Clinical Variant Modeling

Invitae Clinical Variant Modeling harnesses genotype and phenotype data from over 4 million patients to predict the pathogenicity of genetic variants with high accuracy

Background

The goal of clinical genetic testing is to assess risk for or to confirm diagnosis of some hereditary diseases. In the course of genetic testing, distinguishing pathogenic DNA variants from benign ones is a significant challenge. Clinical genetic testing labs continue to encounter many novel rare variants as more individuals undergo genetic testing (Invitae, Data on file), further exacerbating the challenge. The majority of these variants have limited public data associated with them and end up classified as variants of uncertain significance (VUS).

Clinical data are among the most powerful forms of evidence for distinguishing pathogenic from benign variants[1-3]. For example, these data may include observation of a variant in patients with a clearly defined disease, variant segregation in affected individuals, or observation of *de novo* occurrence of a variant. However, there are often challenges with incorporating clinical data into variant classification. For instance, a complete relevant medical history is not always provided at the time of genetic testing, which makes it challenging to differentiate between affected individuals with missing data and those who are unaffected. Second, many hereditary diseases include symptoms that can be associated with common sporadic diseases, such as cancer and cardiovascular diseases, limiting our ability to identify a molecular cause and establish a genetic etiology. Finally, not all genetic diseases exhibit complete penetrance, which makes it difficult to determine when a variant is not associated with hereditary disease, and therefore benign. These challenges are particularly acute when reviewing clinical data on a case-by-case basis for variant classification, but with increasing access to large sets of clinical health information and genetic testing results we believe they can be overcome.

Our laboratory is well-positioned to leverage clinical data efficiently at scale to improve clinical variant classification. By the end of 2023, we had accumulated genotype data and clinical data for over 4 million individuals who were of diverse racial and ethnic backgrounds and referred to Invitae® for a wide range of clinical genetic testing. This dataset is massive and clinically diverse: it includes over 100 million words of clinical descriptions (e.g., personal and family history, indications for testing) submitted for the tested patients, as well as over 2 million unique variants observed across more than 3,900 genes.

What are clinical variant models?

To maximize the utility of large clinical datasets for improving variant classification and reducing VUS, Invitae has developed an approach called Clinical Variant Modeling. It is essentially a method for learning patterns from clinical data available for millions of patients and precisely applying this as evidence for variant classification using a Bayesian approach. Importantly, this approach takes into account the penetrance of disease, age at testing, potential phenocopies and missing data on the test requisition forms (eFigure 1).

Clinical Variant Modeling is composed of two distinct, but connected, sequential machine learning (ML) steps. The first step involves estimating the probability that a patient tested at Invitae is affected with a specific genetic condition. This probability is termed a **Patient Score**, which incorporates clinical and demographic information from ordering providers, including reported signs and symptoms, ICD-10 codes, age at testing and family history. This score is estimated by comparing and distinguishing the clinical profile of patients with a positive molecular diagnosis from those with a negative molecular diagnosis. In the second step of clinical variant modeling, a Bayesian inference model learns the distribution of Patient Scores that could be associated with benign or pathogenic variants. The inferred probability that a variant is pathogenic is termed the **Variant Score**.

To ensure that only the best performing clinical variant models are used for variant classification, we calculated the area under the receiver operating characteristics curve (AUROC) for each model to measure the model's performance at distinguishing between benign and pathogenic variants. Only models with an AUROC≥0.8 were selected for further evaluation. Continued refinement of the set is accomplished by a combination of further validation metrics and expert review. Additional steps were done to establish the weighting of the Variant Scores for incorporation as evidence into Sherloc[4], Invitae's variant classification framework. Ultimately, the weight (points) used for variant classification was proportional to the predictive value of the Variant Score. Importantly, clinical variant models (CVMs) learn from clinical data obtained during the course of genetic testing and can be applied to new patients and variants observed by our lab. CVMs are better at making predictions for individuals in our cohort since they understand data patterns and potential biases in the sampling and allow for generalizability of the predictions to our specific patient population. For more detailed methodology, see the online appendix.

Impact of Clinical Variant Modeling on patients

As of March 2024, CVMs have been developed and validated for 11 genetic conditions associated with 17 genes in which they demonstrate high performance in distinguishing known benign from known pathogenic variants (≥0.8 AUROC curve; Table 1, details in the online appendix). Predictions from these CVMs have been used as evidence to resolve over 1,000 unique VUS, impacting nearly 45,000 individuals. While >99% of these reclassifications corresponded to downgrades of VUS to benign or likely benign, the <1% upgrades to pathogenic or likely pathogenic impacted 160 individuals. Of note, 91% (10/11) of variant upgrades were in genes associated with conditions that have established guidelines for screening and treatment[5,6], highlighting the potential to change an individual's medical management and identify at-risk relatives.

Table 1: List of genes with clinical variant models and the estimated impact of these models at initial launch in March 2024. The estimated patient impact analysis was performed in December 2023 and the actual impact may be higher.

