A New Set of *in Silico* Tools to Support the Interpretation of *ATM* Missense Variants Using Graphical Analysis

Check for updates

Luz-Marina Porras,* Natàlia Padilla,* Alejandro Moles-Fernández,[†] Lidia Feliubadaló,[‡§¶] Marta Santamariña-Pena,[‖**††] Alysson T. Sánchez,[‡‡] Anael López-Novo,[‖**] Ana Blanco,[‖**††] Miguel de la Hoya,[§§] Ignacio J. Molina,[¶¶] Ana Osorio,[‖‖***] Marta Pineda,[‡§¶] Daniel Rueda,[†††] Clara Ruiz-Ponte,[‖**††] Ana Vega,[‖**††] Conxi Lázaro,[‡§¶] Orland Díez,[†‡‡‡] Sara Gutiérrez-Enríquez,[†] and Xavier de la Cruz*[§§§]

From the Research Unit in Clinical and Translational Bioinformatics, Vall d'Hebron Institute of Research, Universitat Autònoma de Barcelona, Barcelona; the Hereditary Cancer Genetics Group,[†] Vall d'Hebron Institute of Oncology, Vall d'Hebron Barcelona Hospital Campus, Barcelona; the Hereditary Cancer Program,[‡] Catalan Institute of Oncology, the Program in Molecular Mechanisms and Experimental Therapy in Oncology,[§] Instituto de Investigación Biomédica de Bellvitge (IDIBELL), and the Hereditary Cancer Program,[‡‡] Oncobell Program, Catalan Institute of Oncology, Instituto de Investigación Biomédica de Bellvitge (IDIBELL), Hospitalet de Llobregat, Barcelona; the Centro de Investigación Biomédica en Red de Cáncer,[¶] Madrid; the Fundación Pública Galega Medicina Xenómica,[‖] Santiago de Compostela; the Instituto de Investigación Sanitaria de Santiago de Compostela,** Santiago de Compostela; the Centro de Investigación Biomédica en Red de enfermedades Raras[††] Madrid; the Molecular Oncology Laboratory,[§§] Hospital Clínico San Carlos, IdISSC (Instituto de Investigación Sanitaria del Hospital Clínico San Carlos), Madrid; the Instituto de Biopatología y Medicina Regenerativa,[¶¶] Universidad de Granada y Instituto de Investigación Biosanitaria ibs.GRANADA, Granada; the Familial Cancer Clinical Unit,[‖‖] Human Cancer Genetics Programme, Spanish National Cancer Research Centre, Madrid; the Spanish Network on Rare Diseases,*** Madrid; the Hereditary Cancer Laboratory,[†††] 12 de Octubre University Hospital, i+12 Research Institute, Madrid; the Area of Clinical and Molecular Genetics,[‡‡‡] Vall d'Hebron Hospital Universitari, Vall d'Hebron Barcelona Hospital Campus, Barcelona; and the Institució Catalana de Recerca i Estudis Avançats,[§§§] Barcelona, Spain

Establishing the pathogenic nature of variants in *ATM*, a gene associated with breast cancer and other hereditary cancers, is crucial for providing patients with adequate care. Unfortunately, achieving good variant classification is still difficult. To address this challenge, we extended the range of *in silico* tools with a series of graphical tools devised for the analysis of computational evidence by health care professionals. We propose a family of fast and easy-to-use graphical representations in which the impact of a variant is considered relative to other pathogenic and benign variants. To illustrate their value, the representations are applied to three problems in variant interpretation. The assessment of computational pathogenicity predictions showed that the graphics provide an intuitive view of pre-diction reliability, complementing and extending conventional numerical reliability indexes. When

119-129, 08035 Barcelona,
Spain.
E-mail: sgutierrez@vhio.net or
xavier.delacruz@vhir.org.

applied to variant of unknown significance populations, the representations shed light on the nature of these variants and can be used to prioritize variants of unknown significance for further studies. In a third application, the graphics were used to compare the two versions of the *ATM*-adapted American College of Medical Genetics and Genomics and Association for Molecular Pathology guidelines, obtaining valuable information on their relative virtues and weaknesses. Finally, a server [ATMision (ATM missense *in silico* interpretation online)] was generated for users to apply these representations in their variant interpretation problems, to check the *ATM*-adapted guidelines' criteria for computational evidence on their variant(s) and access different sources of information. *(J Mol Diagn 2024, 26: 17−28; https://doi.org/10.1016/j.jmoldx.2023.09.009)*

The ataxia-telangiectasia mutated (*ATM*) gene encodes a serine/threonine kinase essential in the detection and signaling to repair DNA double-stranded breaks and with an important role in cancer susceptibility.[1] Monoallelic pathogenic germline variants in *ATM* increase the risk of cancer, particularly breast cancer (lifetime absolute risk of approximately 20% by the age of 85 years),[2,3] but also melanoma,[4] prostate cancer, and pancreatic cancer.[5,6] In addition, studies have shown that women aged <45 years with rare *ATM* missense variants may be at increased risk for developing secondary contralateral breast cancer after radiotherapy.[7] Biallelic germline loss-of-function variants cause the disorder ataxia telangiectasia, a pleiotropic neurodegenerative disease[8] with symptoms such as progressive cerebellar degeneration, immunodeficiency, chronic lung disease, cancer predisposition, endocrine abnormalities, segmental premature aging, chromosomal instability, and radiation sensitivity. Given the role of *ATM* in cancer susceptibility, it is ubiquitously included in hereditary cancer multigene testing panels.[9] Identification of carriers of *ATM* disease-causing variants offers patients and families precise clinical management based on personalized prevention. For example, it can improve patient access to enhanced mammography and magnetic resonance imaging−based screening[10] or allow a better tailoring of radiotherapy treatments to patients with cancer.[7,11]

In this context, it is of paramount relevance to accurately assess the deleterious effect of variants on ATM protein function for ultimately providing a classification for clinical use that will lead to the safest patient management possible. However, variant interpretation is difficult in ATM because it is a large, multifaceted protein that phosphorylates >700 substrates, such as protein P53 and checkpoint kinase CHEK2, involved in cell cycle control, DNA repair, cell survival, and other cellular processes.[1] As a consequence, 93% of missense variants (6119 of 6585; ClinVar database, https://www.ncbi.nlm.nih.gov/clinvar) are variants of unknown significance (VUSs), of little clinical value without further research.

Recent results indicate that *in silico* tools may contribute to alleviating this situation by providing fast variant predictions useful in the clinical context.[12,13] These tools, known as pathogenicity classifiers or predictors, use artificial intelligence algorithms to integrate a series of variant descriptors to produce a binary classification (pathogenic or benign) of variants. However, although their accuracy is increasingly better,[13] *in silico* tools still make mistakes (Figure 1). To mitigate the impact of these errors, it is crucial to conduct user analysis, which involves assessing the reliability of the results and verifying whether they make sense or not. However, this task can be difficult and time-consuming because of the intricate nature of artificial intelligence models and the technical expertise required.

This article presents a graphical approach that addresses this problem, allowing professionals to assess variants and their classifications in light of what is known about *ATM* variants. To this end, target variants are represented in terms of properties related to their molecular impact, and populations of pathogenic and benign *ATM* missense variants are used for comparison purposes. The usefulness of this tool is illustrated in three open problems in the clinical classification of these variants. First, we describe how the graphical representation can support prediction assessment. For this application, two *ATM*-specific pathogenicity predictors have been generated, based on two different sets of properties. Second, the proposed representations are used to analyze and prioritize VUSs. Third, this representation is applied to the comparison of the two adaptations to *ATM* of the guidelines of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology[9,14] in their treatment of computational evidence. Finally, this article presents ATMision (ATM missense *in silico* interpretation online), a web portal (http://biotoclin.org/ATMision, last accessed March 1, 2023) where users can access the representations and predictors developed to analyze their variants. Links to other resources are also provided to facilitate the characterization of variants with relevant biological and biomedical information.

