Computational pathology infer clinically relevant protein levels and drug efficacy in breast cancer by weakly supervised contrastive learning

Xiaodong Xie[1] and Hui Liu[1*]

[1*]School of Computer Science and Technology, Nanjing Tech University, Nanjing, 211816, Jiangsu, China.

*Corresponding author(s). E-mail(s): hliu@njtech.edu.cn;

Abstract

Visual inspection of histopathology slides via optical microscope is the routine medical examination for clinical diagnosis of tumors. In recent years, deep learning has demonstrated that computational pathology has potential to infer tumor-related genetic alterations and transcriptomic patterns. In this paper, we propose a weakly supervised contrastive learning framework to infer the protein level of tumor biomarkers from whole-slide images (WSIs) of breast cancer. We conduct contrastive learning-based pre-training on large-scale unlabeled breast cancer WSIs to extract histopathological features, which are then adapted to downstream tasks, including tumor diagnosis, prediction of protein level and drug response. Our extensive experiments show that our method outperforms other state-of-the-art models in tumor diagnostic task, and achieves high performance in predicting clinically relevant protein levels. To show the model interpretability, we spatially visualize the WSIs colored by attention scores of tiles, and find that the regions with high scores are highly consistent with the tumor and necrotic regions annotated by an experienced pathologist. Moreover, spatial transcriptomic data further verified that the heatmap generated by attention scores agree greatly with the spatial expression map of tumor biomarker genes. Our method achieves 0.79 AUC value in predicting the response of breast cancer patients to the drug trastuzumab. These findings show the remarkable potential of deep learning-based morphological feature is very indicative of clinically relevant protein levels, drug response and clinical outcomes.

Keywords: Weakly-supervised learning, Computational pathology, Contrastive learning, Drug response, Protein level, Morphological feature

1 Introduction

Digital pathology makes use of automated microscopy or optical magnification systems to scan and digitize traditional glass pathology slides. Computer algorithms are then employed to perform high-precision multi-field seamless stitching and processing of the images to produce whole-slide images (WSIs). In recent years, computational pathology has advanced rapidly to facilitate the automatic diagnosis of tumors. This not only alleviates the workload of pathologists, but also helps to eliminate the subjective bias from observers.

Deep learning has made significant strides in the field of computer vision, and has also spurred advances in computational pathology. Such methods have demonstrated impressive progress in a variety of challenging clinical tasks [1–4]. For tumor diagnosis and semantic segmentation, Wang et al. [5] developed an enhanced HD-Staging algorithm based on mask-RCNN that can perform nuclear segmentation and cell classification on lung adenocarcinoma images. This algorithm can also be applied to head and neck cancer, breast cancer, and lung squamous cell cancer. Greenwald et al.[6] developed TissueNet to segment cells and identify the boundaries of individual cells in whole-slide images. Eklund et al.[7] created an integrated model based on Inception V3 to assist pathologists in finding, identifying, and grading prostate tumors. Wang et al[8] introduced a novel deep learning model to improve breast cancer histological grading. Additionally, several studies have presented computational pathology methods for inferring genetic alterations and gene expression profiles. For example, Kather and colleagues [9] introduced a deep learning approach for analyzing genetic alterations across various types of cancer. They successfully predicted the presence of CDC27 mutations in colorectal cancer using pathological images. Wang et al.[10] proposed an approach called expression-morphology (EMO) to predict mRNA expression in 17,695 genes from whole-slide images and validated the spatial variability of intratumor expression through spatial transcriptomics profiling. Several recent studies have demonstrated the potential of using weakly supervised algorithms in computational pathology. For instance, our team has developed Histcode [11], which integrated contrastive learning-based pretraining and weakly supervised learning to predict differential gene expression of cancer driver genes. Mahmood et al.[12] developed the CLAM model, which uses an attention mechanism for subtype classification of renal cell carcinoma and non-small cell lung cancer. Schumach et al.[13] developed a model called HE2RNA to predict RNA-seq from pathological images and identify tumors with microsatellite instability in clinical diagnoses. Shamai et al.[14] built convolutional neural network (CNN)-based model to predict PD-L1 expression from H&E images.

Eliceiri et al. [15] developed the DS-MIL algorithm to classify WSIs as tumor or normal. Mahmood et al. [16] developed SISH, an interpretable search pipeline that achieves high speed in searching histology images after training with only slide-level labels.

In recent times, the field of proteomics has demonstrated exceptional promise in the realm of precision medicine, achieved significant advances in the treatment of various forms of cancer [17–19]. For example, Gao et al. [20] utilized proteomics to identify two prognostic biomarkers, PYCR2 and ADH1A, for hepatocellular carcinoma. The reverse-phase protein array (RPPA) [21], a high-throughput and highly sensitive protein microarray that employs antibodies to detect and quantify proteins and their modifications in biological samples, has shown tremendous potential in the discovery of cancer biomarkers, analysis of functional phenotypes, and elucidation of drug mechanisms. However, the prohibitive cost and lengthy turnaround time associated with RPPA hinder its widespread adoption in routine clinical examinations. There is a pressing need for a high-throughput method to quantify protein levels of tumor biomarkers in clinical practice.

It has been assumed that protein profile changes in tumor cells cause functional changes, which can influence tumor cell morphology. So, routine histopathology tissue slides, which are ubiquitously available, can reflect such morphological changes, thereby the clinically relevant proteins could be inferred directly from digitized whole-slide images. In this paper, we introduced a weakly supervised contrastive learning methods to establish the quantitative associations between clinically relevant proteins and histopathological features in breast cancer. Specifically, Whole-slide images were split into tiles and then MoCo v2 was employed to extract features at the tile level. Attentive pooling was used to aggregate tile-level features to generate slide latent representations, which were then applied to various downstream tasks, including tumor diagnosis, quantification of protein levels, prognostic risk assessment, and prediction of response to the drug trastuzumab. Our extensive experiments show that the proposed method outperforms other state-of-the-art models in tumor diagnostic task, and achieves high performance in quantifying the protein levels of tumor biomarkers. To show the model interpretability, we spatially visualize the WSIs colored by attention scores of tiles, and find that the regions with high scores are highly consistent with the tumor and necrotic regions annotated by an experienced pathologist. Moreover, spatial transcriptomic data further verified that the heatmap generated by attention scores agree greatly with the spatial expression map of tumor biomarker genes. Our method achieves 0.79 AUC value in predicting the response of breast cancer patients to the drug trastuzumab. These findings showed our method could elucidate and quantify genotypephenotype links in breast cancer.

