

An interpretable deep learning model for detecting *BRCA1/2* pathogenic variants of breast cancer from hematoxylin and eosin-stained pathological images (#98990)

1

First submission

Guidance from your Editor

Please submit by **8 May 2024** for the benefit of the authors (and your token reward) .



Structure and Criteria

Please read the 'Structure and Criteria' page for general guidance.



Custom checks

Make sure you include the custom checks shown below, in your review.



Author notes

Have you read the author notes on the [guidance page](#)?



Raw data check

Review the raw data.



Image check

Check that figures and images have not been inappropriately manipulated.

If this article is published your review will be made public. You can choose whether to sign your review. If uploading a PDF please remove any identifiable information (if you want to remain anonymous).

Files

Download and review all files from the [materials page](#).

11 Figure file(s)

4 Table file(s)

4 Raw data file(s)

! Custom checks

Human participant/human tissue checks



Have you checked the authors [ethical approval statement](#)?



Does the study meet our [article requirements](#)?



Has identifiable info been removed from all files?



Were the experiments necessary and ethical?



Structure and Criteria

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. BASIC REPORTING
2. EXPERIMENTAL DESIGN
3. VALIDITY OF THE FINDINGS
4. General comments
5. Confidential notes to the editor

 You can also annotate this PDF and upload it as part of your review

When ready [submit online](#).

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your [guidance page](#).

BASIC REPORTING

-  Clear, unambiguous, professional English language used throughout.
-  Intro & background to show context. Literature well referenced & relevant.
-  Structure conforms to [PeerJ standards](#), discipline norm, or improved for clarity.
-  Figures are relevant, high quality, well labelled & described.
-  Raw data supplied (see [PeerJ policy](#)).

EXPERIMENTAL DESIGN

-  Original primary research within [Scope of the journal](#).
-  Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
-  Rigorous investigation performed to a high technical & ethical standard.
-  Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

-  Impact and novelty not assessed. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
-  All underlying data have been provided; they are robust, statistically sound, & controlled.
-  Conclusions are well stated, linked to original research question & limited to supporting results.



The best reviewers use these techniques

Tip

Example

Support criticisms with evidence from the text or from other sources

Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.

Give specific suggestions on how to improve the manuscript

Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).

Comment on language and grammar issues

The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult. I suggest you have a colleague who is proficient in English and familiar with the subject matter review your manuscript, or contact a professional editing service.

Organize by importance of the issues, and number your points

1. Your most important issue
2. The next most important item
3. ...
4. The least important points

Please provide constructive criticism, and avoid personal opinions

I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC

Comment on strengths (as well as weaknesses) of the manuscript

I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.

An interpretable deep learning model for detecting *BRCA1/2* pathogenic variants of breast cancer from hematoxylin and eosin-stained pathological images

Yi Li^{Equal first author, 1, 2}, Xiaomin Xiong^{Equal first author, 1, 2}, Xiaohua Liu³, Yihan Wu², Xiaoju Li⁴, Bo Liu², Yu Li⁴, Bo Lin^{Corresp., 2}, Bo Xu^{Corresp. 1, 2}

¹ School of Medicine, Chongqing University, Chongqing, China

² Chongqing Key Laboratory of Intelligent Oncology for Breast Cancer, Chongqing University Cancer Hospital, Chongqing, China

³ Bioengineering College of Chongqing University, Chongqing, China

⁴ Department of Pathology, Chongqing University Cancer Hospital, School of Medicine, Chongqing University, Chongqing, China

Corresponding Authors: Bo Lin, Bo Xu

Email address: linbo@cqu.edu.cn, xubo731@cqu.edu.cn

Background: Determining the *BRCA1/2* status is crucial for guiding the treatment of breast cancer. However, there is an unmet need for *BRCA1/2* genetic testing among breast cancer patients due to high costs and limited resources. We aimed to establish a Bi-directional Self-Attention Multiple Instance Learning (BiAMIL) algorithm to detect *BRCA1/2* status based on H&E pathological images. **Method:** A total of 319 histopathological slides from 254 patients with breast cancer were included in our study, consisting of two dependent cohorts. After pre-processing the images, 633,484 tumor tiles from the training dataset were used to train our self-developed deep-learning model. The performance of the network was evaluated in the internal and external test sets. **Results:** The results showed that BiAMIL achieved AUC values of 0.917 (95% CI 0.874-0.962) in the training set, 0.819 (95% CI 0.673-0.965) in the internal test set, and 0.817 (95% CI 0.712-0.923) in the external test set. To explore the relationship between *BRCA1/2* pathogenic variants (PV) and human-interpretable features in pathological images, we visualized the heat map of tiles with high-attention scores using Class Activation Mapping (CAM). BiAMIL mainly focused on high-grade tumors and lymphocytic infiltration, the crucial tissue features closely related to *BRCA1/2* PV. Additionally, through unsupervised cell feature clustering analysis, BiAMIL considered morphological features of tumor cell nuclear areas, which are the primary cell features of *BRCA1/2* PV. **Conclusions:** We developed an interpretable deep neural network model based on the attention mechanism to predict the *BRCA1/2* status in breast cancer.

An interpretable deep learning model for detecting *BRCA1/2* pathogenic variants of breast cancer from hematoxylin and eosin-stained pathological images

Yi Li^{1,2*}, Xiaomin Xiong^{1,2*}, Xiaohua Liu³, Yihan Wu², Xiaoju Li⁴, Bo Liu², Yu Li⁴, Bo Lin², Bo Xu^{1,2}

*These authors contributed equally to this work.

¹ School of Medicine, Chongqing University, Chongqing, China

² Chongqing Key Laboratory of Intelligent Oncology for Breast Cancer, Chongqing University Cancer Hospital, Chongqing, China

³ Bioengineering College of Chongqing University, Chongqing, China

⁴ Department of Pathology, Chongqing University Cancer Hospital, School of Medicine, Chongqing University, Chongqing, China

Corresponding Author:

Bo Lin²

Email address: linbo@cqu.edu.cn

Bo Xu^{1,2}

Email address: xubo731@cqu.edu.cn

Abstract

Background: Determining the *BRCA1/2* status is crucial for guiding the treatment of breast cancer. However, there is an unmet need for *BRCA1/2* genetic testing among breast cancer patients due to high costs and limited resources. We aimed to establish a Bi-directional Self-Attention Multiple Instance Learning (BiAMIL) algorithm to detect *BRCA1/2* status based on hematoxylin and eosin (H&E) pathological images.

Method: A total of 319 histopathological slides from 254 patients with breast cancer were included in our study, consisting of two dependent cohorts. After pre-processing the images, 633,484 tumor tiles from the training dataset were used to train our self-developed deep-learning model. The performance of the network was evaluated in the internal and external test sets.

Results: The results showed that BiAMIL achieved AUC values of 0.917 (95% CI 0.874-0.962) in the training set, 0.819 (95% CI 0.673-0.965) in the internal test set, and 0.817 (95% CI 0.712-0.923) in the external test set. To explore the relationship between *BRCA1/2* pathogenic variants (PV) and human-interpretable features in pathological images, we visualized the heat map of tiles with high-attention scores using Class Activation Mapping (CAM). BiAMIL mainly focused on high-grade tumors and lymphocytic infiltration, the crucial tissue features closely related to *BRCA1/2* PV. Additionally, through unsupervised cell feature clustering analysis, BiAMIL considered morphological features of tumor cell nuclear areas, which are the primary cell features of *BRCA1/2* PV.

Conclusions: We developed an interpretable deep neural network model based on the attention mechanism to predict the *BRCA1/2* status in breast cancer.

Keywords: Breast cancer, *BRCA1/2*, Deep learning, Self-attention, Interpretability

Introduction

Breast cancer is the most common cancer worldwide and is the fourth primary cause of cancer-related mortality in women (Sung et al. 2021). The usual treatment for breast cancer involves a variety of therapies, such as chemotherapy, radiotherapy, endocrine therapy, and targeted therapy, which are selected based on the specific molecular subtype of the disease (Asleh et al. 2022; Shubeck et al. 2023). Further genetic testing may be necessary to choose the most appropriate treatment for each patient. One of the genetic alterations clinically associated with breast cancer treatment is *BRCA1/2*, which affects approximately 3-5% of breast cancer patients (Schettini et al. 2021). Recent studies have revealed that *BRCA1/2* mutations can serve as predictive biomarkers for the response to treatment with Poly (ADP-ribose) polymerase inhibitors (PARPi) and chemotherapy based on platinum for breast cancer (Chopra et al. 2020; Tutt et al. 2021).