Materials and Methods

Graphical Representations for Analysis of Variants

There are two key elements in the representations proposed: the populations of pathogenic and neutral variants and the features used to characterize them. The former are obtained as explained in *Variant Data Set*. The features are properties that characterize variants and capture specific aspects of their impact. In this work, two sets of features are tried. The first, Bioinf, is constituted by properties routinely used in
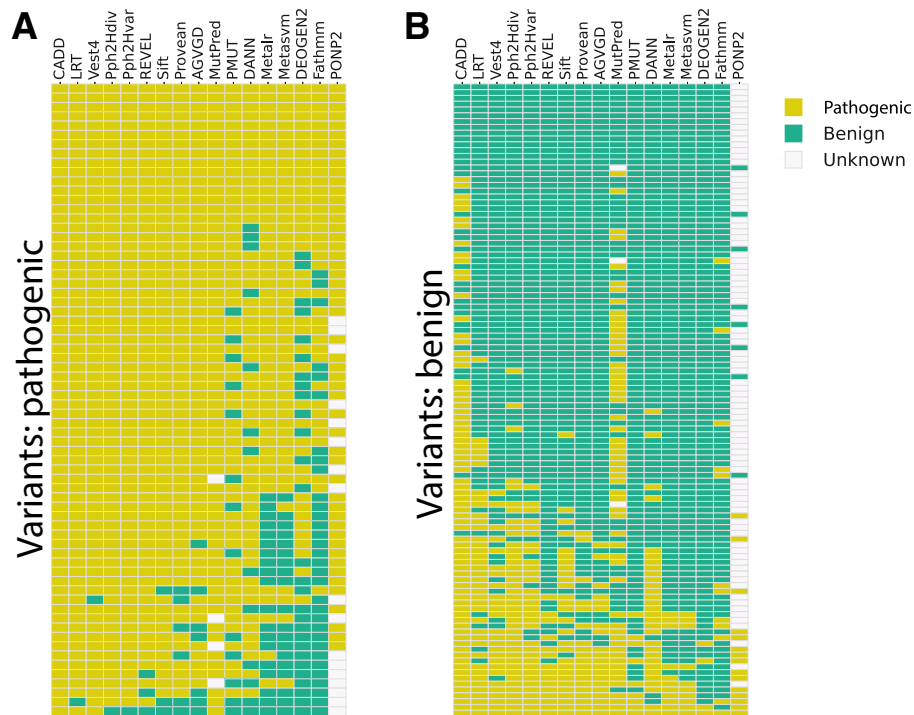
Figure 1   Performance of 17 pathogenicity predictors for *ATM* missense variants. Each column in the colored tables corresponds to an *in silico* predictor, and each row corresponds to a variant in the variant data set. Table cells are colored according to whether the variant is classified as pathogenic (yellow), benign (green), or unknown (white). **A:** Pathogenic variants. **B:** Benign variants.

developing pathogenicity predictors[13] (eg, conservation in multiple sequence alignments and native-mutant differences in hydrophobicity). The second, Metap, is constituted by the score of several pathogenicity predictors. A more detailed description of these properties is provided in the upcoming sections.

For a given set of features (Bioinf or Metap), the graphical representations are obtained as follows.

First, merge the pathogenic and benign populations in a single data set. Second, label the variants with the selected features. For example, for Bioinf, each variant is labeled with five properties (see *Bioinf Features*). And third, do a principal components analysis using the Scikit-learn (*https://scikit-learn.org*, release 1.2.2, last accessed September 22, 2023)[15] implementation (sklearn.decomposition).

Finally, contour plots (kdeplot function, from Seaborn library,[16] *https://seaborn.pydata.org*, release version 0.12, last accessed September 22, 2023) are used to represent the distributions of pathogenic and benign variants in the plane of the first two components. A special symbol is used to highlight the variant of interest.

Variant Data Set

To build the graphical representations in this work, the authors used two data sets of *ATM* missense variants derived previously.[9] The first data set, consisting of 248 variants (35 pathogenic and 213 benign), was only used for generating the representations of variants located in the N-terminal half

of ATM (before protein residue 1960). The second data set, which consisted of 177 variants (68 pathogenic and 109 benign), was employed to represent variants in the protein's C-terminal half (at, or beyond, protein residue 1960). In addition, this second data set was also used for training the pathogenicity predictors presented in this work.

The pathogenic variants were obtained from the literature in the field.[9] The benign variants were obtained using a homology-based model routinely used for developing pathogenicity predictors.[17–20]

A list of the variants, together with the corresponding values of the Bioinf and Metap features, is provided in Supplemental Table S1.

Bioinf Features

The authors used five properties chosen for their good performance in protein-specific pathogenicity predictors.[19,20] Three amino acid indexes were used: van der Waals volume[21] and hydrophobicity[22] changes on mutation and Blosum62[23] matrix elements. The remaining properties are computed from the multiple sequence alignment (MSA) of ATM, retrieved from the Align-GVGD server (*http://agvgd.hci.utah.edu*, last accessed February 1, 2022): Shannon entropy and position-specific scoring matrix elements. Shannon entropy reflects the overall conservation at an alignment column and is equal to $-\Sigma_i p_i.\log(p_i)$, where p_i is the frequency of the natural amino acid i at the variant's locus in the MSA; i varies between 1 and 20. The position-

specific scoring matrix element reflects the conservation of the native residue in an alignment. It is equal to log($f_{nat,i}$/$f_{nat,MSA}$), where $f_{nat,i}$ is the frequency of the native amino acid at locus i of the variant and $f_{nat,MSA}$ is the abundance of the same amino acid throughout the MSA.

Metap Features

The authors used the scores of six known tools: REVEL,[24] CADD,[25] VEST4,[26] Polyphen-2-Hdiv,[17] SIFT,[27] and Align-GVGD.[28]

RF Pathogenicity Predictors

In the applications described in *Graphical Analysis of Variants* and *Contour Plots as Companions for* in Silico *Tools in Variant Classification*, in *Results*, the authors use two in-house random forest (RF) pathogenicity predictors, RF_Bioinf and RF_Metap, that used Bioinf and Metap as discriminant features, respectively, and classify variants as either pathogenic or benign.

These predictors are built with the random forest algorithm, randomForest, implemented in the R package (*https://cran.r-project.org/web/packages/randomForest/index.html*, version 4.7-1.1, last accessed September 22, 2023). They are restricted to variants from the protein's C-terminal half because their training requires a minimum number of cases[20] that is only attained for ATM's C-terminal half, where most pathogenic missense cases in the variant data set happen (68, see *Variant Data Set*). Because of the limited size of this data set, there were certain constraints in training the RF predictors. First, the authors did not incorporate a hyperparameter tuning step; instead, they relied on the default parameters of the randomForest package, which typically yield satisfactory performances.[29] Second, to mitigate overfitting, the authors employed a leave-one-out-cross-validation procedure[18] rather than a traditional training-validation-test scheme to estimate the performance of their models. This approach is commonly followed when working with small data sets.[30]

To enhance the accuracy of their leave-one-out-cross-validation performance estimates, the authors used a more stringent version of this method (Supplemental Figure S1A). Standard leave-one-out-cross-validation divides the available data into a training and a testing data set. The training data set comprises all variants except one, which is set aside to form the test set. The process is then iterated over all the variants in the data set, and the results obtained for the test set variants are used to compute the model's performance. In this work, instead of leaving one variant out at each step, all the variants occurring at the same position are simultaneously set aside. These variants are then scored with the trained random forest. The procedure is iterated along all the locations of the variants in the data set, and the test results are used to compute the final performance of the method. This approach eliminates a major source of data leakage

effects between training and test sets,[31] thus resulting in more realistic performance estimates.

