Nostos Genomics

AION | AI-DRIVEN VARIANT INTERPRETATION PLATFORM

Automated. Interpretable. Accurate

EXECUTIVE SUMMARY

Over **300 million people** around the world are affected by a rare genetic disease. Patients with rare genetic diseases often undergo multiple visits, referrals and batteries of tests to reach a diagnosis, a process known as the "diagnostic odyssey". On average, **patients are diagnosed more than 5 years after the onset** of the first symptoms, leading to poor medical management and high costs for healthcare systems. Genetic tests, such as exome sequencing, can decrease the duration and cost of this diagnostic journey by identifying mutations in the patient's DNA that are causing their disease.

Analyzing a genetic test to find disease-causing mutations in the DNA is a crucial step in genetic testing, known as variant interpretation. Currently it is often a manual process carried out by a specialized variant scientist. While it is important that an expert verifies every case, the current tools used for the process make it time-consuming, costly and prone to variability in the results. Patients must often wait several months to receive their test results. Furthermore, **less than 50%** of patients with rare genetic diseases **receive a clear diagnosis.**

With **AION** we have created a platform that **speeds up the step of variant interpretation** for genetic tests to decrease the costs, time and other limitations associated with this process. For this, we have created a machine learning algorithm for variant interpretation that can analyze variants with high **accuracy** while offering insights into its reasoning process (**white-box approach**), allowing human experts to understand the process through which the algorithm has made a prediction.

Our **performance validation** shows that AION can increase **diagnostic yield by more than 50%** compared against rule-based classification using ACMG guidelines. It identified the patient's disease-causing variant in 96% of cases. Additionally, exomes can be **analysed in less than 2 minutes** leading to a quick turnaround time and opportunity to **increase throughput**. Enabling the analysis of genetic mutations in minimal time, lowering costs and turnaround time for genetic tests is essential to ensure that more patients and health systems can **benefit from genetic diagnostics.**

5000
cases already used for validation

2min
to analyse a whole exome

>50%
increase in diagnostic yield

CURRENT CHALLENGES

Although **~8% of the population has a rare genetic disease**, patients with these disorders face challenges to reach a diagnosis. **Next-generation sequencing** (NGS) has revolutionized disease diagnosis and treatment approaching the point of providing a personal genome sequence for every patient. Typically, exome sequencing of a patient identifies tens of thousands of non-reference coding variants in their DNA, but only one or very few are expected to be significant for the relevant disorder. During an initial filtering stage, usually different methods are employed to shorten the list of variants. However, narrowing down further towards culprit pathogenic variants usually entails a laborious and manual process of seeking gene-phenotype relationships by consulting numerous separate molecular and clinical databases. This process, known as **variant interpretation**, is critical to the success of a genetic test but is constrained by costs and the amount of time that can be spent on this manual step.

Furthermore, due to limited information available to classify genetic variants as disease-causing or benign, variants are frequently reported back as **variants of uncertain significance (VUS)**, with unclear consequences for the patient. As a result, less than 50% of patients undergoing genetic testing receive a clear molecular and clinical diagnosis.

Partial automation of this step would enable exhaustive analysis while decreasing time and costs. For a reliable automated variant interpretation in a clinical setting, the results must be verifiable by a human expert. White-box models allow insight into how input data contributes to output predictions, making them well-suited for clinical use, where transparency and explainability are critical.

OUR SOLUTION

We developed a new variant interpretation tool **AION** that enhances the interpretation of genetic variants for clinical genetic testing by leveraging **functional genomics and artificial intelligence**. Our white-box algorithm automates variant interpretation which performs exhaustive and explainable variant interpretation for diagnostic genetic tests.

"AION combines a classifier and prioritizer and provides laboratories with fast, accurate and interpretable decision support for variant interpretation"

Our unique technology identifies the **consequences of genetic variants on mRNA levels, splicing and protein function** to inform clinical interpretation.

Key features:

- Automated support for the interpretation of candidate disease-causing variants and variants of unknown significance
- Variant annotation with 100+ features reflecting *in silico*, *in vitro* and *in vivo* observations including computational scores (e.g. evolutionary conservation, variant effect predictions) as well as cellular, animal and human data
- Classification of variants according to guidelines from ACMG, supported by expert knowledge and artificial intelligence
- Automated variant prioritization for each case, based on phenotype and variants identified. Our approach offers a confidence score for each prediction
- Support for analysis of gene panels, whole exome sequencing or whole-genome sequencing from individual samples or trios.