2 Results

2.1 Weakly-supervised learning generate accurate tumor diagnosis

We first test our method for tumor diagnostic task on the TCGA-BRCA cohort. The pathology images (n=1,979) of breast cancer are randomly split into training, validation and independent test set by 60%, 20% and 20%. It is worth noting that we used only the slide-level labels to fine-tune the pre-trained model. Our method achieves 0.995 accuracy and 0.996 AUC values. As shown in Figure 2(a-b), we present the ROC curve and confusion matrix on the independent test set. This shows that our weakly supervised learning model can accurately discriminate whether WSI contains tumor cells or not. During the fine-tuned stage for tumor diagnosis, the tile-level embeddings were aggregated to build slide representation (computational histopathological features). As the network was trained to classify tumor and normal tissues, the computational histopathological features show two clusters corresponding to each tissue class by PCA dimension reduction (Figure 2(c)).

The slide-level classification of tumor/normal labels was made by attentive pooling of tile-level features, thereby the derived attention scores are highly indicative of localization of tumor regions. To validate successful spatial localization of tumor tissue by our model, we scale the learned tile-level attention scores to generate heatmaps by spatial deconvolution of tiles to original slide. For objective evaluation, an experienced pathologist was asked to annotate the tumor regions of interest (ROIs), and we visually check whether the high diagnostic area is coincident the annotated regions. As shown in Figure 2(d), the attention heatmaps show remarkable agreement between pathologist annotated ROIs and computational predictions of tumor regions. Visual inspection of tiles would reflect more human-readable pathological feature. So, we showed a few tiles with high or low attention scores, which are verified by pathologist that the high-scored tiles are mostly normal tissue, while high-scored tiles are tumor tissue.

2.2 Histopathological feature effectively predicts protein levels of tumor biomarkers

The protein expression profile of clinically relevant biomarkers produced by reverse-phase protein arrays (RPPAs) allow us to test whether the histopathological feature is predictive of protein levels of tumor biomarkers. As these proteins are enriched in diagnostic biomarkers and therapeutic targets, establishment of the associations between computational pathology and clinically relevant proteins would greatly facilitate clinical diagnosis and translational applications. On the TCGA-BRCA cohort, the protein expression profile contains 223 biomarker proteins of 860 breast cancer patients. We transfer the computational pathological features obtained by contrastive learning to quantitatively predict the levels of biomarker proteins.

We need to make it clear that we used multi-task learning to simultaneously predict the expression level of 223 biomarker proteins. This is quite different from prior methods that train a predictive model per gene or genetic mutant [9, 12, 22]. For evaluation of this regression task, we used Pearson correlation coefficient (r) between the RPPA-measured value and predicted value as the evaluation metric. Similar to previous studies [13], we considered a prediction to be significantly different from the random baseline value if the p-value associated with its coefficient was below 0.05, after applying BenjaminiHochberg (BH) correction to account for multiple-hypothesis testing. On the TCGA-BRCA cohort, almost all proteins yield significant prediction results. As shown in Figure 3(a-b), 215 out of 223 protein (96%) were predicted with a statistically significant correlation under BH correction. The average correlation coefficient reached 0.287, of which 155 proteins obtained correlation coefficient greater than 0.2, and 27 proteins greater than 0.5.

Among the tumor biomarkers, we pay close attention to the proteins that are clinically relevant to the breast cancer, such as estrogen receptor (ER), progesterone receptor (PR) and HER2. Expectedly, our model showed significant predictive power on these biomarkers, as shown in Figure 3 (d). For the estrogen receptor ER-α, positive in 70% of breast cancer and key biomarker for breast cancer diagnosis [23], our method got r=0.632 correlation coefficient. Progesterone receptor (r=0.439) is induced by ER-α and plays an important role in regulating ER-α protein, thereby it functions as an important biomarker for breast cancer treatment and prognosis [24]. Highly expressed PR in luminal A type breast cancer indicates good prognosis [25]. HER2 (r=0.402) overexpression is present in 20-30% of breast cancer and is associated with higher malignancy [26]. The tumor suppressor factor P53 (r=0.432) has mutated in 80% of triple-negative breast cancers [27]. Caveolin-1 (r=0.568) high expression is a biomarker of more malignant tumor and poor prognosis [28]. PD-L1 (r=0.354) is an immune-related biomarker expressed on the surface of several cell types. Its high expression means potential response to immunotherapy, specifically PD-1/PD-L1 immune checkpoint inhibitors [29]. E-cadherin (r=0.541) and N-cadherin (r=0.303) are epithelialmesenchymal transition (EMT)-related biomarkers that confer higher invasion, metastasis and drug resistance [30]. Moreover, we choose one third well-predicted proteins (n=75) to run functional enrichment analysis, and the result was shown in Figure 3(c), from which we found that the well-predicted protein are enriched in the oncogenic, drug-resistance and tumor-driven signaling pathways.

2.3 Contrastive learning significantly improves predictive performance

To verify that the pre-trained encoder by contrastive learning extracted histologic features generalized to various downstream tasks, we compared two popular contrastive learning models (SimCLR and MoCo v2) to the baseline ResNet50 network. The ResNet50 was trained on ImageNet and transferred to the prediction task of protein levels. The contrastive learning used the

ImageNet-trained ResNet50 as backbone network to start the pretraining process. When the contrastive learning-based pretraining finished, the ResNet50 encoder was frozen in the transfer learning for protein levels prediction task. For comprehensive comparison, we also considered fine-tuning the pre-trained ResNet50 encoder together with the appended neural network layers to predict the protein levels. As shown in Figure 4(a), all the model pretrained by contrastive learning outperformed the baseline ResNet50 model. Also, we found that fine-tuned models achieved better predictive power than their respective frozen ones. The experimental result demonstrated the contrastive learning effectively capture expressive and informative features that can be transferred to improve performance of downstream tasks.

We were also interested in the effect of different resolution of the scanned pathology images on the prediction performance. We tested 40x, 20x, 10x and 5x magnification, and the experiment shows that higher magnification achieved better performance in the prediction of protein levels. Figure 4(b) showed the boxplot of correlation coefficients at different resolutions, the 40x, 20x, 10x and 5x magnification yield 0.292, 0.259, 0.246 and 0.209, respectively. For intuitive proof, we visually checked a few exemplar tiles from the different resolution slides, as shown in Figure 4(c-f) and found that higher resolution slides convey more detailed information about the morphological structure and histopathological feature. We speculate that higher magnification enable the model to better capture the molecular phenotypic feature of pathological tissues.