CVM condition model	Gene(s) included	AUROC	Estimated number of unique VUS>LB/B downgrades due to clinical variant models (number of patients impacted)	Estimated number of unique VUS>LP/P upgrades due to clinical variant models (number of patients impacted)	Estimated number of patient reports with VUS reclassific actions	Estimated VUS reduction in patients due to clinical variant models
Lynch syndrome	*MLH1* *MSH2* *MSH6* *PMS2* *EPCAM*	0.88	266 (17,260)	3 (54)	17,945	24%
CASR-related conditions	*CASR*	0.92	45 (2,957)	7 (106)	3,063	45%
DMD-related conditions	*DMD*	0.97	24 (471)	0	486	9%
Hereditary multiple osteochondro-matosis	*EXT1* *EXT2*	0.98	24 (446)	0	446	37%
MEN1-related conditions	*MEN1*	0.96	31 (1,502)	0	1,502	32%
NF1-related conditions	*NF1*	0.96	331 (12,094)	0	12,133	49%
NF2-related conditions	*NF2*	0.99	30 (1,441)	0	1,441	41%
RB1-related conditions	*RB1*	0.98	31 (589)	0	589	19%
SMAD4-related conditions	*SMAD4*	0.94	20 (1,271)	0	1,271	36%
STK11-related conditions	*STK11*	0.97	41 (3,707)	0	3,707	47%
Tuberous sclerosis complex	*TSC1* *TSC2*	0.97	153 (3,070)	0	3,070	16%
Estimated total number			996 (44,808)	10 (160)	44,968	

Clinical variant model case examples

Example 1:
How CVM resolved an *MLH1* VUS to benign

MLH1
c.1633A>G (p.Thr545Ala)

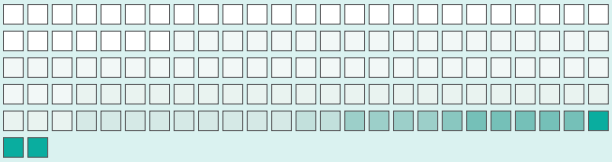
Initial evidence

- Rare in gnomAD
- Reported in colorectal, breast, or ovarian cancer patients
- *In silico* predictors: inconclusive
- RNA analysis: no significant impact on splicing

Invitae initial classification: VUS
ClinVar: VUS (14 submitters)

Clinical variant model evidence

Patient Scores of individuals with *MLH1* c.1633A>G



Patient count: **127** | Mean Patient Score: **0.31**

Contextualizing variant Patient Score distribution

962 benign *MLH1* variants	**693 pathogenic *MLH1* variants**
Patient count: **1,350,215** [Genotype-negative patients]	Patient count: **3,170** [Patients with a molecular diagnosis]
Mean Patient Score: **0.32**	Mean Patient Score: **0.77**

***MLH1* c.1633A>G**
CVM Variant Score: <0.0002
Predicted benign (NPV>0.95)

Invitae final classification: Likely benign

Example 2:
How CVM resolved a *CASR* VUS to pathogenic

CASR
c.494T>G (p.Val165Gly)

Initial evidence

- Five individuals across two families with suspected hypercalcemia and/or hyperparathyroidism; serum calcium measurement data available for one individual only
- Multiple *in silico* predictors: deleterious

Invitae initial classification: VUS
ClinVar: VUS (3 submitters)

Clinical variant model evidence

Patient Scores of individuals with *CASR* c.494T>G

Patient count: **5** | Mean Patient Score: **0.99**

Contextualizing variant Patient Score distribution

706 benign *CASR* variants	**144 pathogenic *CASR* variants**
Patient count: **605,966** [Genotype-negative patients]	Patient count: **546** [Patients with a molecular diagnosis]
Mean Patient Score: **0.22**	Mean Patient Score: **0.82**

***CASR* c.494T>G**
CVM Variant Score: >0.990
Predicted pathogenic (PPV>0.99)

Invitae final classification: Pathogenic

Both the Patient Score and the Variant Score are on a scale from 0 to 1. The higher the Patient Score (represented as darker-shaded boxes), the higher the probability that a patient is affected with the condition based on clinical information alone. The higher the Variant Score, the higher the probability that a variant is pathogenic.

Summary

A major goal at Invitae is to develop and deploy new expert-informed and scalable methods to improve variant classification and reduce uncertainty in genetic testing. As genetic testing is increasingly adopted into healthcare for disease diagnosis and management, the clinical genomics field is increasingly encountering novel rare variants. At the same time, the expansion of genetic testing leads to an ever-growing wealth of data. Leveraging diverse genotype and clinical data from over 4 million patients tested at our laboratory to reduce variants of uncertain significance, we developed Clinical Variant Modeling, a highly accurate Bayesian approach for incorporating clinical evidence at scale. The models described here are only the beginning. We will use this novel approach to improve variant classification for many more genes and diseases tested at Invitae in the future, harnessing information at scale as our real-world datasets continue to grow.

For more information, refer to Clinical Variant Modeling Appendix on page 6

References:

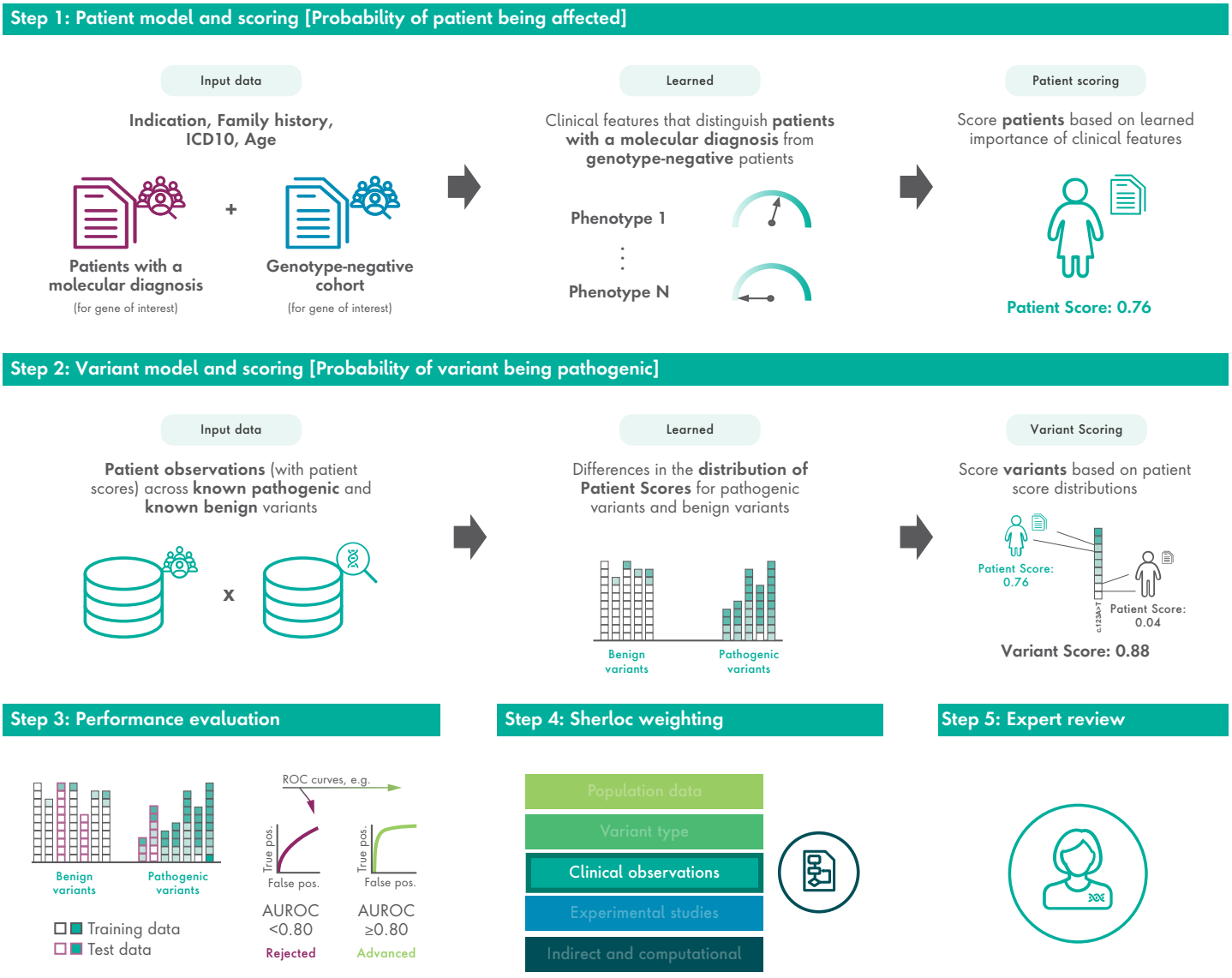
1. Streeten EA, *et al.* KCNQ1 and Long QT Syndrome in 1/45 Amish: The Road From Identification to Implementation of Culturally Appropriate Precision Medicine. *Circ Genom Precis Med.* 2020;13(6):e003133.
2. Salinas V, *et al.* The odyssey of complex neurogenetic disorders: From undetermined to positive. *Am J Med Genet C Semin Med Genet.* 2020;184(4):876–884.
3. Chen E, *et al.* Rates and Classification of Variants of Uncertain Significance in Hereditary Disease Genetic Testing. *JAMA Netw Open.* 2023;6(10):e2339571.
4. Nykamp K, *et al.* Sherloc: a comprehensive refinement of the ACMG–AMP variant classification criteria. *Genet Med.* 2017;19(10):1105–1117.
5. Kastrinos F, *et al.* Gene-Specific Variation in Colorectal Cancer Surveillance Strategies for Lynch Syndrome. *Gastroenterology.* 2021;161(2):453–462.e15.
6. Tőke J, *et al.* Rare diseases caused by abnormal calcium sensing and signalling. *Endocrine.* 2021;71(3):611–617.

APPENDIX

Clinical Variant Modeling

Methods overview:

The development and application of clinical variant models for each gene follow several general steps: **(1)** First, a Patient Score is generated to represent the probability that a patient is affected with the molecular condition of interest. This Patient Score is used for the next step. **(2)** Second, a Variant Score is calculated to represent the probability that a variant is pathogenic based on the distribution of Patient Scores for that variant. **(3)** Next, the performance of CVM is determined by using a holdout set of known phenotype-genotype relationship data points. **(4)** Models that perform well are then calibrated by measuring positive and negative predictive values (PPV and NPV) from the previous step and then integrated with appropriate weight into Sherloc's variant classification framework. **(5)** A subset of the variant classifications is reviewed by a panel of clinical genomic experts to ensure CVMs are performing as expected. Each one of these general steps is explained in further detail below.