According to guidelines from the National Comprehensive Cancer Network (NCCN) and the American Society of Breast Surgeons (ASBrS), it is recommended that breast cancer patients, especially those diagnosed at a young age (≤ 50 years), with a family history of cancer, or with bilateral breast cancers, should receive *BRCA1/2* genetic testing (Daly et al. 2021; Valencia et al. 2017). However, many patients who meet the above criteria have not been tested for *BRCA1/2* in medical practices due to the test's complexity, time-consuming, and high costs (Grindedal et al. 2017). Therefore, there is a growing need to develop a fast, affordable, and reliable approach for assessing the *BRCA1/2* status.

In recent years, artificial intelligence (AI) has become increasingly important in various aspects of tumor screening, diagnosis, therapeutic evaluation, and prognosis prediction (Bhinder et al. 2021; Lin et al. 2022). Specifically, deep learning (DL), a branch of AI technology, has emerged as a powerful tool for extracting abundant hidden information from digital whole-slide

65 images (WSIs) (Corti et al. 2022; Freeman et al. 2021; Schrammen et al. 2022). Hematoxylin
 66 and eosin (H&E)-stained slides contain enormous information about molecular features, cell
 67 morphology, and tissue structure, which can help reveal changes in molecular biomarkers
 68 (Greenson et al. 2009; Shia et al. 2017). Genetic alterations in tumor cells cause functional
 69 changes that affect the morphology of tumor cells and can also have an impact on the tumor
 70 microenvironment, resulting in genotype-phenotype correlations (Kather et al. 2020). The
 71 morphological features associated with *BRCA1/2* mutations, such as high histological grade, a
 72 high mitotic index, pushing tumor margins, and lymphocytic infiltration, can be reflected in
 73 H&E-stained images. Recent studies have shown that DL enables the detection of genetic
 74 alterations from histopathology images (Cifci et al. 2022). For example, DL models can predict
 75 *BRAF* mutation and microsatellite instability (MSI) in colorectal cancer (Guo et al. 2023),
 76 *BRCA1/2* mutations in breast cancer (Wang et al. 2021), *IDH1* mutation in brain cancer (Jiang et
 77 al. 2021), and *CTNNB1* mutation in hepatocellular carcinoma (Liao et al. 2020) based on WSIs.
 78 Although these techniques demonstrate high precision in identifying genetic changes,
 79 understanding them is still challenging because DL models are often viewed as "black boxes" in
 80 the decision-making process (Vinuesa & Sirmacekc 2021). Several studies have employed
 81 visualization strategies to recognize morphological features identified by DL frameworks.
 82 However, these features often display inaccuracies, inconsistency, and a lack of transparency,
 83 making it difficult to explain the model's predictions (Singh et al. 2020). The main reason is that,
 84 due to the limitations and diversity of training data, the features learned by the models may not
 85 be universal, resulting in inconsistency across different datasets. Furthermore, some visualization
 86 techniques may not fully capture all the subtle features during the prediction process, and the
 87 generated heatmaps do not provide a reasonable medical explanation (Tizhoosh & Pantanowitz

2018). Therefore, there is a need to provide replicable descriptions of these features and evaluate their impact on explaining the DL model, which remains a challenge.

In this study, we developed an interpretable DL network named Bi-directed Self-Attention Multi-Instance Learning (BiAMIL) to detect *BRCA1/2* status from H&E images in breast cancer. To apply human-interpretable features analysis for the model, we visualized heatmaps and analyzed morphological patterns significantly associated with *BRCA1/2* pathogenic variants (PV). Additionally, we identified quantifiable changes in cell features that are highly correlated with *BRCA1/2* PV through unsupervised clustering (Fig. 1).

Materials & Methods

Patient selection

A total of 254 breast cancer patients in two dependent cohorts were enrolled in this retrospective study. The first cohort comprised 152 patients (217 WSIs) who were obtained from Chongqing University Cancer Hospital (CUCH) between May 2017 and December 2022. The second cohort comprised 102 patients (102 WSIs) who were obtained from the Cancer Genome Atlas (<https://gdc.cancer.gov/>). The inclusion criteria included: 1) Diagnosis with primary breast invasive ductal cancer (IDC) and invasive lobular carcinoma (ILC); 2) *BRCA1/2* gene or genomic testing was conducted on blood and/or tumor samples, including the examination of germline and/or somatic variations, confirming the precise records of *BRCA1/2* status; 3) Operative pathological H&E images before antineoplastic treatments were available; The exclusion criteria included: 1) Lack of operative pathological H&E images; 2) Poor quality of H&E-stained images; 3) Bilateral, multifocal, or special invasive breast cancer. The study received approval from the Ethics Committee of Chongqing University Cancer Hospital (Ethics number: CZLS2023213-A), and patient consent was waived for this retrospective analysis. The

CUCH cohort was randomly divided into a training set and an internal test set. The TCGA cohort served as an independent external test set. The training set was utilized for hyperparameter tuning through cross-validation, while the test sets were employed to assess generalization performance.

Genomic DNA samples were extracted from peripheral blood and/or surgical tissue samples. The *BRCA1/2* genetic testing was conducted using next-generation sequencing (NGS) technology. Variants were annotated using the Human Genome Variation Society (HGVS) nomenclature guidelines (<http://varnomen.hgvs.org/>). The biological significance of all reported variants was assessed using the ClinVar database (www.ncbi.nlm.nih.gov/clinvar/). According to the guidelines of the American College of Medical Genetics and Genomics (ACMG) (Richards et al. 2015), the detected variants were classified as pathogenic variants, likely pathogenic variants, variants of uncertain significance, likely benign, or benign. Based on the testing results, only patients who identified with pathogenic and likely pathogenic variants were classified as having PV. The *BRCA1/2* PV group includes *BRCA1* PV and *BRCA2* PV. In this study, we focused on predicting these two binary outcomes, *BRCA1/2* PV and *BRCA1/2* wild type (WT) for breast cancer patients.

Image preprocessing and sample preparation

All samples were fixed using 4% neutral formalin, followed by paraffin embedding, cut into 4 μ m thick sections, and then stained with H&E. The H&E-stained histopathological slides were scanned at a magnification level of 40 $\times$ using a KFBIO KF-PRO-005 digital scanner and saved in SVS format. Experienced pathologists with over one year in breast cancer pathology conducted quality control on the images. Slides of poor quality, especially those affected by tissue folds and blurriness, were excluded. The TCGA cohort exclusively utilized 40 $\times$ FFPE

WSIs labeled as diagnostic while excluding slides with poor image quality (such as pen marks or poor staining). The WSIs were segmented into tiles of 512×512 pixels, with a 50% overlap. Tiles that displayed more than 75% of the background were removed. The observed variations in color can be attributed to differences in raw materials, manufacturing processes, staining methods, and different digital scanners. These variations in color may cause the model to concentrate on color distinctions over the essential tissue morphology for analysis. To resolve this issue, color normalization was applied to all tiles using the structure-preserving color normalization (SPCN) technique (Vahadane et al. 2016). The model facilitated standardizing all tiles to resemble the color pattern of the target (Fig. S1). Various types of data augmentation techniques were employed, including random flipping, random rotation, cropping, and adjustments to brightness, contrast, saturation, and hue.

We used a stratified sampling approach to randomly assign images to the training and internal test set, maintaining an 8:2 ratio at the patient level. This method ensures that images of the same patient are exclusively assigned to either the training set or the internal test set. It guarantees that images from the same patient do not appear in both sets simultaneously.

Segmentation network

We constructed a segmentation network that could automatically identify tumor regions in WSIs. Two experienced pathologists, each with more than three years of experience in breast cancer pathology, annotated tumor areas on 145 WSIs using the open-source software Qupath Image Scope. These annotated areas were then independently reviewed by a pathologist with over ten years of experience. The labeled slides were randomly divided into a training set (80%) and a test set (20%), ensuring that WSIs from the same patient did not appear in both sets. Based on the ResNet 34 architecture, the segmentation network used tiles from tumor and non-tumor

regions as input. The deep convolutional and pooling layers were utilized to extract features from each tile. These features were then passed to the fully connected layer. After the fully connected layer, a softmax function was applied to generate a probability (tumor or non-tumor) for each tile. The results were compared with the manual annotations made by three pathologists to evaluate the performance of the network. These pathologists included junior, medium, and senior pathologists with two, from five to eight, and ten years of experience, respectively. Finally, this segmentation network was utilized to automatically segment tumor regions in the TCGA cohort.