AION WORKFLOW:

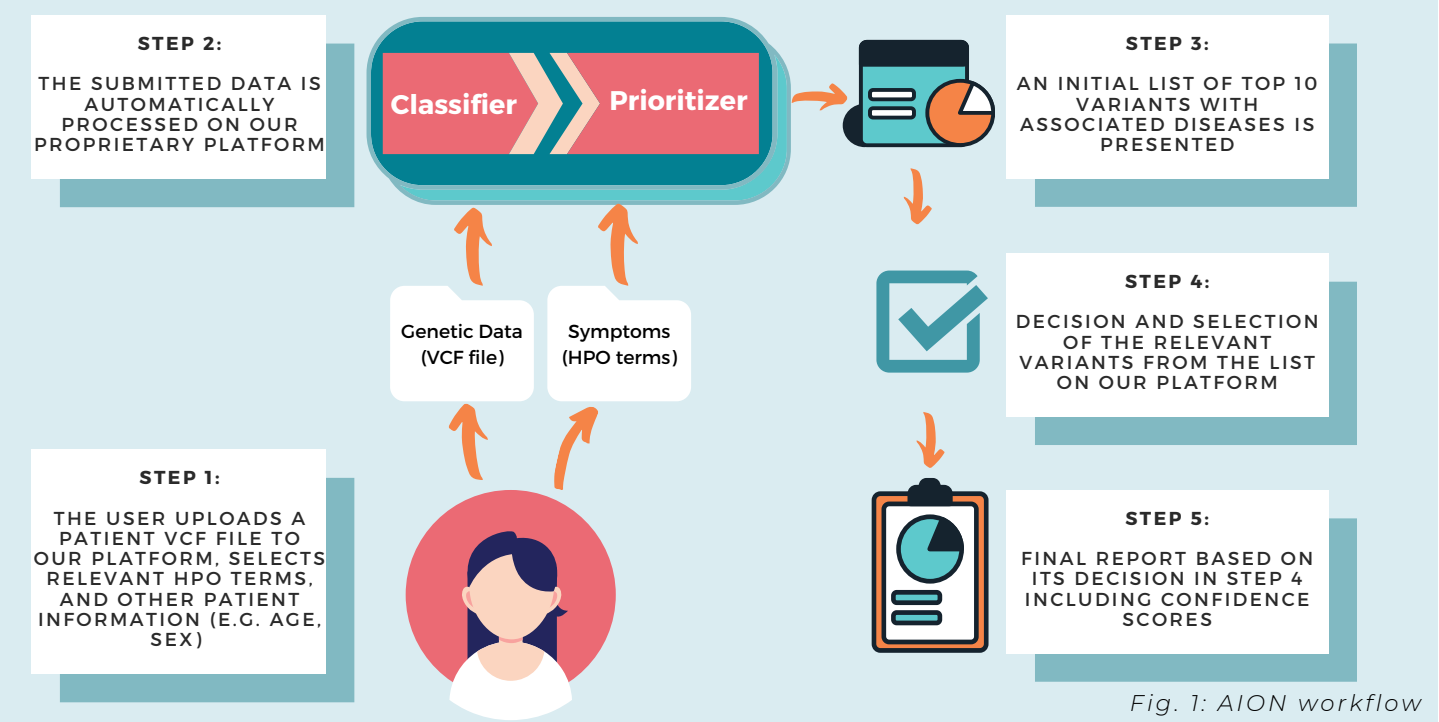


Fig. 1: AION workflow

Variant annotation

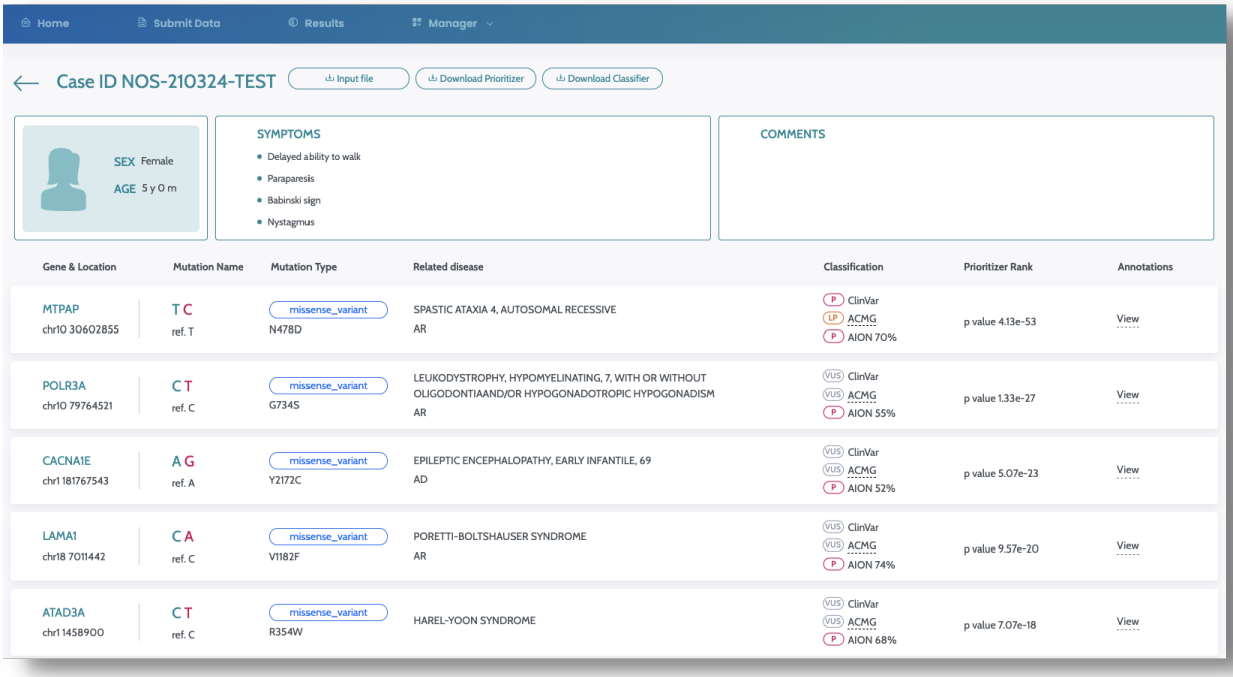
Initially, we annotate the provided VCF file with our in-house annotation pipeline, which includes data from over **25 trusted reference sources of public data**, in addition to licensed and proprietary data. Data sources include ENSEMBL, Uniprot, HGNC, UCSC, gnomAD, ClinVar, ClinGen, OMIM, Orphanet, Monarch, Human Phenotype Ontology and expert-curated data sets from published literature.

Classifier

The annotated VCF file is then used as input for our machine learning algorithm (ML) which **classifies all variants as either benign or pathogeni**c, assigning to each variant a confidence score quantifying the evidence supporting each prediction. Additionally, the interpretation of each genetic variant is supported by predictions from our ML algorithm indicating whether a genetic variant is in an **evolutionarily conserved region** of the genome, in a region **under constraint to variation in humans** and whether the variant is predicted to **cause alterations in mRNA or protein**. Each category is accompanied by a **confidence score**, indicating the evidence available for each prediction.

Prioritizer

The list of annotated variants obtained from the classifier is then entered in the prioritizer along with the descriptors of the patient's **clinical phenotype** (e.g. list of terms from the Human Phenotype Ontology). It then computes for each known disease present in databases of reference (i.e. OMIM, Orphanet, Monarch), the **probability** of observing that specific disease given the combination of the **clinical phenotype and a genetic variant**. In the final step, the prioritizer ranks the probabilities in descending order and returns a list of the **most likely diseases** associated with the patient's clinical phenotype and genetic variants.



