



Original Research

Transcriptional intra-tumour heterogeneity predicted by deep learning in routine breast histopathology slides provides independent prognostic information



Yinxi Wang ^a, Maya Alsheh Ali ^a, Johan Vallon-Christersson ^b,
Keith Humphreys ^a, Johan Hartman ^{c,d,e}, Mattias Rantalainen ^{a,e,*}

^a Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

^b Division of Oncology, Department of Clinical Sciences Lund, Lund University, Lund, Sweden

^c Department of Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden

^d Department of Clinical Pathology and Cancer Diagnostics, Karolinska University Hospital, Stockholm, Sweden

^e MedTechLabs, BioClinicum, Karolinska University Hospital, Solna, Sweden

Received 11 January 2023; Received in revised form 5 June 2023; Accepted 17 June 2023

Available online 23 June 2023

KEYWORDS

Deep learning;
Intra-tumour
heterogeneity;
Prognosis;
Breast cancer;
Histopathology

Abstract **Background:** Intra-tumour heterogeneity (ITH) causes diagnostic challenges and increases the risk for disease recurrence. Quantification of ITH is challenging and has not been demonstrated in large studies. It has previously been shown that deep learning can enable spatially resolved prediction of molecular phenotypes from digital histopathology whole slide images (WSIs). Here we propose a novel method (Deep-ITH) to predict and measure ITH, and we evaluate its prognostic performance in breast cancer.

Methods: Deep convolutional neural networks were used to spatially predict gene-expression (PAM50 set) from WSIs. For each predicted transcript, 12 measures of heterogeneity were extracted in the training data set ($N = 931$). A prognostic score to dichotomise patients into Deep-ITH low- and high-risk groups was established using an elastic-net regularised Cox proportional hazards model (recurrence-free survival). Prognostic performance was evaluated in two independent data sets: SöS-BC-1 ($N = 1358$) and SCAN-B-Lund ($N = 1262$).

Results: We observed an increase in risk of recurrence in the high-risk group with hazard ratio (HR) 2.11 (95%CI:1.22–3.60; $p = 0.007$) using nested cross-validation. Subgroup analyses confirmed the prognostic performance in oestrogen receptor (ER)-positive, human epidermal growth factor receptor 2 (HER2)-negative, grade 3, and large tumour subgroups. The prognostic value was confirmed in the independent SöS-BC-1 cohort (HR = 1.84; 95%CI:1.03–3.3; $p = 3.99 \times 10^{-2}$). In the other external cohort, significant HR was observed

* Corresponding author: Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Nobels väg 12A, Solna 171 65, Sweden.
E-mail address: mattias.rantalainen@ki.se (M. Rantalainen).

in the subgroup of histological grade 2 patients, as well as in the subgroup of patients with small tumours (< 20 mm).

Conclusion: We developed a novel method for an automated, scalable, and cost-efficient measure of ITH from WSIs that provides independent prognostic value for breast cancer.

Significance: Transcriptional ITH predicted by deep learning models enables prediction of patient survival from routine histopathology WSIs in breast cancer.

© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Intra-tumour heterogeneity (ITH) refers to the variability in molecular or morphological phenotypes in different areas of a tumour (spatial heterogeneity) [1]. During the course of tumorigenesis, cell subpopulations harbouring driver mutations result in subclones with selective growth advantage, rendering tumours with potentially increasing invasiveness and metastatic ability. As tumour progression is a continuous and dynamic process, various degrees of neoplastic cell heterogeneity could arise as a function of molecular factors on multiple levels, including somatic mutations, copy number variations, and non-genetic factors such as epigenetic alterations, changes in mRNA expression states, and changes in tumour microenvironment [2]. Intra-tumour clonal heterogeneity has been shown as a prognostic factor across multiple cancer diseases [3] and is currently believed to be one of the factors that enable therapeutic escape [4,5], causing a challenge for the implementation of effective precision medicine.

The underlying genomic heterogeneity orchestrates variations in protein expressions, which has a close relation to treatment efficacy. In breast cancer, oestrogen receptor (ER) and human epidermal growth factor receptor 2 (HER2) status serve as indications for targeted systemic treatments. In tumour cells expressing ER, endocrine therapy is typically considered to effectively target ER-positive cells. However, it has been observed that the presence of ER-negative subclones could impair the clinical response, with a declined response rate in proportion to the degree of ER heterogeneity [6,2,7]. Likewise, while HER2-targeted therapy is a class of systemic treatment for HER2-amplified breast cancer, the treatment efficacy is affected by the percentage of HER2-positive cells. Compared with more homogeneous HER2-positive tumours, a worse prognosis has been observed in regional heterogeneous tumours [8].

Multiple approaches are available to measure intra-tumoural spatial heterogeneity, such as bulk-level DNA-sequencing, as well as single-cell sequencing techniques. Performing whole exome sequencing or whole genome sequencing in multiple regions within a primary tumour can enable the study of genetic heterogeneity in different cancer types [9,10]. Casasent et al. developed topographic single nucleus

sequencing (SNS) to study the spatial resolved genomic lineage between *in situ* cancer and invasive breast cancer, results identified subclones of evolved tumour cells within the duct, and depicted the migration and invasion of tumour cells from *in-situ* regions to surrounding tissues [11]. Navin et al. applied SNS techniques in breast cancer patients and identified three main clonal branches within a single tumour; In addition, a cell from one single clonal expansion was then seeded to establish the distant metastasis [12]. Although SNS or single cell sequencing (SCS) estimates the heterogeneous genomic evolutions on the single tumour cell level, it doesn't capture the spatial localisation and relationships within a tissue. The limitations of SCS have been addressed by spatial transcriptomics techniques [13]. To date, spatial transcriptomic studies have revealed heterogeneous expression profiles of cancer foci as well as interaction patterns between cells with distinct states within the tumour microenvironment [14,15].

Computational and digital pathology are emerging fields where machine learning techniques are used to assist with analysing digitised histopathology slides [16]. In particular, the rapid development of deep learning methodologies has enabled fast and automatic extraction of clinical relevant information from images, bringing about novel opportunities to facilitate cancer research and precision medicine [17].

Using computational pathology, several studies have demonstrated that molecular phenotypes, such as gene expression and DNA mutations, can be predicted from conventional haematoxylin and eosin (HE) stained tissue sections with deep learning models [18–23]. In the case of mRNA gene expression predictions, results have also been validated by spatial transcriptomic profiling [20,21]. Other aspects of ITH have also been modelled using deep learning methodologies, including graph neural networks, and hand-crafted machine learning approaches [24,25]. In [26], deep learning-based cell segmentation and classification were employed to model spatial immune infiltration, enabling classification of tumour regions into immune hot and cold subtypes. This difference with respect to intra-tumour spatial immune activity was found to be correlated with RNA-seq estimated immune infiltration as well as risk for relapse [26]. Comparing computational pathology solutions with molecular assays (single cell profiling or spatial

profiling), the computational approach doesn't require additional sample collection nor expensive profiling, and can be easily scaled to whole transcriptome level as well as large sample sizes with relatively low cost.

In this study, we propose a novel model, Deep-ITH, that employs a deep learning-based computational pathology approach to predict spatial gene expression and uses these predictions to quantify ITH in histopathology whole slide images (WSIs) of HE-stained tissue from surgically resected breast tumours. Our study focuses on the PAM50 gene panel [27] that is well-established for breast cancer subtyping and prognostication. Our main objective was to evaluate to what extent computationally quantified ITH is prognostic (recurrence-free survival) in breast cancer. The method was validated on two external test cohorts to evaluate the generalisability across different studies and medical centres.

2. Materials and methods

2.1. Dataset and cohorts

For training and internal validation of the Deep-ITH model, patients from two data sources were merged: Clinseq-BC ($N = 256$) and the Cancer Genome Atlas (TCGA-BC; $N = 675$) [20]. The Deep-ITH model was subsequently evaluated in two external test sets, SöS breast cancer cohort phase I (SöS-BC-1) ($N = 1358$) [28], and SCAN-B-Lund ($N = 1262$) [29]. Summary of clinicopathological information of each cohort was displayed in the [Supplementary Table 1](#). For further details regarding each cohort, see [30]. For the training set (Clinseq-BC and TCGA-BC), only patients with gene expression data and survival information were included. In addition, to ensure consistency in image quality, only TCGA-BC patients with slides that scanned at 40X magnification were included. In both training and test data, patients that received neoadjuvant treatment were excluded ([Supplementary Fig. 1](#)). For each included patient from Clinseq-BC, SöS-BC-1 and SCAN-B-Lund, the HE-stained formalin-fixed paraffin-embedded (FFPE) histopathology slides were scanned into WSIs with Hamamatsu Nanozoomer XR (Hamamatsu Photonics) at 40X magnification (0.226 mm/pixel). For patients from TCGA-BC, the WSIs scanned at 40X were retrieved from <https://portal.gdc.cancer.gov/>. Only one WSI per patient was included.

