Fundamental Research 4 (2024) 678–689

Contents lists available at ScienceDirect

Fundamental Research

journal homepage: http://www.keaipublishing.com/en/journals/fundamental-research/

Article

Deep learning framework for comprehensive molecular and prognostic stratifications of triple-negative breast cancer

Shen Zhao[a,1], Chao-Yang Yan[b,1], Hong Lv[c,1], Jing-Cheng Yang[a,d,1], Chao You[e,1], Zi-Ang Li[b], Ding Ma[a], Yi Xiao[a], Jia Hu[a], Wen-Tao Yang[c,*], Yi-Zhou Jiang[a,*], Jun Xu[b,*], Zhi-Ming Shao[a,*]

[a] Key Laboratory of Breast Cancer in Shanghai, Department of Breast Surgery, Fudan University Shanghai Cancer Center, Shanghai 200032, China
[b] Institute for AI in Medicine, School of Artificial Intelligence, Nanjing University of Information Science and Technology, Nanjing 210044, China
[c] Department of Pathology, Fudan University Shanghai Cancer Center, Shanghai 200032, China
[d] Greater Bay Area Institute of Precision Medicine, Guangzhou 511466, China
[e] Department of Radiology, Fudan University Shanghai Cancer Center, Shanghai 200032, China

ARTICLE INFO

Article history:
Received 5 December 2021
Received in revised form 9 June 2022
Accepted 16 June 2022
Available online 29 June 2022

Keywords:
Triple-negative breast cancer
Deep learning
Digital pathology
Patient stratification
Online platform

ABSTRACT

Triple-negative breast cancer (TNBC) is the most challenging breast cancer subtype. Molecular stratification and target therapy bring clinical benefit for TNBC patients, but it is difficult to implement comprehensive molecular testing in clinical practice. Here, using our multi-omics TNBC cohort (N = 425), a deep learning-based framework was devised and validated for comprehensive predictions of molecular features, subtypes and prognosis from pathological whole slide images. The framework first incorporated a neural network to decompose the tissue on WSIs, followed by a second one which was trained based on certain tissue types for predicting different targets. Multi-omics molecular features were analyzed including somatic mutations, copy number alterations, germline mutations, biological pathway activities, metabolomics features and immunotherapy biomarkers. It was shown that the molecular features with therapeutic implications can be predicted including the somatic *PIK3CA* mutation, germline *BRCA2* mutation and PD-L1 protein expression (area under the curve [AUC]: 0.78, 0.79 and 0.74 respectively). The molecular subtypes of TNBC can be identified (AUC: 0.84, 0.85, 0.93 and 0.73 for the basal-like immune-suppressed, immunomodulatory, luminal androgen receptor, and mesenchymal-like subtypes respectively) and their distinctive morphological patterns were revealed, which provided novel insights into the heterogeneity of TNBC. A neural network integrating image features and clinical covariates stratified patients into groups with different survival outcomes (log-rank $P < 0.001$). Our prediction framework and neural network models were externally validated on the TNBC cases from TCGA (N = 143) and appeared robust to the changes in patient population. For potential clinical translation, we built a novel online platform, where we modularized and deployed our framework along with the validated models. It can realize real-time one-stop prediction for new cases. In summary, using only pathological WSIs, our proposed framework can enable comprehensive stratifications of TNBC patients and provide valuable information for therapeutic decision-making. It had the potential to be clinically implemented and promote the personalized management of TNBC.

1. Introduction

Breast cancer has become the most commonly diagnosed cancer according to the recent global cancer estimates [1]. Triple-negative breast cancer (TNBC) encompasses a subset of breast cancers that lack expression of the estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2). It accounts for 15% to 20% of newly diagnosed breast cancer cases [2,3]. Compared with hormone receptor-positive or HER2-positive breast cancer, TNBC has highly aggressive biological behavior and is associated with a poor prognosis [3,4].

Much effort has gone into the molecular characterization of TNBC in the past decades [5–7]. Our research team has performed multi-omics profiling of a large cohort of TNBC and characterized its genomic, transcriptomic, germline, metabolic and microenvironmental alterations [8–11]. A following umbrella clinical trial (NCT03805399) demonstrated that the molecular stratification and tailored target therapy can bring clinical benefit for patients with refractory TNBC [12]. Despite these en-

* Corresponding authors.
E-mail addresses: yangwt2000@163.com (W.-T. Yang), yizhoujiang@fudan.edu.cn (Y.-Z. Jiang), xujung@gmail.com (J. Xu), zhimingshao@fudan.edu.cn (Z.-M. Shao).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.fmre.2022.06.008

couraging findings, the wide implementation of comprehensive molecular testing is still difficult in either clinical trials or routine practice due to its high cost, long turnaround time and complicated technological process. Thus, for better clinical translation of our research findings, there is a need for an inexpensive, fast and convenient approach to realize the molecular stratification of TNBC patients.

In recent years, the advent of deep learning has revolutionized the image analysis field and it has been increasingly used in medical image analysis. Certain molecular alterations can influence the biological behavior of tumor cells and was subsequently reflected on pathological morphology [13,14]. The application of deep learning algorithm to hematoxylin and eosin (H&E) stained whole slide images (WSIs) can capture these morphological patterns to infer the corresponding molecular alterations. Based on this rationale, Kather et al. explored the detection of microsatellite instability in intestinal cancer [15,16]. Woerl et al. and Sirinukunwattana et al. built digital pathology-based classifiers for the molecular subtyping of bladder cancer and colorectal cancer respectively [17,18]. Kather et al. and Fu et al. performed pan-cancer studies and found that numerous molecular features can be inferred from WSIs [19,20]. These studies demonstrated the feasibility of cancer molecular stratification based on digital pathology and deep learning, but they had some limitations. First, most studies incorporated manual annotation of tumor tissue by pathologists, which required pathological expertise and led to much additional manual operation and lowered scalability and reproducibility of the workflow. Kather et al. used a tumor detector, but it was only suitable for targets associated with the tumor tissue [16]. Second, only the two pan-cancer studies by Kather et al. and Fu et al. examined a relatively broad class of molecular features [19,20]. Nevertheless, there were still some other important molecular features that were not analyzed including tumor metabolic features and germline variant spectrum. Third, none of these studies packaged their models or workflows into an easy-to-use tool, which greatly limited the potential clinical use.

In this study, using the multi-omics TNBC cohort from Fudan University Shanghai Cancer Center (FUSCC) and the TNBC cases from The Cancer Genome Atlas (TCGA), we developed a deep learning-based image analysis framework and explored the predictions of a wider range of molecular features and patient prognosis from pathological WSIs. Based on our findings, we built an online platform tool for real-time one-stop prediction and potential clinical implementation.

2. Materials and methods

2.1. Data source

Our study comprised two cohorts. The discovery cohort was the multi-omics TNBC cohort from FUSCC, which was established as described in our previous study [8]. The paraffin-embedded H&E stained tumor slides were collected and scanned. 425 digital WSIs (from 425 cases) were determined for analysis. The external validation cohort consisted of the TNBC cases from the TCGA dataset. 143 diagnostic slides (from 143 cases) were used which were collected from different centers. The patient clinicopathological characteristics were outlined in Table S1. The somatic mutation, mRNA expression and copy number data were available for both cohorts, while the germline mutation and metabolomics data were only available for the FUSCCTNBC cohort (Fig. S1).

2.2. Deep learning-based framework

A deep learning-based framework was established to develop models for predicting multi-omics molecular features and patient relapse risk from pathological WSIs (Fig. 1a). In brief, the scanned WSIs of tumor slides were first tessellated into 256 × 256 pixel image tiles at 20 × magnification which is consistent with the previous studies [21–23]. The background tiles were filtered out and structure-preserved color normalization was performed (Fig. S2, see "Supplementary Methods") [24]. These steps generated several hundred to more than twenty thousand of non-background color-normalized tiles per WSI (Fig. S3). Then, the framework contained two separate convolutional neural networks (CNNs) in series. The first was a tile-level tissue type classifier. It served as a filter for tiles of certain tissue types which were associated with the prediction target. The second CNN was trained based on these tiles for predicting different targets. Three-fold cross-validation was adopted for model training and validation in the FUSCCTNBC cohort. The best models were applied to the TCGATNBC cohort for external validation (Fig. 1b).

Detailed methods for data collection, image preprocessing, all experiments and analyses, and platform establishment are described in Supplementary Methods.

2.3. Statistical analysis

Deep learning experiments were performed in Python with Pytorch [25]. For tissue type classification, the accuracy was measured by the weighted F1 score [26]. For molecular feature prediction, the accuracy was measured by receiver operating characteristics (ROC) analysis using the area under the curve (AUC) with 95% confidence interval (CI) [27]. Continuous variables were compared between groups using the Mann-Whitney U test. False discovery rates (FDRs) were calculated to account for multiple testing using the R function "p.adjust" [28]. All tests were two-sided, and an FDR of < 0.1 was considered statistically significant. Relapse-free survival (RFS) was defined as the time from diagnosis to the first relapse, a contralateral breast cancer or death due to any cause. Survival curves were plotted using the Kaplan-Meier method and survival differences were compared between groups using the log-rank test and Cox proportional hazards models.