Because of the imbalance between pathogenic and benign variants in the training set, the resampling program SMOTE (smotefamily in R) (Supplemental Figure S1B) is used to keep to a minimum the impact of imbalance on predictor performance.[20] To smooth out statistical fluctuations, this process is repeated 100 times, ending up with 100 versions of the same predictor (eg, 100 versions of a random forest trained with Bioinf features) (Supplemental Figure S1C). To classify a new variant, these 100 tools are applied; if ≥ 50 of them call it pathogenic, the variant is classified as pathogenic; otherwise, it is classified as benign.

Reliability Computation

The aim is to produce a reliability index related to the accuracy of predictions (ie, an index for which high and low values correspond to high and low accuracies, respectively). The procedure proposed here is inspired by the observation according to which the values of prediction scores are related to prediction accuracy.[32,33] In the present case, this translates to the following principle: the closer a prediction is to the boundary between the target classes, the less reliable it is. This principle is embodied in the following procedure, designed for an arbitrary pathogenicity predictor with M input features. It is divided into two parts: clustering and reliability assignment.

Clustering
First, normalize each feature. Second, build an M-dimensional lattice using all input features (internode spacing, 0.1). Third, obtain the pathogenicity prediction for each lattice node. Fourth, do a K-means clustering of these nodes using the Scikit-learn[15] implementation (sklearn.cluster.KMeans).

For each final cluster, two quantities are computed, r_p and r_b. The r_p is the fraction of pathogenic nodes in the cluster: $r_p = n_p/(n_p + n_b)$, where n_p and n_b are the number of pathogenic and benign predictions for the nodes inside the cluster, respectively. The r_b is the fraction of benign nodes in the cluster: $r_b = n_b/(n_p + n_b)$.

Reliability Assignment
For a variant of interest, find the cluster with the closest centroid. Reliability will be equal to the r_p or r_b of the cluster, depending on whether the variant's prediction was pathogenic or benign, respectively.

The relationship between reliability and prediction accuracy was checked for RF_Bioinf and RF_Metap. To this end, the predictions of these tools for the variant data set were partitioned using a threshold of reliability, $R = 0.6$ (which guarantees a minimum sample size for accuracy computations). For $R \leq 0.6$, the accuracies of RF_Bioinf and RF_Metap were 58% and 40%, respectively; for $R > 0.6$, the accuracies of RF_Bioinf and RF_Metap were

98% and 97%, respectively. In summary, high and low reliabilities correspond to high and low accuracies, respectively, as desired.

In ATMision, the website where these results are available (see *ATMision Web Portal*), a modified procedure was employed in the case of *ATM*-adapted American College of Medical Genetics and Genomics and Association for Molecular Pathology guidelines of Feliubadaló et al.[9] This new procedure considers that the scores of only two predictors are used. The clustering was performed on a two-dimensional lattice with an internode spacing of 0.001. Here, r_p was equal to: $r_p = n_p/(n_p + n_b + n_{dis})$, where n_p and n_b have the same meaning as before, and n_{dis} is the number of nodes for which the two predictors in the guidelines are discordant. The r_b is defined analogously.

In Silico Pathogenicity Predictors for Benchmarking

The authors used 17 pathogenicity predictors. The scores for 14 of them were retrieved from dbNSFP4,[34] version 4.0: REVEL,[24] PROVEAN,[35] VEST4,[26] Polyphen-2-Hvar and Polyphen-2-Hdiv,[17] SIFT,[27] MetaSVM and MetaLr,[36] DEOGEN2,[37] CADD,[25] MutationAssessor,[38] FATHMM,[39] LRT,[40] and GenoCanyon.[41] For the remaining tools, the predictions from the corresponding websites were retrieved: Align-GVGD[28] (*ATM* multiple sequence alignment option: human to sea urchin, *http://agvgd.hci.utah.edu/agvgd_input.php*, last accessed February 1, 2022), PON-P2[42] (*http://structure.bmc.lu.se/PON-P2*, last accessed February 1, 2022), and PMut[33] (*http://mmb.irbbarcelona.org/PMut*, last accessed February 1, 2022).

Performance Assessment of *in Silico* Pathogenicity Predictors

The authors used four standard performance measures[43,44]: sensitivity, specificity, accuracy, and Matthews correlation coefficient (MCC).

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{1}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{2}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FN) \cdot (TN + FP) \cdot (TP + FP) \cdot (TN + FN)}} \tag{4}$$

where true positive (TP) and false negative (FN) are the numbers of correctly and incorrectly classified pathogenic variants, respectively; and true negative (TN) and false positive (FP) are the numbers of correctly and incorrectly classified benign variants, respectively.

ATMision Web Portal

The tools presented are available at the website ATMision (*http://biotoclin.org/ATMision*). At a technical level, to provide easy access to the data, they were stored in a PostgreSQL database and then a website was built on top of it using the Django framework version 4.0.1 (*http://www.djangoproject.com*, last accessed March 1, 2023). Finally, the information on each variant is completed with a series of links to external databases of interest. A description of the web portal structure and options is provided below in *Results*.

Results

Graphical Analysis of Variants

The authors present a family of graphical representations to support the interpretation and classification of *ATM* missense variants. On the basis of the use of contour plots, these representations allow the authors to locate a variant, or set of variants, relative to the populations of known pathogenic and benign variants. The following sections present the application of these plots to three problems in variant interpretation: reliability analysis of *in silico* evidence for variant classification; VUS prioritization; and comparison of classification guidelines. For the first two applications, two in-house predictors, RF_Bioinf and RF_Metap, are used. These tools have been developed specifically for this work, following the standards in the field (explained in *Materials and Methods*), and have a competitive performance compared with standard *in silico* pathogenicity predictors (Supplemental Figure S2).

Contour Plots as Companions for *in Silico* Tools in Variant Classification

The authors encode the populations of pathogenic and benign variants in the C-terminal region of the ATM protein (native residue locations, $\geq$1960) using two-dimensional contours (Figure 2, A and B). These contours are produced after a principal component analysis of the known variants in this region (see *Materials and Methods*), performed after labeling the variants with a given set of properties. In the final graphic, a specific symbol shows the variant of interest.