Prediction network

The CUCH cohort consisted of 217 WSIs from 152 patients, each with one or two slides. The dataset was randomly divided into a training set (633,484 tiles from 174 WSIs) and an internal test set (112,096 tiles from 43 WSIs) through stratified sampling. We developed the BiAMIL network using a Multi-Instance Learning (MIL) framework and attention mechanisms. The BiAMIL architecture comprised three main components: the feature extraction module, the bi-directional self-attention module, and the classification module. The feature extraction module was based on the ResNet 34 model (Sun 2016). A bag containing N image tiles was fed into the feature extractor, resulting in a feature matrix with dimensions of $N \times 1000$. The feature matrix was then passed into the attention module for aggregation. In the bi-directional self-attention module, the original data features were transformed into embedding vectors using a feature embedding layer. Two attention heads were then designed: a high-risk PV head and a low-risk PV head. Specifically, the attention distribution among tiles was adjusted by employing the softmax function, converting the $N \times 1000$ feature matrix into a $1 \times N$ attention weight. The weights obtained from the two attention heads were used to multiply the original feature matrix. Two sets of attention-weighted discriminative feature vectors were generated, each set having a

dimensionality of 1000. These two 1000-dimensional integrated feature vectors were merged and inputted into the classification module. The classification module consisted of a three-layer Multilayer Perceptron (MLP) with 1000, 64, and 2 neurons in each layer, respectively, utilizing the Tanh function as the activation function. The prediction probability for each instance bag was ultimately output through the softmax function. To optimize the model, we employed a 5-fold cross-validation by randomly dividing the training set into five balanced subsets (Fig. S2). Using the Adam optimizer, the optimal hyperparameters were identified as a bag of $N = 35$, a learning rate of 0.001, a batch size of 18, and total epochs of 20 (with the learning rate decreasing by 0.1 every ten epochs). At last, the network was evaluated on the internal and external test sets.

Tissue features visualization and cell features quantitative analysis

To further understand the critical histological features that contribute the most to the prediction of the BiAMIL model, we selected the top 20% of tiles based on their attention scores in predicting *BRCA1/2* PV. Attention-based visualization techniques were then employed to visually represent these tiles. Firstly, we calculated the weight values for each tile. Secondly, we identified the top 20% of tiles in WSIs with high attention weights and performed Smooth Grad-CAM to identify the regions within each tile that the neural network utilized to generate predictions.

Cell features, including cell morphology, nuclear morphology, and staining features, were extracted from the tumor regions using QuPath (Bankhead et al. 2017). The k -means clustering method was applied to partition cells into different clusters based on their cell features. The resulting clusters were then projected onto a 2-D UMAP projection (Dorrity et al. 2020). The difference in cell features between *BRCA1/2* PV and WT was evaluated by the Wilcoxon rank-sum test.

Statistical analysis

We evaluated the performance of the model at the slide level. The model generated a probability value for each bag, and we assigned a final probability to each slide by calculating the average probability across all bags in the slide. The performance of the model was evaluated using several measures, including the area under the receiver operating characteristic curve (AUROC), accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and the F1 score. The most relevant cell features for discriminating between *BRCA1/2* PV and *BRCA1/2* WT were identified using the Wilcoxon rank-sum test. All statistical analyses were performed using scikit-learn 0.24.2 in Python and R (version 4.1.1), with a p-value of less than 0.05 was considered statistically significant.

Results

Clinical characteristics of patients

A total of 254 patients from two independent cohorts were included in our study based on the inclusion and exclusion criteria. Specifically, patients from the CUCH cohort (n=152) were randomly assigned to a training set and an internal test set, while patients from the TCGA cohort (n=102) served as an external test set. The detailed clinicopathological characteristics of the patients in the training, internal test, and external test sets are provided in Table 1. In the training set, 62 (50.8%) of patients were ER and PR positive, 12 (9.8%) were HER2 positive, and 48 (39.4%) were negative for ER, PR, and HER2. In the internal test set, 16 (53.4%) of patients were ER and PR positive, 5 (16.6%) were HER2 positive, and 9 (30.0%) were negative for ER, PR, and HER2. In the external testing group, 70 (68.6%) of patients were ER and PR positive, 11 (10.8%) were HER2 positive, and 18 (17.6%) were negative for ER, PR, and HER2. In the training group, 29 patients had *BRCA* PV, with 17 (58.6%) for *BRCA1* and 12 (41.3%) for

BRCA2. The internal test set consisted of seven patients with *BRCA* PV, including 4 (57.1%) with *BRCA1* and 3 (42.9%) with *BRCA2*. In the external test group, among the 26 patients with *BRCA* PV, there were 13 (50.0%) with *BRCA1*, 11 (42.3%) with *BRCA2*, and 2 (7.7%) with both *BRCA1* and *BRCA2*.

Performance of the segmentation network

We developed a segmentation network to accurately distinguish tumor regions from WSIs. The segmentation model demonstrated high performance in identifying tumor regions, achieving an AUC of 0.960 (95% CI, 0.959-0.961). The accuracy, sensitivity, and specificity were 0.888 (95% CI, 0.887-0.890), 0.859 (95% CI, 0.856-0.861), and 0.908 (95% CI, 0.906-0.909), respectively. Additionally, compared with three pathologists at different levels (junior, medium, and senior), the performance of our segmentation model is almost equal to that of a medium-level pathologist (Fig. 2A). The confusion matrix is presented in Fig. 2B. Finally, typical examples of the segmentation model output are presented in Fig. 2C, where the tumor regions are highlighted in red.

Performance of BiAMIL network

To demonstrate that focusing on the most relevant and important regions within WSIs can improve the prediction accuracy of *BRCA1/2* PV, we developed the BiAMIL model, which adds an attention structure to ResNet 34. Color-normalized tiles were fed into a pre-trained ResNet 34 model to extract a 1000-dimensional feature vector. These feature vectors were then stacked and input into an attention-based Multiple Instance Learning (MIL) framework to predict the probability of *BRCA1/2* PV (Fig. 3). BiAMIL achieved AUC values of 0.917 (95% CI 0.874-0.962) in the training set, 0.819 (95% CI 0.673-0.965) in the internal test set, and 0.817 (95% CI 0.712-0.923) in the external test set (Fig. 4 and Table S1). Comparatively, ResNet 34 achieved

AUC values of 0.874 (95% CI, 0.700-0.978) in the training set, 0.783 (95% CI 0.624-0.941) in the internal test set, and 0.684 (95% CI, 0.559-0.810) in the external test sets (Fig. 4 and Table S2). The BiAMIL model outperformed the ResNet 34 model in terms of accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and the F1 score (Fig. 5). Due to hormone receptor-positive and triple-negative breast cancer accounting for 80% of breast cancer cases and the high occurrence of *BRCA1/2* in triple-negative breast cancer, we conducted a subgroup analysis of BiAMIL's performance in hormone receptor-positive and triple-negative breast cancer. We obtained similar results in Table 2. These findings demonstrate that the BiAMIL model is a more accurate and reliable approach for estimating the *BRCA1/2* status compared to the traditional ResNet 34 model.

Interpretability analysis of tissue features

To explore the relationship between pathological tissue features and *BRCA1/2* PV, we conducted a post-hoc explanation of the BiAMIL model using Class Activation Mapping (CAM). The model generated heatmaps for the top 20% of tiles based on attention scores, with the areas that contributed the most displayed in red. Overlaying the original image tiles with the CAM analysis results, we observed that the regions with high-grade tumors and lymphocytic infiltration exhibited the most significant contributions in our model. In contrast, areas associated with the stromal matrix demonstrated comparatively lower contributions. These findings indicate that BiAMIL identifies the known morphological features of *BRCA1/2* PV (Fig. 6). Subsequently, we used t-distributed Stochastic Neighbor Embedding (t-SNE) to perform cluster analysis on the features generated by BiAMIL, where patients in the same class remained clustered and clearly separated from those in the other class (Fig. 7).

Interpretability analysis of cell features

We employed the Qupath software to automatically identify tumor cells in regions that significantly contributed to the BiAML model. Pathologists reviewed and modified the tumor cells delineated in each WSI, extracting relevant features such as cell shape, color, and other characteristics (Fig. 8A). The cells were then clustered into six groups using the *k*-means algorithm based on single-cell features, which corresponded to cell solidity, nuclear eccentricity, nuclear area, nuclear roundness, nuclear staining, and cytoplasmic staining (Fig. 8B). Notably, the features related to nuclear size and shape were particularly enriched in Cluster 2 (Fig. 8C, Fig. S3). To further investigate nuclear morphological differences between the *BRCA1/2* PV and WT groups, we analyzed the changes in nuclear features within Cluster 2. Specifically, compared to the WT group, tumor cells in the *BRCA1/2* PV group exhibited a larger nuclear area (180.00 versus 168.75, $p < 0.001$) and a higher nuclear cell area ratio (0.34 versus 0.36, $p < 0.001$) (Fig. 8D). At the level of cell feature interpretability, nuclear size and the nuclear cell area ratio of tumor cells emerged as the most distinctive features distinguishing the *BRCA1/2* PV group from the WT group. We showed examples of high tumor cell nuclear areas and low tumor cell nuclear areas (Fig. 8E), each associated with *BRCA1/2* PV and WT, respectively.