3. Results

3.1. Tile-level tissue type classification

To select certain types of image tiles for the following predictions, we developed a tile-level tissue type classifier. We manually annotated tissue regions of tumor, stroma, immune infiltrates, normal breast gland and necrosis/hemorrhage on 20 WSIs and divided the annotated areas into 256 × 256 pixel tiles (Fig. S4). We used 100,066 tiles for training and 15,034 tiles for validation (Fig. 2a). Tiles were fed into the pretrained ResNet-18 network for fine-tuning. The best model achieved a weighted F1 score of 0.96 in the validation set. For all five tissue types, the classification accuracy achieved near 90% (Fig. 2b). The internal representations of image tiles were visualized through the tSNE method (Fig. 2c). Tissue types aggregated in separate clusters, which indicated that the deep features learned by our model can discriminate these five tissue types. The tissue type segmentation results further verified the validity of our classifier (Fig. 2d).

This tissue type classifier was applied to all image tiles from our two cohorts to categorize them into the five tissue types. For the following prediction tasks, tiles of certain types were sampled which were associated with the prediction target (Table S2).

3.2. Predictions of multi-omics molecular features

In this section, we aimed to develop CNN models for predicting multi-omics molecular features from WSIs. To select an efficient neural network architecture with suitable deep learning hyperparameters, we used the prediction of somatic *PIK3CA* mutation, which was a known therapeutic target, as a benchmark task (see "Supplementary Methods") [29,30]. We compared the difference in the prediction accuracy measured by ROC analysis and average training time among four commonly used CNN architectures with different settings. The ResNet-18 network with all layers trainable showed relatively high accuracy with moderate

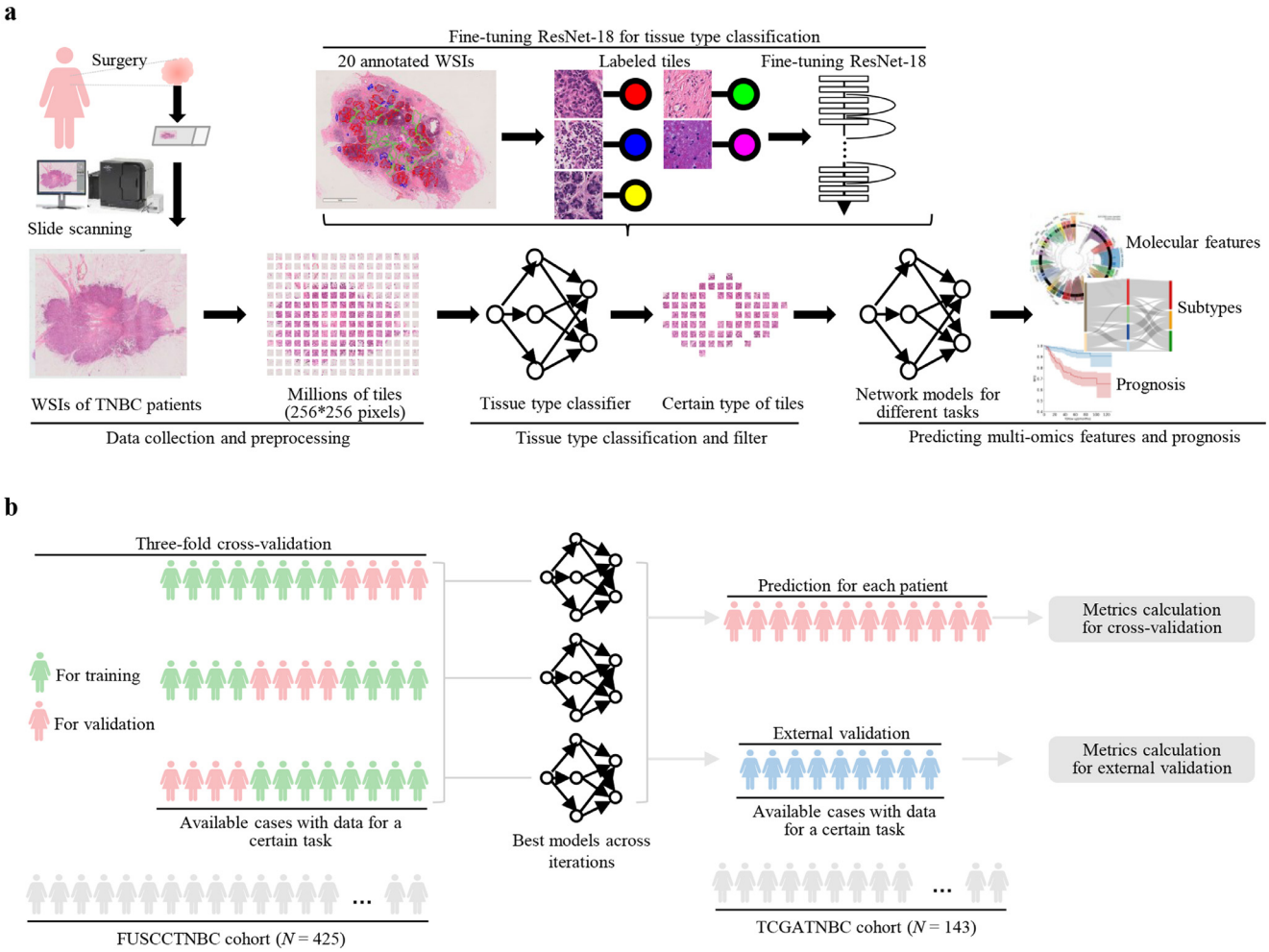

Fig. 1. Study overview. (a) Deep learning-based prediction framework. The framework architecture comprises two separate convolutional neural networks (CNNs) in series. The first is a tissue type classifier, which is developed based on 20 WSIs with pixel-level tissue type annotations. It enables automatic sampling of certain types of image tiles which are associated with the prediction target. The second CNN is trained based on the sample tiles for different targets. (b) Data usage and modeling strategies. Models are trained and validated in the FUSCCTNBC cohort through three-fold cross-validation. For each fold, the model that shows the highest prediction accuracy in the validation set is determined. The prediction scores generated by the three best models for their corresponding validation set are concatenated. Thus, each patient obtains a prediction score, but none of the patients has ever been a member of the training set and validation set at the same time. Based on this prediction score, performance metrics for cross-validation are calculated. Then, all of the three models are applied to the patients from the TCGATNBC cohort. Each patient obtains three prediction scores and the average is used as the final prediction. Based on this final prediction, performance metrics for external validation are calculated. Abbreviations: TNBC, triple-negative breast cancer; WSIs, whole slide images; FUSCC, Fudan University Shanghai Cancer Center; TCGA, The Cancer Genome Atlas.

training time and therefore was selected for the following tasks (Fig. S5; Table S3).

We first developed CNN models to predict somatic mutations with a frequency of ≥ 4% in the FUSCCTNBC cohort. The cross-validation AUC ranged from 0.66 to 0.80 (Fig. 3a; Table S4). *PIK3CA* mutation can be predicted with an AUC of 0.78 (95% CI: 0.71-0.86). Second, we aimed to predict the recurrent focal copy number alterations. We concentrated on the top five frequent focal amplifications and deletions and the AUCs were between 0.57 and 0.68 (Fig. 3b; Table S5). Third, we explored the predictions of germline mutations and found that the *BRCA2* mutation can be predicted with an AUC of 0.79 (95% CI: 0.67-0.91) (Fig. 3c; Table S6). Fourth, we examined whether certain biological states or processes can be inferred from pathological WSIs. The gene set enrichment analysis (GSEA) scores of ten cancer-related hallmark gene sets were calculated and used as the labels (see "Supplementary Methods") [31,32]. The AUCs ranged from 0.69 to 0.82 (Fig. 3d; Table S7). Fifth, using the metabolomics data, we calculated the abundance scores for certain category metabolites or metabolic pathways and developed models to predict these scores [33]. The accuracy was highest for predicting the abundance of fatty acid metabolites with an AUC of 0.70 (95% CI: 0.64-0.76) (Fig. 3e; Table S8). Sixth, since immunotherapy has been shown great potential as an effect therapy for TNBC, we tried to predict the biomarkers that were associated with immunotherapy response (see "Supplementary Methods") [34–36]. The AUCs ranged from 0.70-0.87. The stromal tumor-infiltrating lymphocytes (sTILs) and PD-L1 immune cell score, which were commonly used biomarkers, can be predicted with AUCs of 0.87 (95% CI: 0.84-0.91) and 0.74 (95% CI: 0.69-0.79) respectively (Fig. 3f; Table S9).