The resulting representations for Bioinf (Figure 2A) and Metap (Figure 2B) show common motifs: an overlap region between populations, the maxima of each population (around which contour lines concentrate), and the external regions where only one variant type is present. By seeing how a variant relates to these regions, users can complete and visually confirm standard, reliability-based prediction assessments. For example, the benign variant D2987E is correctly classified by RF_Bioinf as benign, with a high reliability (1.0); its location nearby the core of the benign
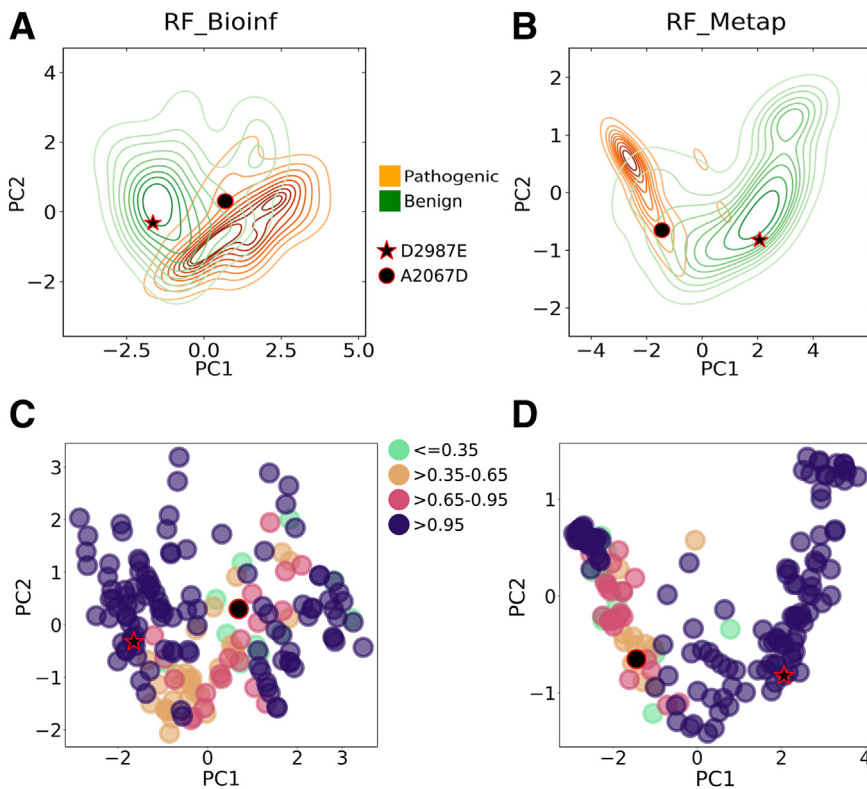
Figure 2 Contour representations for prediction and classification analysis. **A** and **B:** Contour plots representing populations of pathogenic (orange) and benign (green) variants. Data are obtained by performing a principal component (PC) analysis on the variant data set, with variants labeled with Bioinf and Metap features (see *Bioinf Features* and *Metap Features*). PC1 and PC2 indicate the first and second principal components, respectively. Specific symbols are used for the examples. **C** and **D:** Variants in the variant data set are colored according to the reliability of their prediction. The variant predictions were obtained with RF_Bioinf and RF_Metap tools, respectively. The horizontal and vertical axes are the same as in **A** and **B. C** and **D:** The distribution of colors matches the main features in the **A** and **B** contour plots, respectively, showing the relationship between *in silico* prediction reliability and the contour representations.

population (Figure 2A) supports the *in silico* prediction. On the contrary, A2067D, a pathogenic variant incorrectly classified as benign, with low reliability (0.02), falls in a region where both variant types are present, reinforcing the suspicions raised by the low reliability. A comparable analysis can be performed using the contour plot for Metap (Figure 2B) and the RF_Metap predictions for D2987E and A2067D that, in this case, coincide with those from RF_Bioinf. Here, the reliabilities are 1.0 and 0.18, respectively, and are consistent with the locations of the variants in the contour plots.

In general, there is a qualitative agreement between *in silico* prediction reliabilities and the main features of the contour plots (Figure 2, C and D): highly reliable predictions predominate in contour regions (maxima and outer sides), where pathogenic or benign variants predominate, and less reliable predictions predominate in the overlap region. This agreement supports the value of the contour plots as a tool for visually checking prediction reliability. The graphical analysis, although related, is not equivalent to using reliability indexes from pathogenicity predictors because it includes the behavior of pathogenic and benign populations.

VUS Prioritization

The contour plots are used to explore the structure of the population of *ATM* missense VUSs, focusing on those cases

located in the C-terminal domain of the protein (beyond residue 1960): 1572 variants were retrieved from ClinVar (November 2021) (*https://www.ncbi.nlm.nih.gov/clinvar*). These variants are represented in the plots corresponding to Bioinf (Figure 3A) and Metap (Figure 3B) features. Considering that VUSs are variants whose pathogenic nature is difficult to determine, one would expect them to cluster at the boundary between pathogenic and benign populations. Interestingly, this is not the case; for both representations, VUSs spread over most of the plot. Some fall in intermediate regions, but others fall in predominantly pathogenic or benign regions, indicating that the population of VUSs is heterogeneous.

On the basis of this result, one may want to prioritize these variants for further study (eg, focusing on the clearest pathogenic or benign cases). A simple procedure is proposed based on the correspondence (Figure 2, C and D) between the contour plots' predominantly pathogenic and benign regions and high prediction reliabilities. First, the VUSs are assessed with RF_Bioinf; then, those cases with reliability indexes below a certain threshold (the authors tried 0.65 and 0.95) are eliminated; finally, the remaining variants are plotted. Figure 4, A and C, shows how increasingly stringent reliability thresholds deplete the overlapping regions, leaving only the variants that populate the desired regions. A similar result is observed with the Metap representation (Figure 4, D and F) after assessing the VUSs with RF_Metap and applying the reliability
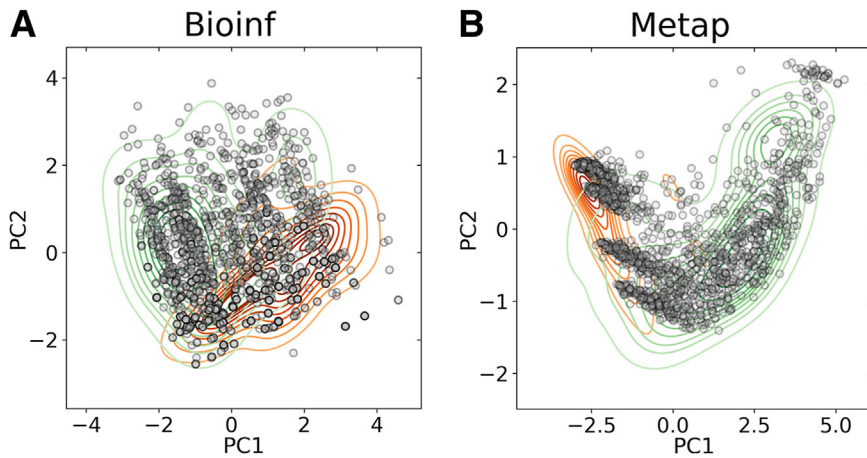
Figure 3 Representation of 1572 *ATM* missense variants of unknown significance (VUSs) relative to the populations of pathogenic and benign variants. **A** and **B:** The contour plots obtained using Bioinf and Metap properties. The 1572 VUSs (gray circles) were retrieved from ClinVar (November 2021) (https://www.ncbi.nlm.nih.gov/clinvar, last accessed February 15, 2023) and belong to the C-terminal half of the protein (see *VUS Prioritization*). The pathogenic and benign variant populations are shown with orange and green contour plots, respectively. PC, principal component.

thresholds. Supplemental Table S2 contains the final list of selected variants.

Comparison of the *ATM*-Adapted Guidelines

The two versions of the *ATM*-adapted guidelines[9,14] follow different criteria to accept or discard computational evidence, and users may want to understand the differences. The contour plots provide an intuitive approach for their comparison.