Discussion

Accurately identifying *BRCA1/2* status is crucial for clinical decision-making, as it facilitates the selection of appropriate therapeutic agents and enables effective patient management. However, due to costs and resource constraints, *BRCA1/2* genetic testing is not widely adopted in clinical practice. In this study, we developed a DL model to directly detect the *BRCA1/2* status of breast cancer from histopathology images. Furthermore, we explored the interpretability of the model from the perspectives of tissue and cell features using visualization and clustering techniques, providing medical interpretability for our model.

Wang et al. (Wang et al. 2021) proposed a traditional DL algorithm to predict *BRCA1/2* mutations in breast cancer through H&E pathological images. The study assumed equal contributions of all tiles to predicting *BRCA1/2* mutation status. This assumption might be ineffective and potentially lower the accuracy of the predictions. In practice, each WSI comprises hundreds to thousands of tiles, but most do not significantly contribute to the final prediction. Conversely, only a few key tiles play a significant role in the prediction. Our DL model, based on the attention mechanism (Mobadersany et al. 2021; Yao et al. 2020), autonomously identified the contribution of each tile to the overall WSI-level prediction, allowing the model to concentrate more accurately on the most vital regions within the WSI. This mechanism enables the network to capture crucial information more effectively within the WSI, ultimately improving prediction accuracy. Compared to the ResNet 34 model, our results demonstrated better performance in terms of AUC, sensitivity, specificity, etc.

The interpretability of DL models is a challenge in medical applications (Teng et al. 2022; Yao et al. 2020). These models should provide medical interpretability to enable clinicians to understand better, validate, and trust them in clinical routine (Ibrahim et al. 2020; Liu et al. 2019). To explore this issue, we visualized the important tiles identified by the attention mechanism during the decision-making process, aiming to gain a better understanding of the morphological characteristics related to *BRCA1/2* PV. By this approach, we found that the model primarily captures the features of high-grade tumors and lymphocytic infiltration, which correlate with the predicted *BRCA1/2* PV. Previous studies have reported that the morphological characteristics primarily associated with *BRCA1/2* mutations include the absence of gland formation, a high mitotic index, nuclear pleomorphism, increased necrotic cells, and lymphocytic infiltrates (Lakhani et al. 1998; Larsen et al. 2014). The tissue features extracted by our model

were consistent with previous research findings. More importantly, we conducted further analysis of cell features such as cell morphology, nuclear morphology, and staining to provide a more comprehensive interpretation of the pathological features. Cell morphology, functionality, and genetic characteristics are crucial in predicting *BRCA1/2* mutations (Alizadeh et al. 2020; Roy et al. 2011). Cell morphological features include cell shape, size, structure, etc. *BRCA1/2* mutations may exhibit cellular morphological abnormalities, such as irregular shapes and size variations. Observing these cellular morphological features can provide valuable insights for predicting *BRCA1/2* PV. Therefore, the features of tissues and cells are related to *BRCA1/2* PV, which improves the medical interpretability of the features extracted by our model when determining the *BRCA1/2* status.

There are some limitations in our study. Firstly, the model was built from a relatively small sample size in this retrospective study. To improve the accuracy and applicability of the model, it would be beneficial to collect larger samples from multiple centers. Secondly, the model only focused on assessing *BRCA1/2* status and did not include other genes in the homologous recombination-deficient (HRD) pathway. It is possible that other gene mutations may also contribute to the typical pathological image features. Finally, further investigation is required to establish the correlation between gene-related features extracted from pathological images and the effectiveness of targeted treatment outcomes.

Conclusions

In conclusion, we developed a DL model to detect *BRCA1/2* status directly from histopathological images. The interpretability of this model was explored through the characteristics of pathological tissues and cell features. In the future, further optimization and

validation on larger and more diverse datasets may allow it to serve as a pre-screening tool, providing clinical value in selecting breast cancer patients for *BRCA1/2* genetic testing.

Acknowledgements

We thank all members of the Xu laboratory for technical support and discussion.

References

- Alizadeh E, Castle J, Quirk A, Taylor CDL, Xu W, and Prasad A. 2020. Cellular morphological features are predictive markers of cancer cell state. *Comput Biol Med* 126:104044. 10.1016/j.compbio.2020.104044
- Bankhead P, Loughrey MB, Fernandez JA, Dombrowski Y, McArt DG, Dunne PD, McQuaid S, Gray RT, Murray LJ, Coleman HG, James JA, Salto-Tellez M, and Hamilton PW. 2017. QuPath: Open source software for digital pathology image analysis. *Sci Rep* 7:16878. 10.1038/s41598-017-17204-5
- Bhinder B, Gilvary C, Madhukar NS, and Elemento O. 2021. Artificial Intelligence in Cancer Research and Precision Medicine. *Cancer Discov* 11:900-915. 10.1158/2159-8290.CD-21-0090
- Cifci D, Foersch S, and Kather JN. 2022. Artificial intelligence to identify genetic alterations in conventional histopathology. *J Pathol* 257:430-444. 10.1002/path.5898
- Corti C, Cobanaj M, Marian F, Dee EC, Lloyd MR, Marcu S, Dombrowschi A, Biondetti GP, Batalini F, Celi LA, and Curigliano G. 2022. Artificial intelligence for prediction of treatment outcomes in breast cancer: Systematic review of design, reporting standards, and bias. *Cancer Treat Rev* 108:102410. 10.1016/j.ctrv.2022.102410
- Daly MB, Pal T, Berry MP, Buys SS, Dickson P, Domchek SM, Elkhanany A, Friedman S, Goggins M, Hutton ML, Cgc, Karlan BY, Khan S, Klein C, Kohlmann W, Cgc, Kurian AW, Laronga C, Litton JK, Mak JS, Lcgc, Menendez CS, Merajver SD, Norquist BS, Offit K, Pederson HJ, Reiser G, Cgc, Senter-Jamieson L, Cgc, Shannon KM, Shatsky R, Visvanathan K, Weitzel JN, Wick MJ, Wisinski KB, Yurgelun MB, Darlow SD, and Dwyer MA. 2021. Genetic/Familial High-Risk Assessment: Breast, Ovarian, and Pancreatic, Version 2.2021, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw* 19:77-102. 10.6004/jnccn.2021.0001
- Dorrity MW, Saunders LM, Queitsch C, Fields S, and Trapnell C. 2020. Dimensionality reduction by UMAP to visualize physical and genetic interactions. *Nature communications* 11:1-6.
- Freeman K, Geppert J, Stinton C, Todkill D, Johnson S, Clarke A, and Taylor-Phillips S. 2021. Use of artificial intelligence for image analysis in breast cancer screening programmes: systematic review of test accuracy. *BMJ* 374:n1872. 10.1136/bmj.n1872
- Greenon JK, Huang SC, Herron C, Moreno V, Bonner JD, Tomsho LP, Ben-Izhak O, Cohen HI, Trougouboff P, Bejhar J, Sova Y, Pinchev M, Rennert G, and Gruber SB. 2009. Pathologic predictors of microsatellite instability in colorectal cancer. *Am J Surg Pathol* 33:126-133. 10.1097/PAS.0b013e31817ec2b1
- Grindedal EM, Heramb C, Karsrud I, Ariansen SL, Maehle L, Undlien DE, Norum J, and Schlichting E. 2017. Current guidelines for BRCA testing of breast cancer patients are insufficient to detect all mutation carriers. *BMC Cancer* 17:438. 10.1186/s12885-017-3422-2
- Guo B, Li X, Yang M, Jonnagaddala J, Zhang H, and Xu XS. 2023. Predicting microsatellite instability and key biomarkers in colorectal cancer from H&E-stained images: achieving state-of-the-art predictive performance with fewer data using Swin Transformer. *J Pathol Clin Res* 9:223-235. 10.1002/cjp.2312
- Ibrahim A, Gamble P, Jaroensri R, Abdelsamea MM, Mermel CH, Chen PC, and Rakha EA. 2020. Artificial intelligence in digital breast pathology: Techniques and applications. *Breast* 49:267-273. 10.1016/j.breast.2019.12.007
- Jiang S, Zanazzi GJ, and Hassanpour S. 2021. Predicting prognosis and IDH mutation status for patients with lower-grade gliomas using whole slide images. *Sci Rep* 11:16849. 10.1038/s41598-021-95948-x
- Lakhani SR, Jacquemier J, Sloane JP, Gusterson BA, Anderson TJ, van de Vijver MJ, Farid LM, Venter D, Antoniou A, Storer-Isser A, Smyth E, Steel CM, Haite N, Scott RJ, Goldgar D, Neuhausen S, Daly PA, Ormiston W, McManus R, Scherneck S, Ponder BA, Ford D, Peto J, Stoppa-Lyonnet D, Bignon YJ,

