Deep Learning for Detecting BRCA Mutations in High-Grade Ovarian Cancer based on an Innovative Tumor Segmentation Method from Whole-Slide Images

Raphaël Bourgade, Noémie Rabilloud, Tanguy Perennec, Thierry Pécot, Celine Garrec, Alexis F. Guedon, Capucine Delnatte, Stéphane Bézieau, Alexandra Lespagnol, Marie de Tayrac, et al.

TITLE

Deep Learning for Detecting *BRCA* Mutations in High-Grade Ovarian Cancer based on an Innovative Tumor Segmentation Method from Whole-Slide Images.

AUTHORSHIP

Raphaël Bourgade[1], Noémie Rabilloud[2], Tanguy Perennec[3], Thierry Pécot[4], Céline Garrec[5], Alexis F. Guédon[6], Capucine Delnatte[5], Stéphane Bézieau[5], Alexandra Lespagnol[7], Marie de Tayrac[7], Sébastien Henno[8], Christine Sagan[1], Claire Toquet[1], Jean-François Mosnier[1], Solène-Florence Kammerer-Jacquet[2,8], Delphine Loussouarn[1]

1: Department of Pathology, University Hospital of Nantes, 9 Quai Moncousu, 44093 Nantes cedex 01, France.

2: LTSI—UMR 1099 Inserm, University of Rennes, 2 avenue du Professeur Léon Bernard, 35043 Rennes, France

3: Department of Radiation Oncology, Institut de Cancérologie de l'Ouest Nantes, boulevard Professeur Jacques Monod, 44800 Saint-Herblain, France

4: Biosit, UAR 3480 CNRS - US 18 Inserm, University of Rennes, 2 avenue du Professeur Léon Bernard, 35043 Rennes, France

5: Department of Medical Genetics, University Hospital of Nantes, 9 Quai Moncousu, 44093 Nantes cedex 01, France.

6: Sorbonne University, National Institute of Health and Medical Research, Pierre Louis Institute of Epidemiology and Public Health, 184 rue du faubourg Saint-Antoine, 75012 Paris, France

7: Department of Molecular Genetics and Genomics, University Hospital of Rennes, 2 Rue Henri le Guilloux, 35000 Rennes, France

8: Department of Pathology, University Hospital of Rennes, 2 Rue Henri le Guilloux, 35000 Rennes, France

Corresponding author : Raphaël Bourgade

- tel : 0786252329
- mail : raphael.bourgade@gmail.com
- postal address : 9 Quai Moncousu, 44093 Nantes cedex 01, France.

ABSTRACT

BRCA1/2 genes play a crucial role in repairing DNA double-strand breaks through homologous recombination. Their mutations represent a significant proportion of homologous recombination deficiency and are a reliable effective predictor of sensitivity of high-grade ovarian cancer (HGOC) to poly(ADP-ribose) polymerase inhibitors. However, their testing by next-generation sequencing is costly, time-consuming, and can be affected by various preanalytical factors. In this study, we present a deep learning classifier for *BRCA* mutational status prediction from HES-stained whole-slide images (WSI) of HGOC. We constituted the OvarIA cohort composed of 867 HGOC patients with known *BRCA* somatic mutational status coming from two different pathology departments. We first developed a tumor segmentation model according to dynamic sampling and then trained a visual representation encoder with momentum contrastive learning on the predicted tumor tiles. We finally trained a *BRCA* classifier on over a million tumor tiles in multiple-instance learning with an attention-based mechanism. The tumor segmentation model trained on 8 WSI obtained a dice score of 0.915 and an intersection-over-union of 0.847 on a test set of 50 WSI while the *BRCA* classifier achieved the state-of-the-art AUC of 0.739 in 5-fold cross-validation and 0.681 on the testing set. An additional multiscale approach indicates that the relevant information for predicting *BRCA* mutations is more located in the tumor context than in the cell morphology. Our results suggest that *BRCA* somatic mutations have a discernible phenotypic effect which could be detected by deep learning and could be used as a pre-screening tool in the future.

INTRODUCTION

Epithelial ovarian cancer (EOC) is a highly lethal gynecologic malignancy and a significant cause of female cancer-related deaths[1]. The most common histological subtype of EOC is high-grade serous (HGSOC). Because of their silent course, high-grade ovarian cancers (HGOC) are frequently diagnosed at an advanced stage and are of poor prognosis. The standard treatment for patients with diagnosed advanced ovarian cancer consists of cytoreductive surgery and a combination of paclitaxel and platinum-based chemotherapy[2]. Despite standard therapy, 70% of the patients relapse within the subsequent 3 years[2].

HGOC is characterized by defects in the DNA damage repair pathway known as homologous recombination. The *BRCA*1 and *BRCA*2 genes play a crucial role in repairing DNA double-strand breaks through homologous recombination[3]. Mutations in these genes are found in 17-25% of patients with HGSOC, the most common type of EOC[4,5]. In 2005, two studies demonstrated that in vitro, poly(ADP-ribose) polymerase inhibitors (PARPi) are effective at killing *BRCA*-deficient cancer cells, but have no effect on non-*BRCA*-deficient cancer cells. This phenomenon, known as synthetic lethality, first described by Hartwell[6], occurs because PARPi inhibits the repair of DNA single-strand breaks through the base excision repair system. This mechanism leads to the accumulation of double-strand breaks that can be repaired by homologous recombination in *BRCA*-competent cells but not in *BRCA*-deficient cells, causing cell death.

Germline and somatic *BRCA*1/2 mutations are known predictive markers of response to PARPi targeted therapy, both in relapse and first-line settings[7–13]. *BRCA*1/2 somatic mutational status analysis is thus essential for selecting appropriate treatment.

Its characterization with next-generation sequencing (NGS) is widely available in developed countries, using DNA extracted from formalin-fixed paraffin-embedded (FFPE) tumor tissue from either archival or current tumor samples. However, NGS performed on FFPE tissue sections may be challenging, due to pre-analytical conditions or amount of cancer cells in the sample, which could affect molecular results[14]. Indeed, approximately 5% of FFPE NGS

analyses fail and must be repeated with no molecular results in 3% of cases[15]. Moreover, this analysis is quite expensive, time-consuming and not available or accessible to all patients. Additionally, the growing demand for personalized medicine has led to an increasing use of cancer molecular genetics platforms, which can further strain resources and increase costs. This potentially leads to delays in therapeutic management, as patients may have to wait for diagnostic results before they can receive the appropriate treatment.

Digital pathology has enabled the acquisition of high-resolution images known as whole-slide images (WSI) for the application of computational techniques[16]. With the progress of computer vision and the development of convolutional neural networks (CNN), machine learning algorithms are able to extract information from hematoxylin-eosin-safran (HES)-stained WSI. It has the significant potential to improve the standard clinical process in pathology.