2.4 Computational histopathological feature indicates patient prognosis

After aggregation of tile-level features by attentive pooling strategy, we obtained the latent representation of a whole slide image. Following the feature aggregation, we further test whether the computational histopathological feature is indicative of the prognosis of breast cancer patients. For this purpose, a Cox proportional hazards regression model was developed to estimate the survival risk (see Method Section). According to the prognostic risk score, the breast cancer patients (n=860) of TCGA-BRCA cohort were stratified to high-risk group and the low-risk group, and the survival analysis curve was shown in Figure 5(a). The result showed that the 5-year survival time of the high-risk group was significantly lower than that of the low-risk group (HR=3.63, p-value$<$0.05), and the concordance index (C-index) of the survival model was 0.64. Moreover, we compared the histopathological difference of the high-risk and low-risk pathology images. As shown in Figure 5(b-d), it can be found that the high-risk slide contains more tumor regions (high attention score regions in the heatmap) and conveys more invasive histologic feature, compared to the low-risk slide. These results show that our method automatically capture histologic feature that is indicative of patient prognosis, which is different from other methods based on manually annotated ROIs.

2.5 Computational pathology predicts trastuzumab treatment outcome

We went further to explore the capacity of computational pathology in predicting drug response of breast cancer. The Yale trastuzumab response cohort [31] allows us to run this exploratory experiment. This cohort included the patients with a pre-treatment breast core biopsy with HER2 positive invasive breast carcinoma who then received neoadjuvant targeted therapy with trastuzumab±pertuzumab prior to definitive surgery. A case was designated as responder if the pathological examination of surgical resection specimens did not report residual invasive, lympho-vascular invasion or metastatic carcinoma, and otherwise non-responders. After removal of the H&E images without enough detectable foreground tissue by redefined area threshold, we obtained 75 cases (34 responders and 41 non-responders) from the Yale trastuzumab response cohort.

As the HER2 status is used as the clinical biomarker for trastuzumab response of breast cancer, we used only the RPPA-level of HER2, rather than the clinical treatment outcomes, to fine-tune the model pretrained by contrastive learning. Thereafter, the fine-tuned model was directly used to predict the trastuzumab treatment response. This would empower our method to approach clinical practice and output actionable suggestion. As shown in Figure 6(a-c), our method achieved 0.79 AUC value for the prediction of responder and non-responders, which is much better than 0.68 obtained by the previous study [31]. Of note, our model was trained by unannotated slides, but achieved comparable performance to previous model trained using pathologist-annotated slides with invasive tumor cells area. Again, we visualized a heatmap generated by the learned attention scores, and compared the highly scored regions to the tumor regions annotated by the pathologists of Yale University. Figure 6(d) demonstrated high consistence between them, and the tiles randomly picked out from high- and low -scored regions agree remarkably with the regions marked as normal and tumor. These results support the feasibility of image-based biomarkers to predict trastuzumab treatment and the ability of deep learning model to recognize morphological changes related to treatment results.

2.6 Spatial transcriptomics validates spatial localization of tumor biomarkers

Spatial transcriptomics measures the RNA abundance at a high spatial resolution enable us to evaluate our model capacity in spatial localization of tumor biomarkers. For this purpose, we used the RPPA-measured protein levels of two typical biomarkers, HER2 and P53, to fine tune our model. Next, we generated the heatmaps using the learned tile-level attention scores. Correspondingly, we visualize the spatial expression of her2 and p53 genes. As shown in Figure 7, these heatmaps shows that the highly-scored regions were notably consistent with the high expression region of the spatial transcription

for both HER2 and P53. This finding showed that our method effectively performed spatial deconvolution of tumor biomarkers based on conventional H&E staining sections.

3 Materials and methods

3.1 Whole slide images

All digital slides of fresh frozen tissue stained with hematoxylin and eosin (H&E) were obtained from TCGA via the Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov/). From the TCGA-BRCA project, we collected 1,979 WSIs of 1,094 breast cancer patients, comprising 1,580 tumor slides and 399 normal slides. The pathology diagnosis provided by TCGA database were used as ground truth labels for classification task.

An independent cohort came from Yale trastuzumab response cohort was used to evaluate model ability in predicting drug response. The Yale cohort contained 75 FFPE WSIs of 75 cases of breast cancer (34 responder and 41 non-responders). A case was designated as responder if the pathological examination of surgical resection specimens did not report residual invasive, lympho-vascular invasion or metastatic carcinoma, and otherwise non-responders. Note that only the slides with a magnification greater than 20* were included in this study.

3.2 RPPA dataset

All RPPA data were downloaded from The Cancer Proteome Atlas (TCPA)[32]. The database collects the RPPA-assayed protein levels of samples mainly came from TCGA (The Cancer Genome Atlas) tumor tissue sample sets. The RPPA data contains more than 200 proteins covering most cancer signaling pathways such as PI3K, MAPK, mTOR, TGF-b and WNT pathways. In total, there are 893 RPPA samples came from TCGA-BRCA cohort. After removal of sample without matched WSIs, we obtained 860 matched samples that are used to train protein level prediction model.

3.3 Spatial transcriptomic data

We acquired the spatial transcriptomic data of a breast cancer specimen using the Visium Spatial Gene Expression protocol from https://www.10xgenomics.com/. To explore the spatial map of gene expression, we utilized the 10x Loupe Browser, a desktop application that offers interactive visualization capabilities for various 10x Genomics solutions. This enabled us to visualize the spatial landscape of specific biomarker genes experssion upon the pathology image.

3.4 Preprocessing of whole slide images

Due to the ultra-high resolution of pathological images, which can reach up to gigapixels, they are not immediately suitable for input into a deep learning model. As a result, we divided each whole slide image into small squares known as tiles or patches. The Python package OpenSlide was utilized to read a slide into memory, and the Otsu algorithm was then employed to differentiate between tissue and background regions. Following segmentation, the tissue area was divided into 256x256 pixel tiles.

3.5 Contrastive learning for feature extraction

Given a large scale of unlabeled tiles, we leveraged the contrastive learning-based pretraining to learn an encoder to produce latent representations for the tiles. In this study, we used MoCo v2 for the preraining. We also tested another contrastive learning method, SimCLR, and found that MoCo v2 achieved better performance in downstream tasks.