We investigated which morphological patterns held clues for the molecular features that can be predicted through visualization analysis (see "Supplementary Methods"). We visually reviewed the representative tiles from the true positive patients with the highest prediction confidence. The prediction for somatic *PIK3CA* mutation and germline *BRCA2* mutation displayed evident correlation with different patterns. Tiles indicating somatic *PIK3CA* mutation contained large tumor cells with abundant cytoplasm and evident nucleolus. By con-

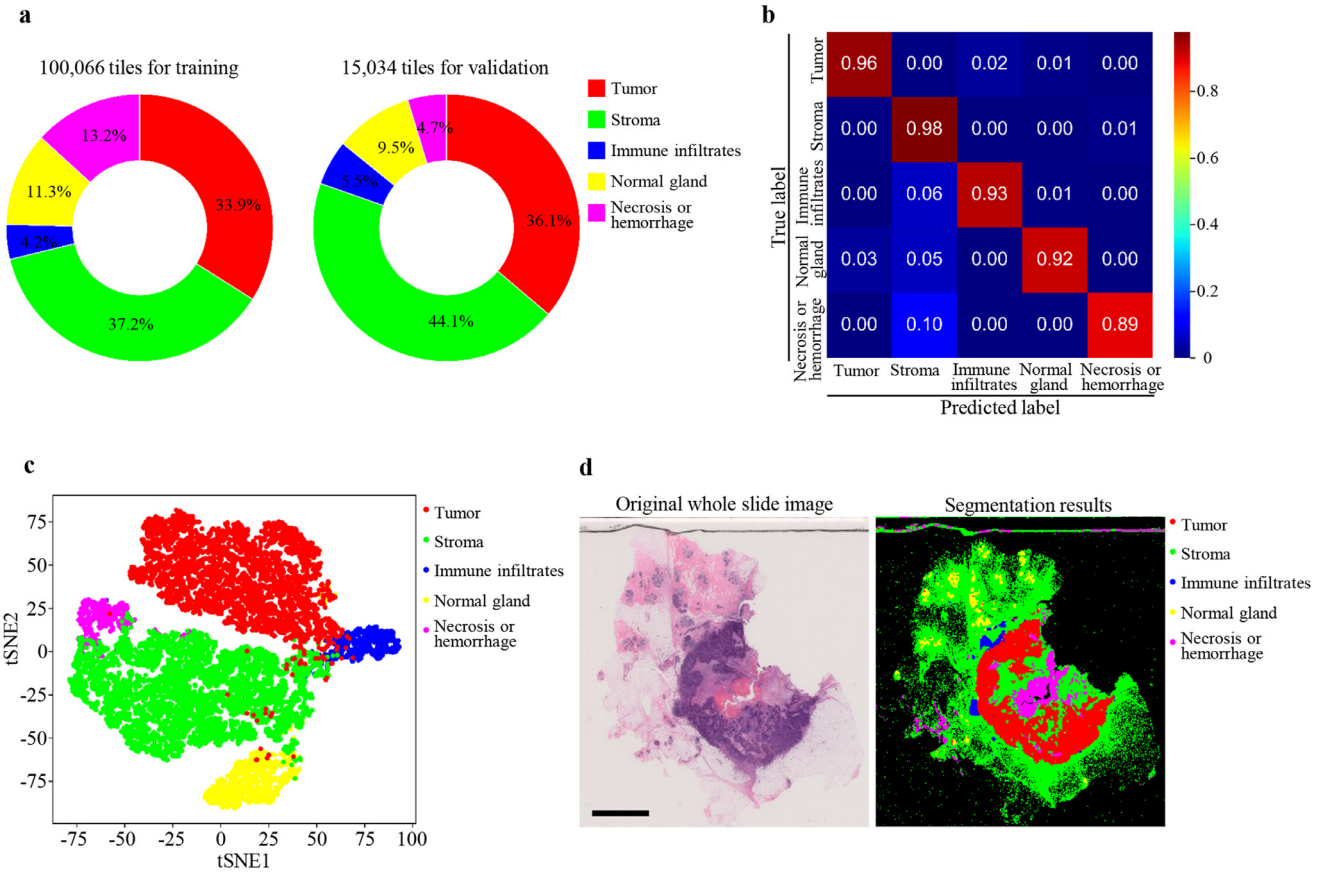

Fig. 2. Development of convolutional neural network models for automatic tissue type classification on whole slide images. (a) Training and validation sets for the development of tissue type classifier. (b) Confusion matrix of tissue type classification results in the validation set. (c) Visualization of the internal representation of image tiles by applying the tSNE method to the deep features extracted by the classification model. (d) One example of tissue segmentation results from the validation set. Whole slide image scale bar, 4 mm.

trast, those indicating *BRCA2* mutation were characterized by densely distributed tumor cells (Fig. 3g-h). This analysis indicated the biological mechanisms of how these alterations influenced cancer growth and progression.

3.3. Predictions of TNBC molecular subtypes

Molecular subtyping of TNBC has deconstructed the molecular portraits of this heterogeneous disease. Subtyping-based treatment strategy has been demonstrated to be effective in clinical trials [12]. In this section, we explored the use of deep learning algorithm to infer the TNBC molecular subtypes from WSIs. We first developed models to identify each of the four subtypes. The AUC was 0.84 (95% CI: 0.80-0.88) for identifying the BLIS subtype, 0.85 (95% CI: 0.79-0.90) for the IM subtype, 0.93 (95% CI: 0.90-0.96) for the LAR subtype and 0.73 (95% CI: 0.65-0.81) for the MES subtype (Fig. 4a-d; Table S10). We next integrated the prediction scores from the four models and compared them with the subtype probability scores from our initial molecular subtyping which is determined by mRNA-based k-means clustering (see “Supplementary Methods”) [8]. The comparison of score distributions indicated a high level of concordance in most cases (Fig. 4e). In all four subtypes, the cosine similarity between the probability scores and our prediction scores was significantly higher than the similarity between the probability scores and random prediction (Fig. 4f).

For each subtype, we presented the representative tiles (Fig. 4g-j). Both the BLIS and IM subtypes had densely distributed tumor cells. Significant immune cell infiltrates discriminated the IM subtype from the BLIS subtype. The LAR and MES subtypes showed discrete distribution of tumor cells. The LAR subtype was characterized by the large tumor cells with evident cell borders and nucleolus, while the MES subtype seemed associated with the presence of abundant mesenchymal tissue around the tumor cells.

3.4. Improvement of relapse risk assessment

We hypothesized that integration of the information held by the pathological WSIs and traditional prognostic factors can improve the accuracy of prognostic evaluation. On the basis of the DeepSurv network architecture [37], we compared three modeling strategies: a) based on traditional clinicopathological prognostic factors, b) based on the tiles’ deep features extracted by our tissue type classifier and c) integrating the clinicopathological prognostic factors and tiles’ deep features (Fig. 5a, see “Supplementary Methods”). RFS was used as the endpoint. The median length of follow-up was 64.9 months (interquartile range, 48.1-79.5 months) in the FUSCCTNBC cohort and 24.0 months (interquartile range, 13.5-53.0 months) in the TCGATNBC cohort. According to univariate Cox analysis, tumor T category, N category and radiotherapy were identified as clinical prognostic factors which were significantly associated with RFS (Table S11). Radiotherapy was not used in the subsequent modeling because its negative prognostic effect might be attributed to selection bias [38]. We used concordance index (c index) to evaluate the models’ prediction accuracy. The integrated models and image-based models significantly outperformed