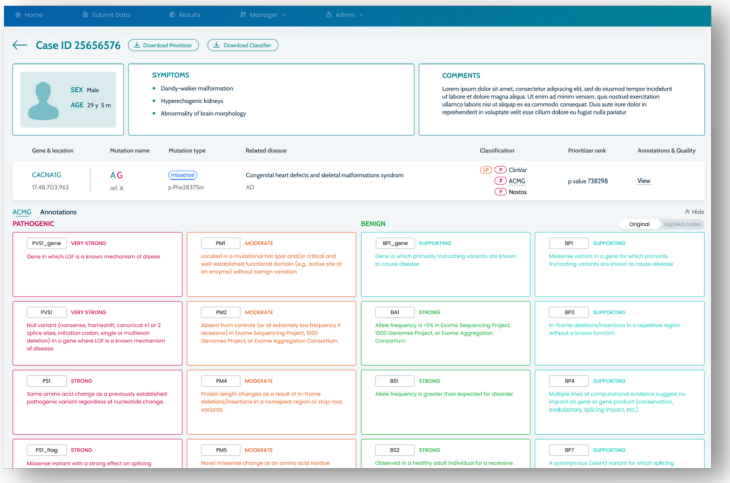
AION: Prioritized list of variants

Results

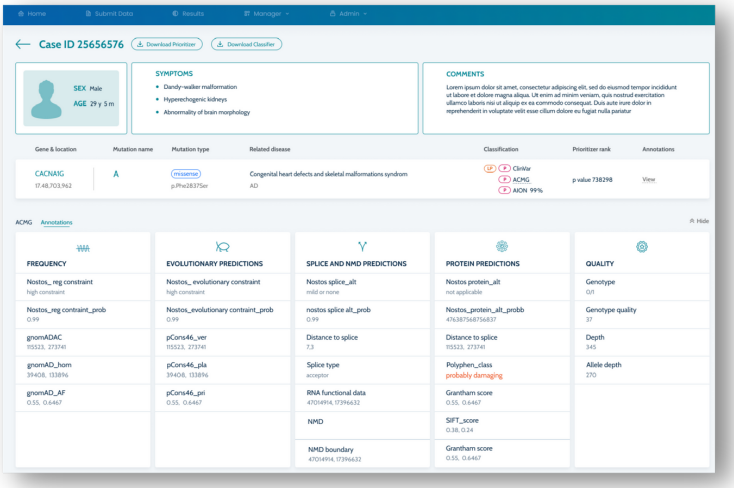
AION provides a list of **prioritized variants** including two-gold star classifications from ClinVar and automated classifications according to the **ACMG** guidelines as well as interpretations from our **AI algorithm**.

Our analysis offers a **white-box approach** to inform the clinical variant scientist of the confidence of the prediction based on the information available, as well as the **factors contributing to the classification** of a variant as pathogenic or benign. The white-box approach allows a transparent interpretation of the results, where the user can see precisely, and without the need for any decoding tools, the patterns in the data supporting the interpretation.

All the results from our proprietary pipeline (incl. the **100+ annotations**) can be easily downloaded as a CSV file, fetched via API or directly stored on a cloud service.



AION: ACMG classification of a variant



AION: Detailed overview if a genetic variant is in an evolutionarily conserved region / region under constraint in humans and if variant is predicted to cause alterations in mRNA or protein

EXAMPLE

A woman with early-onset breast cancer undergoes genetic testing to evaluate whether the cause of cancer may be genetic. A novel missense variant is identified in the *BRCA1* gene, which has never been observed in large cohorts of healthy individuals and has never been classified previously. **Genetic variant:** chr17:41197772:A>C, NM_007294:c.5515T>G, Leu1839Val

No **interpretation is available in ClinVar** and **no reference** to this variant was found in **published literature**.

ACMG classification: fulfils criteria for PM1, PM2

- PM1: Mutational hot spot and/or critical and well-established functional domain. → BRCT2 on amino acids 1756-1855 on BRCA1 protein (Uniprot: P38398)
- PM2: Variant is absent from a large general population or a control cohort (>1,000 ethnicity-matched individuals) →Variant is absent from gnomAD (>140,000 population controls)

Criteria for classification as pathogenic or benign are not met.

ACMG: The variant is classified as a variant of uncertain significance (VUS).

AION classification: inclusion of functional genomics data and ML analysis

- Data from functional genomics assay including effects on mRNA levels and protein function are available for this variant. Functional data validated with **96.7% sensitivity, 98.2% specificity** for clinical variant interpretation (Findlay et al. 2018).
- Prediction from ML model, which uses functional genomics data, in combination with other annotations, to predict that the **variant identified is pathogenic due to severe effects at the molecular level.**
- The clinical and functional interpretations of this variant are accompanied by **high confidence scores, indicating that there is strong evidence available to support these predictions.** (Low confidence scores would flag the results for further analysis by the variant scientist, to include additional information such as variant segregation in the clinical interpretation of this variant.)