In addition to computer-aided diagnosis, whose goal is to reproduce human tasks, such as the detection of lymph node metastases in breast cancer[17,18], detection of mitosis[19], or gastric cancer diagnosis[20], artificial intelligence (AI) algorithms using WSI can predict patient survival and specific molecular features in the lung[21], prostate[22], brain[23,24], colorectal[25], gastrointestinal[26], and breast cancer[27,28].

In this study, we present CNN classifiers for *BRCA* 1/2 mutational status prediction from HES-stained HGOC WSI. As we assumed that the relevant information resides in the tumor, we developed a tumor segmentation model by using an innovative approach based on an efficient sampling strategy with a small annotated training dataset. Finally, we studied the impact of the resolution field on the final prediction through a multiscale approach.

MATERIAL AND METHODS

OvarIA *BRCA* cohort

The OvarIA cohort consists of two independent cohorts of patients diagnosed with HGOC from biopsy or surgery and who underwent tumor *BRCA* mutational status testing by NGS in Department of Medical Genetics at the University Hospital of Nantes and Rennes. NGS library preparation was performed with *BRCA* MASTR Dx Kit (Multiplicom Agilent Technologies) for Nantes patients and with Advanta™ Solid Tumor NGS Library Prep Assay (Standard BioTools) for Rennes patients. The sequencing was performed with Illumina MiSeq for Nantes patients and Illumina NextSeq 500/550 for Rennes patients (Illumina, San Diego, California, US). The whole OvarIA cohort consists of 867 patients: 169 with *BRCA* mutation (*BRCA*-mut) and 698 with no *BRCA* mutation (*BRCA*-wt). The first cohort included 551 patients (122 *BRCA*-mut and 429 *BRCA*-wt) and the second cohort included 316 patients (47 *BRCA*-mut and 269 *BRCA*-wt), respectively from the Pathological Departments of the University Hospital of Nantes and Rennes. All variants of uncertain significance (VUS) and uninterpretable results have been discarded. One FFPE block per patient was collected and HES-stained in each Pathological Department. All HES-stained slides were finally reviewed and characterized by two gynecopathologists in the Pathological Department of the University Hospital of Nantes. This study only included high-grade serous and high-grade endometrioid subtypes and excluded mucinous subtypes. In decreasing order, the samples concern peritoneum, ovary, fallopian tube, and metastases site (omentum, node, colon, rectum, ileum, lung, liver, uteri). The slides were anonymized and then digitized on the same scanner at 20x magnification (0.46 µm/pixel) using a NanoZoomer S60 Digital slide scanner (Hamamatsu, Japan) at the Micropicell platform (IRS-UN unit, Nantes). The WSI were saved in ndpi format and converted to tiff format using openslide[29] and pyvips library.

TCGA cohort

The diagnostic slides from the cancer genome atlas (TCGA) were downloaded from the genomic data commons portal (https://portal.gdc.cancer.gov/) in svs format and constitute an external validation cohort. Any slide with tissue folds, blurred artifacts, or pen marks were excluded. A final dataset of 103 FFPE and H&E-stained samples of HGSOC with *BRCA* status

available (21 *BRCA*-mut and 82 *BRCA*-wt) were retained, converted in tiff format, and downscaled to 20x magnification (0.5 µm/pixel).

TUMOR SEGMENTATION

Proposed pipeline

During the study design, we hypothesized that the relevant information for predicting *BRCA* somatic mutations could be mostly found in the tumor tissue which we have defined as the carcinomatous cells with the exclusion of stroma. We thus trained an ovarian cancer segmentation model to predict the tumor pixels. During training, 8 HES-stained slides were randomly selected from the first cohort (e.g. Nantes) and then pixel-annotated by a first gynecopathologist. To assess the robustness of our algorithm, 50 HES-stained validation slides were randomly selected from both cohorts (e.g. Nantes and Rennes) in equal proportions and were then pixel-annotated by a second gynecopathologist. Because of the different HES-staining protocols, we applied the Vahadane stain normalization algorithm during preprocessing to match the color spectrum distribution of the two medical centers[30]. The validation was assessed using the dice coefficient (DC) and the intersection-over-union (IoU) score as evaluation metrics.

Data annotations

The 8 randomly selected training slides were annotated by a first gynecopathologist in less than 2 hours, using the opensource software QuPath (version 0.3.2)[31]. All the carcinomatous areas were delineated with the exclusion of necrosis. The polygonal annotations were then exported using the javascript object notation format, including X and Y coordinates corresponding to the annotated regions. These coordinates were then encoded into a binary tumor mask for strongly-supervised training.

Training process

Pre-processing techniques, such as tiling the region of interest into smaller patches, are often necessary when working with WSI due to their large size. One limitation of this approach is the necessity to predetermine certain hyperparameters, such as the size of the tile and the resolution level. This can significantly lengthen the pre-processing time as different combinations of these hyperparameters must be evaluated to determine the optimal combination. To address this issue, we used the *Deepflash2*, a python open-source library developed by Griebel et al.[32] in order to implement a more dynamic training process. Instead of tiling the WSI into fixed tiles by using a sliding window approach, the slides are saved to zarr files, a file storage format for chunked, compressed, and N-dimensional arrays. This format allows a memory-efficient loading of the slides which are then randomly sampled during training according to a probability density function. Each region of the WSI is weighted, determining the probability of being sampled at each epoch. One advantage of this dynamic approach is to use flexible tile shape and scaling factor at runtime. Although this sampling method is in itself a data augmentation process, we used additional operations such as flipping, rotating, mirroring, and color changes to increase diversity during training. Powered by the *Segmentation PyTorch models* package, *Deepflash2* allows the use of different state-of-the-art architecture (*U-Net*[33], *UNet++*[34], *LinkNet*[35], *DeepLabv3+*[36]) and encoder (*ResNet*[37], *Inception*[38], *EfficientNet*[39], *ResNeSt*[40]) combinations.

Hyperparameter optimization

Hyperparameters have been set according to a grid search evaluated through 5-fold cross-validation. The best score was obtained using 512 x 512 pixels tile size, a downscaling factor of 2, and a sampling weight of 1 for the predicted tumor pixels, 0.7 for ovarian non-tumor tissue determined by the subtraction of the tumor mask and the whole tissue mask determined by the Otsu's method, and 0.1 for background regions (Figure 1). We used a U-Net architecture and an EfficientNet-B4 as an encoder to perform segmentation. We used a cross-entropy loss and a Ranger optimizer. We trained the model in mixed precision with a 10^{-3} weight decay.