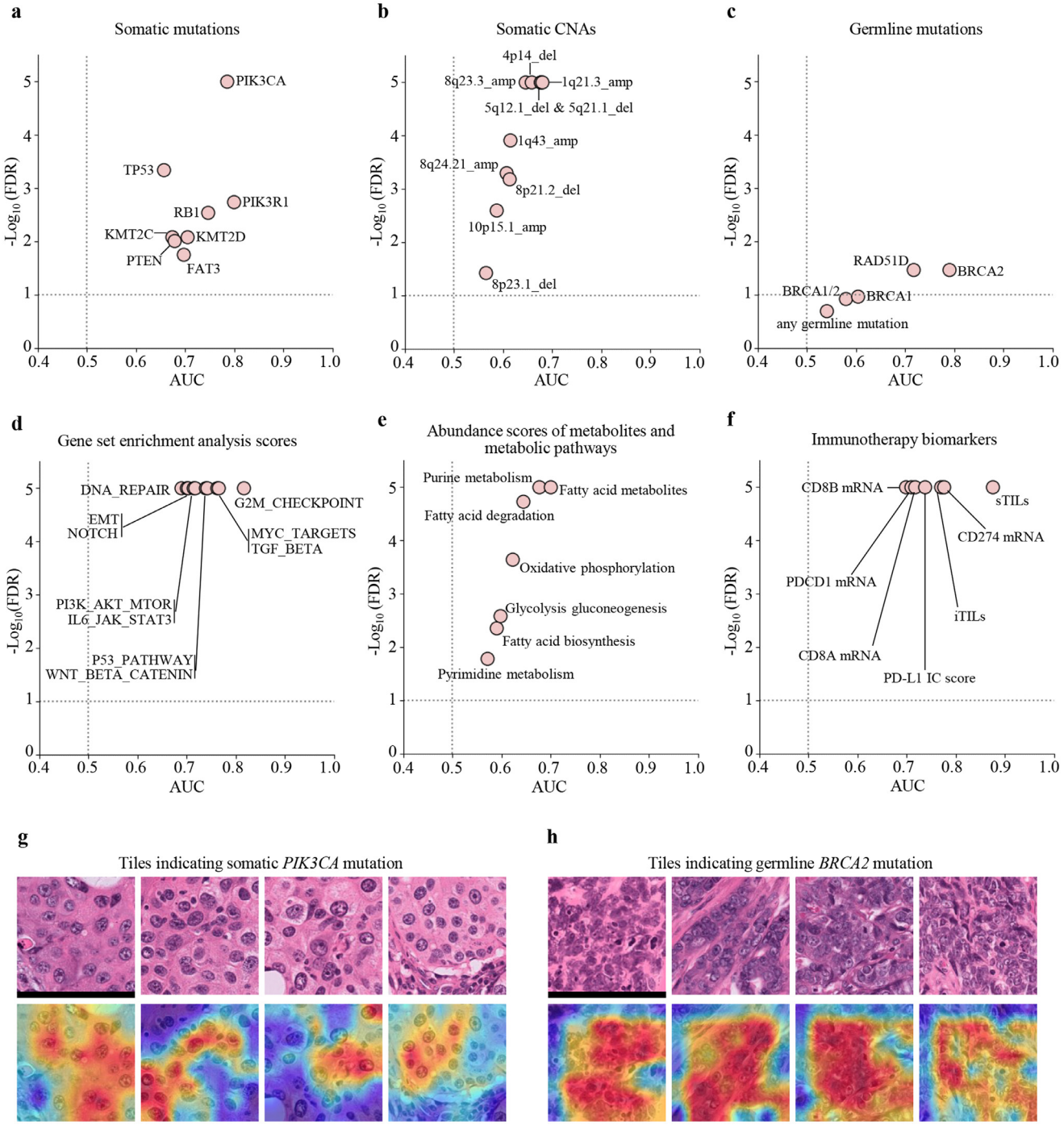

Fig. 3. Predictions of multi-omics molecular features based on digital whole slide images using deep learning. (a)-(f) illustrate the models' prediction accuracy. Patient-level prediction scores are used for ROC analysis and are compared between the positive and negative groups according to the true labels. The prediction accuracy is measured with the AUC values and FDR-corrected *P* values. FDR values smaller than 10^{-5} are set to 10^{-5}. (a) Predictions of somatic mutations. The genes mutated in at least 4% of cases in the FUSCCTNBC cohort are predicted. (b) Predictions of focal copy number alterations. The top five frequent focal amplifications and deletions are predicted. (c) Predictions of germline mutations. The top three frequent germline mutations, cases with either *BRCA1* or *BRCA2* mutation, and those with any germline mutation are predicted. (d) Predictions of hallmark gene set enrichment analysis (GSEA) scores. Ten cancer-related hallmark gene sets are selected from Molecular Signatures Database. The enrichment score is calculated for each sample using the single sample GSEA method. (e) Predictions of metabolic abundance scores. Six metabolic pathways and fatty acid metabolites are predicted. (f) Predictions of immunotherapy biomarkers. Stromal and intratumoral tumor-infiltrating lymphocytes, PD-L1 expression (measured as the PD-L1 immune cell scores and CD274 mRNA expression), PD-1 expression (measured as the PDCD1 mRNA expression) and CD8 expression (measured as the CD8A and CD8B mRNA expression) are predicted. (g)-(h) Tiles from the true-positive cases with the highest prediction score for (g) somatic *PIK3CA* mutation and (h) germline *BRCA*2 mutation. Class activation maps highlight the subregions that contribute most to the prediction. Tile scale bar, 128 μm. Abbreviations: AUC, area under the curve; FDR, false discovery rate; CNAs, copy number alterations; amp, amplification; del, deletion; EMT, epithelial mesenchymal transition; sTILs, stromal tumor-infiltrating lymphocytes; iTILs, intratumoral tumor-infiltrating lymphocytes; IC, immune cell.

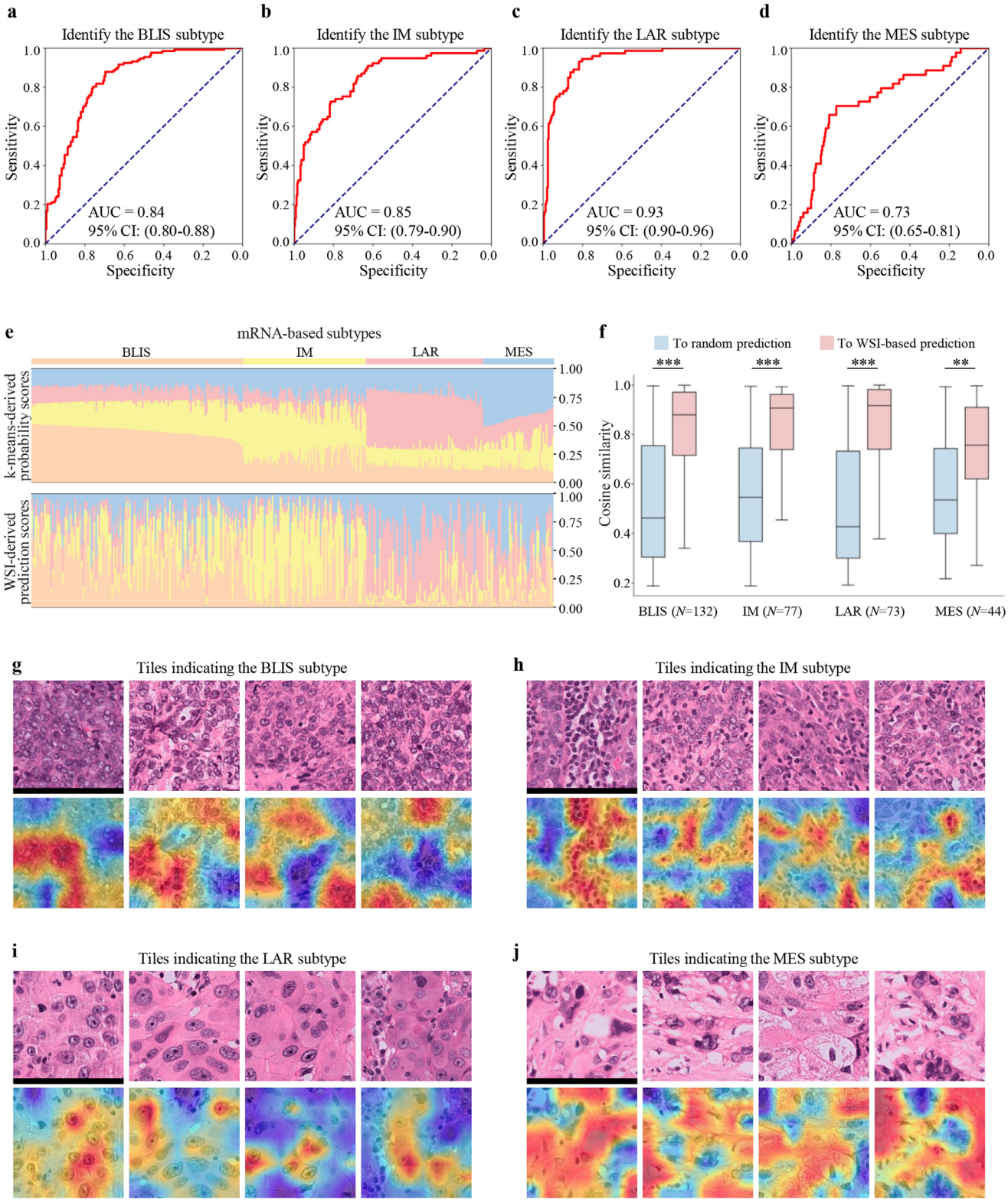

Fig. 4. Image-based identification of TNBC molecular subtypes using deep learning. (a)-(d) ROC curves for identifying the four TNBC subtypes based on WSIs. (e) Comparison between the mRNA expression-based molecular classification results and our prediction results. Each vertical line displays the distribution of classification scores of a sample. The upper panel shows the probability scores derived from k-means clustering on mRNA data, which was used in our previous study to determine the molecular subtype for each sample (see "Supplementary Methods"). The lower panel presents the scores generated by our prediction models. Samples are sorted by the mRNA expression-based molecular subtypes and the corresponding probability scores. (f) Cosine similarity between our prediction results and the k-means clustering results. It is compared with the similarity between random prediction results and the k-means clustering results. The middle line, box's bounds, and whiskers of the boxplot represent the median, 25%–75% interquartile range (IQR), and 1.5 times the IQR from the upper and lower quartiles, respectively. *P* values are calculated using the Mann-Whitney U test. ***, $P < 0.001$; **, $P < 0.01$. (g)-(j) Tiles from the cases of each molecular subtype with the highest prediction score. Class activation maps highlight the subtype-specific discriminative subregions. Tile scale bar, 128 μm. Abbreviations: AUC, area under the curve; CI, confidence interval.

