Congenica

Enabling genomic medicine

Rare Disease Artificial Intelligence validation program

August 2020

Contents

Rare Disease Artificial Intelligence validation program

Congenica A.I.

Congenica has developed novel, explainable artificial intelligence (A.I.) to accelerate clinical scientists in determining causal pathogenic variants for rare disease cases.

Congenica's patent-pending A.I. provides prediction of casual pathogenicity of a patient's genomic variants (SNVs) based on user-inputted phenotype/s. Unlike other "black box" systems, Congenica A.I. explains how dozens of factors were weighed and contributed to the decision, offering transparency for users.

SNVs previously classified as "VUS" can be given to the A.I. to suggest an explained reclassification as benign, likely benign, likely pathogenic or pathogenic.

Congenica A.I. is trained to infer causal pathogenicity by using a large and unique, real-world, expert-generated rare disease interpretations available to Congenica and curated from a range of industry-leading data sources. It breaks down information silos and closes information gaps while protecting patient privacy to streamline clinical interpretation and diagnostic decisions. Congenica's prediction model is monthly updated to include new available knowledge and decisions.

About the program

Our free of charge Validation Program assists partners with re-analyzing diagnosed and undiagnosed cases while validating the effectiveness of Congenica A.I. and identifying opportunities for improvement.

We're interested in both solved and unsolved cases from each partner. For solved cases, we'll work with partners to validate the efficacy of Congenica A.I. for their diagnostic setting. For unsolved cases, we'll be able to provide predicted casual variants identified by our system, providing evidence-based recommendations that enable partners to focus their interpretation efforts and potentially increase the yield of their interpretation process.

Validation Program partners will have access to the benefits from a novel A.I. technology ahead of open market release, assess/confirm the accuracy of Congenica A.I. on cases that have been already solved, receive free predictions of unsolved cases that could help in increasing the interpretation yield. The validation program will see a small number of selected organizations collaborate with Congenica to test the beta version of Congenica A.I. during the summer of 2020.

Program objectives

The dual goals for our Validation Program are to both generate new insights into potentially causal variant/s for collaborators to validate Congenica A.I. on solved cases and to provide evidence to assist partners' interpretation of unsolved cases.

Process

Data access for existing partners & cases

Existing partners (Congenica customers) will give permission for Congenica to use cases that they've already uploaded to Congenica. Depending on the type of cases provided, we'll be able to provide different outputs to partners. The members of our Validation Program shall select cohorts of both solved cases to validate Congenica A.I. for their laboratories & unsolved cases to highlight variants prioritized by the A.I.

- **solved cases** to validate performance: either WES or WGS with the specific patient's phenotypes where one or more pathogenic variants that have been associated with the patient's phenotypes. Note that only cases uploaded after 1st June 2020 can be used. This is a requirement to assure that the prediction is not performed on cases included in the training set.

- **unsolved cases**, either WES or WGS either WES or WGS with the specific patient's phenotypes where no causal variants were reported.

Data access for new partners & cases

New partners not currently using Congenica software will share with Congenica pre-permissioned VCFs containing. Depending on the type of cases provided, we'll be able to provide different outputs to partners. The members of our Validation Program shall select cohorts of both solved cases to validate Congenica A.I. for their laboratories & unsolved cases to highlight variants prioritized by the A.I.

- **solved cases** to validate performance: either WES or WGS with the specific patient's phenotypes where one or more pathogenic variants that have been associated with the patient's phenotypes.

- **unsolved cases,** either WES or WGS either WES or WGS with the specific patient's phenotypes where no causal variants were reported.

Prediction generation

The VCF files supplied will be shown to Congenica A.I., within our secure Amazon Web Services (AWS) cloud environment. Congenica A.I. will then generate predictions of pathogenicity for every variant presented, with their contribution to phenotype, and it will select only a small set of these variants (the number depends on the specific submitted case).

The following data/information are necessary to carry out the prediction:

- Genomic coordinates on GRCh37 (i.e., chromosome, start, ref_allele, alt_allele) ideally in VCF format but csv/tsv are also accepted
- ACMG classification (i.e., Pathogenic, Likely pathogenic, Uncertain Significance, Likely benign, Benign) where available
- Transcript used to make the interpretation (RefSeq/Ensembl) where available
- Phenotypes defined as HPO terms
- For solved cases the contribution to phenotype for pathogenic variants (i.e., Full, Partial, None)
- Zygosity of variant (heterozygous/homozygous)
- Results will be available for review by a secure web-portal that will be accessible via provided credentials. In addition, the list containing only the causative pathogenic variants can be provided if requested.

Partner review
On receipt of the results, the Partner will have the opportunity to verify the efficacy of Congenica A.I. for solved cases and further investigate any likely pathogenic variants for the unsolved cases. Congenica will then solicit feedback on the predictions, including:

- Whether and how the information provided by Congenica A.I. can enable more efficient interpretation
- Whether the information provided is sufficient to solve a case
- For solved cases, potential time-saving due to this information
- For unsolved cases, whether variants identified by Congenica A.I. led to new diagnoses
- any other feedback and recommendations for future product improvements.

Deliverables

Congenica will return to the partner their predictions of each variant via a secure password protected web-portal. In addition, the list containing only the causative pathogenic variants can be provided if requested.

A prediction will include:

- Predicted probability for each pathogenic level (i.e., Pathogenic, Likely pathogenic, Uncertain Significance, Likely benign, Benign)
- The predicted pathogenic level which is associated to the most likely decision (i.e., Pathogenic, Likely pathogenic, Uncertain Significance, Likely benign, Benign)
- Predicted level of contribution to phenotype (i.e., Partial, Full, None)
- Overall confidence score that will enable the ranking of the predicted variants

Joint Responsibilities

Participation in this trial program allows partners one-off access to our A.I. variant predictions for your research purposes. In return, Congenica will publish a qualitative and quantitative description of the validation exercises, including for example the number of additional cases resolved, or the time saved in analyzing hard to solve cases.

After receiving the Partner's feedback all sensitive information/data not contained in the Congenica database will be permanently deleted.

Upon the completion of this Data Analysis, Congenica & Discovery will review results together and identify appropriate next steps, which can include additional technical, commercial, and/or marketing agreements and activities.

Data Governance

Congenica software is GDPR, HIPAA, and ISO 27001 compliant. Partners will permission Congenica access to their genomic data in VCF format for the purposes of this program. Data will be transferred securely, either via Congenica sFTP or partner's protocol with specifics determined during sign up process.

Upon completion of the program, Congenica will permanently delete all genomic data files that had been provided from the partner. Congenica reserves the right to preserve partner's feedback and other communications related to this program and to publish in marketing literature anonymized stories of successful validation exercises, for example the number of new cases resolved, or the time saved in analyzing hard to solve cases.

Contact information

For further information please contact us using the details below.

Congenica Support
support@congenica.com

Program lead – Sandro Morganella, Senior Machine Learning Engineer
sandro.morganella@congenica.com