RNA-seq data were available for patients from Clinseq-BC and TCGA-BC. Clinical data for patients from Clinseq, SöS-BC-1 and SCAN-B-Lund were retrieved from the Swedish National Breast Cancer Registry whereas for patients from TCGA-BC, the clinical data were extracted from the TCGA database as well as pathology reports (dbGAP projectID:5621).

The Clinseq-BC and TCGA-BC were used to train and validate the Deep-ITH model while the external test sets (SöS-BC-1 and SCAN-B-Lund) were used only once to provide a final evaluation of the generalisability of the Deep-ITH model.

2.2. Generation of spatial predictions

Each WSIs were tiled into image patches that covered an original region of 1196×1196 pixels (0.271×0.271 mm), under 40X magnification, and were subsequently down-scaled by a factor of 2, so that final tiles were 598×598 pixels in size. Overlaps between adjacent patches were 25% for slides from SöS-BC-1%, and 50% percent for slides from the other three cohorts as previously described [28]. The coordinate information was preserved during tiling. Deep Convolutional Neural Network (CNN) models that were previously optimised [20] for predicting expression levels of the PAM50 gene panel [31] were utilised to obtain spatially resolved gene expression predictions for each tile in each WSI in all data sets. Next, tile level predictions were mapped back to their original locations within a WSI using the coordinate information, resulting in a predicted spatial gene expression map. The predictions in overlapping regions were calculated as an average across the overlapping tiles. Spatial heterogeneity features were calculated based on the unit of tiles.

2.3. Extraction of texture features

For each WSI and for each predicted spatial gene (PAM50; 50 genes) expression map, eleven heterogeneity features were extracted ([Supplementary Table 2](#)). These features were selected with previously identified clinical relevance [32], aiming to describe the spatial heterogeneity per gene expression pattern. The features consist of four major types that belong to either second-order statistical method: grey-level co-occurrence matrix ('GLCM') [33], or higher order statistical methods including grey-level size zone matrix ('GLSZM') [34], grey-level run length matrix ('GLRLM') [35], neighbourhood grey tone-difference matrix ('NGTDM') [36]. Second-order metrics compute the statistical measurements of intensity values between pairs of neighbouring pixels, which according to the current implementation, describes the relationship between gene expression predictions of neighbouring image tiles; Higher-order methods compute the combination among > 3 pixels, which in the current scenario models the gene expression patterns that stretch across the tumour region with a size of > 3 image tiles. Together, these features capture the frequency and type of spatial variability in the gene expression patterns associated with individual genes within a tumour section,

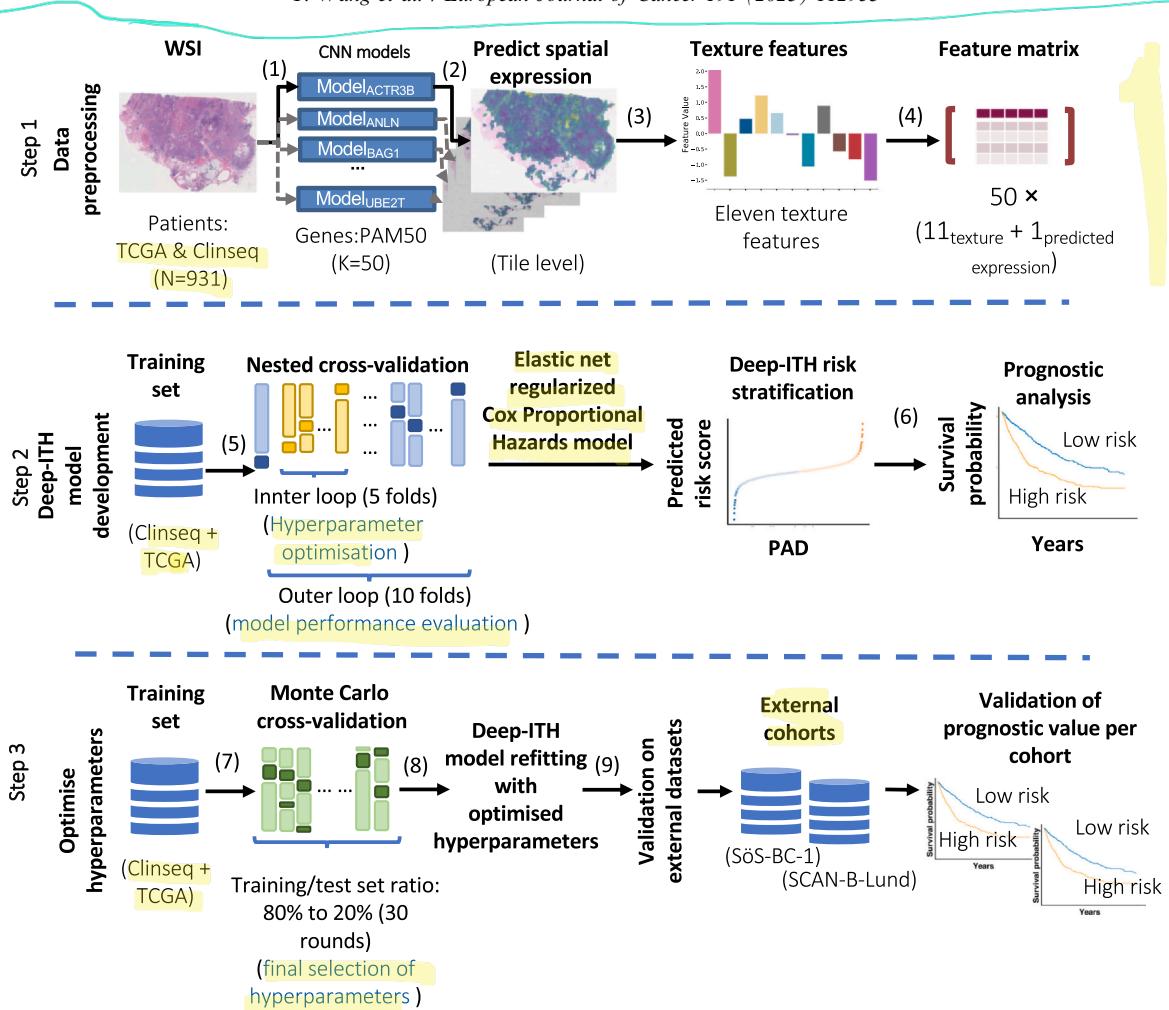


Fig. 1. Overview of the data preprocessing, model optimisation and validation workflow. Step 1 shows the general data preprocessing pipeline. (1) Deep learning models that were previously optimised to predict expression in the PAM50 gene panel were applied on each WSI. (2) Spatial predictions of gene expression levels were generated per gene and WSI. (3) Next, 11 texture features were extracted from spatially resolved predicted gene expression patterns to quantify the level of ITH in each slide. (4) Together with the predicted slide-level gene expression from the deep learning model, for each patient, and each individual gene from the PAM50 gene panel, we collected 12 features, resulting in a total of 600 features per patient. Step 2 depicts the process of development and initial evaluation of prognostic value in the training data set. (5) In the training data set, a Cox proportional hazards model with elastic net regularisation was fitted using a nested cross-validation approach. (6) Model-predicted Deep-ITH scores were applied to stratify patients into Deep-ITH low- and high-risk groups (in cross-validation) by the median risk score among all patients. Step 3 shows the final fitting of the model in the training data, and subsequent external validation procedure. (7) After confirming the prognostic value of risk stratification (in (6)), the final combination of hyperparameters was determined in a 30-round Monte-Carlo CV using the training data set. (8) The final Deep-ITH model was refitted on all training data. (9) In the end, the model was applied to evaluate its generalisability on two external test cohorts. WSI, whole slide image; PAD, patient ID.

thus, providing measurements of intra-tumour gene expression heterogeneity. All texture features were extracted with PyRadiomics v.3.0.1. For each spatial gene expression prediction, we firstly discretised the continuous value to a fixed bin count of 4, and then computed features with default parameters associated with the package.