Feliubadaló et al[9] score variants with two *in silico* pathogenicity predictors; when they disagree, computational evidence is rejected. Three aspects deserve mention in the contour plots corresponding to this criterion (Figure 5, A
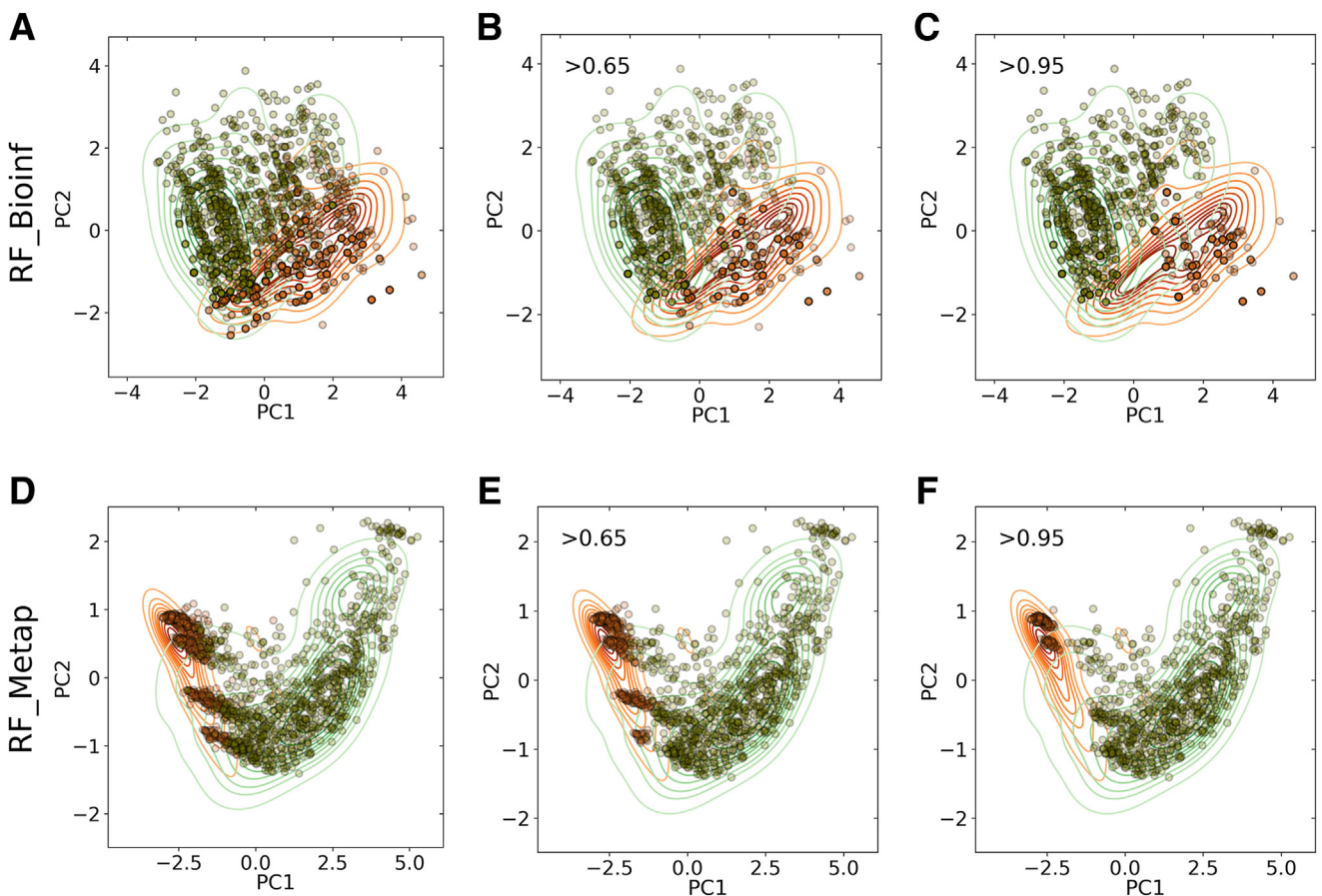
Figure 4 Application of contour plots for the prioritization of missense variants of unknown significance (VUSs) in *ATM*. **A** and **D:** Pathogenicity predictions for the set of 1572 missense VUSs are obtained using RF_Bioinf (**A**) and RF_Metap (**D**). Variants predicted as pathogenic and benign by these tools are shown with orange and green circles, respectively. Their location is shown in the contour plots for the Bioinf (**A**) and Metap (**D**) input features. **B** and **E:** The resulting plots after eliminating those variants with $R \leq 0.65$. **C** and **F:** Equivalent for $R \leq 0.95$. PC, principal component.

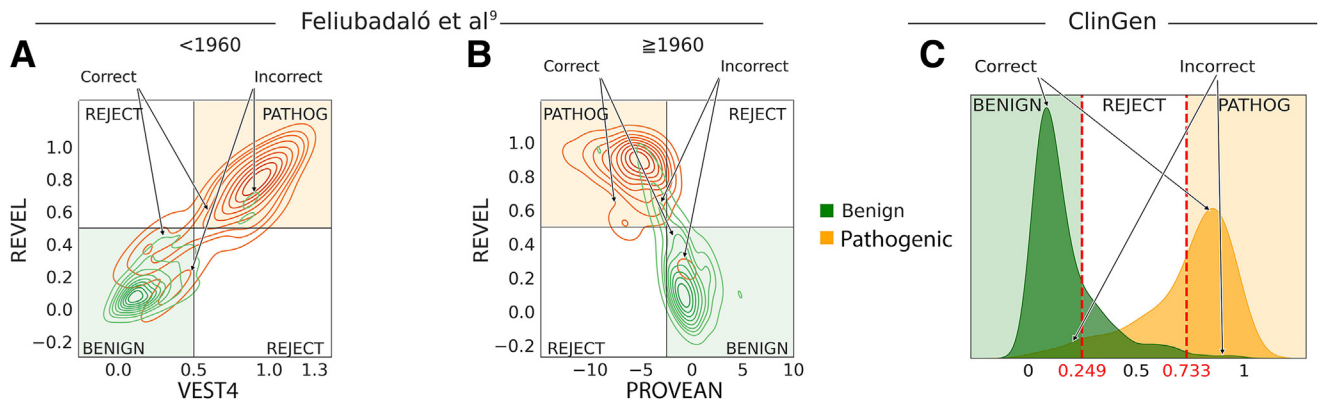**Figure 5** Visual representation of the criteria for computational evidence in the *ATM*-adapted American College of Medical Genetics and Genomics and Association for Molecular Pathology guidelines.[9,14] The *ATM*-adapted guidelines of Feliubadaló et al[9] use two *in silico* pathogenicity predictors whose contribution is accepted when they agree and rejected otherwise. The chosen predictors depend on the variant's location in the protein sequence: REVEL + VEST4 (before residue 1960) and REVEL + PROVEAN (at, or after, residue 1960). **A** and **B:** The combination of colored areas and contour lines shows how these criteria work and their accuracy. Colored and white squares correspond to the acceptance and rejection regions of the criteria, respectively. The overlap between squares and contour lines reflects accuracy. Contour lines traversing a square of the same color indicate variants that will be correctly classified; if the colors are opposite, the classification will be incorrect. **Arrows** indicate these two situations. Finally, contour lines traversing a white square indicate variants for which computational evidence will be rejected. **C:** The same scheme is applied to ClinGen's *ATM*-adapted guidelines,[14] which use a single classifier, REVEL, for the *in silico* classification of variants. Here, a single histogram is enough to illustrate the guideline's criterion. The acceptance and rejection regions are colored as before, and the **arrows** indicate the potential classification successes and failures of the acceptance regions. Pathog, pathogenic.

and B). First, contour lines traversing an acceptance region of the same color correspond to variants that will be correctly classified. Second, contour lines traversing an acceptance region of a different color (eg, green lines traversing the orange region) correspond to variants that will be incorrectly classified. And third, contour lines traversing the rejection region correspond to variants for which computational evidence will be disregarded.