Formally, contrastive learning learns to extract feature by minimizing the distance between the representations of positive pairs (e.g. image and its augmentation) and maximizing the distance between representations of negative pairs (e.g. different images). Suppose an tile x has its augmentation x. We use a CNN network as the backbone encoder f to obtain their embedding h and h, where $h = f(x)$. Subsequently, h goes through a projection head g composed of two fully connected layers to produce vector z and z, where $z = g(h)$. The contrastive loss in a minibatch was defined as follows:

$$L = -log \frac{\exp(sim(z_i, z_i^{'}))}{\sum_{k=1}^{K} \exp(sim\,(z_i, z_k))} \tag{1}$$

in which $sim()$ is the similarity function, and K is the number of negative samples cached in memory bank. We adopted a simple similarity defined as below:

$$sim\,(z_i, z_j) = \frac{z_i^T z_j}{\tau |z_i||z_j|} \tag{2}$$

where τ is an adjustable temperature parameter. Once the pretraining was finished, we discarded the projection head and used the trained encoder to obtain the feature of tiles.

We used ResNet50[33] as the backbone network, and loaded the pretrained weights on ImageNet. The SGD optimizer was used, learning rate of the backbone network was set to 0.03, the weight decay rate was 0.0001, and the momentum was 0.9.

3.6 Feature aggregation

For downstream tasks, we need to aggregate individual tile features into slide-level features. Instead of traditional max-pooling or average-pooling, we employ gated-attention pooling[34]. Suppose a slide has N tiles

S=$\{p_1, p_2, ...p_N\}$, we got the feature h_i using the pretrained encoder for tile p_i, the gated-attention pooling is essentially the instance-level weighted average pooling:

$$S = \sum_{i=1}^{N} a_i h_i \tag{3}$$

and

$$a_i = \frac{exp(w^T(\tanh(Vh_i) \bullet sigm(Uh_i)))}{\sum_{j=1}^{N} exp(w^T(\tanh(Vh_j) \bullet sigm(Uh_j)))} \tag{4}$$

where U and V are trainable parameter matrices, $\bullet$ denotes element-wise multiplication, $sigm()$ is a sigmoid non-linear activation function. The introduction of learnable parameters that are updated iteratively during the training process makes the model highly flexible and adaptable to various downstream tasks. Furthermore, after training, the attention weights reflect the significance of tile-level features for the downstream task, rendering the model interpretable.

3.7 Tumor diagnosis

We formulated the tumor diagnosis as the binary classification task. The ground truth labels (tumor or normal) came from TCGA-BRCA cohort. We employed a fully-connected layer plus a softmax layer as the prediction model that took the slide feature as input. The cross-entropy was used as the loss function for the classification task.

3.8 Protein level prediction

The protein level estimation for 223 tumor biomarkers was formulated as multi-task regression task. Taking as input the slide features, we adopted a multi-task learning model with a fully-connected layer and a output layer. The output layer had 223 nodes and each of them correspond to the level of a protein. In our practice, we have also tested multiple hidden layers architecture, and found that a single hidden layer could achieve superior performance in the regression task. The mean squared error (MSE) was used as loss function:

$$L = \frac{1}{M} \sum_{i=1}^{M} (y_i - y_i')^2 \tag{5}$$

Where y_i is the measured level of protein i by RPPA assays, $y_i^{'}$ is the predicted level, and M is the total number of proteins.

3.9 Drug response prediction

The Yale trastuzumab response cohort was used to evaluate the model capacity to predict drug response. As a case was designated as responder or non-responders, we formualted the prediction of drug response from pathological features as binary classification problems, and the cross-entropy loss function was used.

3.10 Prognosis model

We use slide-level features to predict prognostic risk for each breast cancer patient. Given the slide feature x, time period T, and an event indicator E, we use a Cox proportional hazard regression model based on deep learning to estimate prognostic risk. The hazard function is defined as:

$$h(t|x) = \lim_{\triangle t \to 0} \frac{Pr(t \leq O \leq t + \triangle t | O \geq t;\ x)}{\triangle t} \tag{6}$$

It estimates the instantaneous death rate of individual x at time t. The Cox proportional hazard regression models it as:

$$h(t|x) = h_0\,(t) \exp(\theta^T x) \tag{7}$$

where h_0 represents the baseline risk function at time t, which is estimated by Breslow estimation method. Our prognostic model was constructed and trained using the Pycox Python package. During training, the follow-up status and survival time are used to compute censored data and risk set. To estimate the parameters θ , we minimize the negative log partial likelihood:

$$L(\theta) = -\sum_{i:E_i=1} (\theta^T x_i - log \sum_{j \in R(t_i)} exp(\theta^T x_j)) \tag{8}$$

$R(t_i)$ is the risk set of time t_i, which represents the set of patients who may be at risk at time t, and $E_i = 1$ indicates the death event.

4 Discussion and Conclusion

There are increasing studies showing that solid tumor exhibit significant intra-tumor and inter-tumor heterogeneity. The heterogeneity leads to variations in the clinical treatment of different patients, pressing significant challenges to tumor clinical treatment. Precision medicine, which emphasizes personalized diagnosis and treatment, has been proposed to address this challenge. Proteomics plays an extremely important role in oncology research by identifying key biomarkers for treatment and prognosis. With the rapid advancement of digital pathology, computational pathology has been applied to predict gene mutations and microsatellite instability. However, few studies have focused on proteomics. Changes in protein expression levels reflect changes in gene expression that correspond to changes in cell and tissue phenotypes. In this paper, we investigate the prediction of cancer protein markers in conjunction with computational pathology. In fact, our experiments have demonstrated that changes in underlying molecular expression patterns are more reflected in pathological morphologies.

While comprehensive molecular and genetic tests are difficult to implement at scale, tissue sections stained with hematoxylin and eosin are ubiquitously available. We hypothesized that these routine tissue sections contain

information about established and candidate biomarkers and that molecular biomarkers could be inferred directly from digitized whole-slide images (WSIs). The rationale for this hypothesis is that genetic changes in tumor cells cause functional changes, which can influence tumor cell morphology. In addition to such first-order genotypephenotype correlations, genetic changes in tumor cells can influence the tumor microenvironment, resulting in higher-order genotypephenotype correlations.

The powerful feature extraction capability of deep learning makes it possible to infer molecular patterns from phenotypic features of pathological tissues. However, traditional deep learning algorithms rely on a large amount of manual annotations, while self-supervised contrastive learning has shown excellent performance on a variety of tasks. Our study shows that task performance can be significantly improved by pre-training using comparative learning and fine-tuning for specific downstream tasks. Contrastive learning can automatically capture latent features in unlabeled data, which avoids significant human and financial consumption.

