Presenter
Sandro Morganella
Head of Artificial Intelligence - Congenica

Congenica A.I.
A High Scalable Machine Learning Framework for Improving the Diagnostic Yield in Rare Diseases

Background

Rare genetic conditions impact more than 400 million people globally with each individual affected by at least one of the ~7,000 recognized rare diseases. Despite the technological advance in DNA sequencing, the characterization of the one or two actionable variants amongst the hundreds of thousands of mutations in an individual's genome remains a very complex task and hence the identification of an accurate diagnosis for the patient can take several years. One of the main challenges for the clinicians is the lack of enough evidence for delivering a clear classification resulting in variants of uncertain significance (VUS).

Methods

Here, we present an innovative high scalable machine learning framework that can accurately predict the pathogenicity of the variants, reclassify VUS, and identify the diagnostic variants by using an ad-hoc computational framework that combines patient phenotypes with the information contained in the Online Mendelian Inheritance in Man (OMIM) database. Our model is trained using a unique and well curated set of data generated by expert users of the Congenica platform. The training data includes ~35,000 variants observed in more than 10,000 unique patients and spanning more than 50 rare disease phenotypes.

Results

We validated our pathogenicity model on ClinVar 3 and 4 stars variants reviewed by expert panels showing a compelling 95.5% accuracy in predicting the correct pathogenicity. In addition, we showed that we were able to correctly reclassify 92.8% of variants that were previously considered VUS by our clinical team. Finally, we retrospectively assessed our ability to detect the diagnostic variants in 73 singletons, showing that we returned the causative variant as top candidate in 63% of cases and in the top 10 in 85% of cases. We have also analysed 126 singletons obtained from two external research institutes showing that in 48% of cases the diagnostic variant was at the top of the list and in 77% of cases in the top 10.

Conclusions

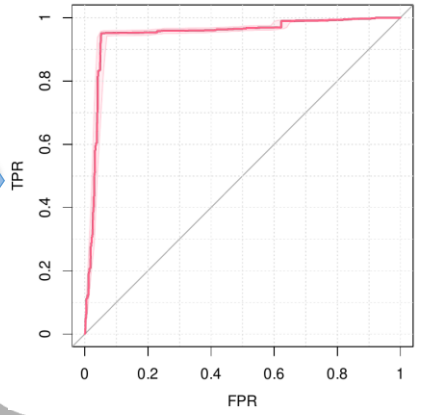
The presented results strongly suggest that by using our machine learning-based decision support framework into clinical settings would help clinicians in increasing the diagnostic yield and improve the diagnosis and treatment even for the most complex cases.

Our Unique Training Set

- **>35,000 unique variants**
- **>10,000 unique patients**
- **>50 rare disease phenotypes**

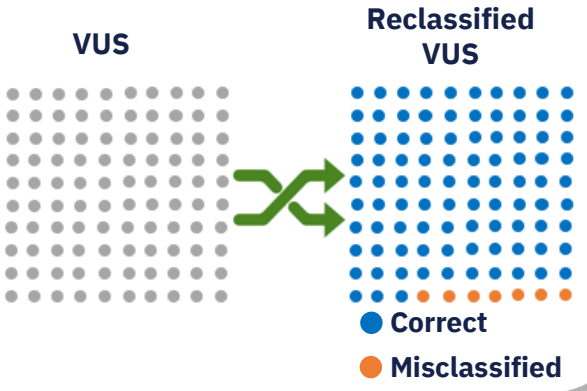
Congenica AI

95.5% Accuracy in Predicting Variant Pathogenicity



Results obtained on 8,449 ClinVar 3- and 4-stars variants reviewed by expert panels

92.8% of VUS Correctly Reclassified



Reducing the Interpretation Time via Precise Variants Prioritization