2.4. Development of a prognostic model based on ITH

We applied an elastic net penalised Cox regression model [37] to generate a prognostic score (Deep-ITH score) based on intra-tumour spatial gene expression heterogeneity.

Fig. 1 provides a conceptual overview of the three steps in our study. In the first step, a preprocessing pipeline was applied to generate a predicted gene expression heatmap for each WSI (one for each patient) and each gene in the PAM50 gene panel. Next, eleven texture features were extracted (see above) from individual heatmap; In addition, a slide-level predicted gene expression value was also included (mean predicted expression across all tiles within a WSI). This resulted in a total of 600 features for each WSI (corresponding to each patient), which were included into the time-to-event model.

In the second step, a nested cross-validation (CV) approach was applied to at the same time optimise

hyperparameters and to obtain an estimate of the prognostic performance. The outer CV was based on 10 folds, where patients were randomly split into training and test (9 folds for training, 1 for test). The training split in the outer CV loop was subsequently used in the inner CV (5 folds, random splitting on patient level) to optimise hyperparameters (alpha, lambda). The shrinkage penalty (lambda) was chosen empirically while the relative weight between ℓ_1 and ℓ_2 penalties (alpha) was evaluated by grid search (0.0001, 0.001, 0.005, 0.1, 0.2, 0.3, 0.5, 0.7, 0.9). The hyperparameter combination that yielded the best C-index in the inner CV was selected, and the corresponding model was used to predict on the held out test set in the outer CV. To account for batch differences, features from the training set were first normalised to zero mean and unit variance before fitting the model, and the test set features were normalised against the training set using the same estimated values.

For each individual in the test set in the outer CV, we obtained a risk score defined as the sum of products between feature values and coefficients in the penalised Cox regression model. After aggregating predicted risk scores across all outer CV rounds, we re-stratified patients into Deep-ITH low and high-risk groups using the median risk score, and subsequently compared the survival outcomes between the dichotomised groups.

After initial confirmation of prognostic performance based on nested CV (see above), the final model was refitted (including hyperparameter optimisation) using a 30-round Monte-Carlo CV, again in the training data (Clinseq-BC and TCGA-BC) with a training/test set ratio of 80–20%. The same grid search strategy was employed as described above with the objective of maximising the C-index.

2.5. Prognostic performance evaluation in external test data sets

The generalisability of the Deep-ITH model was evaluated in two external test sets, namely SöS-BC-1 and SCAN-B-Lund (Fig. 1, step 3).

The final model was applied to dichotomise the external test cohorts using the median value of predicted Deep-ITH score in each cohort. To allow for potential calibration issues, prognostic performance was evaluated in the cohorts separately.

All statistical analyses performed in the test sets were pre-specified at the start of the study, with the exception of the cluster analysis of the features (see below). The two external test data sets were only used to evaluate model performance, and they were never used to inform any aspect of model development or optimisation.

2.6. Statistical analysis

We compared recurrence-free survival (RFS) rates between identified Deep-ITH low- and high-risk

patient groups (see above) to evaluate whether ITH could serve as an independent prognostic factor using the Deep-ITH model. The primary analysis was performed across all patients in each data set. Additional subgroup analyses were performed in the ER-positive and HER2-negative subgroup, and also separately by tumour grade, tumour stage (size), and lymph node status.

A recurrence event was considered if locoregional or distant relapses, metastasis, contralateral tumours or death was detected. Patients that died without any reported metastasis were assumed to have an unidentified metastasis before death [38,39]. The time-to-event interval was defined as the diagnostic date to the date when the first recurrence event was observed or the last follow-up date.

Furthermore, as an alternative end-point to RFS, overall survival (OS) was also evaluated.

Univariable and multivariable Cox proportional hazards (Cox PH) regression models were used to estimate unadjusted and adjusted hazard ratios (HRs) and corresponding 95% CI, respectively. In the univariable Cox PH model, only the Deep-ITH risk group was included as a covariate to generate an unadjusted HR. The same approach was used to calculate an unadjusted HR for each individual texture or predicted gene expression feature selected by the Cox PH elastic net model. In the univariate analyses of the features, the continuous features were first standardised to 0 mean and unit standard deviation prior to the Cox PH regression analysis. Univariable Cox PH was also performed on features that were dichotomised by their median to estimate the corresponding HRs. P-values were corrected for multiple testing using Benjamini–Hochberg's FDR-based approach [40]. In the multivariable Cox regression model, age, tumour size, lymph node status, ER status, HER2 status, and grade were included as covariates. Age was modelled as continuous value; tumour size was coded as ≥ 20 mm or < 20 mm according to the diameter of the tumour; Lymph node status was determined based on whether there exists lymph node metastases; ER status was dichotomised based on a 10% cut-off for positively stained cells with immunohistochemical staining; HER2 status was assessed using fluorescent *in situ* hybridisation (FISH) or chromogenic *in situ* hybridisation (CISH) together with immunohistochemistry assay, a sample was labelled as HER2 positive if there's amplification in FISH/CISH or, when no such result available, if it was graded 3+ through the immunohistochemistry assay. Patients with missing data in the above risk factors were excluded in the multivariable Cox regression analysis. The proportional hazards (PH) assumption was examined and corrected for by stratification on variables that violated the assumption.

2.7. Further investigation of selected features

Feature (predictor) selection is an intrinsic part of elastic net regularised models, where features having non-zero coefficients are included (i.e. selected).

To investigate the covariability between selected features, we applied hierarchical clustering. The features were first normalised and then clustered based on Euclidean distance in an agglomerative manner. We adopted ‘average’ as the method for cluster distance computation based on cophenetic correlation coefficient [41]. The number of clusters was determined by consensus clustering. The relative area change under the cumulative distribution function between k and (k-1) clusters was computed and the optimal number of clusters was selected where the relative change exhibits the most rapid decrease. The result from hierarchical clustering was visualised as a dendrogram. A feature correlation heatmap was computed based on the Spearman’s rank correlation.

In addition, differences in feature means between the groups were tested (*t*-tests) and feature distributions were visualised using grouped boxplots. All texture features were log-transformed prior to *t*-test analysis and visualisation. Results were adjusted for multiple testing using Benjamini–Hochberg’s approach.

2.8. Software

The gene expression prediction was performed with Python (v.3.6.6) with packages OpenCV (v. 3.4.1), Open-Slide (v.3.4.1 and API v.1.1.1) [42], Keras (v. 2.2.4) with Tensorflow backend [43] (v.1.12). The texture features were extracted using Python package PyRadiomics v.3.0.1 [44]. The Cox Proportional Hazards regression model with elastic net penalty was estimated using scikit-survival v. 0.14.0 [45]. The model optimisation procedure was implemented with methods provided in scikit-learn (v. 0.23.2) [46]. The survival analysis was carried out using R (v.3.6.3) with R packages (‘survminer’ and ‘survival’). Other statistical analyses were performed with SciPy (v.1.5.2) [47] and multiple test adjustment was performed in statsmodels (v.0.11.1) [48]. Consensus clustering was performed with R package ‘ConsensusClusterPlus’. Dendrogram coloured with identified clusters from the hierarchical clustering was generated using ‘Scipy’. Figures were plotted with the Python package ‘matplotlib’ [49] except for the heatmap with ‘seaborn’ [50].

3. Results

3.1. Study overview

In a previous study, we developed a novel deep learning-based modelling approach that provides spatial prediction of mRNA expression directly from HE-stained

WSIs [51]. These deep CNN models were optimised in a weakly supervised manner using tiled image patches as predictors and slide-level gene expression as responses. In the current study, we predict tile-level gene expression using the previously developed models, and use the predicted expression patterns to quantify ITH. Based on the predicted heterogeneity measurements, we implemented a prognostic risk stratification model and applied it to stratify breast cancer patients into Deep-ITH low- and high-risk groups; The prognostic (risk of recurrence) performance was evaluated in independent test data.

Spatial gene expression values from the PAM50 gene set were first predicted by deep-learning models for each gene, and for each WSI. Next, eleven texture features were extracted to quantify the heterogeneity of gene expression patterns from the spatial expression predictions for each gene. Subsequently, all texture features together with slide level predicted gene expression values were used to optimise a Cox PH regression model with elastic net penalisation, in order to enable us to calculate a risk score (Deep-ITH score) for each WSI and patient. Patients were then dichotomised into two risk groups based on the predicted scores and their differences in RFS were compared to confirm the prognostic value of the Deep-ITH model. The final model was validated in two external cohorts to evaluate generalisability and prognostic performance in real world clinical data.