Fifteen epochs with a batch size of 32 have been performed using the one-cycle policy with a 10^{-3} max learning rate.

Post-processing

In order to remove all the cutting artifacts localized in the tumor and over-predicted by the algorithm, we calculated the intersection of the predicted tumor mask with the tissue mask determined with Otsu's method.

Model assessment

The performance of the models was assessed on a testing set of 50 WSI. As the task involved semantic segmentation and the tumor ratio within the images exhibited high variability, both the DC and the IoU were calculated. These metrics evaluate the similarity between the predicted binary mask (PM) and the ground truth (GT) delineated by the gynecopathologist.

These metrics are defined as

$$Dice = \frac{2 \times TP}{(TP + FP) + (TP + FN)} \qquad IoU = \frac{TP}{(TP + FP + FN)}$$

Code Latex :

DC $=\frac{2 \times TP}{(TP+FP)+(TP+FN)}$

$IoU=\frac{TP}{(TP+FP+FN)}$

The true positives (TP) are the sum of all matching tumor pixels of the PM and the GT while the false positives (FP) and false negatives (FN) are the sum of tumor pixels that only appear in the PM or in the GT. We finally averaged the DC and the IoU over the 50 testing WSI.

***BRCA* CLASSIFICATION**

Proposed pipeline

After segmentation, we trained a *BRCA* classifier by randomly splitting the OvarIA cohort into a development set (n=80%) and a testing set (n=20%). We trained five models using cross-validation on the development set and tested each of them on the testing set. To determine the most optimal scale for predicting *BRCA* mutation, we performed three different training schemes relying on the multiple instance learning (MIL) paradigm, and using momentum contrast (MoCo). MoCo is an unsupervised learning technique that aims to learn useful representations from unlabeled data. It involves training a model by maximizing agreement between two differently augmented views of the same input tile. Additionally, MoCo incorporates a momentum-based moving average of the model's parameters during training to stabilize and improve the learning of representations.

Data pre-processing

The WSI were first tiled into non-overlapping patches of 512 x 512 pixels at 20x magnification (0.46 µm/pixel), according to the tumor mask predicted by our segmentation algorithm. To optimize computation, a maximum random subset of 5000 tiles with at least 50% of their surface covered by the tumor mask was selected from each slide. The first 2000 tiles randomly selected in this subset were retained for the classification, while the following 3000 were used for learning MoCo representations. If the tumor-predicted mask consisted of less than 2000 tiles, these tiles were all retained for classification and no tile was used to train MoCo representations. After tiling, 7 slides have been removed from the OvarIA dataset, because they contained no 512 x 512 pixels tiles with at least 50% of tumor area (Figure 2). This process generated 599 697 tumor tiles for self-supervised learning. However, because of their likely low impact during training, we decided to discard all slides with less than 200 predicted tumor tiles, (i.e. 85 slides) for a final total of 775 WSI and 1 040 149 tumor tiles used for

classification (details in Table 1). We used the MoCo repository available at (https://github.com/facebookresearch/moco) to train our model by randomly considering the following transformations: horizontal flip, grayscale, color jitter and gaussian blur. According to their popularity in feature extraction, we benchmarked different ResNet depths (18, 34, 50, and 101 layers). A first feature extraction has been conducted according to the imagenet parameters then a second after finetuning each model on their final layers. In both cases, the best performance was observed with the ResNet-50 which was thus chosen as an encoder and trained from scratch for 200 epochs on 4 GPU Nvidia RTX A4500 20 Go. We used the SGD optimizer with a momentum of 0.9, a weight decay of 1e-4 and a learning rate schedule with a cosine decay starting at 3e-2.

Training process

Due to the unbalanced dataset, we performed a stratified 5-fold cross-validation on the development set. As the impact of each tumor tile on prediction is unknown, we decided to use MIL technique with the attention-based model proposed by Ilse et al.[41]. After encoding, the feature vector of each tile is mapped to an attention score by a multi-layer perceptron. This attention score determines how much a given tile will contribute to the slide representation. The higher the score, the more the tile contributes to the final prediction. The slide representation is obtained by the sum of the vector of individual features for all tiles, weighted by the attention score. A second linear layer is then used to output the final prediction from the slide-level vector representation by minimizing a binary-weighted cross-entropy loss, according to the imbalanced classes.

Evaluation criteria

Several classification metrics such as area under the ROC Curve (AUC) but also precision, recall, and F1-score have been used to evaluate the model performance. Precision is the proportion of TP predictions among all positive predictions while recall is the proportion of TP predictions out of all positive instances. Their combination determines the F1-score which

calculates their harmonic mean. It is particularly beneficial in our case since the classes are imbalanced. All of these metrics were calculated by their mean over the five validation folds of the development set. Each of these five models was then applied to the remaining testing set and their classification metrics were also averaged.

External validation cohort

We performed an additional validation on an external dataset of EOC from TCGA. We selected 103 high-quality FFPE and H&E-stained slides. We then performed the same preprocessing steps as before, for a total of 111 727 tumor tiles. Their representations were provided by MoCo and each of our five previous models trained in cross-validation were evaluated on this external dataset. The final result is the AUC and the standard deviation averaged over the five models.

Optimal observation field

To determine the most optimal observation field for predicting *BRCA* mutation, we conducted three additional approaches with different scaling. *Baseline approach* : a maximum of 2000 tiles of 512 x 512 pixels were randomly sampled in each WSI. This first “high-level” approach provides more contextual information. *Approach 2* : each of these tiles was divided into four tiles of 256 x 256 pixels (i.e maximum of 8000 tiles per WSI). Intuitively, this “low-level” approach brings more information in MIL, but less in the tumor spatial conformation. *Approach 3* : these four tiles were finally randomly reassembled into a new mixed tile of 512 x 512 pixels (i.e maximum of 2000 tiles per WSI) (Figure 3). This approach provides more information in MIL than the first approach, but with degraded information on the spatial configuration of the tumor.

Computational configurations:

All analyses have been done under Python 3.7 using the Pytorch framework. The computation tasks were performed on an Intel Xeon Gold 6140 2.30 GHz processor with NVIDIA V100 Tensor Core GPU 32Go and 32 Go memory RAM. We used CUDA V10.2.89 Toolkit.

RESULTS

Tumor segmentation

The first step of this study was to develop a deep-learning model for the automatic segmentation of HGOC. We used 58 HES-stained WSI randomly selected from the OvarIA cohort and splitted into a training set of 8 WSI from the first center (e.g. Nantes) and a validation set composed of 50 WSI from the two centers (e.g. Nantes and Rennes) randomly sampled in equal proportion. For a better assessment of the real performance of the model, the training set and the validation set were annotated by two different gynecopathologists. The similarity between the PM and the GT was computed by averaging the DC and the IoU over the 50 WSI and no significant difference was observed between the 2 cohorts. The segmentation model finally achieved an overall DC of 0.915 (± 0.05) and an overall IoU of 0.847 (± 0.079).