In summary, we developed a self-supervised contrastive learning framework for tumor diagnosis from WSIs, for predicting tumor protein markers, and for predicting patient outcomes. In addition, our method has been applied to trastuzumab drug response prediction and spatial transcriptome data. Our extensive experiments show that pre-training based on contrastive learning effectively improves downstream tasks, including tumor diagnosis and protein marker prediction. We explore the interpretability of the model by using the tile-level attention scores learned from the model. We believe that our study will yield illuminating insights into computational pathology.

Declarations

- Funding
  This work was supported by National Natural Science Foundation of China (No. 62072058).
- Conflict of interest
  The authors declare no competing interests.

- Data Availability
  The whole slide images and corresponding labels of the TCGA-BRCA cohort from the TCGA database are available at https://portal.gdc.cancer.gov/. All RPPA data from The Cancer Proteome Atlas (TCPA) database is available at https://www.tcpaportal.org/tcpa/. The whole slide images and drug response data from Yale trastuzumab response cohort are available at TCIA database https://wiki.cancerimagingarchive.net/. The spatial transcriptomic data of a breast cancer specimen from 10X genomics are available at https://www.10xgenomics.com/. All other data supporting the findings of this study are available from the corresponding author upon reasonable request. Source data are provided with this paper.
- Code availability

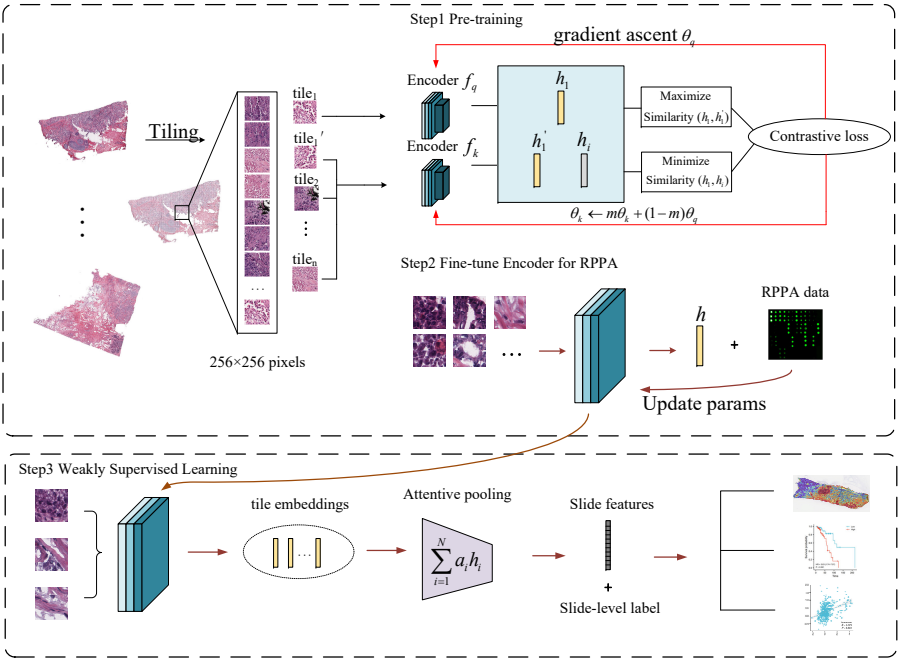

Fig. 1: Illustrative flowchart of our model framework. There were three steps, step1 shows slide proprocessing and the contrastive learning process, step2 shows fine-tuning backbone based on RPPA data, step3 shows the weakly supervised learning process and transfer learning to downstream tasks.

- Authors' contributions
  H.L. and X.X. conceived the study and designed the experiments. X.X. performed the experimental analysis. H.L. and X.X. prepared the manuscript. H.L. supervised the research.

References

[1] LeCun Y, H.G. Bengio Y: Deep learning. Nature **521**, 436–444 (2015)

[2] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation, 234–241 (2015)

[3] Hochreiter, S., Schmidhuber: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)

[4] Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using RNN

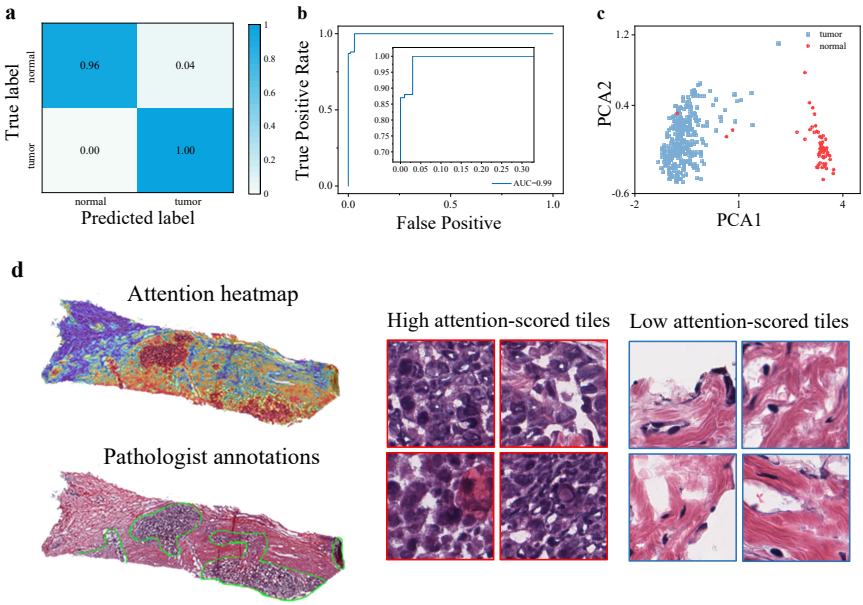

Fig. 2: **(a-b)** The ROC curve and confusion matrix on the independent test set. **(c)** The learned slide-level feature space is visualized by PCA. **(d)**The comparison between the spatial location of the tumor area displayed by the attention mechanism and the annotation area of the pathologist. The heatmap (upper left) is generated by the spatial deconvolution of the block and the original slide (lower left), and each tile is colored according to its normalized attention score. The area circled by the green line is the tumor necrosis area marked by an experienced pathologist (middle left). The right column shows some representative tiles with the highest and lowest attention scores.

encoder–decoder for statistical machine translation, 1724–1734 (2014)

[5] Wang, S., Rong, R., Yang, D.M., Fujimoto, J., Yan, S., Cai, L., Yang, L., Luo, D., Behrens, C., Parra, E.R., Yao, B., Xu, L., Wang, T., Zhan, X., Wistuba, I.I., Minna, J., Xie, Y., Xiao, G.: Computational Staining of Pathology Images to Study the Tumor Microenvironment in Lung Cancer. Cancer Research **80**(10), 2056–2066 (2020)

[6] Greenwald, N.F.e.a.: Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. Nature Biotechnology **40**(4), 555–565