The distribution of grade, ER, PR, HER2, and tumour size were not different between two external test sets ($p > 0.05$, chi-square test). In comparison, there were a larger number of cases with positive lymph node involvement in the SCAN-B-Lund cohort, and the proportion of subtype was also different ($p < 0.05$, chi-square test) (Supplementary Table 1).

3.2. Spatially resolved ITH can be used to stratify breast cancer patients with distinct survival outcomes

Patients from the Clinseq-BC ($N = 256$) and TCGA-BC ($N = 675$) cohorts were merged into the single training set. A nested CV procedure (see Methods) were applied to the training data to optimise the prognostic model while simultaneously providing an estimate of the prognostic performance. Patients were re-stratified into low- and high-risk Deep-ITH groups according to the median Deep-ITH score. A distinct difference in RFS can be observed from the Kaplan-Meier curves (Fig. 2A). The HR between Deep-ITH high- and low-risk groups was 2.11 (95% CI: 1.22–3.60, $p = 0.007$) (Fig. 2B), adjusting for established risk factors including age, tumour size, lymph node status, HER2 status, ER status, and grade. In addition, we further compared the prognostic value of the Deep-ITH model in ER-positive (ER+), HER2 negative (HER2-) subgroups, and in patients with different histological grades, tumour size as well as lymph node status. The results are

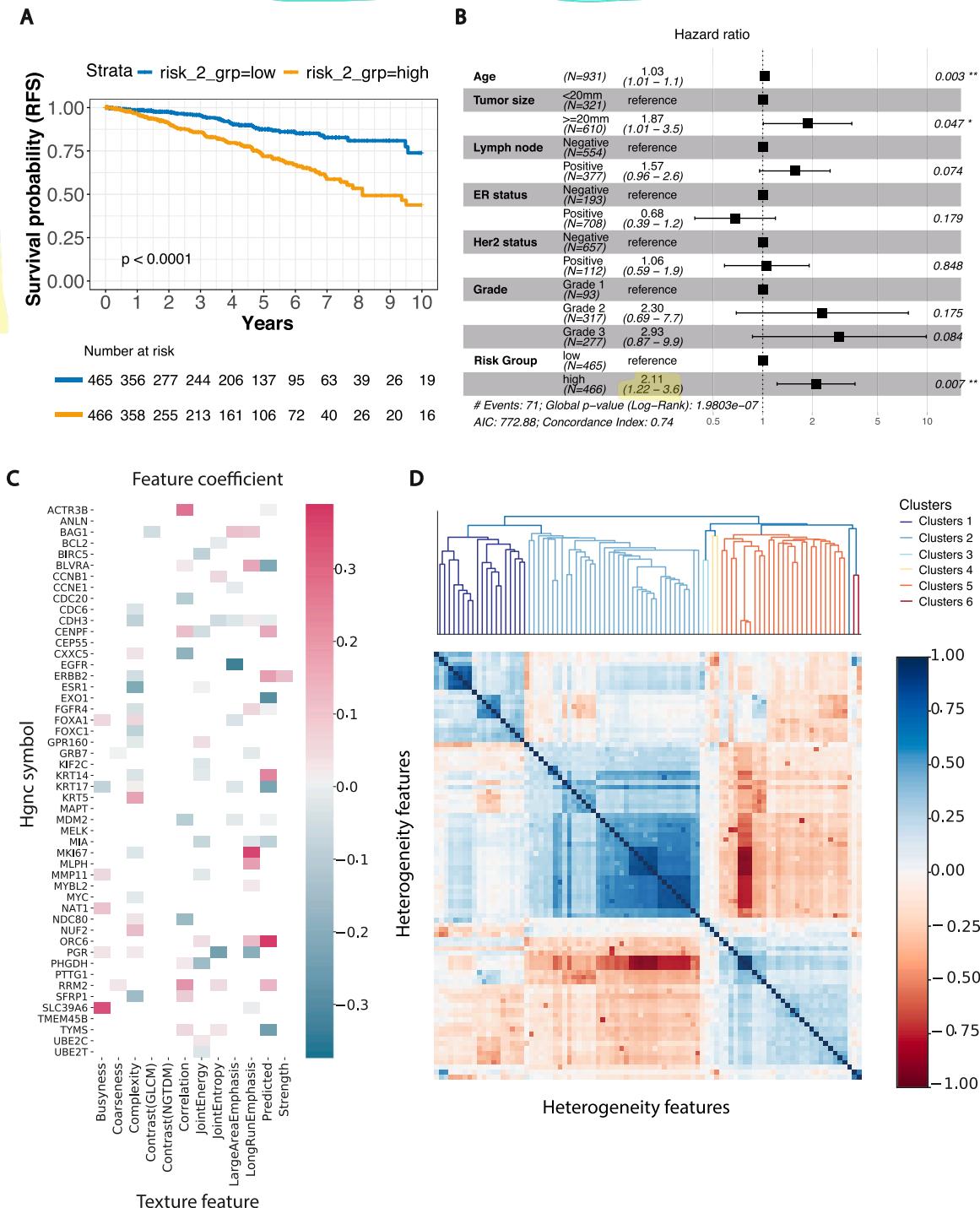


Fig. 2. Summary of model performance with nested cross-validation in the training set. A. Kaplan-Meier curves of recurrence-free survival between patients dichotomised by the Deep-ITH score. Deep-ITH low-risk group revealed a higher survival rate compared with high-risk group ($p = 7.5 \times 10^{-9}$, log-rank test). B. Forest plot of multivariable Cox proportional hazard regression analysis among dichotomised cases. The HR between Deep-ITH low- and high-risk groups was estimated to be 2.11 (95% CI: 1.22–3.60, $p = 0.007$). C. Heatmap of feature importance. Colour indicates the coefficient in the final model, x-axis shows the name of the genes, y-axis shows the name of feature associated with the individual gene. D. Hierarchical clustering and feature correlation heatmap identified six clusters among all features selected from the final model. (For the heatmap with a full list of feature names with higher resolution, please refer to Supplementary Fig. 4). HR, hazard ratio; CI, confidence interval.

summarised in Supplementary Table 3. The proposed restratification was also confirmed to be an independent prognostic factor in several subgroups, including ER

+(HR = 2.12, 95%CI: 1.11–4.07, $p = 0.0236$), HER2-(HR = 2.43, 95%CI: 1.31–4.48, $p = 0.0046$), grade 3 (HR = 2.31, 95%CI: 1.08–4.96, $p = 0.0319$), and

tumours with a size > 20 mm (HR = 2.20, 95%CI: 1.16–4.17, p = 0.0155). For the full list of subgroup analysis, please refer to **Supplementary Table 3**. The prognostic value was also investigated with OS as end-point, with similar findings, and an adjusted HR of 2.40 (95%CI: 1.18–4.9, p = 0.015, **Supplementary Fig. 2 A, B**).

After the successful confirmation of prognostic performance in the training data using nested CV, a final model was optimised using all data from the training set, and this model was subsequently used to evaluate performance in independent test sets (see below). Out of the 600 features included in the elastic-net regularised time-to-event model, 90 features with non-zero coefficients (i.e. included in the model) were included in the model (see **Supplementary Fig. 2C** and **Supplementary Table 4**). 45 out of 50 genes from the PAM50 panel have at least one feature included in the final prognostic model. There were five texture features associated with gene *CDH3*, ranking second were *KRT17*, *RRM2* and *PGR*, each having four features selected. The texture ‘Complexity’ was selected 18 times, followed by ‘LongRunEmphasis’ and ‘JointEnergy’, both were selected by 12 times, ranking fourth was ‘Correlation’ which appeared 11 times in the final mode (**Fig. 2C**). The effect size associated with each feature varies for different gene-texture feature pairs. For instance, the predicted slide level expression of *ORC6*, ‘LongRunEmphasis’ of *MKI67* and ‘Busyness’ of *SLC39A6* had the largest effect sizes associated with poor prognosis, whereas the ‘Long-AreaEmphasis’ for *EGFR*, predicted expression of *EXO1* as well as *TYMS* exhibited the largest protective effect sizes among the selected features. The underlying similarity and potential functional relationship between features was further investigated through hierarchical clustering (**Fig. 2D**). In total, six ITH subgroups were identified by consensus clustering analysis (**Supplementary Fig. 3**), indicated by the colour in the dendrogram. A high-resolution heatmap with a full list of features can be found in **Supplementary Fig. 4**. We observed three clusters that each included two gene expression features: ‘*KRT14*’ and ‘*KRT17*’ ‘*FGFR4*’ and ‘*ERBB2*’ ‘*BLVRA*’ and ‘*MDM2*’. There were another two texture feature-related clusters and one cluster that contains both gene expression and texture features.