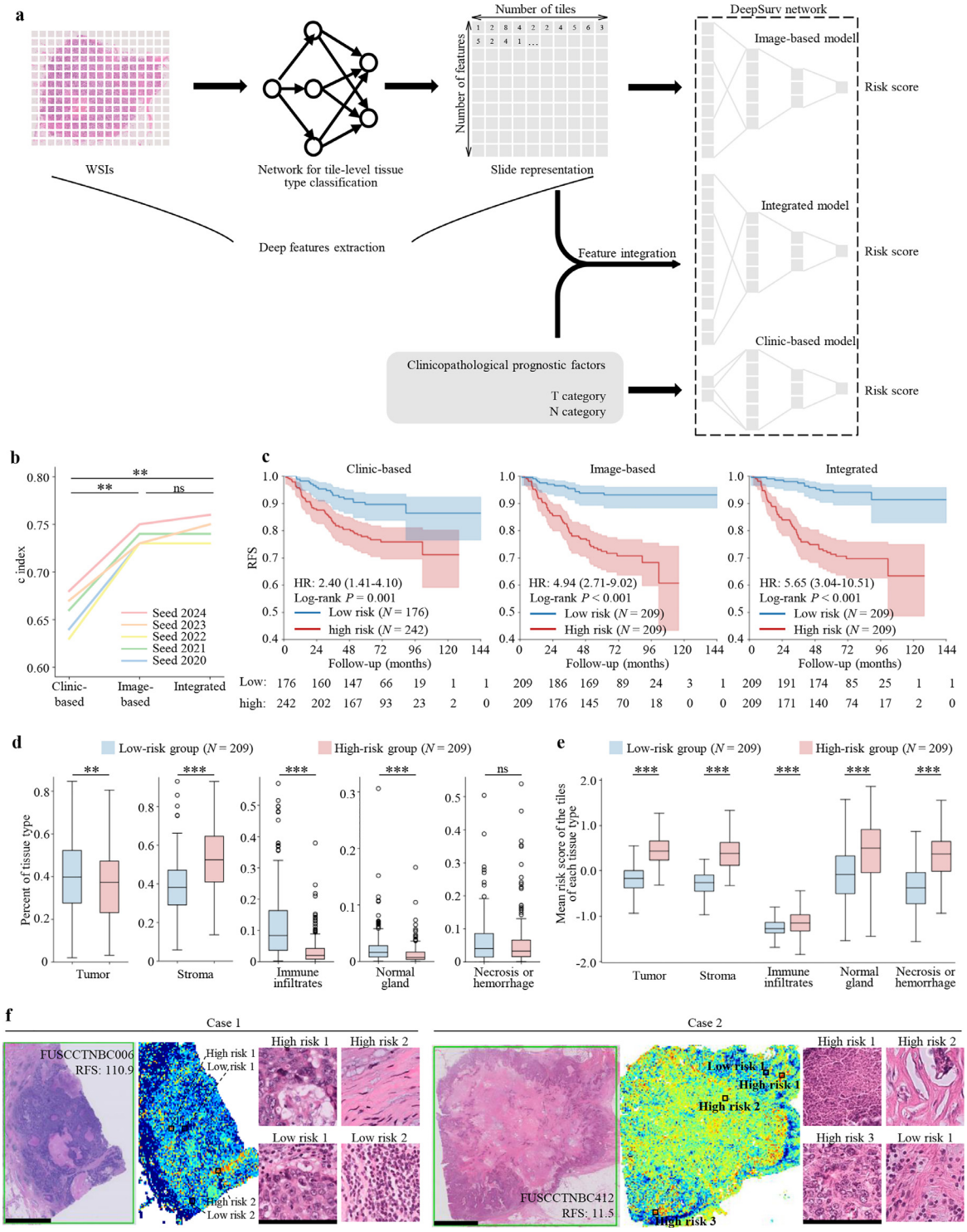

Fig. 5. Pathological whole slide images improve prognostic evaluation of TNBC. (a) Diagram for predicting relapse risk using three strategies. (b) Comparison of prediction accuracy of the clinic-based, image-based and integrated prognostic models. *P* values are calculated using the Mann-Whitney U test with false discovery rate (FDR)-correction for multiple testing. **, FDR < 0.01; ns, not significant. (c) Kaplan-Meier curves of the high- and low-risk groups classified by the prognostic models. (d)-(e) Comparison of the (d) tissue type proportions and (e) the mean risk score of each tissue type between the high- and low-risk groups according to the image-based model. The middle line, box's bounds, and whiskers of the boxplot represent the median, 25%–75% interquartile range (IQR), and 1.5 times the IQR from the upper and lower quartiles, respectively. *P* values are calculated using the Mann-Whitney U test. ***, $P < 0.001$; **, $P < 0.01$; ns, not significant. (f) Two examples of spatial risk prediction heatmap with representative tiles indicating high or low relapse risk. The left case is a patient with long RFS. While either the tiles of immune infiltrates (low risk 2) or tumor with abundant adjacent lymphocytes (low risk 1) are associated with low relapse risks, tiles of tumor (high risk 1) or stroma (high risk 2) with few immune infiltrates are correlated with high relapse risks. By contrast, the right case is a patient whose tumor relapsed after surgery within 1 year. The tiles of necrosis (high risk 1) and densely-distributed tumor cells (high risk 3) are associated with high relapse risks. The stroma tiles can indicate either a high or a low risk according to the abundance of immune infiltrates (high risk 2 and low risk 1). WSI scale bar, 4 mm. Tile scale bar, 128 μm. Abbreviations: WSIs, whole slide images; c index, concordance index; HR: hazard ratio; RFS, relapse-free survival.

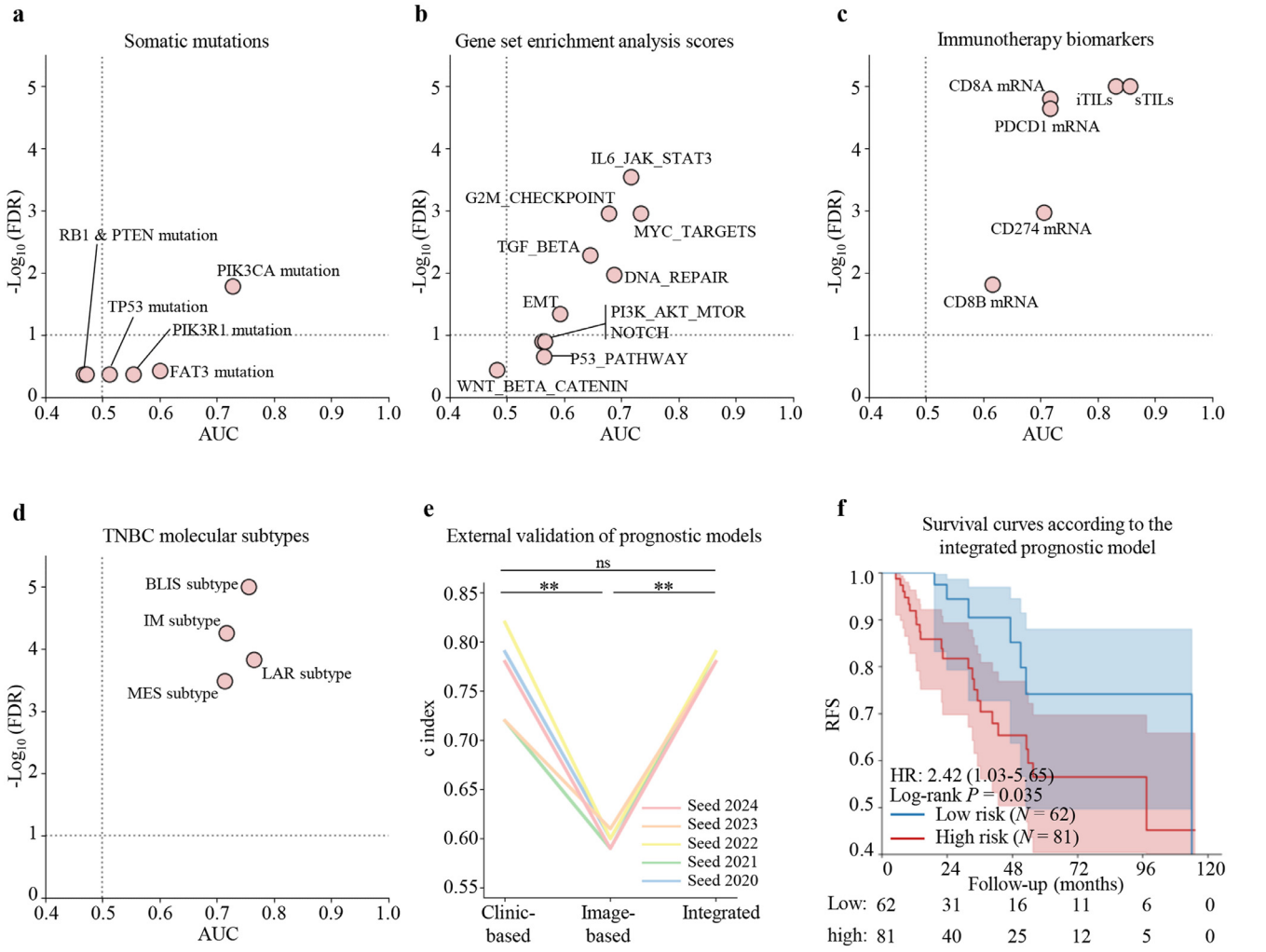

Fig. 6. External validation in the TCGATNBC cohort. (a)-(d) External validation of the models for predicting (a) somatic mutations, (b) gene set enrichment analysis scores, (c) immunotherapy biomarkers and (d) TNBC molecular subtypes. (e) External validation of the models for predicting patient relapse risk. *P* values are calculated using the Mann-Whitney U test with false discovery rate (FDR)-correction for multiple testing. **, FDR < 0.01; ns, not significant. (f) Kaplan-Meier curves of the high- and low-risk groups in the TCGATNBC cohort discriminated by the integrated prognostic model. Abbreviations: AUC, area under the curve; FDR, false discovery rate; EMT, epithelial mesenchymal transition; sTILs, stromal tumor-infiltrating lymphocytes; iTILs, intratumoral tumor-infiltrating lymphocytes; IC, immune cell; HR, hazard ratio; RFS, relapse-free survival.