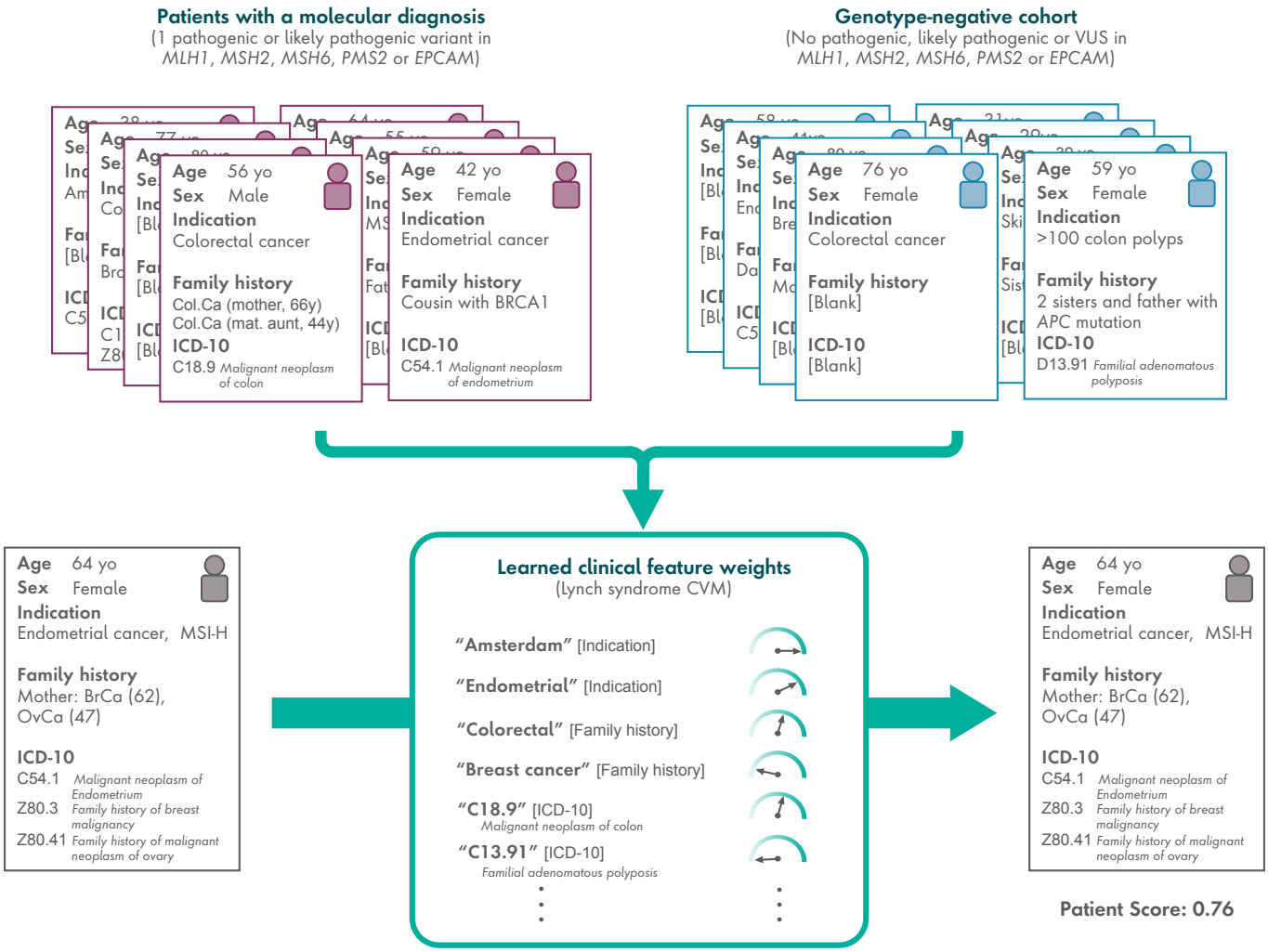
eFigure 1. Methods overview of CVMs. See details in the text below.

Step 1: Patient Score generation

CVMs leverage clinical data to predict the pathogenicity of variants for a given genetic condition in a stepwise manner. By leveraging details found in the clinician-reported data from the test requisition form (e.g., personal health history; family health history; age; sex; patient's race, ethnicity, and ancestry; ICD-10 codes; and clinical area of the test ordered), we first train a model to learn the clinical picture that distinguishes patients with a positive molecular diagnosis for the condition of interest from genotype-negative controls (i.e., patients without VUS, LP, or P variants in the condition of interest and who do not have a molecular diagnosis in another condition; eFigure 1, Step 1 and eTable 1). For each patient, a Patient Score is generated (scale from 0 to 1), which is the probability that a given patient is affected with the genetic condition of interest based on the clinical information alone (eFigure 1, Step 1). Based on what is learned, we apply this information to other patients who have VUS in the condition of interest, provided they do not have a current molecular diagnosis in another gene. We do this by scoring the other patients' clinical profiles on how similar they look to positive or negative cases.

CVM condition model	Gene(s) included	Genotype-negative patients (n)	Patients with a molecular diagnosis (n)	Known benign variants (labels, n)	Known pathogenic variants (labels, n)
Lynch Syndrome	*MLH1* *MSH2* *MSH6* *PMS2* *EPCAM*	1,350,215	20,521	962 1571 2126 1002 51	693 726 898 408 7
CASR-related conditions	*CASR*	605,966	546	706	144
DMD-related conditions	*DMD*	486,110	3,732	1871	766
Hereditary multiple osteochondro-matosis	*EXT1* *EXT2*	49,129	993	171 225	321 136
MEN1-related conditions	*MEN1*	1,155,756	1,046	669	358
NF1-related conditions	*NF1*	1,476,630	10,537	3070	3281
NF2-related conditions	*NF2*	10,330	294	538	129
RB1-related conditions	*RB1*	337,230	830	925	342
SMAD4-related conditions	*SMAD4*	28,191	396	569	157
STK11-related conditions	*STK11*	559,563	448	779	169
Tuberous sclerosis complex	*TSC1* *TSC2*	1,022,619	2,510	1535 4301	394 725

eTable 1. The number of genotype positive and genotype negative patients, as well as the number of known benign and pathogenic variants used for training and testing each clinical variant model.

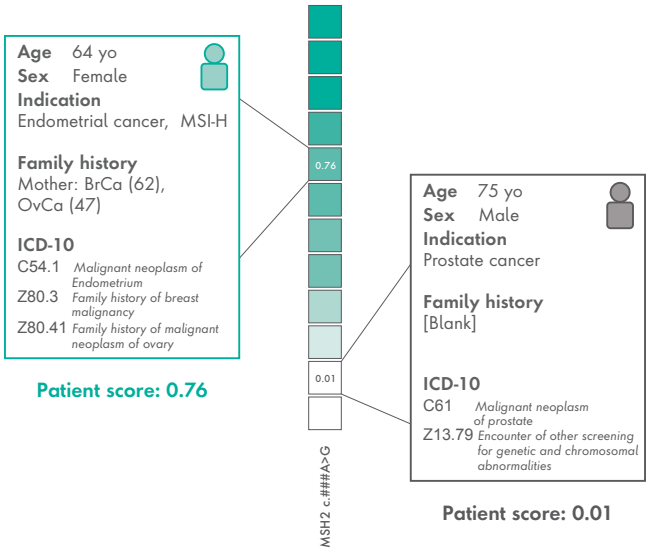


eFigure 2. Generation of the Patient Score. (A) First, for a given molecular disease, which may be defined by a single gene (e.g., Neurofibromatosis type I) or multiple genes (e.g., Lynch syndrome), clinical-related patient information is gathered for all patients with a molecular diagnosis for the condition as well as for all patients who have only benign variation in the genes of interest and no other molecular diagnosis (i.e., genotype-negative cohort). An ML model learns the appropriate evidence weights for pieces of clinical information that distinguish individuals with molecular diagnosis from genotype-negative individuals. (B) Next, the learned evidence strengths for the clinical symptoms seen in the cohorts are then applied to each individual in the entire cohort to generate a Patient Score for each individual.