- Struwing JP, Spurr NK, Bishop DT, Klijn JG, Devilee P, Cornelisse CJ, Lasset C, Lenoir G, Barkardottir RB, Egilsson V, Hamann U, Chang-Claude J, Sobol H, Weber B, Stratton MR, and Easton DF. 1998. Multifactorial analysis of differences between sporadic breast cancers and cancers involving BRCA1 and BRCA2 mutations. *J Natl Cancer Inst* 90:1138-1145. 10.1093/jnci/90.15.1138
- Larsen MJ, Thomassen M, Gerdes AM, and Kruse TA. 2014. Hereditary breast cancer: clinical, pathological and molecular characteristics. *Breast Cancer (Auckl)* 8:145-155. 10.4137/BCBCR.S18715
- Liao H, Long Y, Han R, Wang W, Xu L, Liao M, Zhang Z, Wu Z, Shang X, Li X, Peng J, Yuan K, and Zeng Y. 2020. Deep learning-based classification and mutation prediction from histopathological images of hepatocellular carcinoma. *Clin Transl Med* 10:e102. 10.1002/ctm2.102
- Lin B, Tan Z, Mo Y, Yang X, Liu Y, and Xu B. 2022. Intelligent oncology: The convergence of artificial intelligence and oncology. *Journal of the National Cancer Center*.
- Liu Z, Wang S, Dong D, Wei J, Fang C, Zhou X, Sun K, Li L, Li B, Wang M, and Tian J. 2019. The Applications of Radiomics in Precision Diagnosis and Treatment of Oncology: Opportunities and Challenges. *Theranostics* 9:1303-1322. 10.7150/thno.30309
- Mobadersany P, Cooper LAD, and Goldstein JA. 2021. GestAltNet: aggregation and attention to improve deep learning of gestational age from placental whole-slide images. *Laboratory Investigation* 101:942-951. 10.1038/s41374-021-00579-5
- Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL, and Committee ALQA. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 17:405-424. 10.1038/gim.2015.30
- Roy R, Chun J, and Powell SN. 2011. BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nat Rev Cancer* 12:68-78. 10.1038/nrc3181
- Schrammen PL, Ghaffari Laleh N, Echle A, Truhn D, Schulz V, Brinker TJ, Brenner H, Chang-Claude J, Alwers E, Brobeil A, Kloor M, Heij LR, Jager D, Trautwein C, Grabsch HI, Quirke P, West NP, Hoffmeister M, and Kather JN. 2022. Weakly supervised annotation-free cancer detection and prediction of genotype in routine histopathology. *J Pathol* 256:50-60. 10.1002/path.5800
- Shia J, Schultz N, Kuk D, Vakiani E, Middha S, Segal NH, Hechtman JF, Berger MF, Stadler ZK, Weiser MR, Wolchok JD, Boland CR, Gonen M, and Klimstra DS. 2017. Morphological characterization of colorectal cancers in The Cancer Genome Atlas reveals distinct morphology-molecular associations: clinical and biological implications. *Mod Pathol* 30:599-609. 10.1038/modpathol.2016.198
- Singh A, Sengupta S, and Lakshminarayanan V. 2020. Explainable Deep Learning Models in Medical Image Analysis. *J Imaging* 6. 10.3390/jimaging6060052
- Sun KHXZSRJ. 2016. Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition* 770-778.
- Teng Q, Liu Z, Song Y, Han K, and Lu Y. 2022. A survey on the interpretability of deep learning in medical diagnosis. *Multimed Syst* 28:2335-2355. 10.1007/s00530-022-00960-4
- Tizhoosh HR, and Pantanowitz L. 2018. Artificial Intelligence and Digital Pathology: Challenges and Opportunities. *J Pathol Inform* 9:38. 10.4103/jpi.jpi_53_18
- Vahadane A, Peng T, Sethi A, Albarqouni S, Wang L, Baust M, Steiger K, Schlitter AM, Esposito I, and Navab N. 2016. Structure-Preserving Color Normalization and Sparse Stain Separation for Histological Images. *IEEE Trans Med Imaging* 35:1962-1971. 10.1109/TMI.2016.2529665
- Valencia OM, Samuel SE, Viscusi RK, Riall TS, Neumayer LA, and Aziz H. 2017. The Role of Genetic Testing in Patients With Breast Cancer: A Review. *JAMA Surg* 152:589-594. 10.1001/jamasurg.2017.0552
- Vinuesa R, and Sirmacekc B. 2021. Interpretable deep-learning models to help achieve the Sustainable Development Goals. *Nature Machine Intelligence*
- Wang XX, Zou C, Zhang Y, Li XQ, Wang CX, Ke F, Chen J, Wang W, Wang D, Xu XY, Xie L, and Zhang YF. 2021. Prediction of BRCA Gene Mutation in Breast Cancer Based on Deep Learning and Histopathology Images. *Frontiers in Genetics* 12. ARTN 661109 10.3389/fgene.2021.661109
- Yao J, Zhu X, Jonnagaddala J, Hawkins N, and Huang J. 2020. Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks. *Med Image Anal* 65:101789. 10.1016/j.media.2020.101789

Figure 1

Workflow for detecting *BRCA1/2* status using a deep learning algorithm.

Firstly, we collected and processed the H&E images, including scanning the images, delineating and segmenting the tumor regions, and normalizing the color of the tiles. Secondly, we input the color-normalized tiles into the prediction network to deduce the *BRCA1/2* status. Finally, we generated attention heat maps using visualization techniques to explore the medical interpretability of the model.

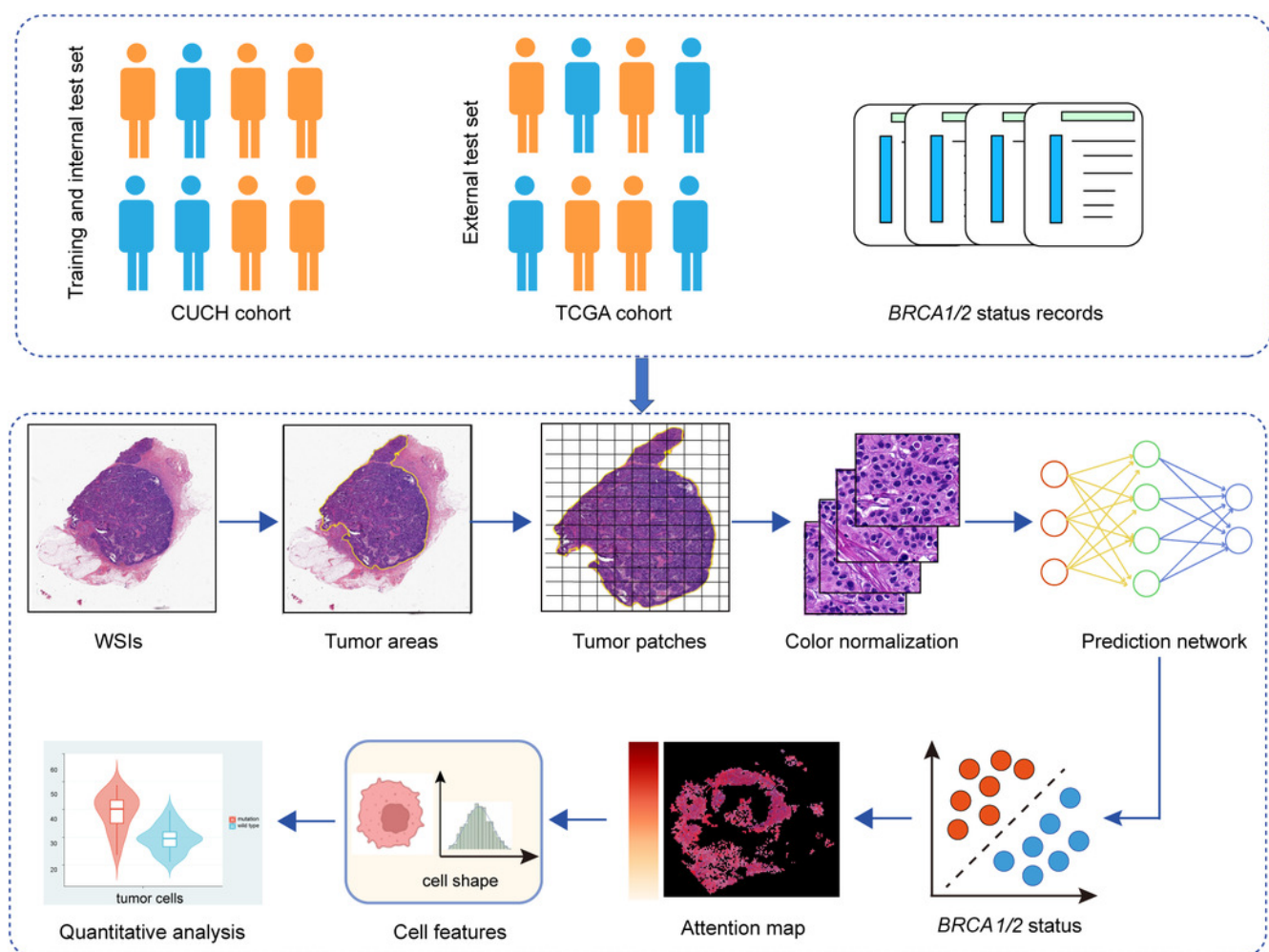


Figure 2

Performance of the segmentation network .