the clinic-based models (FDR = 0.009 for both). The difference in c index between the integrated and image-based models was not significant (FDR = 0.094) (Fig. 5b; Table S12). Under all three modeling strategies, patients can be categorized according to the risk score into high- and low-risk groups with significantly different RFS. The hazard ratio between groups was highest in the integrated models (Fig. 5c). The output risk scores of both image-based and integrated models were still of prognostic value after the stratification for tumor stage, which indicated that our models can capture prognostic information from WSIs independent of clinical prognostic factors (Tables S13, S14).

We explored how pathological WSIs informed patient prognosis. Based on the image-based model, we analyzed the difference in the tissue type proportions and the average risk of each tissue type between the high- and low-risk groups. Compared with the low-risk group, cases of the high-risk group had more tiles of stroma but less of tumor, immune infiltrates and normal gland (Fig. 5d). The mean risk score of each tissue type was also higher in the high-risk group than in the low-risk group (Fig. 5e). In Fig. 5f, we presented two representative cases. The prediction heatmap spatially deconstructed the WSIs. High- and low-risk score tiles can be identified on a same slide.

3.5. External validation in the TCGATNBC cohort

We applied our framework to the TCGATNBC cohort for external validation without re-training the neural network models. The somatic *PIK3CA* mutation can be predicted with an AUC of 0.73 (95% CI: 0.59-0.87), even though the mutation rate of *PIK3CA* was lower in the TCGATNBC cohort (10.7%) than in the FUSCCTNBC cohort (17.6%). The models for other mutations did not achieve a comparable accuracy (Fig. 6a; Table S15). The high accuracy and generalizability of the models for *PIK3CA* mutation may be attributed to its confirmed and consistent biological effect across cohorts and relatively high mutation frequency (Fig. S6) [19,39,40]. The AUCs for predicting the GSEA scores were between 0.48-0.73 (Fig. 6b; Table S15). Most immunotherapy biomarkers can be inferred using models developed from the FUSCCTNBC cohort (Fig. 6c; Table S15). For the TNBC molecular subtypes, the AUC was 0.76 (95% CI: 0.68-0.83) for the BLIS subtype, 0.72 (95% CI: 0.62-0.81) for the IM subtype, 0.76 (95% CI: 0.65-0.88) for the LAR subtype and 0.71 (95% CI: 0.60-0.83) for the MES subtype (Fig. 6d; Table S15).

We also validated our DeepSurv-based prognostic models in the TCGATNBC cohort. The integrated models exhibited good and stable accu-

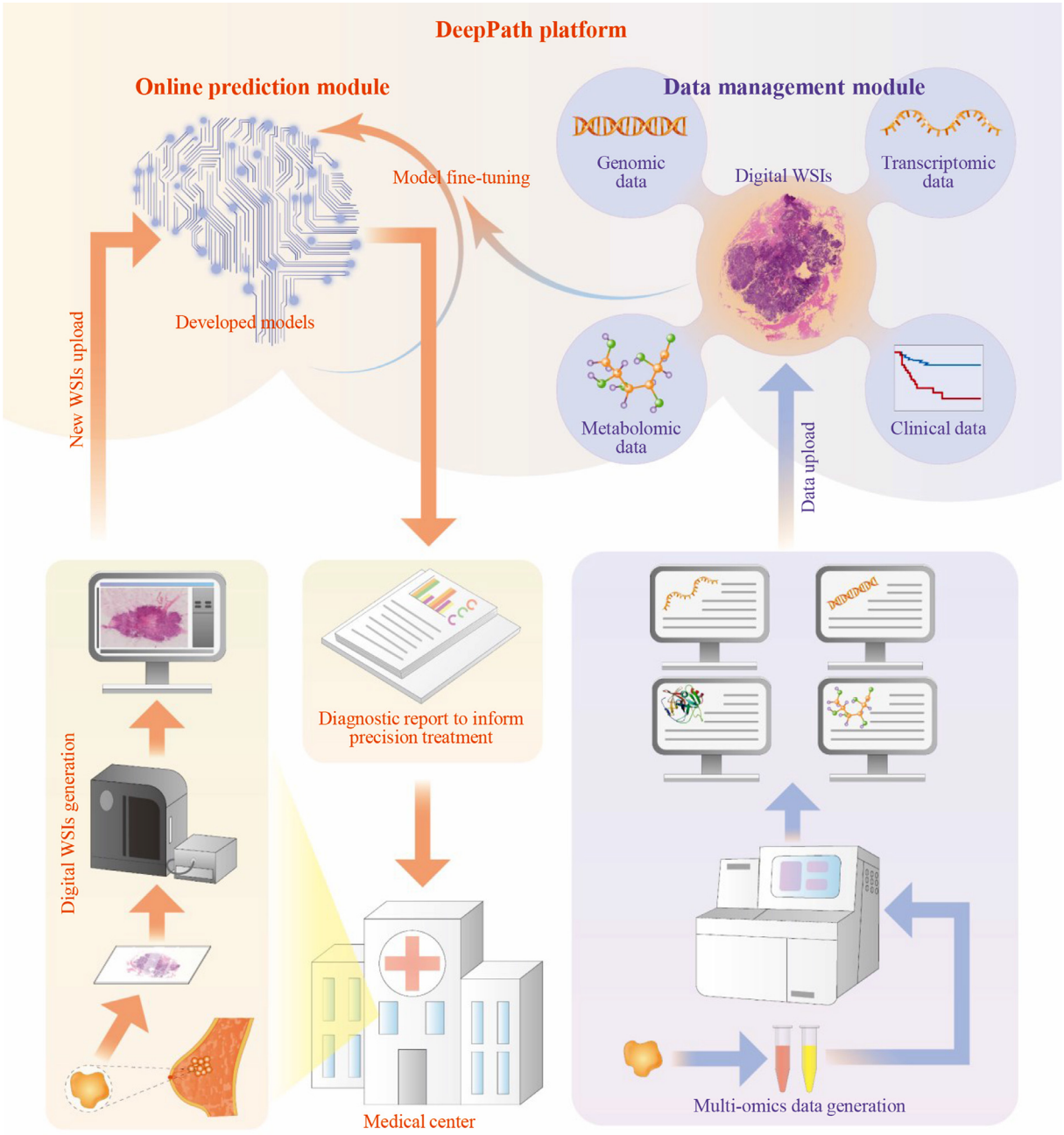

Fig. 7. Diagram of the DeepPath platform for data management and online prediction. Abbreviations: WSIs, whole slide images.

racy. An evident drop in accuracy was observed for the image-based models, whereas the accuracy for the clinic-based models was even higher in the external validation cohort than in the discovery cohort (Fig. 6e-f; Table S16).

3.6. Online platform for data management and real-time prediction

To enable further validation of our findings and promote their clinical translation, we developed an online platform called DeepPath for data management and real-time one-stop prediction (http://fuscc-pathology-ai-model.oss-cn-shanghai.aliyuncs.com/#/user/login, see "Supplementary Methods").

The DeepPath platform comprises two modules (Fig. 7). The first is a data management and sharing module. All WSIs of the FUSCCTNBC cohort have been uploaded and can be viewed under 40× magnification. The clinical and multi-omics information is presented in company with the WSI of the corresponding patient. The second is an online prediction module. We have modularized and deployed our complete prediction framework and several validated models on the platform. It supports online real-time prediction for newly uploaded WSIs. At present, the DeepPath platform requires a registered account to log in. We have provided a video to illustrate its functions (Video S1).

4. Discussion

In this study, based on our multi-omics TNBC cohort, we collected pathological WSIs, devised a deep learning-based framework, and developed neural network models for predicting molecular features, subtypes and patient relapse risk from WSIs. For the clinical translation of our findings, we integrated our framework and validated models into

an online platform tool, which can be implemented in clinical practice and realize real-time one-stop prediction.

The sampling of certain types of image tiles is an important step in either the model development or the model application to new cases. Most previous studies relied on manual annotations of tumor tissue by pathologists, which required pathological expertise and led to much additional manual work [15,17-19, 41]. A few studies used a tumor detector [16], but the limitation of this method was that it was not suitable for targets associated with the non-tumor tissue like sTILs and PD-L1 expression. In our study, we developed a CNN model to classify image tiles into five tissue types and incorporated this classifier in our framework. It enabled automatic selection of one or several types of tiles that were potentially associated with a given prediction target. It can also reduce manual operation and ensure the framework's reproducibility and scalability.