3.3. Validation of the heterogeneity-based prognostic risk stratification in external cohorts

To assess the generalisability of the Deep-ITH model, we evaluated the optimised model in two independent cohorts, namely (SöS-BC-1) (N = 1358) and SCAN-B-

Lund (N = 1262), using univariate as well as multivariable Cox proportional hazard models.

In SöS-BC-1, we observed a significantly different RFS rate between the low and the high Deep-ITH risk groups (p = 0.00075, log-rank test, **Fig. 3A**), and a comparable HR (multivariable analysis, Cox Proportional Hazards model) compared with the training data set (HR = 1.84, 95% CI: 1.03–3.3, p = 0.04, **Fig. 3B**). Subgroup analyses were carried out as in the training data set, and the prognostic values remained significant in the ER+ and HER2- subgroups, with HRs of 1.97 (95% CI: 1.06–3.66, p = 0.0327), 1.9 (95% CI: 1.01–3.56, p = 0.046) (**Supplementary Table 5**), respectively. The adjusted HR for OS in SöS-BC-1 was 1.80 (95%CI: 0.99–3.3, p = 0.055, **Supplementary Fig. 5 A, B**).

We also observed a significant (p < 0.05) difference in RFS rate between low and high Deep-ITH risk groups in the SCAN-B-Lund cohort (p = 0.017, log-rank test, **Fig. 3C**). In the multivariable analysis (Cox Proportional Hazards model), a violation of the proportional hazards assumption was observed for the ‘tumour size’ variable, and corrected by stratification on tumour size. This variable was thus excluded in the output of the multivariable analysis. The adjusted HR for risk stratification in the full SCAN-B-Lund cohort did not meet the significance criteria (p < 0.05; HR = 1.19, 95% CI: 0.81–1.7, p = 0.375, **Fig. 3D**). Subgroup analysis revealed a significant prognostic risk stratification in patients with grade 2 tumours (HR = 1.74, 95% CI: 1.01–3.0, p = 0.0451), and in patients with tumour size < 20 mm (HR = 1.76, 95% CI: 1.03–2.99, p = 0.0388, **Supplementary Table 5**). For more information regarding the subgroup analysis, please refer to **Supplementary Table 5**. The effect sizes also exhibited the same direction for the estimation of HRs in ER+ and the HER2- subgroups. Considering OS as end-point, the adjusted HR in SCAN-B-Lund was 1.14 (95%CI: 0.75–1.75, p = 0.535, **Supplementary Fig. 5 C, D**).

3.4. Marginal prognostic value of ITH features

To understand the relationship between patient prognosis and all individual texture features that were included in the final model, we computed univariable Cox regression for each feature in the training and two test cohorts (**Supplementary Table 4**). 36 out of the 90 features exhibited a significant association with outcome (RFS) in the training set (**Fig. 4A**), whereas in the SöS-BC-1 (**Fig. 4B**) and SCAN-B-Lund (**Fig. 4C**) test cohorts, 52 and 23 features were significantly associated with RFS, respectively. In particular, 13 features were found to be significant in all three cohorts, including six features that quantify the level of ‘Busyness’ of textures (*SLC39A6_Busyness*, ‘*NAT1_Busyness*’, ‘*KRT17_Busyness*’, ‘*PGR_Busyness*’, ‘*FOXA1_Busyness*’ and

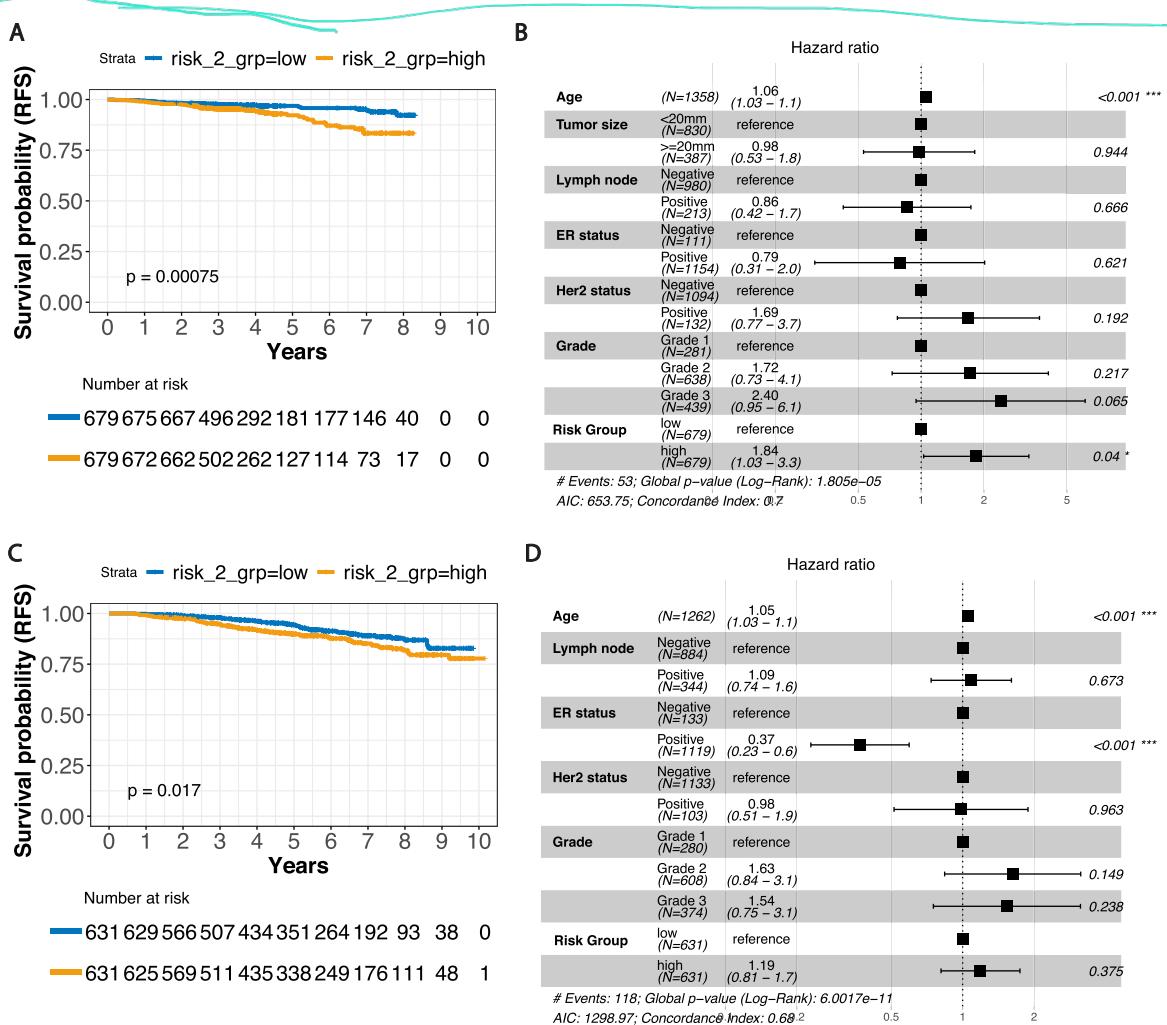


Fig. 3. Validation of Deep-ITH model in external test sets. A. Kaplan-Meier curves of recurrence-free survival for patients dichotomised by Deep-ITH score from SöS-BC-1 cohort ($p = 0.00075$, log-rank test). B. Forest plot of multivariable Cox proportional hazard regression analysis among dichotomised cases from SöS-BC-1 cohort (HR: 1.84; 95% CI: 1.03–3.3, $p = 0.04$). C. Same analysis as in A, with patients from SCAN-B-Lund cohort ($p = 0.017$, log-rank test). D. Same analysis as in B, with patients from SCAN-B-Lund cohort (HR: 1.19; 95% CI: 0.81–1.7, $p = 0.375$). HR, hazard ratio; CI, confidence interval.