(A) Receiver Operating Characteristic (ROC) curve for the performance of the segmentation network versus three pathologists (senior pathologist, medium pathologist, junior pathologist) in identifying tumor areas. (B) Confusion matrices for the segmentation network. (C) Original WSIs and representative maps identified by the segmentation network. Left: original WSI images; Right: corresponding segmentation maps.

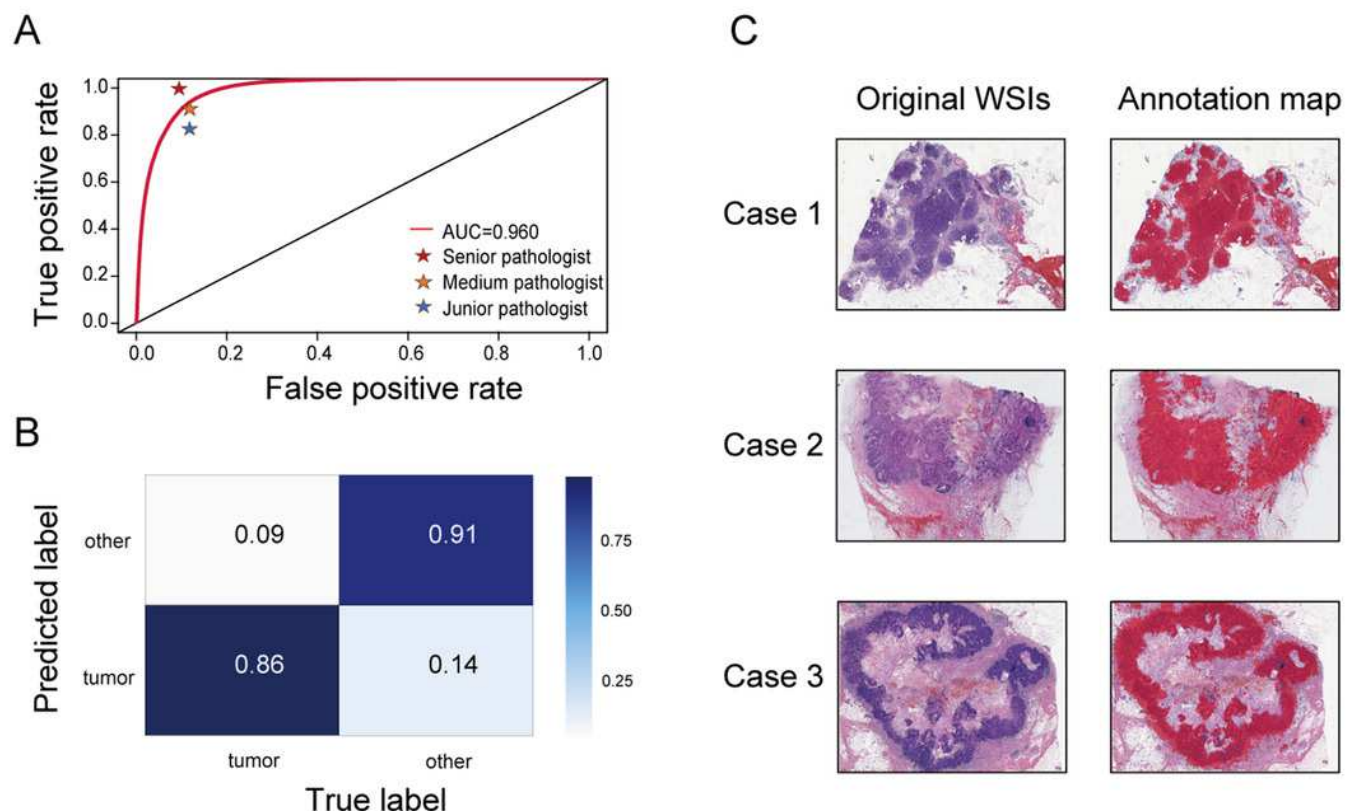


Figure 3

Architecture of the BiAMIL model.

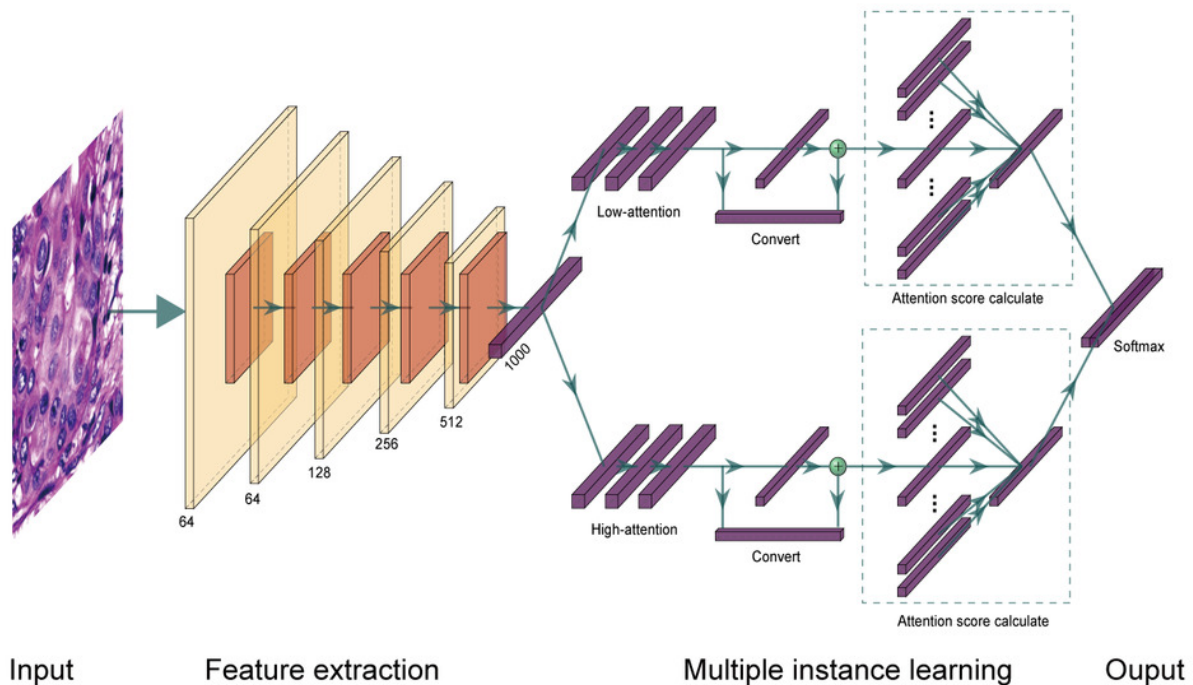


Figure 4

Evaluation of the prediction network performance.

(A-C) Receiver operating characteristic (ROC) curves of different models on the training, internal test, and external test sets.

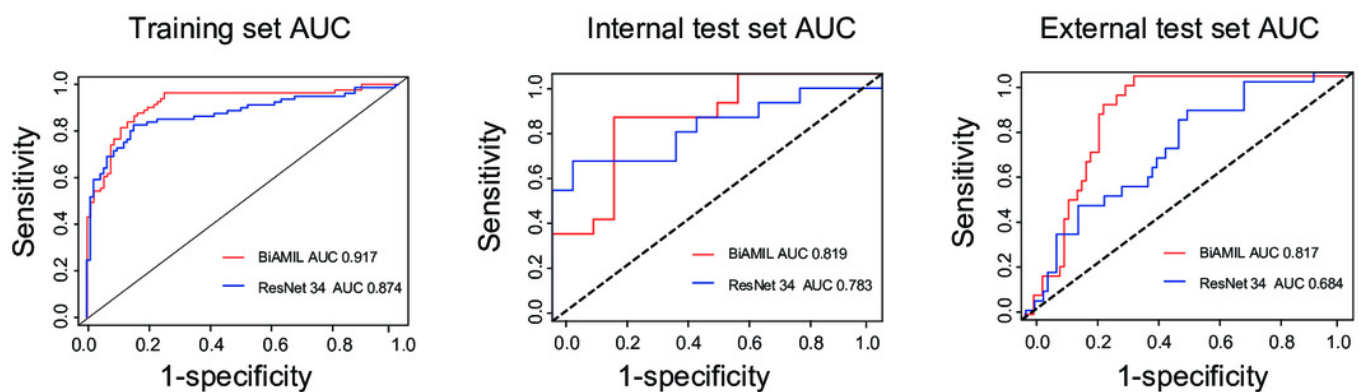


Figure 5

Performance comparison between BiAMIL and ResNet 34.