The recent studies by Kather et al. and Fu et al. tried to predict numerous molecular features using the TCGA dataset [19,20]. While this pan-cancer and multi-center dataset provided adequate training data, the CNN models may undesirably pay attention to the morphological difference across cancer types or tissue submitting sites and miss subtler cancer type-specific morphological patterns [42,43]. By contrast, our study focused on TNBC which is the most challenging breast cancer subtype to treat. Despite the relatively low patient number, several models developed based on our single-center dataset outperformed those trained in the pan-cancer and multi-center setting, including those for predicting clinically relevant information. For example, our model for predicting the *PIK3CA* mutation achieved higher accuracy than the previously published one (AUC: 0.73 in TCGA external validation vs. 0.63 in TCGA cross-validation[19]). Our integrated model also outperformed the previously reported one in prognostic evaluation (c index: 0.78 in TCGA external validation vs. 0.70 in TCGA cross-validation [20]). Further studies are needed to systematically compare these two modeling strategies.

Our multi-omics TNBC dataset allowed comprehensive predictions of a wider range of molecular features than those analyzed in the previous studies [8,19,44]. As far as we are concerned, our study is the first to investigate the predictions of germline variant spectrum and metabolic features from pathological WSIs. The prediction of germline *BRCA2* mutation achieved a cross-validation AUC of 0.79 (95% CI: 0.67-0.91). However, since only 6 patients harbored this mutation in the FUSCCTNBC cohort, our models should be further validated in an independent cohort.

Immunotherapy has shown great promise in the treatment of TNBC [34–36]. TILs and PD-L1 expression are extensively studied biomarkers to distinguish the patients who are likely to benefit from immunotherapy [34,45]. Our study provided an accurate and automatic evaluation approach for TILs. In addition, the CD274 (encoding PD-L1) mRNA expression and PD-L1 protein expression evaluated through immunohistochemistry can be inferred from the H&E stained WSIs. Other immunotherapy biomarkers including the CD8 and PDCD1 (encoding PD-1) expression also seemed predictable. Therefore, it may be possible in future studies to establish a comprehensive tool to predict the immunotherapy response from WSIs.

Molecular subtyping of TNBC is important for understanding the molecular essence of this heterogeneous disease and for guiding individualized treatment. We previously devised a surrogate assay based on immunohistochemistry for the TNBC molecular subtyping [46]. This method incorporated immunohistochemical staining and manual evaluation of four markers, which still resulted in additional cost, turnaround time, much manpower and poor reproducibility. By contrast, making H&E stained slides is already included in clinical routine. The implementation of our deep learning-based classifier only takes extra work of slide scanning and our modularized framework can ensure the reproducibility. In addition, we, for the first time, revealed the distinctive morphological features of the four TNBC subtypes, which provided new insight into the heterogeneity of TNBC.

In the relapse risk evaluation, our data verified that the tumor morphological patterns held additional prognostic information beyond clinicopathological characteristics. Further analysis indicated that both the proportions of different tissue types and the morphological features within certain tissue types were associated with the relapse risk. Interestingly, the image-based models showed comparable accuracy to the integrated models in the FUSCCTNBC cohort, but their accuracy obviously dropped in external validation. Although color normalization was performed, the inherent dataset differences in image properties including resolution (micrometer per pixel) and image pyramid levels and in tiles' tissue type distribution between the samples of the FUSCCTNBC and TCGATNBC cohorts were not fully resolved and may be the possible cause for the model performance drop (Fig. S7) [47,48]. Future work may tackle this problem using the domain adaption approach [18,49]. By contrast, despite an unsatisfactory accuracy in the FUSCCTNBC cohort, the clinic-based models had better performance in the external validation cohort. These results implied that the image-based models had relatively strong fitting ability but poor generalizability compared with the clinic-based models. Future studies with larger sample size are required to compare these three modeling strategies and perhaps improve the accuracy and robustness of the prognostic models.

Deep learning has achieved remarkable success in medical image analysis [50,51]. In the pathology area, several studies have explored the predictions of cancer molecular features and patient prognosis based on WSIs, but few of the research findings were translated into a tool and implemented in clinical trials for prospective validation. The major challenges included the requirement of deep learning infrastructure and inexperience of clinicians with the deep learning technology [52]. To overcome these obstacles, we developed an online platform called DeepPath, on which we modularized and deployed our prediction framework. The platform can realize real-time one-stop prediction for certain clinically relevant information of TNBC. Users only need to upload the tumors' WSIs and select a prediction target. The highly automatic and modularized framework will be executed on a cloud server and output the prediction results in a few minutes. Although some improvement is needed including the data transmission speed and the number of WSIs or tasks that can be parallel processed, the platform can enable further validation and optimization of our algorithm on a larger scale and can be implemented for the molecular and prognostic stratification of patients with a lower cost and shorter turnaround time in larger multi-center clinical trials and perhaps in the future routine practice.

Our study has some limitations. First, there is still room for optimization of the prediction models for some relevant targets. We plan to make full use of our DeepPath platform to collect more WSIs and fine-tuning the existing models. Another strategy may be the use of spatially annotated data from spatial profiling, which can overcome the effects of intratumoral heterogeneity and provide more accurate label data [53,54]. Second, before the clinical application, our algorithm should be prospectively validated in clinical trials to build the evidence that it can be used as a predictive tool to indicate treatment response and patient prognosis.

5. Conclusion

We developed a deep learning-based framework and a clinically applicable online platform to predict molecular features and patient prognosis of TNBC from pathological WSIs. Our work may contribute to more efficient patient stratifications for personalized management of TNBC and shed light on the application of artificial intelligence guided precision treatment in clinical trials and future routine practice.

Ethical approval

This study was approved by the FUSCC Ethics Committee (ethics approval number: 050432-4-1911D) and was conducted in accordance

with the principles of the Declaration of Helsinki. All patients provided written informed consent.

Data sharing statement

For the FUSCCTNBC cohort, WSI data have been uploaded to the DeepPath platform and will be available for download when the manuscript is published. Sequence data have been deposited in the NCBI Gene Expression Omnibus (GEO: GSE118527) and Sequence Read Archive (SRA: SRP157974). For the TCGATNBC cohort, all data are publicly available at https://portal.gdc.cancer.gov/ and https://cbioportal.org/. All other data supporting the findings of this study are available on request from the corresponding author ZMS.

Declaration of competing interest

The authors declare that they have no conflicts of interest in this work.

Acknowledgments

This work was supported by grants from the National Key Research and Development Program of China (2021YFF1201300 and 2021YFF1201305) and the National Natural Science Foundation of China (82103039, 81572583, 81922048, 91959207, U1809205, 92159301, 61771249 and 62171230)

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.fmre.2022.06.008.