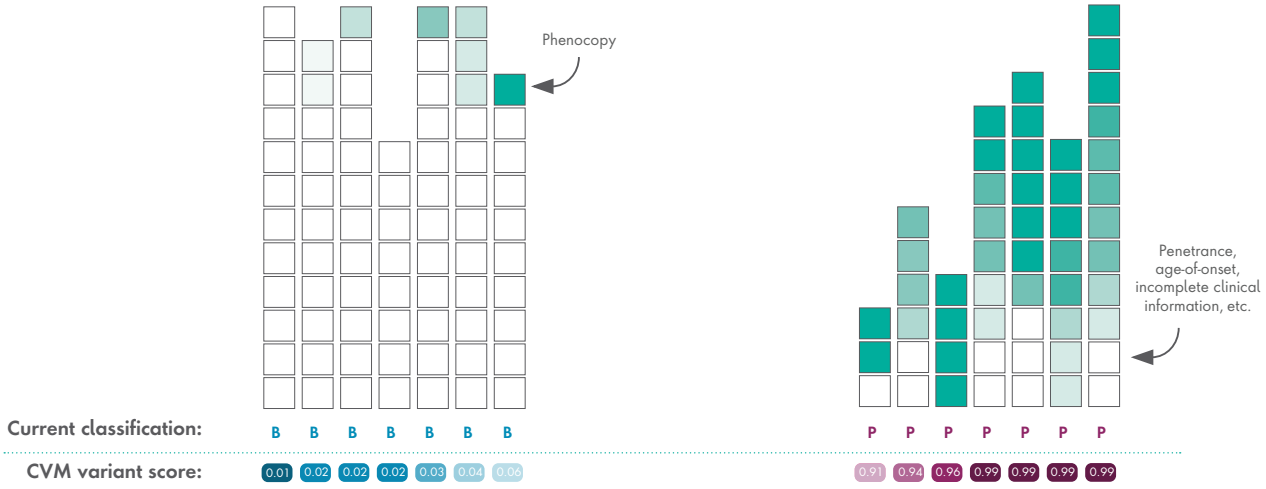
Step 2: Variant Score generation

Next, using a set of known pathogenic and benign variants for the condition, called labels (eTable 1), a second model learns the distribution of Patient Scores that are typical for pathogenic and benign variants (eFigure 3A, B). Based on what is learned, we can score VUS based on how similarly their distribution of Patient Scores look to pathogenic and benign variants. A Variant Score is generated for each variant (scale from 0 to 1), which is the probability that a given variant is pathogenic. The higher the Variant Score, the higher the probability that the variant is pathogenic (eFigure 3C).

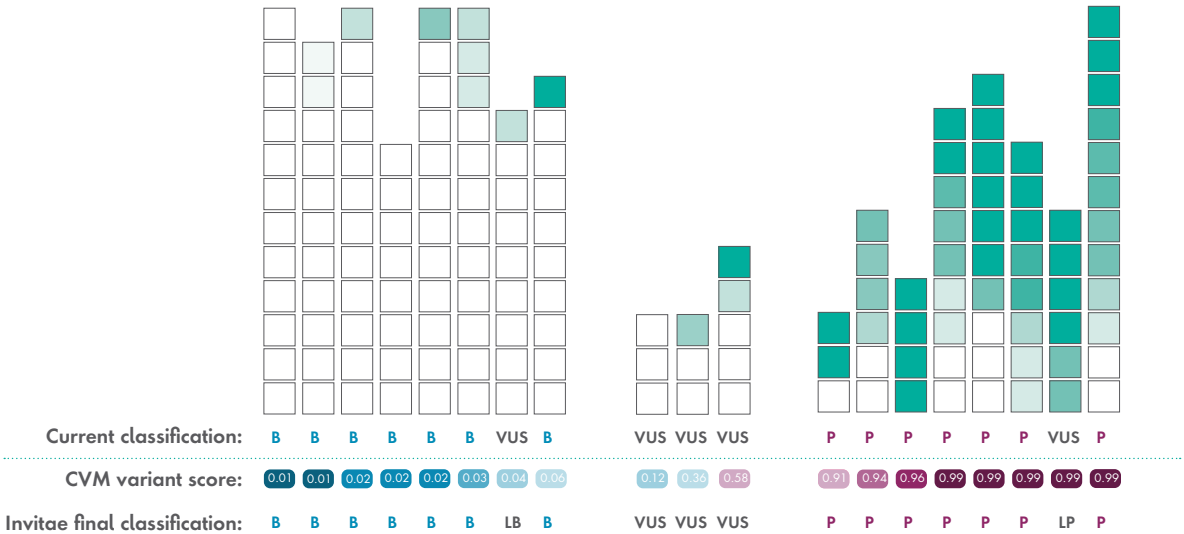
eFigure 3A.