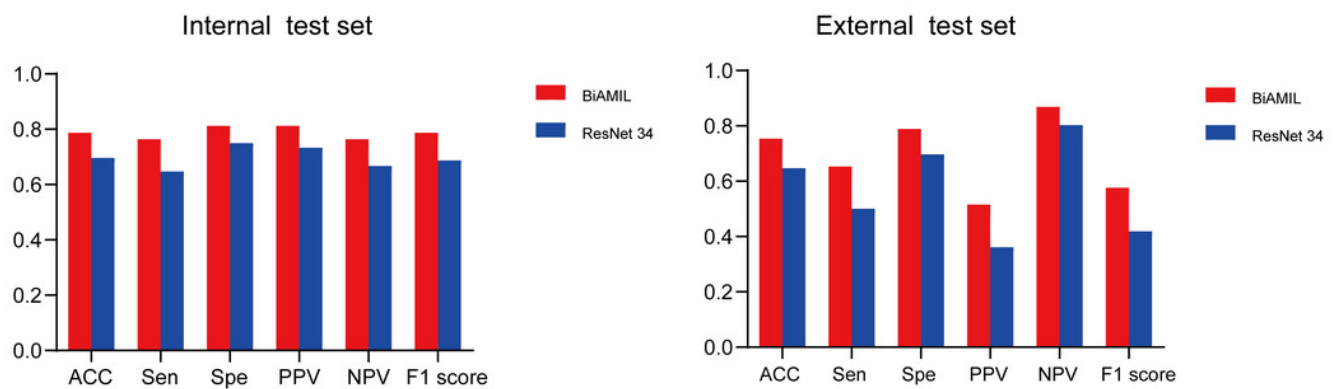


Figure 6

Visualization of the important regions in H&E images for the detection of *BRCA1/2* PV.

The highly predicted *BRCA1/2* PV tiles are primarily characterized by high-grade tumors and lymphocytic infiltrations. In each panel, the first row displays the original high-attention tiles of WSI, while the second row shows the corresponding heatmap. The red areas on the heatmap represent regions with large contributions from the model.

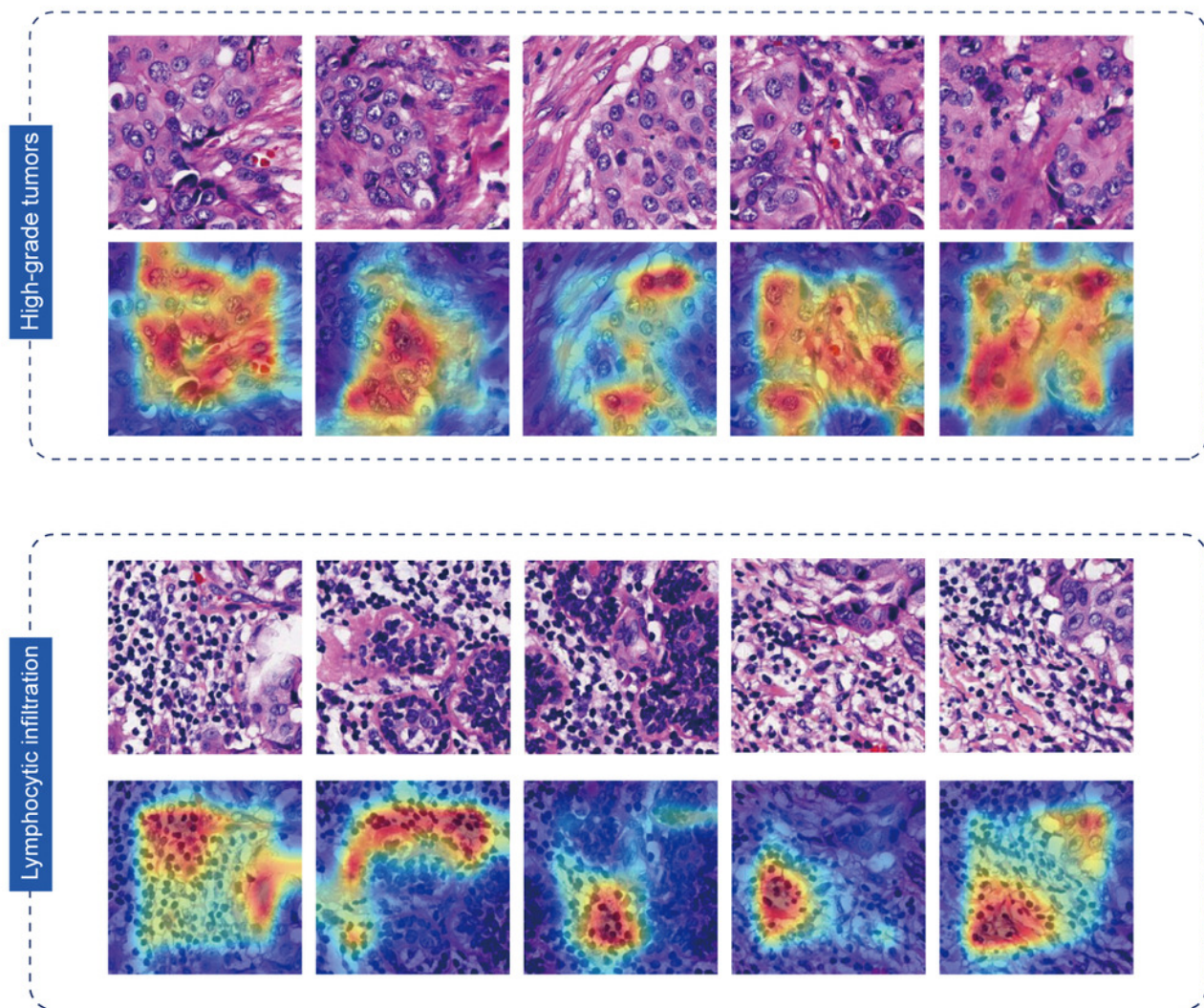


Figure 7

Visualization of the classification features learned by BiAMIL using the t-SNE algorithm.