References

[1] H. Sung, J. Ferlay, R.L. Siegel, et al., GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J. Clin. 71 (2021) 209–249.
[2] K.R. Bauer, M. Brown, R.D. Cress, et al., Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: A population-based study from the California cancer Registry, Cancer 109 (2007) 1721–1728.
[3] C. Criscitiello, H.A. Azim Jr., P.C. Schouten, et al., Understanding the biology of triple-negative breast cancer, Annals of oncology: Official journal of the European Society for Medical Oncology 23 (Suppl 6,) (2012) vi13-18.
[4] L.A. Carey, C.M. Perou, C.A. Livasy, et al., Race, breast cancer subtypes, and survival in the Carolina breast cancer study, JAMA 295 (2006) 2492–2502.
[5] B.D. Lehmann, J.A. Bauer, X. Chen, et al., Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies, J. Clin. Invest. 121 (2011) 2750–2767.
[6] M.D. Burstein, A. Tsimelzon, G.M. Poage, et al., Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer, Clin. Cancer Res. 21 (2015) 1688–1698.
[7] Y.R. Liu, Y.Z. Jiang, X.E. Xu, et al., Comprehensive transcriptome analysis identifies novel molecular subtypes and subtype-specific RNAs of triple-negative breast cancer, Breast Cancer Res. 18 (2016) 33.
[8] Y.Z. Jiang, D. Ma, C. Suo, et al., Genomic and transcriptomic landscape of triple-negative breast cancers: Subtypes and treatment strategies, Cancer Cell 35 (2019) 428–440 e425.
[9] D. Ma, S.Y. Chen, J.X. Ren, et al., Molecular features and functional implications of germline variants in triple-negative breast cancer, J. Natl. Cancer Inst. (2020). https://doi.org/10.1093/jnci/djaa175.
[10] Y. Gong, P. Ji, Y.S. Yang, et al., Metabolic-pathway-based subtyping of triple-negative breast cancer reveals potential therapeutic targets, Cell Metab. 33 (2021) 51–64 e59.
[11] Y. Xiao, D. Ma, S. Zhao, et al., Multi-omics profiling reveals distinct microenvironment characterization and suggests immune escape mechanisms of triple-negative breast cancer, Clin. Cancer Res. 25 (2019) 5002–5014.
[12] Y.Z. Jiang, Y. Liu, Y. Xiao, et al., Molecular subtyping and genomic profiling expand precision medicine in refractory metastatic triple-negative breast cancer: The FUTURE trial, Cell Res. 31 (2021) 178–186.
[13] J. Shia, N. Schultz, D. Kuk, et al., Morphological characterization of colorectal cancers in The Cancer Genome Atlas reveals distinct morphology-molecular associations: Clinical and biological implications, Mod. Pathol. 30 (2017) 599–609.
[14] E.A. Rakha, M. Alsaleem, K.A. ElSharawy, et al., Visual histological assessment of morphological features reflects the underlying molecular profile in invasive breast cancer: A morphomolecular study, Histopathology 77 (2020) 631–645.
[15] A. Echle, H.I. Grabsch, P. Quirke, et al., Clinical-grade detection of microsatellite instability in colorectal tumors by deep learning, Gastroenterology 159 (2020) 1406–1416 e1411.
[16] J.N. Kather, A.T. Pearson, N. Halama, et al., Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer, Nat. Med. 25 (2019) 1054–1056.
[17] A.C. Woerl, M. Eckstein, J. Geiger, et al., Deep learning predicts molecular subtype of muscle-invasive bladder cancer from conventional histopathological slides, Eur. Urol. 78 (2020) 256–264.
[18] K. Sirinukunwattana, E. Domingo, S.D. Richman, et al., Image-based consensus molecular subtype (imCMS) classification of colorectal cancer using deep learning, Gut 70 (2021) 544–554.
[19] J.N. Kather, L.R. Heij, H.I. Grabsch, et al., Pan-cancer image-based detection of clinically actionable genetic alterations, Nat. Cancer 1 (2020) 789–799.
[20] Y. Fu, A.W. Jung, R.V. Torne, et al., Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis, Nat. Cancer 1 (2020) 800–810.
[21] N. Naik, A. Madani, A. Esteva, et al., Deep learning-enabled breast cancer hormonal receptor status determination from base-level H&E stains, Nat. Commun. 11 (2020) 5727.
[22] M.Y. Lu, T.Y. Chen, D.F.K. Williamson, et al., AI-based pathology predicts origins for cancers of unknown primary, Nature 594 (2021) 106–110.
[23] P. Mobadersany, S. Yousefi, M. Amgad, et al., Predicting cancer outcomes from histology and genomics using convolutional networks, Proc. Natl. Acad. Sci. U. S. A. 115 (2018) E2970–E2979.
[24] G. Ramakrishnan, D. Anand, A. Sethi, Fast GPU-enabled color normalization for digital pathology, 2019 International Conference on Systems, Signals and Image Processing (IWSSIP), 2019.
[25] A. Paszke, S. Gross, F. Massa et al., PyTorch: An Imperative Style, High-Performance Deep Learning Library. (2019). https://dl.acm.org/doi/10.5555/3454287.3455008.
[26] Y. Sasaki, The truth of the F-measure. (2007). https://www.researchgate.net/publication/268185911.
[27] S. Xu, W. Xu, Fast implementation of DeLong's algorithm for comparing the areas under correlated receiver operating characteristic curves, IEEE Signal Process Lett. 21 (2014) 1389–1393.
[28] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing, Journal of the Royal Statistical Society: Series B (Methodological) 57 (1995) 289–300.
[29] S.B. Kim, R. Dent, S.A. Im, et al., Ipatasertib plus paclitaxel versus placebo plus paclitaxel as first-line therapy for metastatic triple-negative breast cancer (LOTUS): A multicentre, randomised, double-blind, placebo-controlled, phase 2 trial, Lancet Oncol. 18 (2017) 1360–1372.
[30] P. Schmid, J. Abraham, S. Chan, et al., Capivasertib plus paclitaxel versus placebo plus paclitaxel as first-line therapy for metastatic triple-negative breast cancer: The PAKT trial, J. Clin. Oncol. 38 (2020) 423–433.
[31] S. Hanzelmann, R. Castelo, J. Guinney, GSVA: Gene set variation analysis for microarray and RNA-seq data, BMC Bioinf. 14 (2013) 7.
[32] A. Liberzon, C. Birger, H. Thorvaldsdottir, et al., The molecular signatures database (MSigDB) hallmark gene set collection, Cell Syst. 1 (2015) 417–425.
[33] A.A. Hakimi, E. Reznik, C.H. Lee, et al., An integrated metabolic atlas of clear cell renal cell carcinoma, Cancer Cell 29 (2016) 104–116.
[34] P. Schmid, S. Adams, H.S. Rugo, et al., Atezolizumab and nab-paclitaxel in advanced triple-negative breast cancer, N. Engl. J. Med. 379 (2018) 2108–2121.
[35] P. Schmid, J. Cortes, L. Pusztai, et al., Pembrolizumab for early triple-negative breast cancer, N. Engl. J. Med. 382 (2020) 810–821.
[36] E.A. Mittendorf, H. Zhang, C.H. Barrios, et al., Neoadjuvant atezolizumab in combination with sequential nab-paclitaxel and anthracycline-based chemotherapy versus placebo and chemotherapy in patients with early-stage triple-negative breast cancer (IMpassion031): A randomised, double-blind, phase 3 trial, Lancet 396 (2020) 1090–1100.
[37] J.L. Katzman, U. Shaham, A. Cloninger, et al., DeepSurv: Personalized treatment recommender system using a Cox proportional hazards deep neural network, BMC Med. Res. Methodol. 18 (2018) 24.
[38] H. Yoon, M.P. Mehta, K. Perumal, et al., Atypical meningioma: Randomized trials are required to resolve contradictory retrospective results regarding the role of adjuvant radiotherapy, J. Cancer Res. Ther. 11 (2015) 59–66.
[39] D. Chakravarty, J. Gao, S.M. Phillips, et al., OncoKB: A precision oncology knowledge base, JCO Precis Oncol 2017 (2017). https://doi.org/10.1200/PO.17.00011.
[40] N. Coudray, P.S. Ocampo, T. Sakellaropoulos, et al., Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning, Nat. Cancer 24 (2018) 1559–1567.
[41] S. Foersch, M. Eckstein, D.C. Wagner, et al., Deep learning for diagnosis and survival prediction in soft tissue sarcoma, Ann. Oncol. (2021). https://doi.org/10.1016/j.annonc.2021.06.007.
[42] N. Coudray, A. Tsirigos, Deep learning links histology, molecular signatures and prognosis in cancer, Nature Cancer 1 (2020).
[43] F.M. Howard, J. Dolezal, S. Kochanny, et al., The impact of site-specific digital histology signatures on deep learning model accuracy and bias, Nat. Commun. 12 (2021) 4423.
[44] H. Qu, M. Zhou, Z. Yan, et al., Genetic mutation and biological pathway prediction based on whole slide images in breast carcinoma using deep learning, NPJ Precis. Oncol. 5 (2021) 87.
[45] R. Salgado, S. Loi, Tumour infiltrating lymphocytes in breast cancer: Increasing clinical relevance, Lancet Oncol. 19 (2018) 3–5.

[46] S. Zhao, D. Ma, Y. Xiao, et al., Molecular subtyping of triple-negative breast cancers by immunohistochemistry: Molecular basis and clinical relevance, Oncologist 25 (2020) e1481–e1491.
[47] J. Quiñonero-Candela, *Dataset shift in machine learning*, neural information processing series, MIT Press, Cambridge, Mass, 2009 pp. xv, 229 p.
[48] K. Stacke, G. Eilertsen, J. Unger, et al., Measuring domain shift for deep learning in histopathology, IEEE J. Biomed Health Inform. 25 (2021) 325–336.
[49] Y. Ganin, E. Ustinova, H. Ajakan, et al., Domain-adversarial training of neural networks, Journal of Machine Learning Research 17 (2016) 2096-2030.
[50] A. Madabhushi, M.D. Feldman, P. Leo, Deep-learning approaches for Gleason grading of prostate biopsies, Lancet Oncol. 21 (2020) 187–189.
[51] M. Bilal, S.E.A. Raza, A. Azam, 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). https://doi.org/10.1016/s2589-7500(21)00180-1.
[52] H. Yoshida, T. Kiyuna, Requirements for implementation of artificial intelligence in the practice of gastrointestinal pathology, World J. Gastroenterol. 27 (2021) 2818–2833.
[53] K. AbdulJabbar, S.E.A. Raza, R. Rosenthal, et al., Geospatial immune variability illuminates differential evolution of lung adenocarcinoma, Nat. Med. 26 (2020) 1054–1062.
[54] P.L. Stahl, F. Salmen, S. Vickovic, et al., Visualization and analysis of gene expression in tissue sections by spatial transcriptomics, Science 353 (2016) 78–82.

Shen Zhao (BRID: 07893.00.52913) is a M.D. graduate of Fudan University, majoring in oncology.

Zhi-Ming Shao is a Ph.D. recipient of the National Science Fund for Distinguished Young Scholars, and a distinguished professor of Chang Jiang Scholars Program by the Ministry of Education of China. He is also a director of Key Laboratory of Breast Cancer in Shanghai, and a professor of oncology at Shanghai Medical College, Fudan University.