BRCA mutation prediction

After segmenting the tumor over the 860 slides, we discarded 85 slides with less than 200 predicted tumor tiles. We then trained several *BRCA* classification models on the 775 remaining slides following three sampling approaches. In order to assess the robustness of our model, we next validated the *BRCA* classifier on an external independent test set. We used 103 FFPE and H&E-stained slides from the TCGA cohort with 21 slides *BRCA*-mut and 82 slides *BRCA*-wt. All the results are presented in Table 2.

In order to increase their interpretability and according to their better results, the baseline models trained in 5-fold cross-validation have been more precisely evaluated and achieved a mean precision of 0,714 (± 0.02), a mean recall of 0,731 (± 0.023) and a mean F1-score of 0,722 (± 0.009) on the internal testing set.

DISCUSSION

*BRCA*1/2 mutations have been shown to be effective predictors of the sensitivity of HGOC to PARPi. Their identification is crucial for the selection of appropriate therapeutic options. Currently, somatic *BRCA*1/2 mutation testing is conducted using NGS on FFPE tissue samples, which can be challenging due to the poor quality of the DNA and its high level of fragmentation. Additionally, tumor samples are often highly heterogeneous and can be contaminated with DNA from normal tissue. Several studies have demonstrated the effectiveness of CNN-based models for predicting specific molecular features in various types of cancer, including lung[21], prostate[22], brain[23,24], colorectal[25], and gastrointestinal cancers[26], as well as breast cancer[27,28]. These studies have shown that CNN-based models can accurately predict genomic alterations, such as mutations in specific genes, based on morphological features extracted from WSI.

In this paper, we present a deep learning-based model for predicting *BRCA*1/2 mutation status using an innovative tumor segmentation method. To our knowledge, our model has achieved the best performance hitherto in predicting BRCA mutations in HGOC from WSI. Furthermore, it is also the first model that has undergone evaluation on an external validation cohort.

Tumor segmentation is a classical task in computational pathology. It allows the identification of regions of interest useful for complementary classification tasks[42]. However, pixel-level segmentation requires precise annotations which are not always available. Indeed, annotation

is a costly and time-consuming task that requires a pathologist's expertise. Moreover, there are still few published annotation guidelines which leads to high inter-individual variability and bad interoperability of annotation protocols[43]. In their paper about ovarian cancer segmentation, Ho et al. have adopted a Deep Interactive Learning method to reduce the time of manual annotation[44]. They used a breast cancer segmentation model, pre-trained to segment high-grade invasive ductal carcinoma from triple-negative breast cancer images. They then finetuned it to segment HGSOC, thanks to the similarity of their morphological features. The training process is composed of multiple iterations of segmentation, assessment then correction of the prediction generated by the previous model. They thus initially predicted the HGSOC over 60 randomly selected WSI with their pre-trained breast cancer model and finally trained their own ovarian model after three iterations and the annotation of 25 WSI in 3,5 hours. They then evaluated their model on 14 WSI and achieved an IoU of 0,74 for their best model. However, this approach ideally requires more than twenty WSI in training which may be difficult to obtain in rare diseases. Furthermore, this technique requires a pre-trained model on a tumor of more similar morphology.

A key improvement of our method is the use of a dynamic training process which considerably reduces annotation and pre-processing time. Thanks to the Deepflash2 library intersection and efficient sampling, we were able to train a segmentation model with less than 2 hours of manual annotations on 8 WSI. Our model was then evaluated on 50 WSI and obtained an overall DC of 0.941 (± 0.05) and an IoU of 0.846 (± 0.079). A particularity of our segmentation model is that, despite the few areas of stroma in the tumor mask labeled by the pathologist in the training set, we observed that, for the large majority of WSI, only the tumor pixels were predicted and not the stroma, as shown in Figure 2B. In some cases, this could be explained by a higher contrast of the tumor than the stroma with the rest of the non-tumor tissue. The difference between tumor mask delineated by the second gynecopathologist for the testing set and predicted tumor mask, explains some DC<0.7. Indeed, it is sometimes difficult to exclude all the stromal tissue manually, especially for post-chemotherapy tumors or tumors with papillary or micropapillary architecture. This difference can however be reduced by post-

processing techniques, such as prediction smoothing. Nevertheless, as we seek to predict the somatic mutation status of *BRCA*, the objective of this segmentation step was to extract the tumor pixels while excluding as many non-tumor pixels as possible. Another reason for the decrease of the DC is the localized artifacts within the tumors. These artifacts were bypassed by the gynecopathologist while they are systematically removed during the post-processing with the intersection of the predicted mask and the tissue mask obtained by Otsu's method.

Several studies have focused on the use of deep learning for predicting gene mutations from routine FFPE and H&E-stained histology slides. Coudray et al. showed that six mutated genes in lung adenocarcinoma can be predicted with an AUC ranging from 0.733 to 0.856[21]. Other studies have demonstrated an AUC of 0.71 for the prediction of SPOP mutations in prostate cancer images[22], and AUC ranging from 0.71 to 0.89 for the prediction of CTNNB1, FMN2, TP53, and ZFX4 mutations in hepatocellular carcinoma[45].

To our knowledge, there are only two studies that have investigated the prediction of *BRCA* mutations in ovarian cancer[44,46]. Ho et al. recently conducted a study in which they aimed to predict *BRCA* mutations in HGSOC[44]. They included 609 HGSOC (119 *BRCA*-mut and 490 *BRCA*-wt) and used a similar approach by first training a tumor segmentation model, and then training an additional classifier to predict *BRCA* mutations on the automatically extracted cancer patches. The authors performed an unbalanced sampling with a maximum of 1000 tiles for non-mutated patients and 5000 tiles for mutated patients. They used a fixed tile size of 224 x 224 pixels and trained three *BRCA* classification models at various magnifications (5x, 10x, 20x). The slide-level score was then calculated by averaging all the patch scores of a WSI. They obtained an AUC ranging from 0.49 to 0.67 on the validation set and between 0.40 and 0.43 on the testing set. Another study, conducted by Nero et al. also investigated the prediction of *BRCA* mutations in 664 ovarian cancer patients comprising 233 cases of somatic *BRCA*1/2 mutations[46]. In contrast to Ho et al. they did not focus on the tumor and tiled the whole tissue slide. Each patch was then embedded into a low-dimensional feature representation before being weighted by an attention branch. They finally achieved an AUC

of 0.7 on the training set and 0.55 on the testing set, improved at 0.59 by manually identifying areas of interest on a subset of 285 WSI.