'MMP11_Busyness'), they demonstrated a consistent harmful effect to patient survival across all three cohorts. Another three features associated with 'LongRunEmphasis' ('ORC6_LongRunEmphasis', 'CDH3_LongRunEmphasis', 'MLPH_LongRunEmphasis') also demonstrated an harmful effect, similar with one 'LargeAreaEmphasis' related feature ('BAG1_LargeAreaEmphasis'). Higher levels of three predicted gene expression features were also found to be associated with worse patient survival outcomes ('FGFR4', 'RRM2', 'ORC6').

Furthermore, the univariable prognostic effect was also examined in analysis of dichotomised features (Supplementary Fig. 6, Supplementary Table 6). In the training cohort, 44 out of the 90 features had a significant (FDR adjusted p-value < 0.05) association with RFS, whereas in the two test cohorts, there were 51 and 6 features, respectively. Three of the dichotomised features

were found to have a significant (FDR adjusted p-value < 0.05) prognostic value in all three cohorts, namely, 'RRM2_Coarseness_1', 'RRM2_1' and 'ORC6_1'.

3.5. Differences in intra-tumour characteristics of Deep-ITH low- and high-risk tumours

To characterise how ITH features are associated with recurrence risk, we compared the distribution of included features between Deep-ITH low and high-risk groups. The analysis was performed on log-transformed features.

In the training cohort, the mean of 84 out of 90 features were significantly different between the two risk groups (t -test, adjusted p-value < 0.05). Similarly, in both SöS-BC-1 and SCAN-B-Lund cohorts, 77 features were found to have different means (Supplementary Table 7). 41 genes were found to have at least one

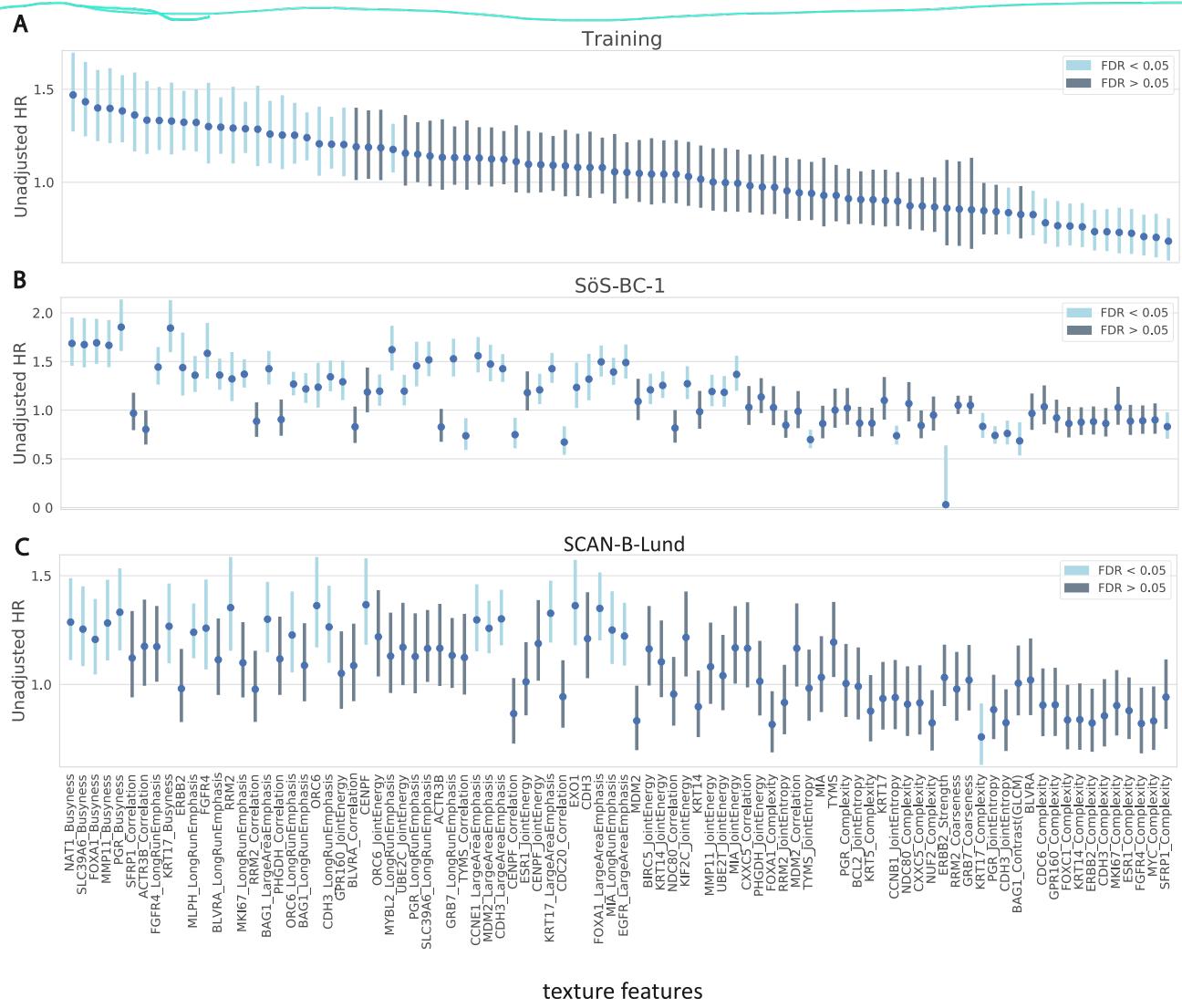


Fig. 4. Plot of unadjusted HR estimates for each feature and each cohort. Marginal HRs estimated from the single variable Cox regression analysis for features selected in the final Deep-ITH model, ranked in decreased HRs. A. In the training set, 37 features demonstrated individual prognostic value with recurrence-free survival. B. The corresponding HRs in the SöS-BC-1 cohort were displayed in the same order, with 52 significant features. C. same as B, with results from the SCAN-B-Lund cohort and 23 significant features. Light Blue indicates FDR-adjusted $p < 0.05$. HR, hazard ratio.

feature significantly associated with Deep-ITH low- and high-risk groups across all three datasets. For *BIRC5* and *UBE2T*, the related features were only found significant in the training cohort, and *KIF2C* was found to be significant in the training set and SCAN-B-Lund cohort, but not in the SöS-BC-1 cohort.

To visualise the distribution of heterogeneity features for patients within each risk group, grouped boxplots were generated. When examining the differences between the means of two groups, we found that in the training set (Supplementary Fig. 7), ‘MMP11_Busyness’, ‘SLC39A6_Busyness’ and ‘NAT1_Busyness’ exhibited the most significant differences, reflected by smallest p values; Besides, all three features had lower mean values in the Deep-ITH low-risk group compared with the high-risk group. For the SöS-BC-1 cohort (Fig. 5), two

complexity-related features (‘ESR1_Complexity’, ‘SFRP1_Complexity’) and PGR_JointEntropy were found to be the most significantly different, with higher mean values in the Deep-ITH low-risk group in comparison to the high-risk group; Similarly, in the SCAN-B-Lund cohort (Supplementary Fig. 8), two complexity related features (‘ESR1_Complexity’, ‘SFRP1_Complexity’) ranked highest; both had higher mean values in the Deep-ITH low-risk group; whereas the third most significant different feature ‘PHGDH_Correlation’, in contrast, had a higher mean value in the high-risk group.