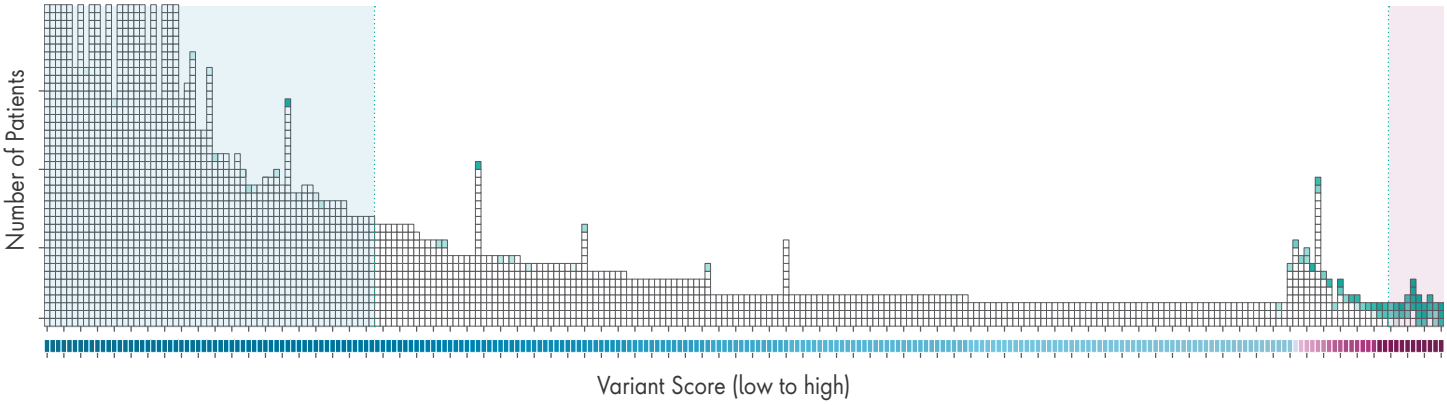
eFigure 3B.



eFigure 3C.



eFigure 3. Generation of the Variant Score. (A) Example distribution of Patient Scores for 12 patients with *MSH2* c.2210+1G>A. Each box represents a patient. Patients with low Patient Scores are shown in white and patients with high Patient Scores are shown in teal. An example patient with a high Patient Score, based on the clinical evidence, is shown on the left and an example patient with a low Patient Score, based on clinical evidence, on the right. (B) Example distribution of Patient Scores for known benign and pathogenic variants in *MSH2*. A second ML model calculates a Variant Score for each variant based on the distribution of Patient Scores. In the example, the known benign variants, which predominantly have individuals with low Patient Scores, all have low Variant Scores, while the known pathogenic variants on the right, which have more enrichment for patients with high Patient Scores, all have high Variant Scores. (C) Variant Scores are generated for VUS in *MSH2* and can be compared to the Variant Scores of known benign and known pathogenic variants.



eFigure 4. Example of CVM results for NF1. Example of CVM results for *NF1*. In this cell plot, each stack of boxes represents a single genetic variant, and each box represents a single patient. Each box is shaded according to the individual's Patient Score, the inferred probability that the patient is affected with the condition. **White** represents a low Patient Score; teal a high Patient Score. In the strip below, each box is shaded according to the Variant Score resulting from the stack of observations from that variant. **Blue** indicates a low Variant Score; **purple**, a high Variant Score. *Note that the y-axis is zoomed in to enable visualization of the patient boxes on the right side. The actual stacks of patient boxes on the left extend much higher than what is shown, as these are higher frequency variants.*

Step 3: Evaluation of CVM's performance

All CVMs are carefully screened to ensure high performance in distinguishing pathogenic and benign variants (e.g., AUROC≥0.8) by testing each model against a set of known pathogenic and benign variants that the model has not seen before (i.e., 20% holdout set; see Table 1, AUROC values for model performance). Models passing our high-performance threshold are then calibrated, followed by evaluation by our clinical genomics experts before implementation. To date, Clinical Variant Modeling has demonstrated high accuracy for over 600 genes and conditions, and the 11 most promising and impactful of these models are initially being rolled out through a careful process with our clinical genomics experts (Table 1).

Step 4: Calibration of CVMs and integration into the Sherloc framework

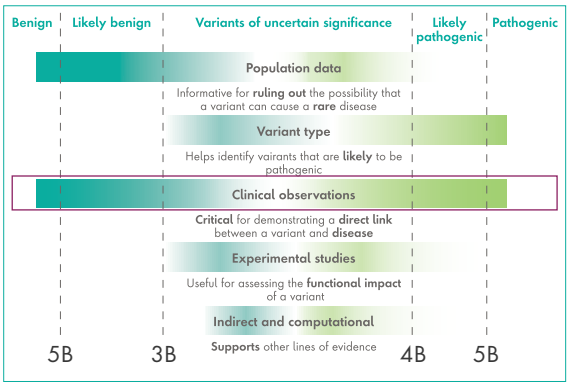
To integrate the predictions from the CVMs into the Sherloc variant classification framework[1], two tiers of predictions were established based on predictive performance thresholds, as measured in negative and positive predictive values (NPV and PPV; eFigure 5). The benign tier was defined as strong benign evidence, which is enough evidence to classify the variant as likely benign without evidence to the contrary. The pathogenic tier was insufficient to reach a likely pathogenic classification on its own but could reach the definitive classification with the addition of the variant being

absent or within the expected pathogenic range in gnomAD or another piece of pathogenic evidence. The predictive performance thresholds for those tiers were respectively defined as (1) strong benign ≥95% NPV, (2) strong pathogenic ≥99% PPV. The third and final tier corresponded to predictions that fell between a 99% PPV and below 95% NPV, which were deemed insufficiently certain to be assigned a weight within the Sherloc scoring system[1] for the first release of CVMs.

