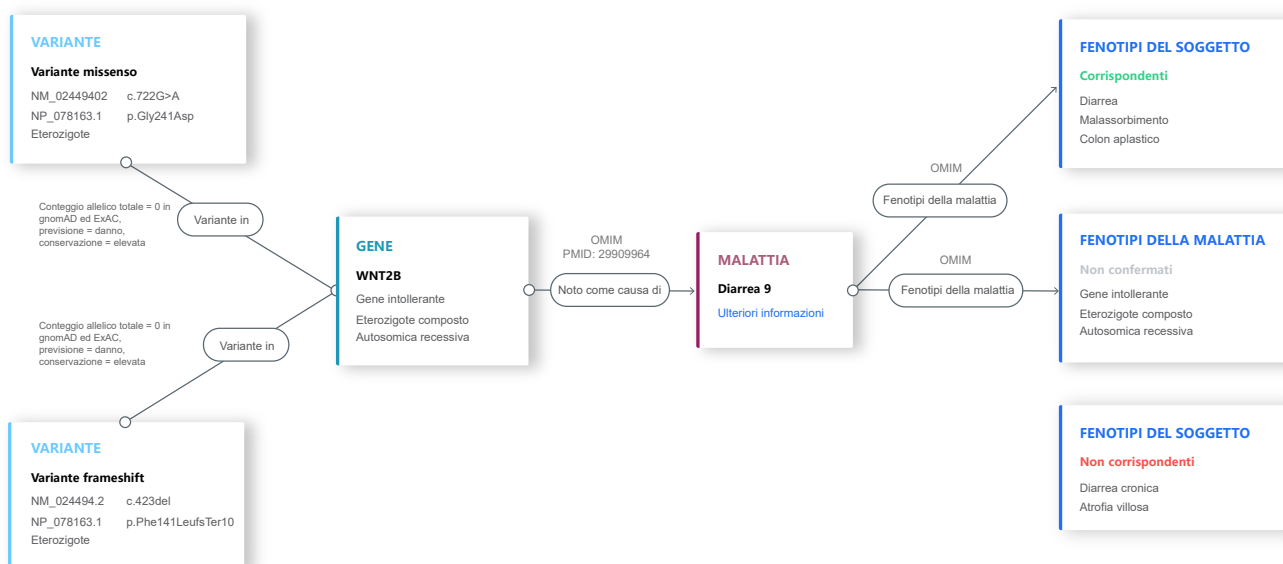


Figura 2: grafico delle evidenze del software Emedgene. Le varianti con un punteggio alto vengono visualizzate con un grafico delle evidenze valutate automaticamente, che consentono a un ricercatore specializzato nell'interpretazione delle varianti di eseguire un esame rapido e semplice, facilitato da collegamenti alla letteratura e ai database pertinenti.