In our study, we combined both a tumor segmentation step and the use of an attention mechanism during classification. This allowed us to achieve the state-of-the-art AUC of 0.739 (± 0.024) in 5-fold cross-validation, an AUC of 0.681 (± 0.014) over the internal testing set, and an AUC of 0.631 (± 0.03) over the TCGA dataset. Nevertheless, most of our patients were recruited before 2019 and only underwent *BRCA*1/2 mutation testing, which constitutes one limitation of our study. Indeed, *BRCA*1/2 mutations represent one of several causes of homologous recombination deficiency (HRD), which is characterized by genomic instability and also defines a target for PARPi. Tumors with similar morphological characteristics to those of *BRCA*-mutated tumors may also harbor mutations in other genes in the HRD pathway. It would be interesting to replicate this work on a cohort of tumors with HRD, as was done by Lazard et al. in breast cancer, where they achieved an AUC of 0.86[47]. They also developed an innovative visualization technique that allows for the automatic extraction of new morphological features related to HRD.

In their paper, Ho et al. have also studied the impact of resolution by training a deep multi-magnification network (DMMN). This model fuses morphological features from both high and low levels by the concatenation of patches at 20x, 10x, and 5x magnification. We decided to confirm this assumption by conducting three different approaches at the same resolution (20x magnification). The first approach consisted in tiling the tumor area in 512 x 512 pixels tiles while the majority of similar studies tile in 256 or 224 pixels. This approach reduces the total number of tiles but allows to get more contextual information about the tumor and its architecture. These tiles were then divided into four tiles of 256 x 256 pixels. This step brings four times more information during MIL, but reduces the contextual information, as the tiles are encoded independently from each other in the CNN. This approach led to lower AUC of 0.721 (± 0.014) for 5-fold cross-validation, AUC of 0.654 (± 0.011) over the testing set and AUC of 0.624 (± 0.024) over TCGA dataset. Finally, we performed a last approach by randomly reassembling the four previous 256 x 256 pixels tumor tiles into a unique 512 x 512 pixels

mixed tumor tile. This approach brings as much information as the baseline, but reduces the tumor context by impairing its spatial architecture. This approach achieved the most disappointing results compared to the two others, with an AUC of 0.697 (± 0.016), for 5-fold cross-validation, AUC of 0.635 (± 0.015) over the testing set, and AUC of 0.619 (± 0.024) over TCGA dataset. This multiscale approach suggests that the relevant information for predicting *BRCA* mutations is preferentially located in the tumor context than in the cell morphology.

It could be interesting to compare this multiscale approach with the use of vision transformer[48]. It is a model introduced in 2021 where the input image is split into fixed-size patches before being linearly embedded with their positional information and fed to a transformer encoder with a self-attention module. This module allows a global interpretation by capturing distant semantic relevances in an image while convolution is only focused on neighboring pixels. However, the first vision transformer usually required large datasets to achieve similar or better performance than CNN on various tasks. This is the reason why Guo et al. in their paper, used a swin Transformer to predict microsatellite instability and key biomarkers in colorectal cancer from H&E stained images[49]. It is a hierarchical transformer whose representations are computed with a shifted windows. This architecture allows them to achieve a state-of-the-art AUC of 0.9 for MSI prediction using the MCO dataset composed of 1065 patients. This achievement will probably pave the way for the use of vision transformers on small datasets in computation pathology.

In conclusion, our results suggest that somatic mutations of *BRCA* have a phenotypic impact in HGOC. They also suggest that the relevant information for the prediction of *BRCA* mutational status seems to reside more in the tumor spatial conformation than in the cell morphology. It could also be interesting to investigate the phenotypic pattern of the tumor tiles with the highest probability of being *BRCA* mutated, according to the attention module in a future study. Moreover this work needs to be confirmed on an HGOC cohort with HRD. These AI-based models will probably be used more and more in the diagnostic routine to pre-screen tumors and optimize the preselection of patients for molecular determination. Even if this study

needs to be validated on a larger and multicenter cohort, it paves the way to clinical application with the future implementation of pre-screening tools for a more personalized medicine.

ETHICS APPROVAL

The study was approved by two French ethics committees for the two cohorts: Sud-Ouest and Outre-mer CPP-registration N°21.00545.000010 for Nantes and Ouest II CPP-registration: N°22.02541.000097 for Rennes.

AUTHOR CONTRIBUTIONS STATEMENT

R.B and D.L. performed study concept and design. D.L., S.-F.K.J, C.S. and S.H. generated the patient cohort. D.L and C.S performed analysis and interpretation of data. C.G., S.B., C.D., A.L and M.D.T. performed genomic analyses. J.-F.M. was the scientific supervisor of the tissue biobank. R.B. performed development of AI methodology and statistical analysis. R.B. and T.Per. developed the tumor segmentation model. N.R and T.Pec. pre-trained the feature extractor on their workstation. R.B. developed the *BRCA* classification model. R.B., A.G. and C.T. prepared the figures. R.B. and D.L. wrote the manuscript. All authors read and approved the final paper.

FUNDING STATEMENT

We are grateful to AstraZeneca for their financial support. We also thank the biological resource center for biobanking (CHU Nantes, Hôtel Dieu, Tumorothèque, Nantes, F-44093, France) for its contribution. T.Pec. is funded by a Chan Zuckerberg Initiative DAF grant (2019-198009).

DATA AVAILABILITY STATEMENT

All the TCGA dataset is available at https://portal.gdc.cancer.gov/.

The OvarIA dataset consists of confidential medical data from the Pathological Department of the University Hospital of Nantes and Rennes and is not open to the public. Any information

and requests for resources should be directed to Delphine Loussouarn (delphine.loussouarn@chu-nantes.fr).