Clinical Variant Modeling

Clinical variant modeling score	Sherloc points	Accuracy
Highly predictive benign score	3 B	≥95% NPV
Highly predictive pathogenic score	3 P	≥99% PPV

Sherloc



eFigure 5. Clinical variant modeling is incorporated into the variant classification framework, Sherloc. For the initial March 2024 release, 2 tiers of predictive bins were used to integrate CVM predictions into Sherloc. Variants with CVM predictions with ≥95% NPV were awarded 3 benign points and variants with CVM predictions ≥99% PPV were awarded 3 pathogenic points. This clinical evidence is taken into account in the context of all variant classification evidence in Sherloc. Typical cutoffs for variant classifications are shown at the bottom right: 5 benign points for benign, 3 benign points for likely benign, 4 pathogenic points for likely pathogenic, and 5 pathogenic points for a pathogenic classification. Clinical genomics expert scientists can override these classification thresholds when necessary due to conflicting evidence and other factors.

Step 5: Clinical genomics expert review

To gain further confidence in the prediction outputs of the CVMs, a subset of the variants was selected for thorough review by our expert clinical genomic scientists. All 14 variants with strong pathogenic (≥99% PPV) and a sampling of 163 /1,052 variants with strong benign (≥95% NPV) CVM predictions were included. For each variant, all currently available non-CVM evidence was reviewed to evaluate for any concerning contradictory data. In this review, 93% (13/14) of variants predicted to be pathogenic and 95% (155/163) predicted to be benign by CVM were confirmed by the experts, while the remainder were kept as VUS. Based on this expert review, only one CVM condition model was not launched, leaving a total of 11 CVM models. Of note, our experts chose the most challenging variants with CVM benign predictions to review (e.g., CVM benign predictions for variants with some level of pathogenic evidence or variants in the *PMS2* pseudogene region). In addition, four of the predicted pathogenic variants had at least one ClinVar entry of likely pathogenic or pathogenic, while a fifth variant predicted to be pathogenic by CVM was recently reclassified from a VUS to likely pathogenic at Invitae based on new family segregation data that was obtained after the CVM prediction was generated, but prior to the results being reviewed. Similarly, 15 of the reviewed predicted benign variants had at least one ClinVar entry of likely benign or benign by another submitter.

Orthogonal validations

To gain further confidence in the prediction outputs of the CVM, a concordance analysis was performed comparing CVMs to the evolutionary model of variant effect (EVE), a deep learning model that uses orthogonal data, namely sequence conservation, to predict the pathogenicity of variants in humans[2]. CVM predictions were highly concordant with EVE predictions (90.5%; eTable 2).

		EVE predictions	
		Benign	Pathogenic
CVM predictions	Benign	1530	137
	Pathogenic	79	517

eTable2. CVM model predictions are highly concordant with EVE *in silico* predictions.

For additional confidence in the prediction outputs of the CVM for the Lynch syndrome genes (*MLH1, MSH2, MSH6, PMS2, EPCAM*), we compared the CVM model predictions to functional datasets (e.g., multiplex assays of variant effects or MAVEs) for *MLH1*, *MSH2*, and *PMS2*. The variants that had both CVM model predictions and MAVE predictions showed high concordance (>98%; eTable 3).