Furthermore, to study whether underlying intra-tumoural gene expression heterogeneity can be inferred from morphological changes that are observable through visual inspection, we compared WSIs as well as spatial gene expression patterns for the three genes whose

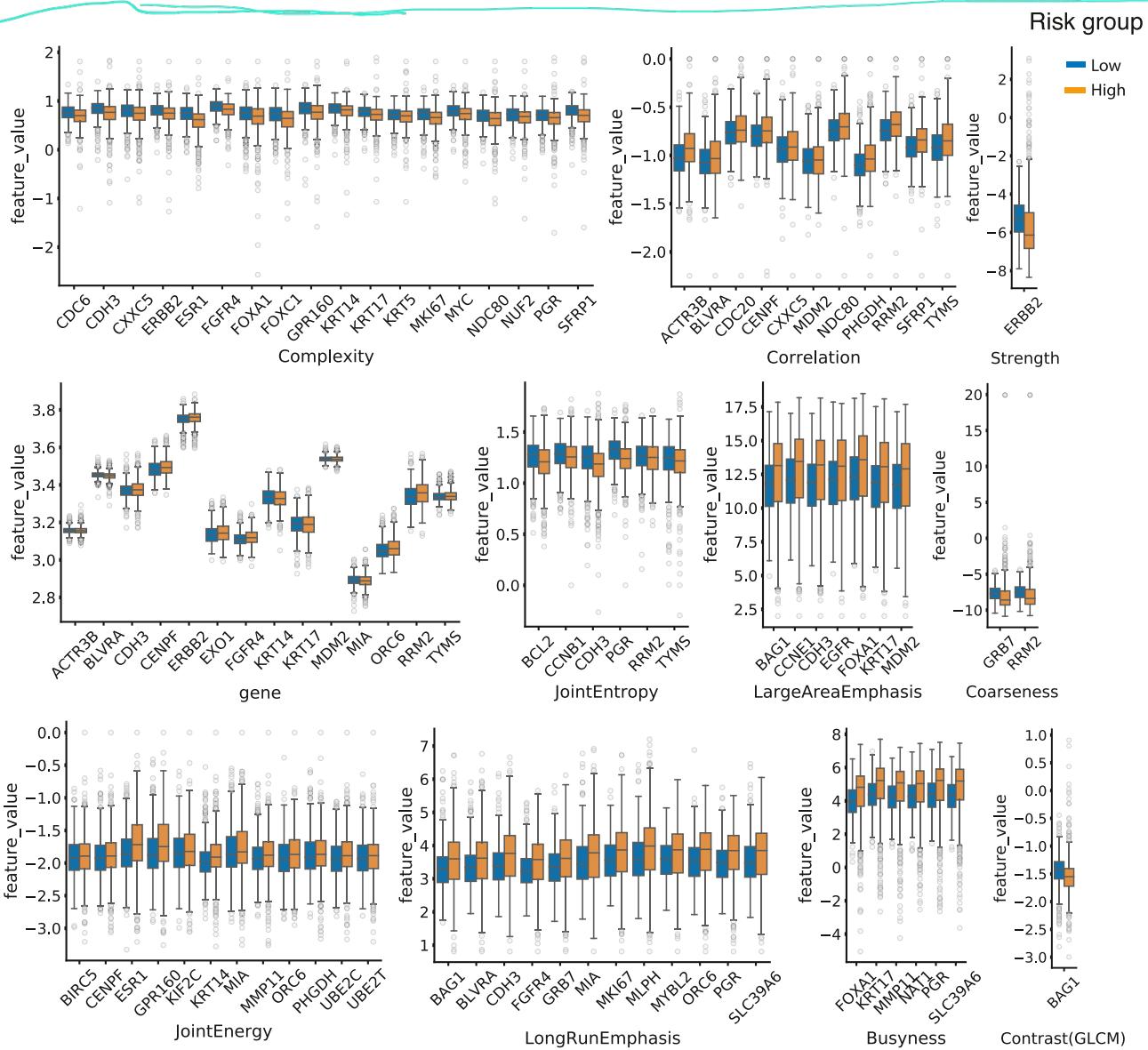


Fig. 5. Visualisation of ITH related features in the external SÖS-BC-1 cohort. Grouped boxplots showed the distributions of each selected feature between Deep-ITH low- and high-risk groups. ITH, intra-tumour heterogeneity.

expression-related texture features (predicted slide level expression of *ORC6*, *MKI67*_LongRunEmphasis and *SLC39A6*_Busyness) had the highest coefficients in the regularised Cox proportional hazard model (Fig. 6). Scrutinising the WSIs or spatial expression heatmaps by eye captured limited information to distinguish prognostic differences between two risk groups, the observation further highlighted the potential utility for a model-based risk stratification strategy.

4. Discussion

Quantification of ITH provides a new type of tumour characteristics with prognostic value, with the potential to further improve prognostic patient stratification that can contribute to guide treatment decisions.

In this study, we developed a novel deep learning-based method to measure ITH from routine HE-stained digital histopathology WSIs. With this method, we implemented a prognostic Deep-ITH score and applied it to stratify breast cancer patients in Deep-ITH low- and high-risk groups. We then evaluated the prognostic (risk of recurrence) performance by survival analysis.

The model was trained with 931 patients and validated in two external cohorts with 1358 and 1262 patients respectively that were completely independent from the training data used to develop the model. In total, 550 texture related, and 50 gene expression-related, features were extracted for analysis, a subset of 90 features were included in the prognostic model based on intrinsic variable selection in the regularised time-to-event model. Our results demonstrate that ITH is

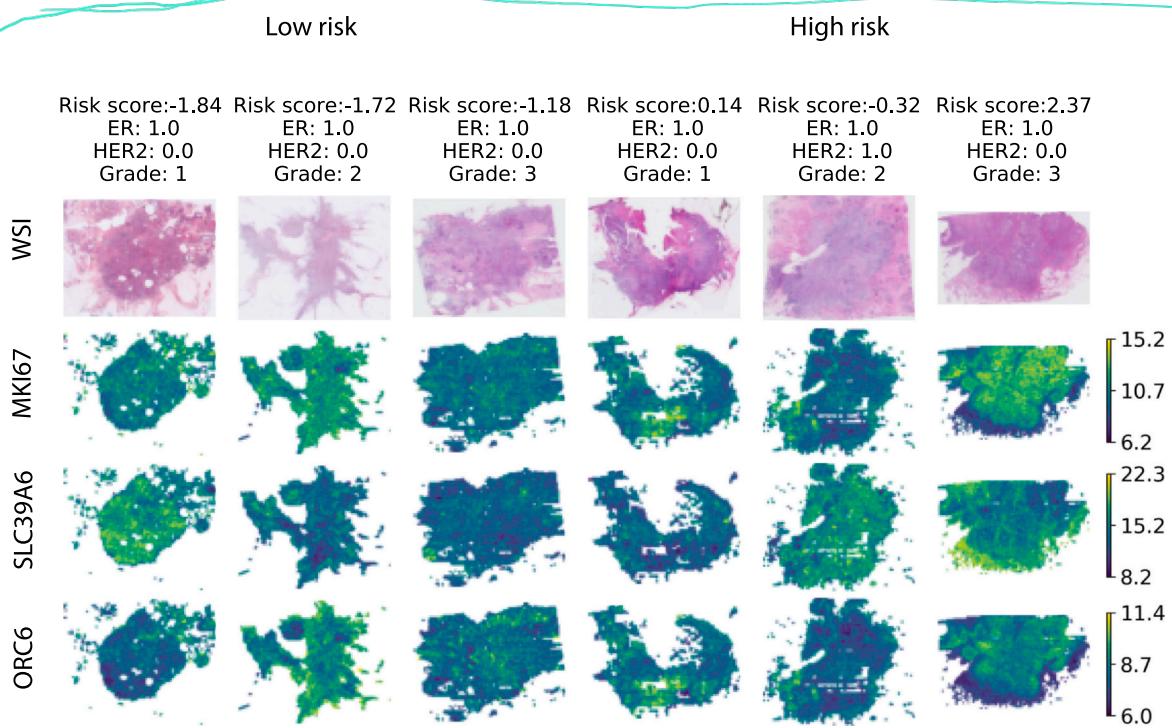


Fig. 6. Visualisation of predicted spatial gene expressions. Predicted spatial gene expressions for three genes (*MKI67*, *SLC39A6*, and *ORC6*) were displayed as heatmaps for six randomly chosen whole slide images (WSIs) from the SöS-BC-1 cohort. The genes exhibited the highest estimated coefficients according to the model. No obvious differences between the two risk groups could be observed by comparing the visual appearance of WSIs or the spatial expression heatmaps by eye.

associated with patient RFS, and thus can be leveraged to identify patients with poor prognosis. The Deep-ITH model stratified all patients in the training cohort into two groups with distinct prognosis, and the differences in RFS remained significant in subgroup analysis of patients with ER+, HER2-, grade 3 tumours and in tumours with size larger than 20 mm.

The findings were subsequently validated in the SöS-BC-1 cohort, with an HR estimate of 1.84 when analysing all patients, and HRs of 1.97 and 1.9 in ER+ and HER2- subgroups, respectively. In the other validation cohort (SCAN-B-Lund), the risk stratification was only significant in the subgroup of patients with grade 2 tumours, as well as in the subgroup of patients with small size tumours. Grade 2 accounts for approximately 50% of all breast cancer cases, and various studies have found high intra- or inter-observer variabilities in tumour grading for this subgroup, resulting in an immediate risk of recurrence and uninformative indication on treatment selection [28,38,52]. The nature of heterogeneity for grade 2 tumours indicates that patients may benefit from a more precise stratification, and our results also verified that, using the degree of ITH, this group can be further dichotomised into two risk levels with distinct differences in recurrence.