REFERENCE

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*. 2021;71(3):209-249. doi:10.3322/caac.21660
2. Ledermann JA, Raja FA, Fotopoulou C, et al. Newly diagnosed and relapsed epithelial ovarian carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol Off J Eur Soc Med Oncol*. 2013;24 Suppl 6:vi24-32. doi:10.1093/annonc/mdt333
3. Caestecker KW, Van de Walle GR. The role of BRCA1 in DNA double-strand repair: past and present. *Exp Cell Res*. 2013;319(5):575-587. doi:10.1016/j.yexcr.2012.11.013
4. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature*. 2011;474(7353):609-615. doi:10.1038/nature10166
5. Pennington KP, Walsh T, Harrell MI, et al. Germline and somatic mutations in homologous recombination genes predict platinum response and survival in ovarian, fallopian tube, and peritoneal carcinomas. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2014;20(3):764-775. doi:10.1158/1078-0432.CCR-13-2287
6. Hartwell LH, Szankasi P, Roberts CJ, Murray AW, Friend SH. Integrating genetic approaches into the discovery of anticancer drugs. *Science*. 1997;278(5340):1064-1068. doi:10.1126/science.278.5340.1064
7. Ledermann J, Harter P, Gourley C, et al. Olaparib maintenance therapy in patients with platinum-sensitive relapsed serous ovarian cancer: a preplanned retrospective analysis of outcomes by BRCA status in a randomised phase 2 trial. *Lancet Oncol*. 2014;15(8):852-

861. doi:10.1016/S1470-2045(14)70228-1

8. Pujade-Lauraine E, Ledermann JA, Selle F, et al. Olaparib tablets as maintenance therapy in patients with platinum-sensitive, relapsed ovarian cancer and a BRCA1/2 mutation (SOLO2/ENGOT-Ov21): a double-blind, randomised, placebo-controlled, phase 3 trial. *Lancet Oncol*. 2017;18(9):1274-1284. doi:10.1016/S1470-2045(17)30469-2
9. Coleman RL, Oza AM, Lorusso D, et al. Rucaparib maintenance treatment for recurrent ovarian carcinoma after response to platinum therapy (ARIEL3): a randomised, double-blind, placebo-controlled, phase 3 trial. *The Lancet*. 2017;390(10106):1949-1961. doi:10.1016/S0140-6736(17)32440-6
10. Moore K, Colombo N, Scambia G, et al. Maintenance Olaparib in Patients with Newly Diagnosed Advanced Ovarian Cancer. *N Engl J Med*. 2018;379(26):2495-2505. doi:10.1056/NEJMoa1810858
11. Ray-Coquard I, Pautier P, Pignata S, et al. Olaparib plus Bevacizumab as First-Line Maintenance in Ovarian Cancer. *N Engl J Med*. 2019;381(25):2416-2428. doi:10.1056/NEJMoa1911361
12. González-Martín A, Pothuri B, Vergote I, et al. Niraparib in Patients with Newly Diagnosed Advanced Ovarian Cancer. *N Engl J Med*. 2019;381(25):2391-2402. doi:10.1056/NEJMoa1910962
13. Coleman RL, Fleming GF, Brady MF, et al. Veliparib with First-Line Chemotherapy and as Maintenance Therapy in Ovarian Cancer. *N Engl J Med*. 2019;381(25):2403-2415. doi:10.1056/NEJMoa1909707
14. Kofanova O, Bellora C, Garcia Frasquilho S, et al. Standardization of the preanalytical phase of DNA extraction from fixed tissue for next-generation sequencing analyses. *New Biotechnol*. 2020;54:52-61. doi:10.1016/j.nbt.2019.07.005
15. Vos JR, Fakkert IE, de Hullu JA, et al. Universal Tumor DNA BRCA1/2 Testing of Ovarian Cancer: Prescreening PARPi Treatment and Genetic Predisposition. *JNCI J Natl Cancer Inst*. 2020;112(2):161-169. doi:10.1093/jnci/djz080
16. Griffin J, Treanor D. Digital pathology in clinical use: where are we now and what is

holding us back? *Histopathology*. 2017;70(1):134-145. doi:10.1111/his.12993

17. Ehteshami Bejnordi B, Mullooly M, Pfeiffer RM, et al. Using deep convolutional neural networks to identify and classify tumor-associated stroma in diagnostic breast biopsies. *Mod Pathol*. 2018;31(10):1502-1512. doi:10.1038/s41379-018-0073-z

18. Campanella G, Hanna MG, Geneslaw L, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med*. 2019;25(8):1301-1309. doi:10.1038/s41591-019-0508-1

19. Veta M, van Diest PJ, Willems SM, et al. Assessment of algorithms for mitosis detection in breast cancer histopathology images. *Med Image Anal*. 2015;20(1):237-248. doi:10.1016/j.media.2014.11.010

20. Ba W, Wang S, Shang M, et al. Assessment of deep learning assistance for the pathological diagnosis of gastric cancer. *Mod Pathol Off J U S Can Acad Pathol Inc*. 2022;35(9):1262-1268. doi:10.1038/s41379-022-01073-z

21. Coudray N, Ocampo PS, Sakellaropoulos T, et al. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. *Nat Med*. 2018;24(10):1559-1567. doi:10.1038/s41591-018-0177-5

22. Schaumberg AJ, Rubin MA, Fuchs TJ. *H&E-Stained Whole Slide Image Deep Learning Predicts SPOP Mutation State in Prostate Cancer*. Pathology; 2016. doi:10.1101/064279

23. Chang P, Grinband J, Weinberg BD, et al. Deep-Learning Convolutional Neural Networks Accurately Classify Genetic Mutations in Gliomas. *AJNR Am J Neuroradiol*. 2018;39(7):1201-1207. doi:10.3174/ajnr.A5667

24. Mobadersany P, Yousefi S, Amgad M, et al. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc Natl Acad Sci U S A*. 2018;115(13):E2970-E2979. doi:10.1073/pnas.1717139115

25. Bilal M, Raza SEA, Azam A, et al. Development and validation of a weakly supervised deep learning framework to predict the status of molecular pathways and key mutations in colorectal cancer from routine histology images: a retrospective study.

Lancet Digit Health. 2021;3(12):e763-e772. doi:10.1016/S2589-7500(21)00180-1

26. Kather JN, Pearson AT, Halama N, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat Med*. 2019;25(7):1054-1056. doi:10.1038/s41591-019-0462-y

27. Farahmand S, Fernandez AI, Ahmed FS, et al. Deep learning trained on hematoxylin and eosin tumor region of Interest predicts HER2 status and trastuzumab treatment response in HER2+ breast cancer. *Mod Pathol*. 2022;35(1):44-51. doi:10.1038/s41379-021-00911-w

28. Wang Y, Acs B, Robertson S, et al. Improved breast cancer histological grading using deep learning. *Ann Oncol Off J Eur Soc Med Oncol*. 2022;33(1):89-98. doi:10.1016/j.annonc.2021.09.007

29. Goode A, Gilbert B, Harkes J, Jukic D, Satyanarayanan M. OpenSlide: A vendor-neutral software foundation for digital pathology. *J Pathol Inform*. 2013;4(1):27. doi:10.4103/2153-3539.119005

30. Vahadane A, Peng T, Sethi A, et al. Structure-Preserving Color Normalization and Sparse Stain Separation for Histological Images. *IEEE Trans Med Imaging*. 2016;35(8):1962-1971. doi:10.1109/TMI.2016.2529665