CVM based predictions for Lynch syndrome genes

		MLH1* MAVE based predictions	
		Benign	Pathogenic
	Benign	61	0
	Pathogenic	1	8

**MLH1* concordance is 98.6%.*

		MSH2* MAVE based predictions	
		Benign	Pathogenic
	Benign	140	1
	Pathogenic	1	12

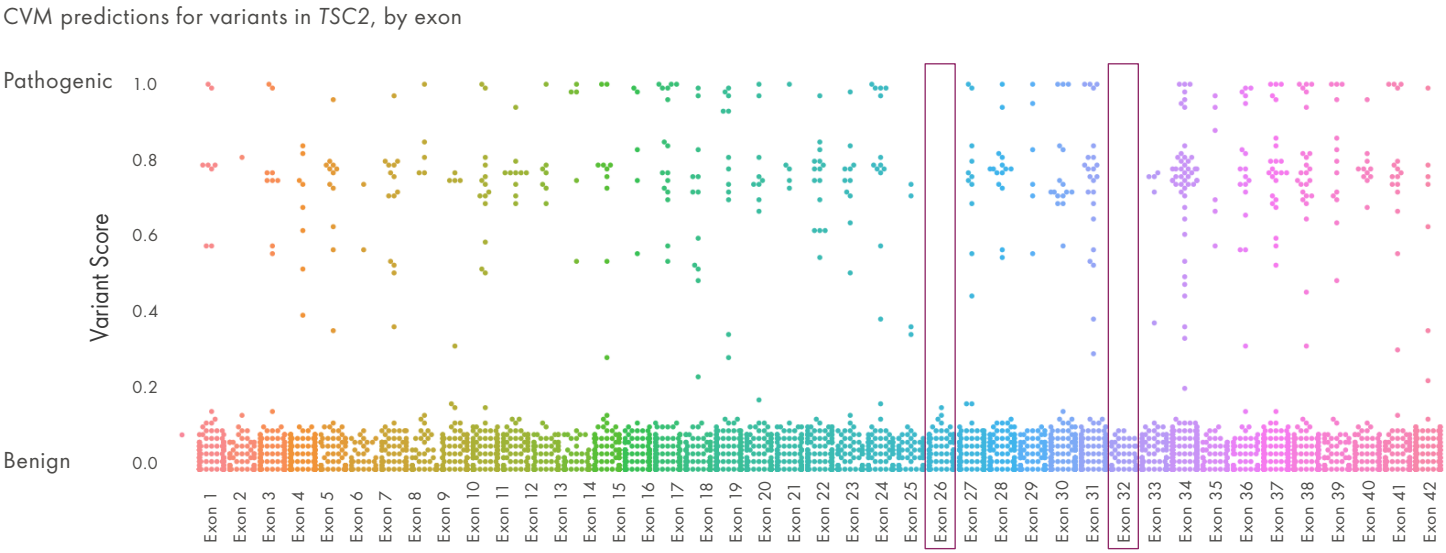
**MSH2* concordance is 98.7%.*

		PMS2* MAVE based predictions	
		Benign	Pathogenic
	Benign	32	0
	Pathogenic	0	10

**PMS2* concordance is 100%*

eTable3. CVM predictions for Lynch syndrome are highly concordant with MAVE functional evidence. Benign CVM predictions (NPV≥95%) and benign MAVE predictions (NPV of ≥80%) as well as pathogenic CVM predictions (PPV≥99%) and pathogenic MAVE predictions (PPV≥80%) were included in the concordance analysis. Given the existence of a saturation MAVE dataset generated by an external group[3], *MSH2* CVM predictions were compared to the saturation dataset. Concordance rate when CVM *MSH2* predictions are compared to an Invitae generated MAVE dataset for *MSH2*, showed similar concordance (data not shown). *MLH1* and *PMS2* CVM predictions were compared to Invitae-generated MAVE datasets.

Another mechanism to assess orthogonal validation for CVM could be performed for *TSC2*, a gene associated with Tuberous Sclerosis. Specifically, exons 26 and 32 of *TSC2* are absent from all known clinically relevant transcripts of the gene[4], (note legacy exon nomenclature). Using *only* clinical evidence, Invitae's CVM for *TSC2* identified genetic variants predicted to be pathogenic in each of the gene's 42 exons, with the exception of exons 26 and 32—a remarkable concordance between the clinical model and independent molecular data (eFigure 6).



eFigure 6. CVM pathogenicity predictions based solely on clinical evidence align with molecular knowledge of *TSC2*. No variants are predicted pathogenic by CVM for exons 26 and 32.

Technical methods/validation:

Dataset definition: First, the Monarch Disease Ontology (MonDO) is used to discover the set of all disease associations for genes tested by Invitae. Each condition is assigned a set of genes and their associated modes of inheritance using the Gene Curation Coalition (GenCC) database.

Cohort identification (patient model): For each condition, positive and negative (control) cohorts of Invitae patients are identified as follows. First, affected individuals are identified as patients who had positive molecular diagnoses in at least one of the included genes. For example, an autosomal recessive association would require either compound heterozygosity or homozygosity for existing pathogenic or likely pathogenic variants. Control cohorts were identified the same way for all conditions–defined as patients who had only benign variants in the included genes.

Data preprocessing and feature selection: Features used for Patient Score modeling include both structured (e.g., ICD-10 codes, age at accessioning, clinical area, etc) and textual (e.g., indication for testing, family history, clinician-reported ancestry) patient information. ICD-10 codes are preprocessed by first abbreviating to the category level and then selecting the codes most enriched in the positive cohort per the chi-square test. For each patient requisition, a string combining both indication for testing and family history is created by concatenating these substrings with special tokens to demarcate the beginning and end of each span of text.

Patient Score model: The Patient Score model follows the following sequence: a pre-trained large language model (UFNLP Gatortron) is fine-tuned to the entire corpus of Invitae's labeled patients across conditions (TL1). For each condition, TL1 is further fine-tuned to that condition's examples (TL2). This model is used to predict a text score for every patient, which is then used as a feature along with other patient information (e.g., age at testing, ICD-10 codes) to train a final model (PS). This model is then used to predict an overall score for each patient, the Patient Score.

Cohort identification (variant model): The variant model utilizes the same variant labels that were used in the genotype filtering for cohorts of genotype-positive and genotype-negative patients. In addition, the set of patients who are informative for interpretation–the bellwether patients–is determined in the following manner. For each gene included in a condition, the annotated mode of inheritance (see dataset definition above), is used to identify patients who could be expected to manifest disease if the variant were causal, who should not have a relevant disease if the variant were benign, and who would not have the disease that would be explained by another known variant in an included gene. For example, the bellwether patients for a variant in a single-gene autosomal dominant model would be those who have the variant of interest and no other non-benign variants in the gene. In a two-gene model, they should also have no non-benign variants in the other gene.

Variant Score model: The Variant Score model is a partially pooled, hierarchical Bayesian inference model for interpreting genetic variants across multiple genes. It estimates parameters like pathogenic rate, penetrance, and gene probabilities based on input tensors representing gene, Patient Scores, and variant labels. The Variant Score, a probability that the variant is pathogenic, is sampled from the posterior predictive distribution of the partially observed Bernoulli variable.

Validation: An estimate of generalization performance was attained by evaluating the model against a holdout set of 20% of labeled variants, which were not used for training. Metrics assessed included area under the receiver operating characteristic (AUROC) curve, average precision (AP), mean squared error (MSE), and classification metrics including F1 score, accuracy, and PPV and NPV. For high-performance models (AUROC≥0.8) and high-performance genes (AUROC≥0.8), variants with a posterior probability of pathogenicity ≤0.05 or ≥0.99 AND with ≥2 affected-appearing observations in unrelated individuals, were nominated for evidence via Sherloc point assignment.

References:

1. Nykamp K, Anderson M, Powers M, *et al.* Sherloc: a comprehensive refinement of the ACMG–AMP variant classification criteria. *Genet Med.* 2017;19(10):1105-1117.
2. Frazer J, Notin P, Dias M, *et al.* Disease variant prediction with deep generative models of evolutionary data. *Nature.* 2021;599(7883):91-95.
3. Scott A, Hernandez F, Chamberlin A, Smith C, Karam R, Kitzman JO. Saturation-scale functional evidence supports clinical variant interpretation in Lynch syndrome. *Genome Biol.* 2022;23(1):266.
4. Ekong R, Nellist M, Hoogeveen-Westerveld M, *et al.* Variants Within TSC2 Exons 25 and 31 Are Very Unlikely to Cause Clinically Diagnosable Tuberous Sclerosis. *Hum Mutat.* 2016;37(4):364-370.