Clinical Genome Resource (ClinGen) expert panel[14] uses a single tool, REVEL, for the *in silico* pathogenicity classification of missense variants. The rejection region corresponds to REVEL scores between 0.249 and 0.733 (Figure 5C).
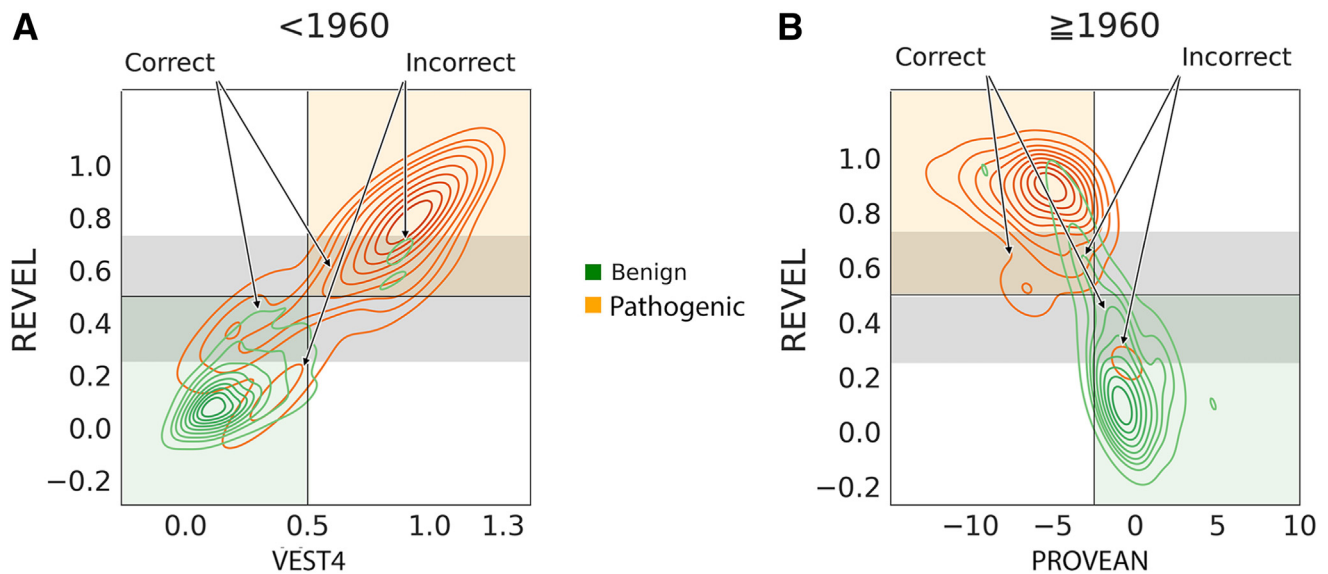

Figure 6 Comparison of the two *ATM*-adapted guidelines.[9,14] **A** and **B:** Compare both versions of the guidelines for variant locations in the protein sequence before and after residue 1960. Over the contour plots corresponding to the criteria of Feliubadaló et al,[9] a gray band identifies ClinGen's[14] rejection region. The following situations may happen. When the band overlaps with a colored acceptance region, ClinGen's[14] criterion will discard *in silico* evidence for the variants inside. Consequently, applying ClinGen's[14] criterion relative to the criteria of Feliubadaló et al[9] results in the elimination of some potential errors and in the loss of a substantial number of correct *in silico* classifications (both cases identified with **arrows**). When the gray band overlaps a white square, both criteria coincide. Finally, outside the gray band, contour lines traversing white squares correspond to cases where *in silico* evidence will be rejected by the criteria of Feliubadaló et al[9] but accepted by ClinGen's[14] criterion.
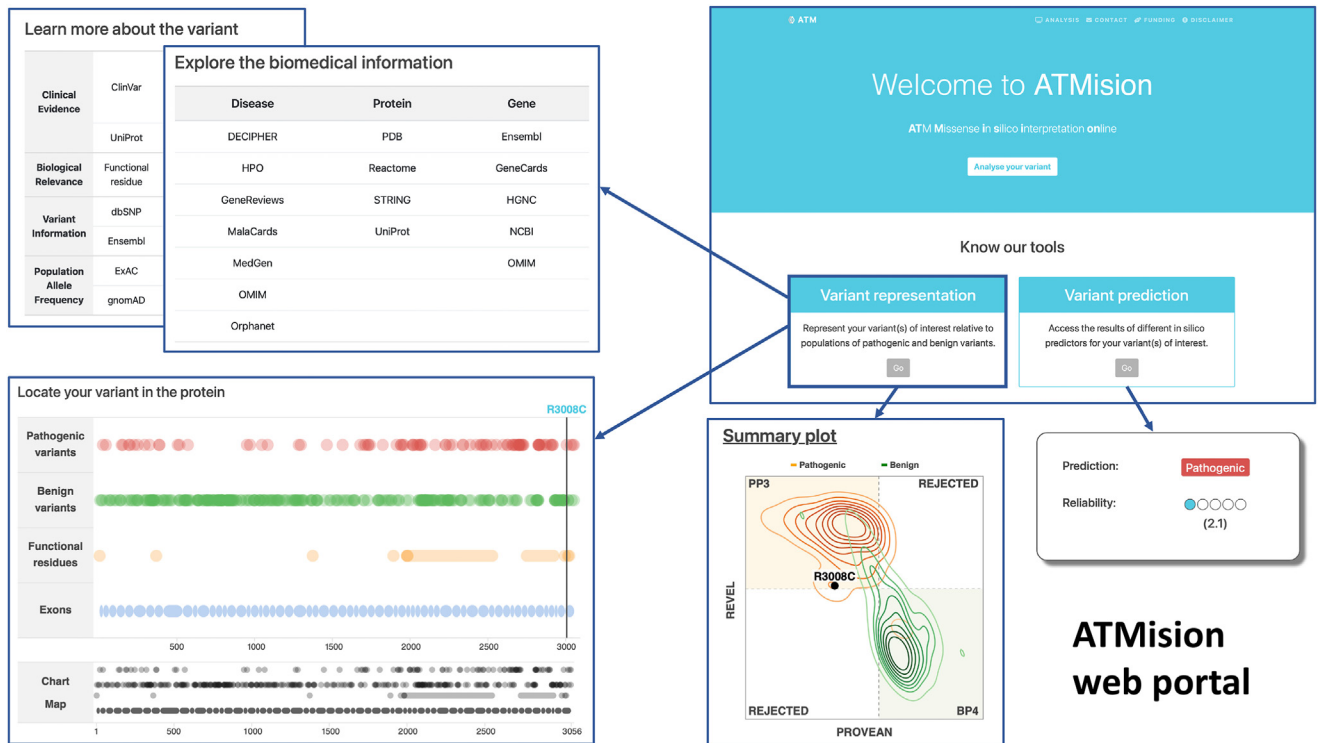
Figure 7 ATMision (ATM missense *in silico* interpretation online) web portal. The figure summarizes the main features of the website. The highlighted variant representation button gives access to the contour plots where users can compare a variant of interest against a population of pathogenic (orange contour) and benign (green contour) variants. The figure also shows the additional information that can be accessed through summary tables (mousing over the numbers, one can access the corresponding record in the external database) and an interactive representation of the ATM protein sequence in which some features of interest are represented. In the variant prediction option, users can access, for their variant(s) of interest, the *in silico* prediction for the tools presented in this work, RF_Bioinf and RF_Metap, and the *in silico* recommendations from the two *ATM*-adapted American College of Medical Genetics and Genomics and Association for Molecular Pathology guidelines.[9,14] This option also gives users additional information (tables and the interactive sequence plot). gnomAD, Genome Aggregation Database; NCBI, National Center for Biotechnology Information; OMIM, Online Mendelian Inheritance in Man.