[7] Strm, P., Kartasalo, K., Olsson, H., Solorzano, L., Delahunt, B., Berney, D.M., Bostwick, D.G., Evans, A.J., Grignon, D.J., Humphrey, P.A., Iczkowski, K.A., Kench, J.G., Kristiansen, G., van der Kwast, T.H., Leite,

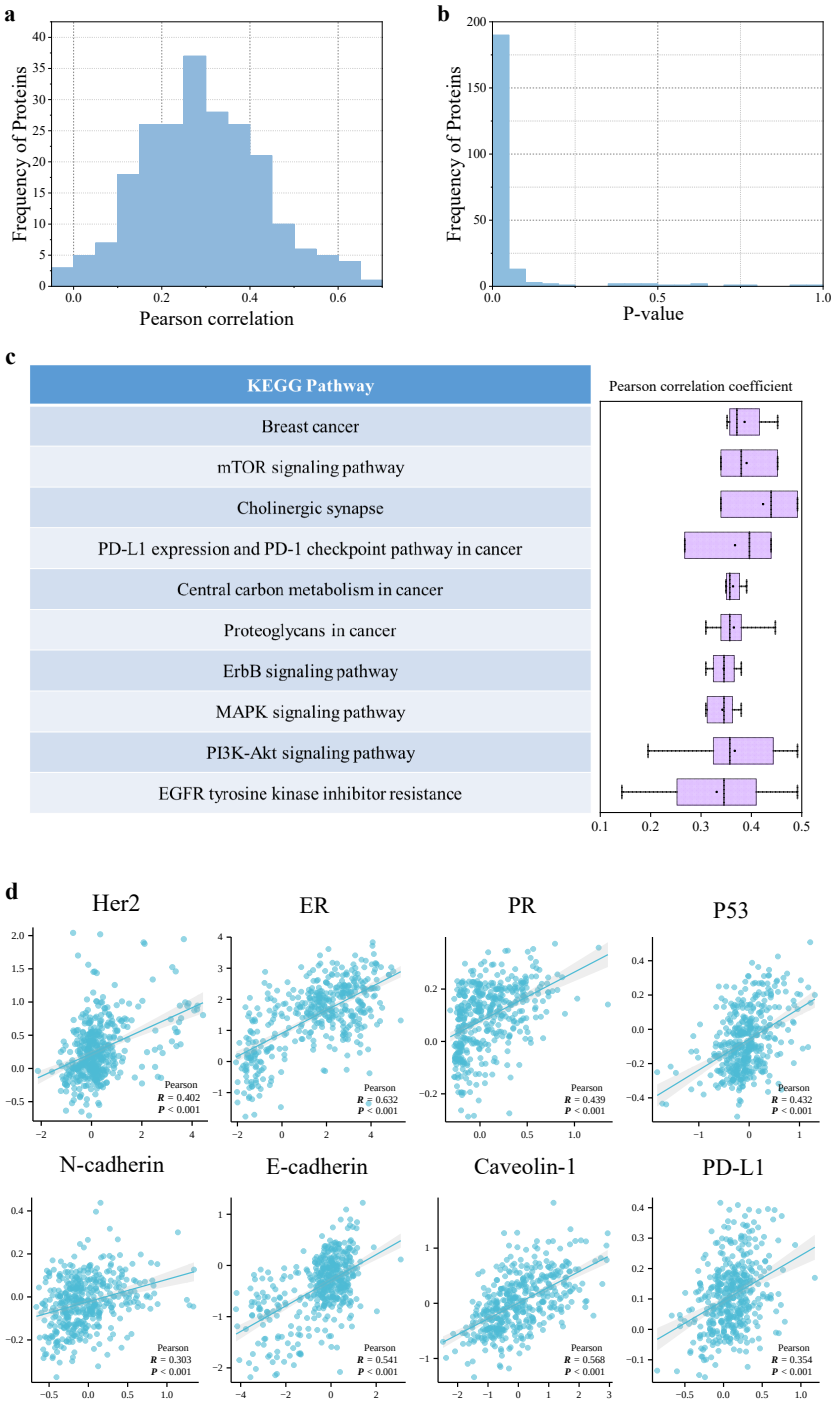

Fig. 3: **(a)** shows the histogram of correlation coefficients between predicted values and true values of 223 proteins; **(b)** shows P-value histogram. **(c)** shows the KEGG pathway analysis of one third well-predicted proteins (left), and the right side shows the box diagram of protein correlation coefficient under this signal pathway. **(d-k)** Some scatter plots of proteins associated with breast

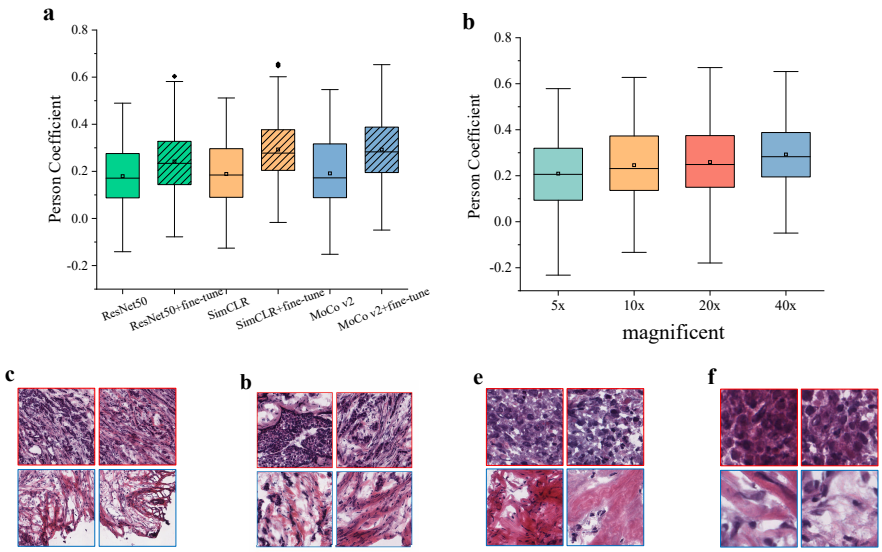

Fig. 4: **(a)** shows the prediction results of ResNet50, SimCLR, MoCo v2 and fine-tuned MoCo v2 respectively. **(b)** shows the prediction results of 5x, 10x, 20x and 40x magnification. **(c-f)** show some tiles with magnification of 5x, 10x, 20x, 40x.