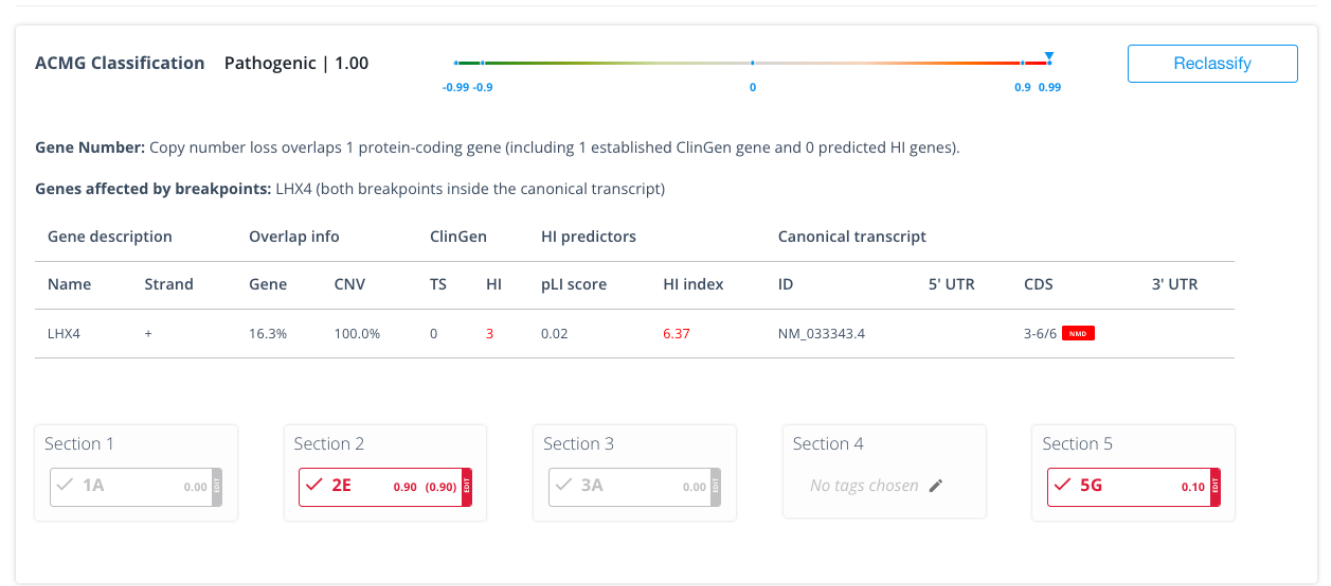


Figura 3: classificazioni ACMG automatizzate. Il software genera automaticamente le classificazioni ACMG per le varianti, comprese le varianti SNV, indel, CNV, SV e mtDNA, consentendo al ricercatore specializzato nell'interpretazione delle varianti di eseguire un esame semplificato.



Figura 4: il software Emedgene dispone di numerose funzioni per semplificare e ottimizzare i flussi di lavoro di interpretazione definiti dall'utente.

Procedure operative standard personalizzabili

Il software Emedgene offre una piattaforma flessibile e personalizzabile che consente ai laboratori di ricerca di massimizzare l'efficienza e di scalare i propri flussi di lavoro di interpretazione specifici. Permette di implementare procedure operative standard (SOP, standard operating procedure) configurabili, creando set di filtri personalizzati predefiniti, flussi di lavoro, report e l'inclusione di ambienti di validazione e produzione.

Descrizione e valutazione specifiche per il laboratorio

Il software Emedgene dispone di un archivio per i dati valutati da un laboratorio, che può includere la descrizione e la valutazione di varianti e geni utilizzate automaticamente per annotare opportunamente i casi e compilare automaticamente i modelli di interpretazione. Il mantenimento e l'aumento di dati valutati si traduce in guadagni significativi in termini di efficienza. Il database valutato è consultabile per variante e per gene, consentendo un accesso facilitato a tutti i casi analizzati in precedenza.

Condivisione dei dati valutati

Grazie alla potenza della collaborazione, i laboratori possono scalare le conoscenze condividendo i dati relativi agli elementi valutati in una rete di laboratori privata. I partner di rete e le autorizzazioni di condivisione dei dati sono granulari e determinati dai laboratori partecipanti. Questa funzione di associazione per la condivisione dei dati consente ai partecipanti di condividere conoscenze più approfondite con partner di fiducia.

Flusso di lavoro basato sulla visualizzazione per l'interpretazione delle CNV

Le visualizzazioni migliorate dell'intero genoma e dell'intero cromosoma accelerano l'interpretazione delle CNV grazie a un pratico accesso con un solo clic ai dati a livello di variante. Il monitoraggio completo della visualizzazione della popolazione e della knowledge base si abbina a una semplice visualizzazione dei dati organizzativi, storici e valutati per fornire una soluzione ottimizzata per l'interpretazione citogenetica da dati NGS o microarray.

Integrazione dell'analisi secondaria DRAGEN

Il software di analisi secondaria DRAGEN è una suite di applicazioni che elabora i dati NGS ed esegue un'identificazione di varianti accurata, completa ed efficiente. Tra gli strumenti partecipanti alla PrecisionFDA Truth Challenge V2, l'analisi secondaria DRAGEN è stata segnalata come la soluzione più accurata (punteggio F1)

e più completa per le varianti valutate in tutte le regioni di riferimento.³ Le prestazioni comprovate della soluzione affrontano le sfide comuni dell'analisi genomica, come l'accuratezza ottimale dell'identificazione di varianti, la compatibilità con un'ampia gamma di saggi e tipi di varianti e l'elaborazione efficiente di set massicci di dati genomici. Il riferimento (grafico) multigenomico acquisisce la diversità genetica, riduce la distorsione etnica e migliora l'accuratezza SNV rispetto ad altre applicazioni. L'analisi secondaria DRAGEN identifica inoltre in modo esaustivo le varianti nelle regioni ad alta omologia.

Il software Emedgene fornisce l'integrazione più completa possibile con l'analisi secondaria DRAGEN per generare una soluzione ottimizzata di analisi secondaria e terziaria. È possibile accedere alle pipeline DRAGEN tramite un'istanza locale, BaseSpace Sequence Hub, Illumina Connected Analytics (ICA) oppure è possibile configurare e accedere alle pipeline direttamente all'interno del software Emedgene per creare una soluzione "plug-and-play" facilmente integrata. Per garantire la massima flessibilità, il software Emedgene è compatibile anche con l'output di formato di identificazione di varianti (VCF, variant call format) di praticamente qualsiasi sistema di identificazione di varianti secondarie, consentendo ai laboratori di affidarsi a una soluzione mono-fornitore per il proprio flusso di lavoro NGS.