Figure 6 compares the two adapted guidelines. The gray central band corresponds to ClinGen's[14] rejection region. This region has one main effect relative to the criterion of Feliubadaló et al[9]: decreasing the number of variants classified because of a reduction in the acceptance region. This effect has one positive consequence: classification errors will decrease when applying ClinGen's[14] criterion. On the negative side, a substantial amount of correct classifications will be rejected: this primarily affects pathogenic variants before protein residue 1960 (Figure 6A) or benign variants above protein residue 1960 (Figure 6B). Another effect of ClinGen's[14] criterion is acceptance, primarily for variants located beyond protein residue 1960 (Figure 6B), of a minor amount of correct classifications discarded by the criteria of Feliubadaló et al.[9]

ATMision Web Portal

The graphical tools and pathogenicity predictors presented in this work have been made available by generating a specific web portal, ATMision (Figure 7) (*http://biotoclin. org/ATMision*).

ATMision allows users to jump directly to the graphical representations presented in the *Results*. A dropdown menu shows the four representation options available

(Supplemental Figures S3 and S4): a plot derived from the Bioinf or Metap features or the criteria of Feliubadaló et al[9] (Supplemental Figure S3) or ClinGen[14] (Supplemental Figure S4). Users can choose to represent one or more variants (Supplemental Figures S3A and S4A, and S3B and S4B, respectively). The first case is designed for an in-depth analysis of the variant: links to a variety of external information sources are accessible from the variant's landing page, as well as an interactive representation of ATM's sequence (Supplemental Figure S5) that includes several features of interest (eg, exon and functional domain locations). Instead of a single variant, users can display up to 500 variants. This option is appropriate for global analyses of variant populations, identifying trends or outliers in quality control checks. Apart from the plot, no other information is provided in this case. In the single variant mode, the variant is labeled for its identification. When plotting several variants, dynamic labels, which appear when mousing over the variant symbols (Supplemental Figures S3B and S4B), replace static labels to avoid crowded figures.

On its landing page, ATMision also allows users to access the RF predictors or to generate the computational evidence corresponding to the two *ATM*-adapted versions of the American College of Medical Genetics and Genomics and the

Association for Molecular Pathology guidelines (Supplemental Figure S6).[9,14] The results are presented on a webpage with prediction information, graphical representations, and additional biomedical information. Again, ATMision gives the option to generate predictions for one or more variants. In the second case, a table with the main results is presented; one can then mouse over it to access the page for a specific variant.

Discussion

This work presents a graphical approach to support and improve the interpretation of missense variants in the gene *ATM*. It is in line with the increasingly widespread view[45,46] that human experts must be able to supervise high-stakes machine learning applications. The graphics are designed to promote and facilitate the role of professionals by providing fast and easy-to-comprehend representations of one or more target variants in the context of populations of pathogenic and benign variants. The main features of the resulting figures are related to the reliability of pathogenicity predictions (Figures 2, C and D). This result supports their use for visually assessing *in silico* predictions for variant classification, thus extending the information provided by standard reliability indexes.[32,33]

The general nature of the representations allows us to use them for purposes other than prediction assessment. For example, these help in the interpretation of the large population of *ATM* VUSs. The number of these variants in genes of clinical interest is a bottleneck in the development of genomic medicine[47] and, with 93% missense VUSs (6119 of 6585 missense variants in ClinVar, February 2023), *ATM* is no exception.[9] In this context, strategies to understand these variants and reduce their numbers are a priority. Here, a preliminary and fast understanding of *ATM* VUSs relative to their molecular impact is obtained with the contour representations, which show (Figure 3) how the VUS population covers the whole pathogenic-benign range. To select the clearest cases for subsequent studies, the reliability indexes of the variants' predictions are used to define acceptance and rejection regions. This strategy allowed the selection of variants in the desired regions (Figure 4, B, C, E, and F, and Supplemental Table S2).

In a third example, the contour plots are used to compare the criteria for using computational evidence in the *ATM*-adapted guidelines.[9,14] It is found that ClinGen's[14] criteria favor a lower misclassification rate at the price of rejecting a substantial number of correct classifications. In contrast, the opposite is true for the criteria of Feliubadaló et al.[9] *A priori*, it is not possible to establish which criteria are preferable; the answer will depend on the application context. For example, if the available evidence on a variant is strong, applying ClinGen's[14] criteria seems preferable as the potential rejection of computational evidence will have a low impact. However, if the available evidence is scarce or

weak, using the more inclusive criteria of Feliubadaló et al[9] may give valuable hints on the nature of the variant.

The tools presented in this work have been developed using the variant data sets in the study by Feliubadaló et al,[9] which are limited to what is available in the literature. As more variants become available, these tools will be updated. *A priori*, a minor impact is expected on the main characteristics of the contour plots because these plots reflect the general behavior of variant populations. There may be some variations, however, in the contour lines representing the less populated regions of the plot. Overall, it is anticipated that the general applicability of contour plots to classification problems will remain the same.

Finally, the results presented are centered on *ATM*; however, the method can also be extended to other genes of interest because of its simplicity. The only requisite is that there are enough classified variants to produce the contour plots.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

During the preparation of this work, the authors used ChatGPT to enhance the language in certain sections of the text. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Acknowledgments

Author Contributions

All authors confirmed they have contributed to the intellectual content of this article and have met the following four requirements: i) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; ii) drafting or revising the article for intellectual content; iii) final approval of the published article; and iv) agreement to be accountable for all aspects of the article, thus ensuring that questions related to the accuracy or integrity of any part of the article are appropriately investigated and resolved.

Disclosure Statement

None declared.

Supplemental Data

Supplemental material for this article can be found at *http://10.1016/j.jmoldx.2023.09.009*.

References

1. Shiloh Y, Ziv Y: The ATM protein kinase: regulating the cellular response to genotoxic stress, and more. Nat Rev Mol Cell Biol 2013, 14:197−210

2. Hu C, Hart SN, Gnanaolivu R, Huang H, Lee KY, Na J, et al: A population-based study of genes previously implicated in breast cancer. N Engl J Med 2021, 384:440−451

3. Breast Cancer Association Consortium, Dorling L, Carvalho S, Allen J, González-Neira A, Luccarini C, et al: Breast cancer risk genes — association analysis in more than 113,000 women. N Engl J Med 2021, 384:428−439

4. Dalmasso B, Pastorino L, Nathan V, Shah NN, Palmer JM, Howlie M, et al: Germline ATM variants predispose to melanoma: a joint analysis across the GenoMEL and MelaNostrum consortia. Genet Med 2021, 23:2087−2095

5. Kaur H, Salles DC, Murali S, Hicks JL, Nguyen M, Pritchard CC, de Marzo AM, Lanchbury JS, Trock BJ, Isaacs WB, Timms KM, Antonarakis ES, Lotan TL: Genomic and clinicopathologic characterization of ATM-deficient prostate cancer. Clin Cancer Res 2020, 26:4869−4881

6. Hannan Z, Yu S, Mamtani R, Reiss KA: Clinical characteristics of patients with pancreatic cancer and pathogenic ATM alterations. JNCI Cancer Spectr 2021, 5:pkaa121

7. McDuff SGR, Bellon JR, Shannon KM, Gadd MA, Dunn S, Rosenstein BS, Ho AY: ATM variants in breast cancer: implications for breast radiation therapy treatment recommendations. Int J Radiat Oncol Biol Phys 2021, 110:1373−1382

8. Shiloh Y: The cerebellar degeneration in ataxia-telangiectasia: a case for genome instability. DNA Repair (Amst) 2020, 95:102950