K.R.M., McKenney, J.K., Oxley, J., Pan, C.-C., Samaratunga, H., Srigley, J.R., Takahashi, H., Tsuzuki, T., Varma, M., Zhou, M., Lindberg, J., Lindskog, C., Ruusuvuori, P., Whlby, C., Grnberg, H., Rantalainen, M., Egevad, L., Eklund, M.: Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. The Lancet. Oncology **21**(2), 222–232. Place: England

[8] Wang, Y., Acs, B., Robertson, S., Liu, B., Solorzano, L., Whlby, C., Hartman, J., Rantalainen, M.: Improved breast cancer histological grading using deep learning. Annals of Oncology **33**(1), 89–98 (2022)

[9] Kather, J.N., Heij, L.R., Grabsch, H.I., Loeffler, C., Echle, A., Muti, H.S., Krause, J., Niehues, J.M., Sommer, K.A.J., Bankhead, P., Kooreman, L.F.S., Schulte, J.J., Cipriani, N.A., Buelow, R.D., Boor, P., Ortiz-Brchle, N., Hanby, A.M., Speirs, V., Kochanny, S., Patnaik, A., Srisuwananukorn, A., Brenner, H., Hoffmeister, M., van den Brandt, P.A., Jger, D., Trautwein, C., Pearson, A.T., Luedde, T.: Pan-cancer image-based detection of clinically actionable genetic alterations. Nature Cancer **1**(8), 789–799

[10] Wang, Y., Kartasalo, K., Weitz, P., cs, B., Valkonen, M., Larsson, C.,

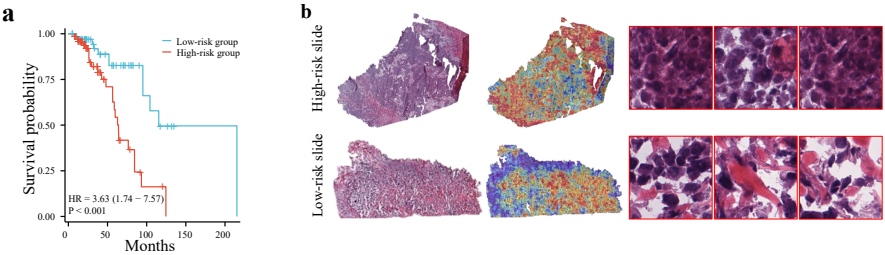

Fig. 5: **(a)** shows the survival curve (P-value<0.05) constructed by our risk score under 40x magnification. **(b)** shows the pathological images of high risk score and low risk score and their attention heatmap and high attention score tiles.

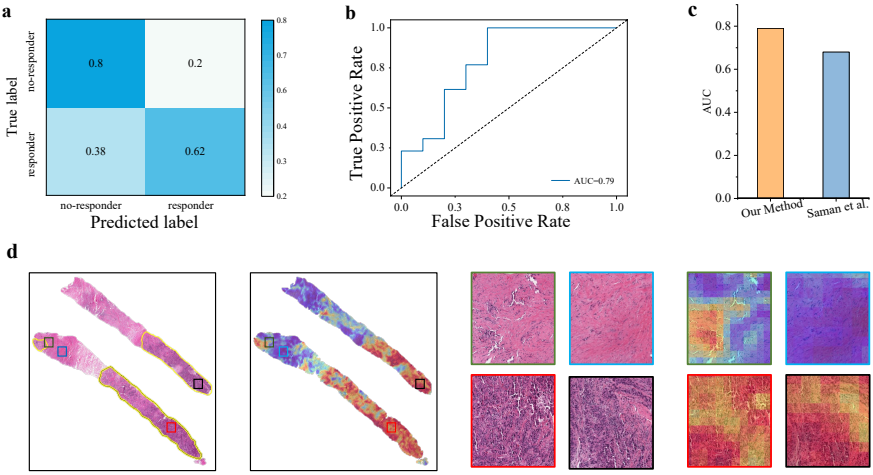

Fig. 6: **(a)** shows the confusion matrix of the predicted results of our model on the test set. **(b)** shows the ROC curve of the predicted results of our model on the test set. **(c)** shows the comparison between AUC value obtained by our method and Yale AUC value. **(d)** shows a comparison of the heatmap generated by the model with pathologist-labeled tumor regions and tiles randomly selected from the high- and low-scoring regions.

Ruusuvuori, P., Hartman, J., Rantalainen, M.: Predicting molecular phenotypes from histopathology images: A transcriptome-wide expression-morphology analysis in breast cancer. Cancer research **81**(19), 5115–5126 (2021)

[11] Huang, H., Zhou, G.W., Liu, X., Deng, L., Wu, C., Zhang, D., Liu,

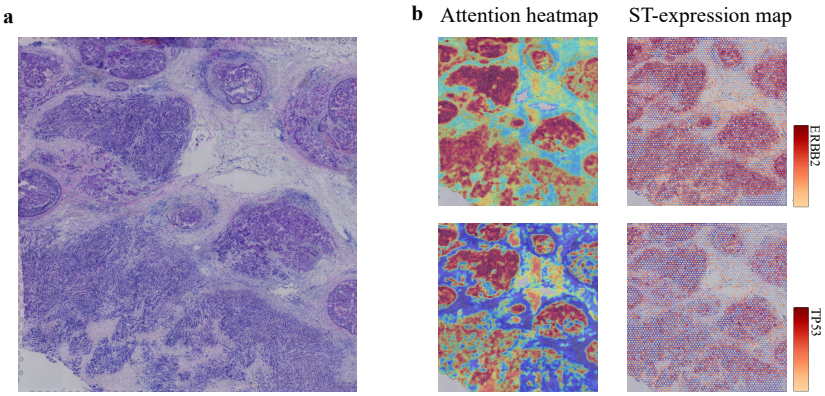

Fig. 7: **(a)**shows the original pathological section. **(b)** shows the comparison between the heatmap generated by our model and the gene high expression region located by the spatial transcriptome.