Creazione di report personalizzati

Gli utenti possono personalizzare, modificare e compilare automaticamente i report seguendo un flusso di lavoro definito dall'utente utilizzando un pratico strumento di editing testuale simile a Microsoft Word. Grazie al supporto di flussi di lavoro modificati e revisionati, i report possono essere inviati per un'ulteriore revisione all'interno del software. Per una maggiore flessibilità, i report possono essere scaricati in formato PDF o JSON, semplificando la condivisione dei dati.

Integrazione con altre piattaforme e sistemi

Rimanere sempre al passo e integrarsi con le nuove tecnologie genomiche, i nuovi strumenti e le nuove metodologie può essere complicato. Il software Emedgene semplifica il processo integrandosi con NovaSeq™ X Series, NovaSeq 6000, NextSeq™ 2000 o altri sistemi di sequenziamento Illumina per l'ottimizzazione dell'analisi NGS. Il software Emedgene rappresenta inoltre un elemento fondamentale nel flusso di lavoro di ricerca sulle malattie genetiche grazie all'integrazione di una soluzione dal DNA al report con le tecnologie di preparazione e sequenziamento delle librerie Illumina.

Il software Emedgene offre efficaci interfacce di programmazione delle applicazioni (API, application programming interface) per consentire l'integrazione con altri sistemi di gestione delle informazioni di laboratorio (LIMS, laboratory information management system) istituzionali. Sono disponibili integrazioni di archiviazione senza spostamento dei dati per BaseSpace Sequence Hub, Illumina Connected Analytics, Amazon S3 per Amazon Web Services, Azure Blob e altre soluzioni per un'integrazione flessibile nelle infrastrutture di information technology esistenti.

Rete privata e sicura

Il software Emedgene supporta i requisiti dell'Health Insurance Portability and Accountability Act (HIPAA) e segue i principi del Regolamento generale sulla protezione dei dati (GDPR, General Data Protection Regulation) dell'Unione europea. Illumina mette in pratica la protezione dei dati GDPR, che richiede alle organizzazioni di implementare misure tecniche e organizzative per la sicurezza dei dati e di integrare le misure di salvaguardia nelle attività di elaborazione dati al fine di proteggere i diritti delle persone, dalla progettazione e per impostazione predefinita. Inoltre, Illumina offre diverse funzioni per supportare i requisiti dell'HIPAA, incluse misure di salvaguardia amministrativa, fisica e tecnica.

Inoltre, il software Emedgene è conforme ai requisiti globali e locali sulla privacy dei dati, tra cui la norma ISO 27001, ed è conforme ai controlli dell'organizzazione di servizi (SOC, service organization control) 1 e 2 di tipo II. Le opzioni di sicurezza avanzate personalizzate includono l'integrazione con la politica di single sign-on dell'organizzazione. È disponibile anche un'opzione per utilizzare la propria chiave e crittografare doppiamente i dati.

Riepilogo

Il software Emedgene offre una soluzione basata sulla XAI e abilitata all'automazione per l'analisi terziaria di malattie rare e dati di ricerca genetica. Aumenta l'efficienza dell'interpretazione delle varianti e consente ai laboratori di scalare le proprie attività. Grazie all'integrazione con le soluzioni di sequenziamento, microarray e software Illumina, compresa l'analisi secondaria DRAGEN per un'identificazione di varianti accurata, completa ed efficiente, il software Emedgene fornisce ai laboratori di ricerca uno strumento potente ed efficace che garantisce una crescita continua.

Maggiori informazioni

[Software Emedgene](#)

Informazioni per gli ordini

Prodotto	N. di catalogo
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
Eventuali articoli aggiuntivi necessari per il laboratorio saranno inclusi nel preventivo redatto.	

Affinché i clienti possano effettuare richieste informate, Illumina propone un'esperienza di valutazione assistita, che consente di lavorare con set di dati di ricerca di esempio disponibili nel software Emedgene o di caricare e valutare i propri set di dati all'interno del software. Per maggiori informazioni, rivolgersi a un rappresentante Illumina.

Bibliografia

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.



Numero verde 1.800.809.4566 (U.S.A.) | Tel. +1.858.202.4566
techsupport@illumina.com | www.illumina.com

© 2024 Illumina, Inc. Tutti i diritti riservati. Tutti i marchi di fabbrica sono di proprietà di Illumina, Inc. o dei rispettivi proprietari.
Per informazioni specifiche sui marchi di fabbrica, visitare la pagina web www.illumina.com/company/legal.html.
M-GL-01057 ITA v2.0