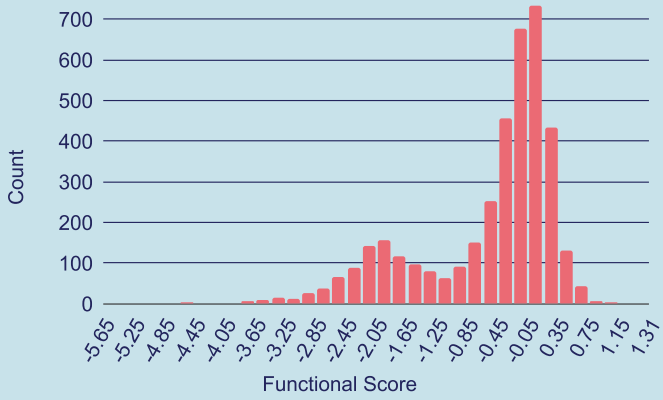


Fig. 2: Distribution of molecular variant effects for variants in the BRCA1 gene. Pathogenic variants are represented in the left peak, whereas the right peak represents benign variants. The variant identified in the patient has a functional score of -1.91, corresponding to a pathogenic variant with clearly damaging effects at the molecular level.

Leveraging functional genomics and artificial intelligence, our unique technology accurately interprets genetic variants in clinical genetic tests while elucidating potential molecular mechanisms of disease.

AION: The variant is classified as pathogenic.

PERFORMANCE & VALIDATION

The performance of AION **was tested on 5,062 exomes** from patients with monogenic disorders. These patients encompassed 17 different types of disorders, including intellectual disability, mitochondrial disorders, metabolic disorders, movement disorders, heart disease, hereditary cancer and sensory disorders among others.

Exomes for patients with a given monogenic disorder were simulated by spiking causal variants for the selected disease into a VCF file from a healthy individual sequenced as part of the 1000 genomes project. **Disease-causing variants were obtained from ClinVar** (release: 29/06/2020), by excluding all variants used as part of our training data (release: 03/10/2019) and extracting 35,027 variants classified as "Pathogenic" or "Likely pathogenic".

To benchmark the performance of our proprietary algorithms embedded in AION, we **compared** the proportion of cases **AION** was able to solve (i.e. identify the correct disease-causing variant) against a **rule-based classification of genetic variants according to the ACMG guidelines**.

Variant interpretation by AION consistently outperformed variant interpretation based on ACMG guidelines. Rule-based ACMG classification identified the patient's disease-causing variant(s) as "Pathogenic" or "Likely pathogenic" in 61% of cases (3108 out of 5062 patients). **AION's algorithms identified the patient's disease-causing variant in 96% of cases** (4845 out of 5062 patients), representing on **average >50% increase in diagnostic yield.**