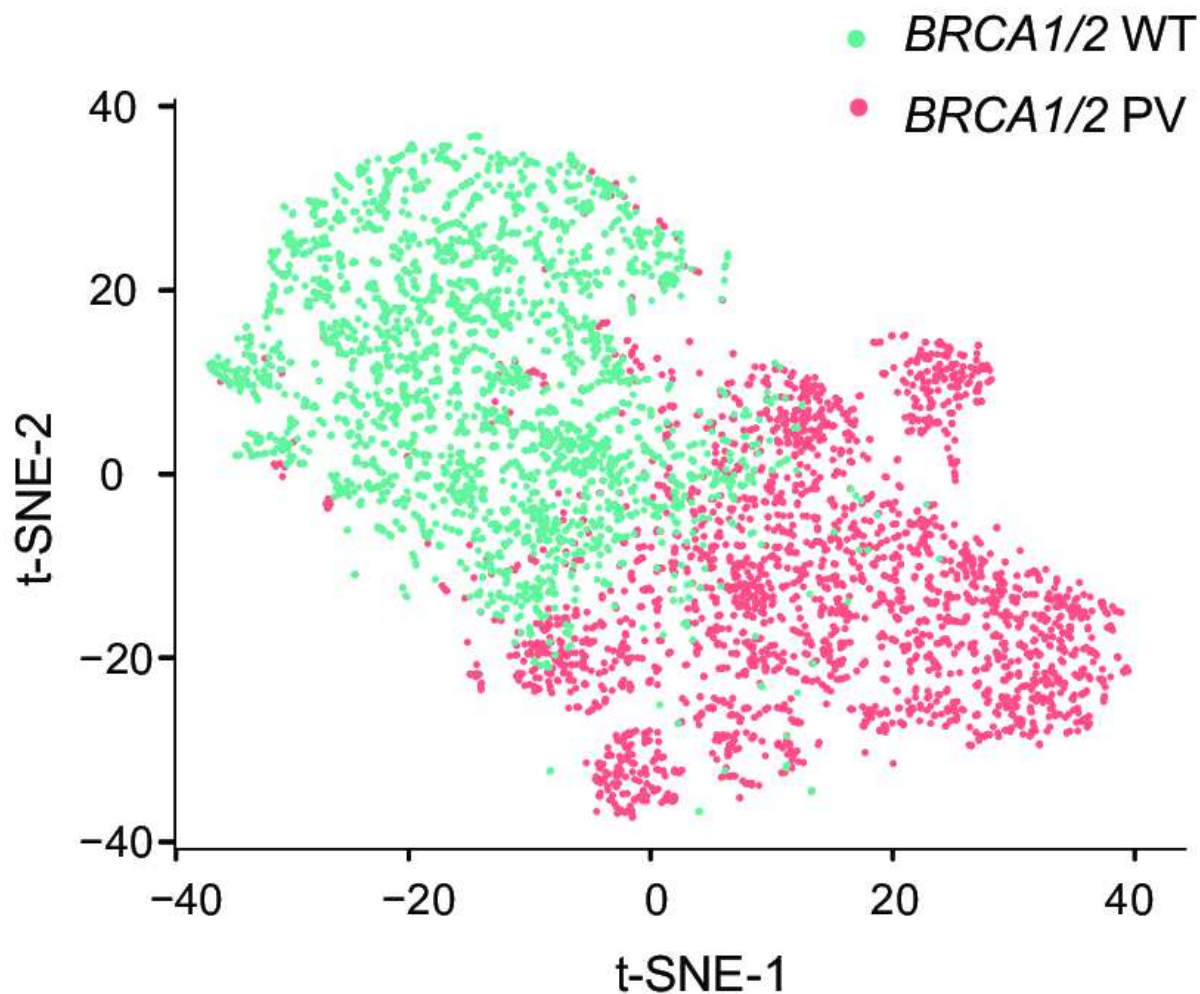
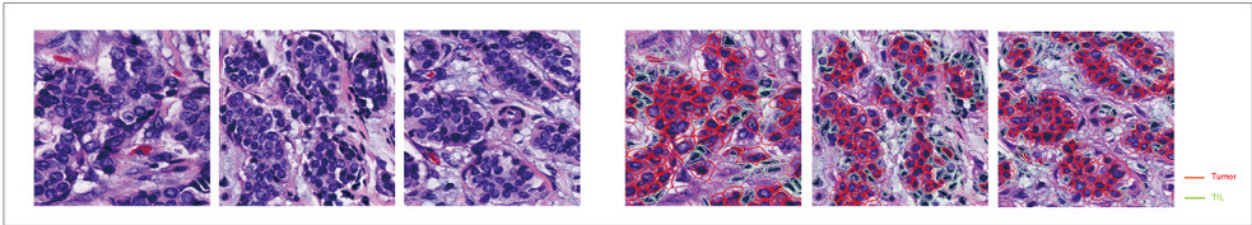


Figure 8

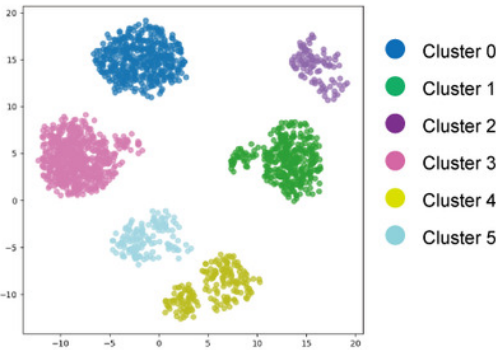
Comparison of cell features between the *BRCA1/2* PV and *BRCA1/2* WT.

(A) The representative H&E images and the cell quantitative analysis masks. Using Qupath for segmentation, tumor cells are indicated in red and TILs in blue. (B) Uniform Manifold Approximation and Projection (UMAP) visualization of cell-type features by clusters. (C) Cluster heatmap of normalized cell features. (D). Comparison of the nuclear features between *BRCA1/2* PV and *BRCA1/2* WT. (E). The mean tumor cell nuclear area of the typical tiles inferred in the model. *P* values were calculated using the two-tailed Mann-Whitney U-test.

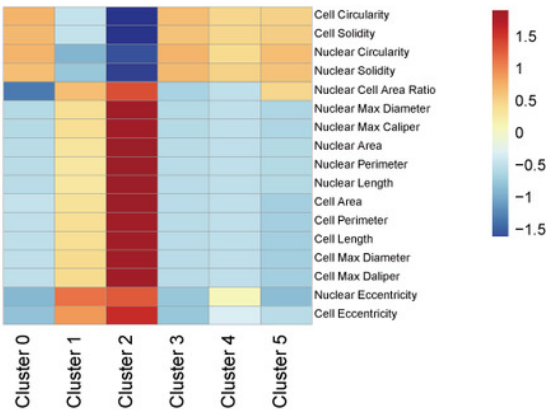
A



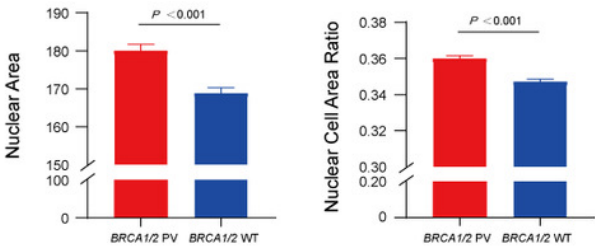
B



C



D



E

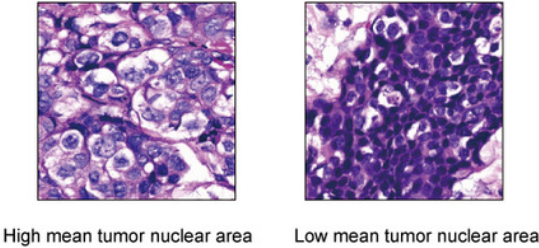


Table 1 (on next page)

Characteristics of patients in the training, internal test, and external test sets.

1

Factors	CUCH cohort		TCGA cohort
	Training set (n=122)	Internal test set (n=30)	External test set (n=102)
Age (years, mean ± SD)	47.89±9.42	49.66±7.66	59.14±12.59
ER status			
Positive	58(47.5%)	16(53.4%)	76(74.5%)
Negative	64(52.5%)	14(46.6%)	23(22.5%)
Missing	0	0	3(2.9)
PR status			
Positive	52(42.6%)	14(46.6%)	68(66.7%)
Negative	70(57.4%)	16(53.4%)	31(30.4%)
Missing	0	0	3(2.9%)
HER2 status			
Positive	28(23.0%)	10(33.4%)	15(28.3%)
Negative	94(77.0%)	20(66.6%)	84(82.4%)
Missing	0	0	3(2.9%)
Subtype			
HR +	62(50.8%)	16(53.4%)	70(68.6%)
HER2+	12(9.8%)	5(16.6%)	11(10.8%)
Triple-negative	48(39.4%)	9(30.0%)	18(17.6%)
Missing	0	0	3(1.2%)
Stage (%)			
I	34(27.9%)	6(20.0%)	22(21.6%)
II	73(59.8%)	17(56.6%)	55(53.9%)
III	11(9.0%)	6(20.0%)	23(22.5%)
IV	4(3.3%)	1(3.4%)	2(2.0%)

2

3 ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2; HR,

4 hormone receptor. PV, pathogenic variants; WT, wild type.

Table 2 (on next page)

The performance of BiAMIL in predicting *BRCA1/2* PV in HR+ and Triple-negative breast cancer.

1								
	Cohorts	AUC (95% CI)	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	F1 score (95% CI)
HR+	Internal test	0.859 (0.667-1.000)	0.812 (0.593-0.971)	0.750 (0.503-0.996)	0.875 (0.693-1.000)	0.857 (0.663-1.000)	0.777 (0.542-0.940)	0.799 (0.574-1.000)
	External test	0.850 (0.713-0.986)	0.785 (0.630-0.941)	0.615 (0.437-0.792)	0.824 (0.679-0.969)	0.444 (0.274-0.614)	0.903 (0.790-1.000)	0.516 (0.339-0.692)
Triple-negative	Internal test	0.761 (0.445-1.000)	0.800 (0.509-1.000)	0.857 (0.612-1.000)	0.666 (0.300-1.000)	0.857 (0.612-1.000)	0.666 (0.300-1.000)	0.857 (0.612-1.000)
	External test	0.740 (0.505-0.975)	0.666 (0.411-0.922)	0.666 (0.411-0.922)	0.666 (0.411-0.922)	0.666 (0.411-0.922)	0.666 (0.411-0.922)	0.666 (0.411-0.922)
2								
3	PV, pathogenic variants.							