31. Bankhead P, Loughrey MB, Fernández JA, et al. QuPath: Open source software for digital pathology image analysis. *Sci Rep*. 2017;7(1):16878. doi:10.1038/s41598-017-17204-5

32. Griebel M, Segebarth D, Stein N, et al. Deep-learning in the bioimaging wild: Handling ambiguous data with deepflash2. Published online November 12, 2021. doi:10.48550/arXiv.2111.06693

33. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, eds. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Lecture Notes in Computer Science. Springer International Publishing; 2015:234-241. doi:10.1007/978-3-319-24574-4_28

34. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In: Stoyanov D, Taylor Z, Carneiro G, et al., eds. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Lecture Notes in Computer Science. Springer International Publishing; 2018:3-11. doi:10.1007/978-3-030-00889-5_1

35. Chaurasia A, Culurciello E. LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation. In: *2017 IEEE Visual Communications and Image Processing (VCIP)*. ; 2017:1-4. doi:10.1109/VCIP.2017.8305148

36. Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y, eds. *Computer Vision – ECCV 2018*. Vol 11211. Lecture Notes in Computer Science. Springer International Publishing; 2018:833-851. doi:10.1007/978-3-030-01234-2_49

37. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE; 2016:770-778. doi:10.1109/CVPR.2016.90

38. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Published online 2016:2818-2826.

39. Tan M, Le QV. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *International Conference of Machine Learning ICML*. Published online 2019:6105-6114.

40. Zhang H, Wu C, Zhang Z, et al. ResNeSt: Split-Attention Networks. *Conference on Computer Vision and Pattern Recognition (CVPR)*. Published online 2022:2736-2746.

41. Ilse M, Tomczak JM, Welling M. Attention-based Deep Multiple Instance Learning. Published online June 28, 2018. doi:10.48550/arXiv.1802.04712

42. Feng R, Liu X, Chen J, Chen DZ, Gao H, Wu J. A Deep Learning Approach for Colonoscopy Pathology WSI Analysis: Accurate Segmentation and Classification. *IEEE J*

Biomed Health Inform. 2021;25(10):3700-3708. doi:10.1109/JBHI.2020.3040269

43. Wahab N, Miligy IM, Dodd K, et al. Semantic annotation for computational pathology: multidisciplinary experience and best practice recommendations. *J Pathol Clin Res*. 2022;8(2):116-128. doi:10.1002/cjp2.256

44. Ho DJ, Chui MH, Vanderbilt CM, et al. Deep Interactive Learning-based ovarian cancer segmentation of H&E-stained whole slide images to study morphological patterns of BRCA mutation. *J Pathol Inform*. 2023;14:100160. doi:10.1016/j.jpi.2022.100160

45. Chen M, Zhang B, Topatana W, et al. Classification and mutation prediction based on histopathology H&E images in liver cancer using deep learning. *Npj Precis Oncol*. 2020;4(1):1-7. doi:10.1038/s41698-020-0120-3

46. Nero C, Boldrini L, Lenkowicz J, et al. Deep-Learning to Predict BRCA Mutation and Survival from Digital H&E Slides of Epithelial Ovarian Cancer. *Int J Mol Sci*. 2022;23(19):11326. doi:10.3390/ijms231911326

47. Lazard T, Bataillon G, Naylor P, et al. Deep learning identifies morphological patterns of homologous recombination deficiency in luminal breast cancers from whole slide images. *Cell Rep Med*. 2022;3(12):100872. doi:10.1016/j.xcrm.2022.100872

48. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. Published online June 3, 2021. doi:10.48550/arXiv.2010.11929

49. Guo B, Li X, Yang M, Jonnagaddala J, Zhang H, Xu XS. 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*. Published online 2023. doi:10.1002/cjp2.312

FIGURE LEGENDS

Figure 1: Original HES slide (A). Tumor mask annotated by a pathologist in red (B). Sampling weighted mask as a combination of the tumor mask and the tissue mask (C). The tumor mask

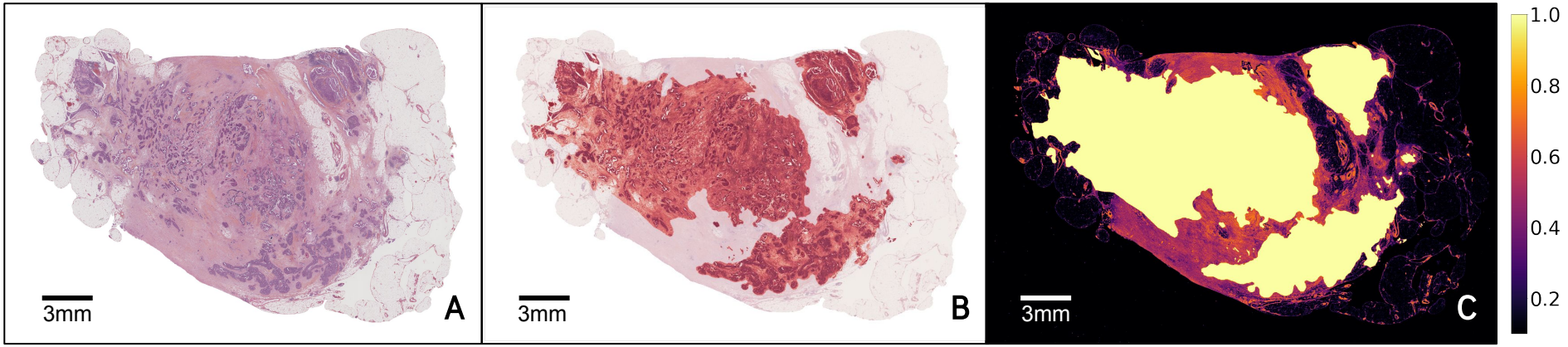
in yellow has a sampling weight of 1 while the tissue mask in pink has a sampling weight of 0.7 and 0.1 for the background.

Figure 2: Original HES slide (A). Predicted tumor mask in purple (B). Tiling of the predicted mask: a maximum of 5000 tiles with more than 50% of their surface covered by the tumor mask are kept (in red). The green tiles include less than 50% tumor pixels and are therefore discarded (C).

Figure 3: Overview of the multiscale approach. First, the tumor pixels are predicted by the ovarian cancer segmentation model. The predicted tumor masks are tiled into 512 x 512 pixels tiles. The resulting tiles are then separately divided into 4 tiles of 256 x 256 pixels and then randomly reassembled into a new mixed tile of 512 x 512 pixels (A). All of these datasets are then embedded into a low-dimensional space using a ResNet-50 as encoder, trained using the momentum contrast technique with N, the number of tiles per WSI (B). The embedded tiles are then scored through an attention mechanism and are finally fed to a decision module to predict BRCA mutation (C).