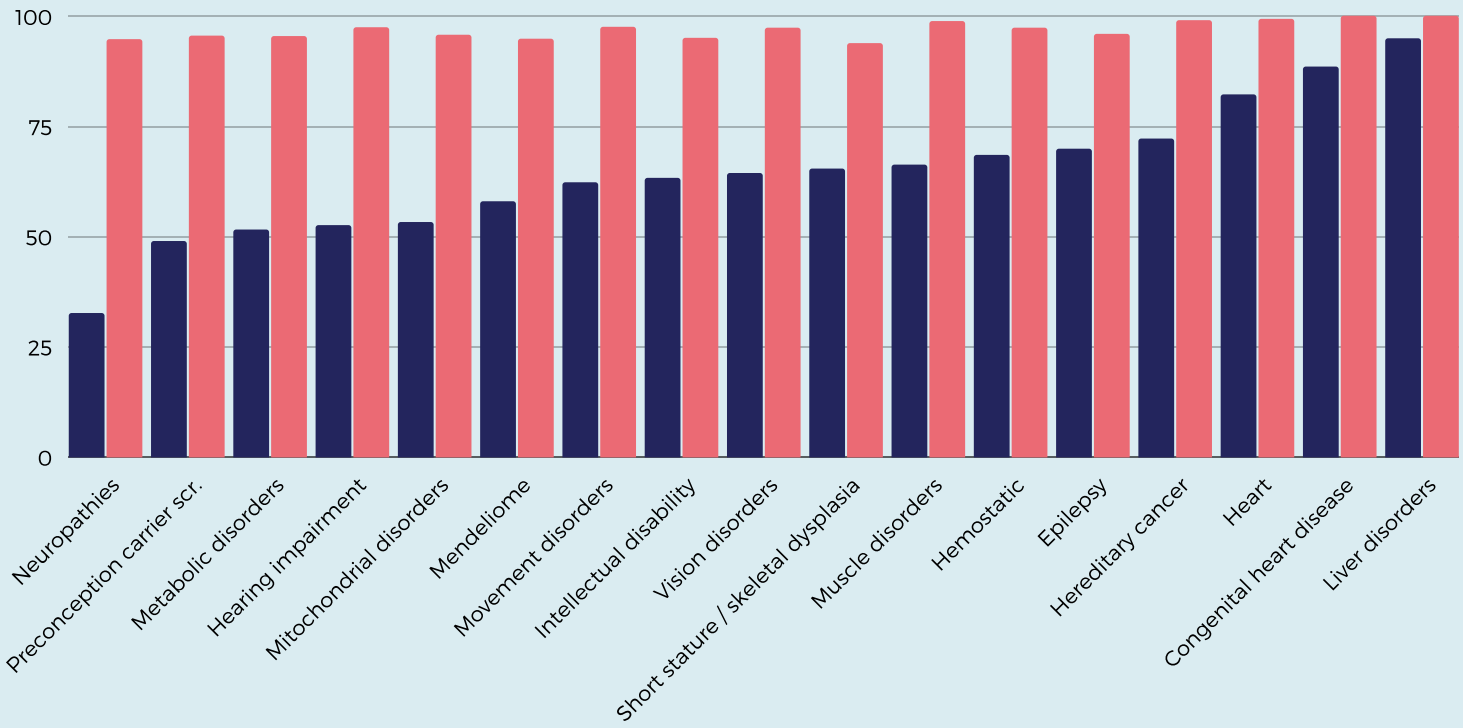


Fig. 3: ***AION*** *vs.* ***Rule-based ACMG*** *solved cases in %*

AION was able to identify the correct disease-causing variant in 95.7% of cases - compared to a rule-based variant classification according to the ACMG guidelines that only identified it correctly in 61.4% of cases.

IMPACT

For people with genetic diseases, a clear molecular and clinical diagnosis is essential to receive appropriate medical care, to make informed medical and personal decisions including family planning and access to medical and social services.

Our automated variant interpretation platform AION can deliver a decrease in time and costs for variant interpretation while increasing the diagnostic yield for genetic testing.

A recent study on the complete costs of genome sequencing in the UK has shown that the clinical interpretation stage for rare disease cases takes on average 11 hours 54 minutes of clinical scientist time per case (Schwarze et al. 2020). It takes AION only 2 minutes to provide strong decision support for a WES and less than 10 minutes for WGS. **This will allow labs to shorten the time, increase throughput and focus on where their expertise matters most.**

Additionally, our validation clearly shows that AION can increase the diagnostic yield for genetic testing. The results indicate that **AION can help identify more than twice as many disease-causing variants compared to a rule-based ACMG classification**. This will help to provide patients with clear diagnoses. At the population level, a higher diagnostic yield for genetic tests leads to earlier testing, shortening the diagnostic odyssey and lowering the costs for the health care systems.

CONCLUSION

Genomics is transforming medicine; we now know that over 4,000 rare disorders are caused by pathogenic mutations and a clear genetic diagnosis can be reached in almost half of patients undergoing genetic testing for clinical entities, which until recently were considered to be multi-factorial. **This better understanding of genetic diseases is paving the way for targeted therapies for several disorders previously considered untreatable.**

Tailoring clinical interventions to each person based on their genetics and individual health risks is the promise of precision medicine. Recent developments in sequencing technologies have revolutionized genomics and those anticipated for the coming years will consolidate genomics as an essential part of medicine. As genetic testing becomes more widespread in health care, there is an increasing need to standardize, collect and connect genomic data so it can be reanalyzed across healthcare systems and throughout a patient's life. **At Nostos Genomics, we are taking the first step towards translating genetic information into impactful clinical interventions.**

Try it out!
team@nostos-genomics.com
www.nostos-genomics.com

Nostos Genomics