9. Feliubadaló L, Moles-Fernández A, Santamariña-Pena M, Sánchez AT, López-Novo A, Porras L-M, Blanco A, Capellá G, de la Hoya M, Molina IJ, Osorio A, Pineda M, Rueda D, de la Cruz X, Diez O, Ruiz-Ponte C, Gutiérrez-Enríquez S, Vega A, Lázaro C: A collaborative effort to define classification criteria for ATM variants in hereditary cancer patients. Clin Chem 2021, 67:518−533

10. González-Santiago S, Ramón Y Cajal T, Aguirre E, Alés-Martínez JE, Andrés R, Balmaña J, Graña B, Herrero A, Llort G, González-del-Alba A; the SEOM Hereditary Cancer Working Group: SEOM clinical guidelines in hereditary breast and ovarian cancer (2019). Clin Transl Oncol 2020, 22:193−200

11. Lazzari G, Buono G, Zannino B, Silvano G: Breast cancer adjuvant radiotherapy in brca1/2, tp53, atm genes mutations: are there solved issues? Breast Cancer 2021, 13:299−310

12. Pejaver V, Byrne AB, Feng B, Radivojac P, Brenner SE, Pejaver V, Byrne AB, Feng B, Pagel KA, Mooney SD, Karchin R, Donnell-luria AO, Harrison SM, Tavtigian SV, Greenblatt MS, Biesecker LG, Radivojac P, Brenner SE; Sequence Variant Interpretation Working Group: Calibration of computational tools for missense variant pathogenicity classification and ClinGen recommendations for PP3/BP4 criteria. Am J Hum Genet 2022, 109:2163−2177

13. Özkan S, Padilla N, Moles-Fernández A, Diez O, Gutiérrez-Enríquez S, de la Cruz X: The computational approach to variant interpretation: principles, results, and applicability. Edited by Lázaro C, Lerner-Ellis J, Spurdle A. In Clinical DNA Variant Interpretation. Theory and Practice. San Diego, CA: Elsevier Inc./Academic Press, 2021. pp. 89−119

14. Couch FJ, Richardson M, Holdren M, Brannan T, Hiraki S, Holdren M, et al: ClinGen Hereditary Breast, Ovarian and Pancreatic Cancer Expert Panel Specifications to the ACMG/AMP Variant Interpretation Guidelines for ATM Version 1.1, Clinical Genome Resource (ClinGen). Bethesda, MD: National Human Genome Research Institute, 2022

15. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M,

16. Duchesnay E: Scikit-learn: machine learning in python. J Mach Learn Res 2011, 12:2825−2830

16. Waskom M: Seaborn: statistical data visualization. J Open Source Softw 2021, 6:3021

17. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR: PolyPhen-2: prediction of functional effects of human nsSNPs. Nat Methods 2010, 7:248−249

18. Riera C, Lois S, de la Cruz X: Prediction of pathological mutations in proteins: the challenge of integrating sequence conservation and structure stability principles. WIREs Comput Mol Sci 2014, 4:249−268

19. Padilla N, Moles-Fernández A, Riera C, Montalban G, Özkan S, Ootes L, Bonache S, Díez O, Gutiérrez-Enríquez S, de la Cruz X: BRCA1- and BRCA2-specific in silico tools for variant interpretation in the CAGI 5 ENIGMA challenge. Hum Mutat 2019, 40:1593−1611

20. Riera C, Padilla N, de la Cruz X: The complementarity between protein-specific and general pathogenicity predictors for amino acid substitutions. Hum Mutat 2016, 37:1013−1024

21. Bondi A: van der Waals volumes and radii. J Phys Chem 1964, 68:441−451

22. Fauchere J, Pliska V: Hydrophobic parameters of amino acid side-chains from the partitioning of N-acetyl-amino-acid amides. Eur J Med Chem Chim Ther 1983, 18:369−375

23. Henikoff S, Henikoff JG: Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci U S A 1992, 89:10915−10919

24. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al: REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. Am J Hum Genet 2016, 99:877−885

25. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M: CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res 2019, 47:D886−D894

26. Carter H, Douville C, Stenson PD, Cooper DN, Karchin R: Identifying Mendelian disease genes with the variant effect scoring tool. BMC Genomics 2013, 14 Suppl 3:S3

27. Kumar P, Henikoff S, Ng PC: Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009, 4:1073−1081

28. Tavtigian SV, Deffenbaugh AM, Yin L, Judkins T, Scholl T, Samollow PB, De Silva D, Zharkikh A, Thomas A: Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. J Med Genet 2006, 43:295−305

29. Probst P, Wright MN, Boulesteix AL: Hyperparameters and tuning strategies for random forest. Wiley Interdiscip Rev Data Min Knowl Discov 2019, 9:e1301

30. Bishop C: Pattern Recognition and Machine Learning. ed 1. New York, NY: Springer US, 2006

31. Frazer J, Notin P, Dias M, Gomez A, Min JK, Brock K, Gal Y, Marks DS: Disease variant prediction with deep generative models of evolutionary data. Nature 2021, 599:91−95

32. Ferrer-Costa C, Orozco M, de la Cruz X: Sequence-based prediction of pathological mutations. Proteins Struct Funct Genet 2004, 57:811−819

33. López-Ferrando V, Gazzo A, De La Cruz X, Orozco M, Gelpí JL: PMut: a web-based tool for the annotation of pathological variants on proteins, 2017 update. Nucleic Acids Res 2017, 45:W222−W228

34. Liu X, Wu C, Li C, Boerwinkle E: dbNSFP v3.0: a one-stop database of functional predictions and annotations for human nonsynonymous and splice-site SNVs. Hum Mutat 2016, 37:235−241

35. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP: Predicting the functional effect of amino acid substitutions and indels. PLoS One 2012, 7:e46688

36. Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, Liu X: Comparison and integration of deleteriousness prediction methods for

nonsynonymous SNVs in whole exome sequencing studies. Hum Mol Genet 2015, 24:2125−2137

37. Raimondi D, Tanyalcin I, FertCrossed JSD, Gazzo A, Orlando G, Lenaerts T, Rooman M, Vranken W: DEOGEN2: prediction and interactive visualization of single amino acid variant deleteriousness in human proteins. Nucleic Acids Res 2017, 45:W201−W206

38. Reva B, Antipin Y, Sander C: Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res 2011, 39:e118

39. Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GLA, Edwards KJ, Day INM, Gaunt TR: Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden markov models. Hum Mutat 2013, 34:57−65

40. Chun S, Fay JC: Identification of deleterious mutations within three human genomes. Genome Res 2009, 19:1553−1561

41. Lu Q, Hu Y, Sun J, Cheng Y, Cheung KH, Zhao H: A statistical framework to predict functional non-coding regions in the human genome through integrated analysis of annotation data. Sci Rep 2015, 5:10576

42. Niroula A, Urolagin S, Vihinen M: PON-P2: prediction method for fast and reliable identification of harmful variants. PLoS One 2015, 10:e0117380

43. Vihinen M: How to evaluate performance of prediction methods? measures and their interpretation in variation effect analysis. BMC Genomics 2012, 13:S2

44. Baldi P, Brunak S, Chauvin Y, Andersen CAF, Nielsen H: Assessing the accuracy of prediction algorithms for classification: an overview. Bioinformatics 2000, 6:412−424

45. Rudin C: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 2019, 1:206−215

46. Cannarssa M: Ethics guidelines for trustworthy AI. Edited by DiMatteo L, Janssen A, Ortolani P, De Elizalde F, Cannarsa M, Durovic M. In The Cambridge Handbook of Lawyering in the Digital Age. Cambridge, UK: Cambridge University Press, 2021. pp. 283−297

47. Shendure J, Findlay GM, Snyder MW: Genomic medicine−progress, pitfalls, and promise. Cell 2019, 177:45−57