FIG1.pdf

Click here to access/download;Figure;FIG1.pdf



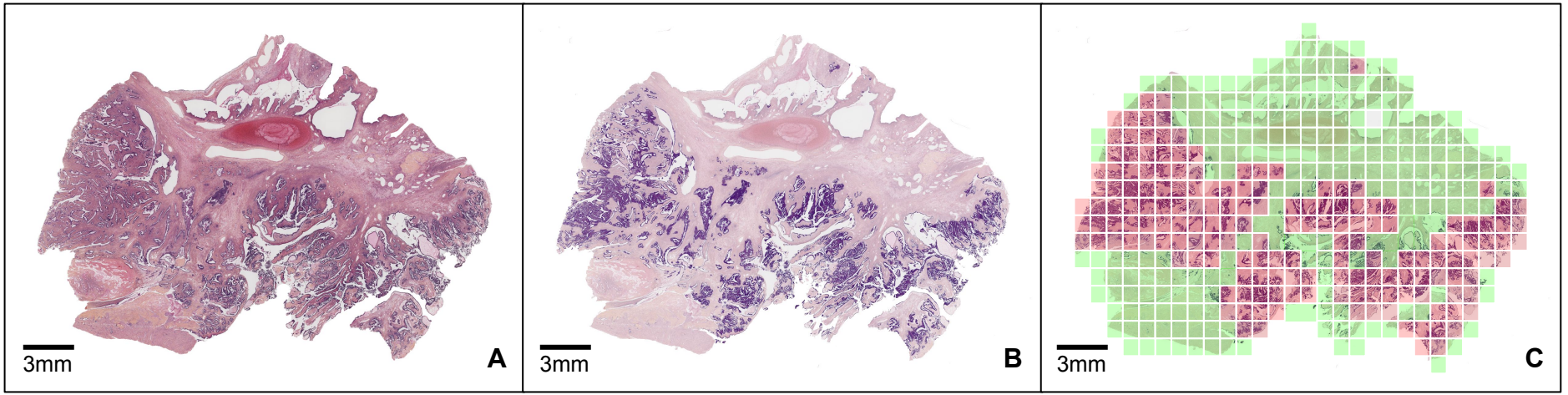
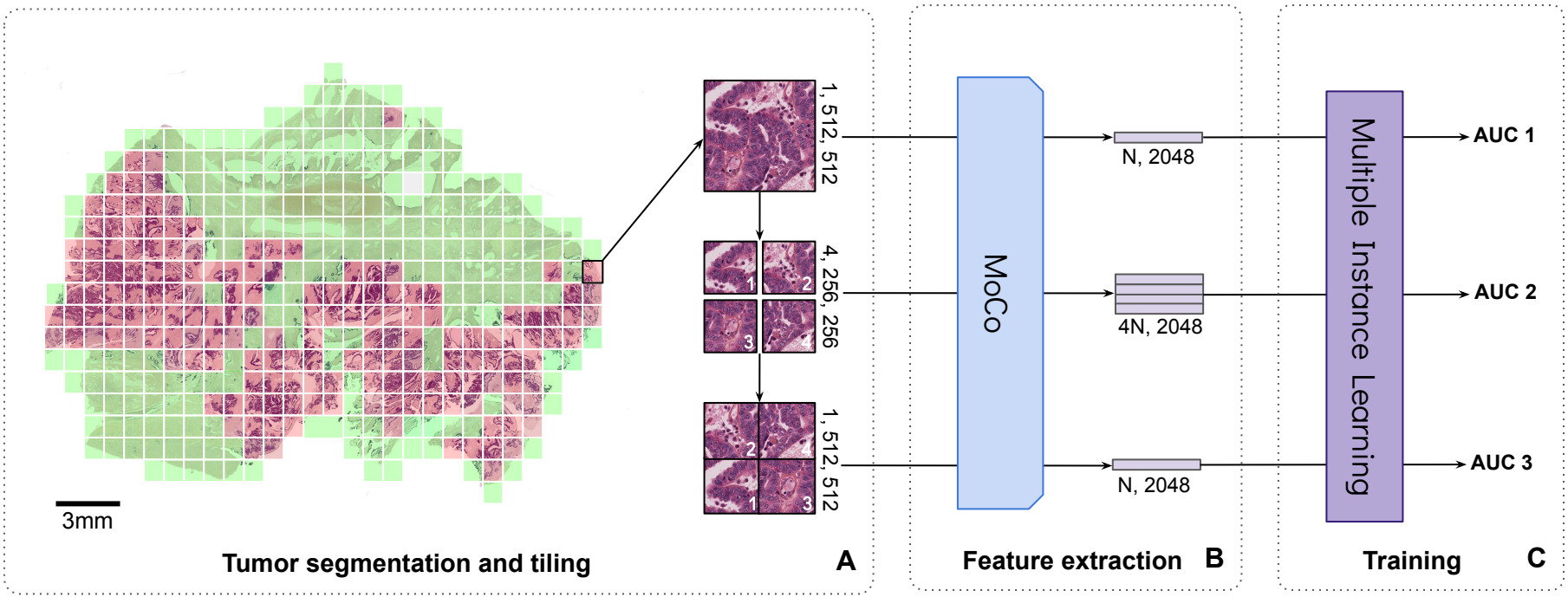


FIG3.pdf

Click here to access/download;Figure;FIG3.pdf



	Dataset 1	Dataset 2	Dataset 3
Source	Nantes	Rennes	TCGA
Number of tumor tiles for self-supervised learning	331 998	267 699	none
Number of tumor tiles for classification : *BRCA*-mut *BRCA*-wt	 129 041 460 832	 53 508 396 768	 20972 90755

Table 1: The number of tumor tiles used to learn MoCo representations and perform the classification according to the MIL paradigm.

	5-fold cross-validation		Internal testing set		TCGA	
	AUC	Sd	AUC	Sd	AUC	Sd
Baseline	0,739	0.024	0,681	0,014	0,631	0,035
Approach 2	0,721	0.014	0,654	0,011	0.624	0.024
Approach 3	0,697	0.016	0,635	0,015	0.619	0.024

Table 2: Summary of performance metrics. Mean and standard deviation (SD) are computed over the five test sets of the cross-validation, the internal testing set, and the TCGA dataset. AUC, area under the (receiver-operating characteristics) curve. The baseline approach relies on 512 x 512 pixels tiles. Approach 2 relies on the division of these tiles into 4 tiles of 256 x 256 pixels. Approach 3 relies on the random reassembling of these 4 tiles into a new mixed tumor tile of 512 x 512 pixels.