H.: Contrastive learning-based computational histopathology predict differential expression of cancer driver genes. Briefings in bioinformatics (2022)

[12] Lu, M.Y., Williamson, D.F.K., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F.: Data-efficient and weakly supervised computational pathology on whole-slide images. Nature Biomedical Engineering **5**(6), 555–570

[13] Schmauch, B., Romagnoni, A., Pronier, E., Saillard, C., Maill, P., Calderaro, J., Kamoun, A., Sefta, M., Toldo, S., Zaslavskiy, M., Clozel, T., Moarii, M., Courtiol, P., Wainrib, G.: A deep learning model to predict RNA-seq expression of tumours from whole slide images. Nature Communications **11**(1), 3877

[14] Shamai, G.e.a.: Deep learning-based image analysis predicts PD-l1 status from h&e-stained histopathology images in breast cancer. Nature Communications **13**(1), 6753

[15] Li, B., Li, Y., Eliceiri, K.W.: Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 14313–14323 (2021)

[16] Chen, C.e.a.: Fast and scalable search of whole-slide images via self-supervised deep learning **6**(12), 1420–1434

[17] Stewart, P.A., Welsh, E.A., Slebos, R.J.C., Fang, B., Izumi, V., Chambers,

M., Zhang, G., Cen, L., Pettersson, F., Zhang, Y., Chen, Z., Cheng, C.-H., Thapa, R., Thompson, Z., Fellows, K.M., Francis, J.M., Saller, J.J., Mesa, T., Zhang, C., Yoder, S., DeNicola, G.M., Beg, A.A., Boyle, T.A., Teer, J.K., Ann Chen, Y., Koomen, J.M., Eschrich, S.A., Haura, E.B.: Proteogenomic landscape of squamous cell lung cancer. Nature Communications **10**(1), 3578

[18] Mun, D.-G., Bhin, J., Kim, S., Kim, H., Jung, J.H., Jung, Y., Jang, Y.E., Park, J.M., Kim, H., Jung, Y., Lee, H., Bae, J., Back, S., Kim, S.-J., Kim, J., Park, H., Li, H., Hwang, K.-B., Park, Y.S., Yook, J.H., Kim, B.S., Kwon, S.Y., Ryu, S.W., Park, D.Y., Jeon, T.Y., Kim, D.H., Lee, J.-H., Han, S.-U., Song, K.S., Park, D., Park, J.W., Rodriguez, H., Kim, J., Lee, H., Kim, K.P., Yang, E.G., Kim, H.K., Paek, E., Lee, S., Lee, S.-W., Hwang, D.: Proteogenomic characterization of human early-onset gastric cancer. Cancer Cell **35**(1), 111–12410

[19] Ge, S., Xia, X., Ding, C., Zhen, B., Zhou, Q., Feng, J., Yuan, J., Chen, R., Li, Y., Ge, Z., Ji, J., Zhang, L., Wang, J., Li, Z., Lai, Y., Hu, Y., Li, Y., Li, Y., Gao, J., Chen, L., Xu, J., Zhang, C., Jung, S.Y., Choi, J.M., Jain, A., Liu, M., Song, L., Liu, W., Guo, G., Gong, T., Huang, Y., Qiu, Y., Huang, W., Shi, T., Zhu, W., Wang, Y., He, F., Shen, L., Qin, J.: A proteomic landscape of diffuse-type gastric cancer. Nature Communications **9**(1), 1012

[20] Qiang Gao, H.Z.e.a.: Integrated proteogenomic characterization of hbv-related hepatocellular carcinoma. Cell **179**(2), 561–57722 (2019)

[21] Coarfa, C., Grimm, S.L., Creighton, C.J.: Reverse-phase protein array: Technology, application, data processing, and integration. Journal of Biomolecular Techniques: JBT **32**(1), 15–29 (2021)

[22] Fu, Y., Jung, A.W., Torne, R.V., Gonzalez, S., Vhringer, H., Shmatko, A., Yates, L.R., Jimenez-Linan, M., Moore, L., Gerstung, M.: Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. Nature Cancer **1**(8), 800–810

[23] Jeselsohn, R., Buchwalter, G., De Angelis, C., Brown, M., Schiff, R.: ESR1 mutationsa mechanism for acquired endocrine resistance in breast cancer. Nature Reviews Clinical Oncology **12**(10), 573–583

[24] Hicks, D.G., Dell'Orto, P., Falzon, M., Hoff, K.D., Levy, Y.Y., McMahon, L.A., Miller, K., Russo, L., Viale, G.: Immunohistochemical performance of estrogen and progesterone receptor antibodies on the dako omnis staining platform: Evaluation in multicenter studies. Applied Immunohistochemistry & Molecular Morphology **25**, 313–319 (2015)

[25] Lim, E., Palmieri, C., Tilley, W.D.: Renewed interest in the progesterone

receptor in breast cancer. British Journal of Cancer **115**, 909–911 (2016)

[26] Mitri, Z., Constantine, T., O'Regan, R.: The her2 receptor in breast cancer: Pathophysiology, clinical use, and new advances in therapy. Chemotherapy research and practice **2012**, 743193 (2012)

[27] Duffy, M.J., Synnott, N.C., Crown, J.: Mutant p53 in breast cancer: potential as a therapeutic target and biomarker. Breast Cancer Research and Treatment **170**, 213–219 (2018)

[28] Qian, X.-L., Pan, Y., Huang, Q.-Y., Shi, Y., Huang, Q.-Y., Hu, Z.-Z., Xiong, L.-X.: Caveolin-1: a multifaceted driver of breast cancer progression and its application in clinical treatment. OncoTargets and therapy **12**, 1539–1552 (2019)

[29] Ghebeh, H., Mohammed, S., Al-Omair, A., Qattan, A., Lehe, C., Al-Qudaihi, G., Elkum, N., Alshabanah, M., Bin Amer, S., Tulbah, A.: Pd-l1 expression and tumor infiltrating pd-1+ lymphocytes associated with outcome in her2+ breast cancer patients. Breast Cancer Research and Treatment **155**(3), 605–614 (2016)

[30] Yeung, K.T., Yang, J.: Epithelialmesenchymal transition in tumor metastasis. Molecular Oncology **11**(1), 28–39 (2017)

[31] Farahmand, S., Fernandez, A.I., Ahmed, F.S., Rimm, D.L., Chuang, J.H., Reisenbichler, E., Zarringhalam, K.: Deep learning trained on hematoxylin and eosin tumor region of interest predicts HER2 status and trastuzumab treatment response in HER2+ breast cancer. Modern Pathology **35**(1), 44–51

[32] Li, J., Lu, Y., Akbani, R., Ju, Z., Roebuck, P.L., Liu, W., Yang, J.-Y., Broom, B.M., Verhaak, R.G.W., Kane, D.W., Wakefield, C., Weinstein, J.N., Mills, G.B., Liang, H.: Tcpa: a resource for cancer functional proteomics data. Nature Methods **10**(11), 1046–1047 (2013)

[33] 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), pp. 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90

[34] Tarkhan, A., Nguyen, T.K., Simon, N., Bengtsson, T., Ocampo, P., Dai, J.: Attention-based deep multiple instance learning with adaptive instance sampling. In: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), pp. 1–5 (2022)