The difference in validation results between the two external validation cohorts could be due to some differences in patient demographics of each cohort. The SCAN-B study excluded patients with poor RNA quality

or too limited amount of available RNA, which could lead to a bias against smaller tumours. The original study [29] found that the distribution of important clinical variables did not differ compared with the background population after the exclusion. However, here we also exclude patients who received neoadjuvant therapy, and we observed a significant difference in lymph node status between two test cohorts. Additionally, systematic differences in the histopathological assessment of critical prognostic factors such as grade, ER and HER2 status, across Swedish healthcare regions may serve as another source of variability in the outcomes [53]. Besides, variability in validation performance could also be attributed to differences in clinical management.

Detailed mechanistic interpretation is beyond the scope of the present study. However, we note that among the most prominent genes, *CDH3* had five related features being selected in the final model; the second frequently picked genes were *KTH17*, *RRM2*, and *PGR*, each having four included features. *CDH3* and *KRT17* are well characterised basal markers whose expressions increased in basal-like breast cancer [54,55], besides, the expression of *CDH3* has been shown to be significantly associated with poor prognosis for both OS and disease-free survival in meta-analysis [56]. The prognostic significance was also reported for *RRM2*, with a higher level indicating inferior survival outcome [57]. *RRM2* is a rate-limiting enzyme for the process of DNA synthesis and repair, and plays an active role in

many critical cellular events. In breast cancer, suppressing *RRM2* expression has resulted in beneficial effects such as reversing the acquired drug resistance, reducing cell motility and tumour growth [58,59]. *PGR*, on the other hand, was one of the strongest prognostic genes whose expression correlated to a better survival [60].

When examining the identified features, the busyness-related pattern from six genes were selected by the model, including a basal-like related gene (*KRT17*), proliferation-related gene (*NAT1*), oestrogen-regulated genes (*FOXA1*, *PGR*, *SLC39A6*) [61,62] and *MMP11*. Overexpression of *MMP11* leads to breakdown of cell basal membranes and extracellular matrices that are important in cancer development and metastasis. Particularly, a heterogeneous expression pattern of *MMP11* was reported to be associated with breast cancer metastasis [63]. The identified busyness of aforementioned genes indicated rapid changes in their expression levels spatially within a tumour, suggesting that tumours with a highly variable gene expression pattern are associated with higher risks of recurrence, reflected by the marginal HR of corresponding features (Supplementary Table 4). As a consequence, higher mean values of busyness levels can also be observed in Deep-ITH high-risk groups for all six genes (Fig. 5).

Aside from busyness-related features, three long-run emphasis (*ORC6*, *CDH3*, *MLPH*) and one large-run emphasis (*BAG1*) related feature were associated with poor prognosis. As the feature depicts elongate patterns where pixel intensities remain unchanged, a higher value in long-run emphasis suggests a more homogeneous tumour expression pattern. While previous studies showed that an increased expression level of these genes are associated with either aggressive breast cancers [64,65] or better prognosis [66,67], our results suggest that their spatial expression pattern might also have a direct association with patient outcomes.

Interestingly, ‘*ESR1*_Complexity’ and ‘*SFRP1*_Complexity’, whose values depict the magnitude of heterogeneous gene expression within a tumour and are most significantly distinctive in both two external cohorts, exhibited protective effects and higher mean values in lower risk groups, suggesting a less heterogeneous expression pattern in Deep-ITH high-risk groups. *ESR1* encodes ER protein whose expression level influences the efficacy of endocrine therapy. A common hypothesis states that the treatment resistance is due to a high ITH in ER protein expression, but the underlying mechanisms vary at different molecular levels and involve the change in expression of abundant ER-dependent transcripts [68,69]. Evidence also showed that the expression level of *ESR1* individually predicted tamoxifen resistance [70]. In our study, *ESR1* expression was not selected as a prognostic feature by the model, and our results suggest that the complexity of *ESR1* expression pattern has a slight protective effect towards patients prognosis. Similarly, *SFRP1* is a tumour

suppressor that regulates the Wnt signalling, but with a controversial prognostic role in breast cancer patients [71–73]. In our current study, a more heterogeneous expression pattern of *SFRP1* is associated with a favourable survival outcome.

While other molecular profiling methods have emerged which can be used to characterise gene expression heterogeneity, challenges remain which limit their clinical utility. Although SCS techniques are under rapid development, they are still costly and do not retain spatial information. ST technique provides spatial information, but is limited in resolution and throughput and remains highly expensive. In contrast, the model-based approach we propose does not require any additional lab-based assay, is highly scalable and cost-effective. Texture features derived from radiology image data have previously been reported to be associated with histological type, intrinsic subtype, routine IHC biomarkers and cancer prognosis [74–76]. However, radiology images are severely limited in resolution, and do not convey any information on the tissue- and cell-morphological level. In this study, we combine some ideas from radiology image analysis, especially with respect to texture metrics, with detailed information available in high-resolution histopathology images. HE slides has the capacity in preserving tumour information on both structural and cellular level, thus, allows a broadened utility of texture features from separating tissue classes [77,78], or cancer subtypes [79] to an in-depth modelling of cellular interactions for the study of lymphocyte infiltration [80] or tumour microenvironment [81]. The findings are in agreement with what we have validated in our current study, that the ITH can be quantified with a selection of texture features.

We have demonstrated that our proposed methodology stratifies patients with an integrated risk score that provides independent prognostic information. The innate capacity of feature selection as well as penalisation of correlated features through elastic net regularisation ensured a principled feature selection and multivariate modelling of all measured features. Importantly, our proposed method and model was developed and validated in relatively large studies, and the validation was performed in fully independent data to demonstrate generalisability.

There is a potential to improve the approach from several aspects. For instance, preserving features with known clinical importance, and adding interaction terms between features could contribute to a better reflection of the intricate oestrogen-related transcriptional changes. Furthermore, fully deep-learning-based approaches that capture relevant ITH features and is optimised through end-to-end training has the potential to provide improved performance in the future. We note that the present study focused on the PAM50 gene panel, which has a known relevance for breast cancer patient stratification. However, the PAM50 gene set

does not exploit information that might be present in a larger gene set, this is something that could be of interest to study in the future.

The study has some limitations. Firstly, we used RFS as the primary end-point, which combines disease progression and OS events [28,38]. Although a disease-specific endpoint would have been desirable, the specific cause of death was not available to us, and therefore an assumption was made that the OS events were disease-related, which is likely to be true in the vast majority of cases. It would be of interest to compare prognostic performance in other survival end-points in future studies. Secondly, our validation was performed through retrospective study materials. Although prospective studies would offer higher degree of evidence, in breast cancer, relatively long follow-up time, ideally more than 10 years, would be required, hence, this is beyond the scope of the present study and will have to be considered in future studies. Thirdly, additional subgroup analysis, such as by treatment types could be pursued, as the prognostic effects of many risk factors can vary depending on hormonal therapy and chemotherapy. However, due to a limited number of events (statistical power), this could not be pursued in the current study. Lastly, based on the generalisability of the method and the associated features in two independent study materials from different hospital sites, we see no indication that the heterogeneity features should be less robust to possible domain shifts compared to end-to-end trained models. However, additional studies would be required to assess the impact of specific factors that might influence the generalisability of the proposed methodology, such as tissue handling, staining or digital image acquisition.

In conclusion, deep learning models that enable prediction of spatial gene expression directly from routine histopathology WSI provide a cost-efficient alternative to single-cell sequencing or ST profiling to measure and quantify ITH—a phenomenon that relates to patient prognosis and the development of treatment resistance. Here we have demonstrated how a deep learning-based approach can be applied to extract breast cancer ITH information with independent prognostic value.

Financial support

This project was supported by funding from the Swedish Research Council under the frame of ERA PerMed (ERAPERMED2019-224 - ABCAP), Swedish Research Council, VINNOVA (SwAIPP project), Swedish Cancer Society, Karolinska Institutet (Cancer Research KI; StratCan), MedTechLabs, Swedish e-science Research Centre (SeRC) - eCPC, Stockholm Region, Stockholm Cancer Society and Swedish Breast Cancer Association.

CRediT authorship contribution statement

Yinxi Wang: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation,

Visualization, Writing – original draft, Writing – review & editing. **Maya Alsheh Ali:** Investigation. **Johan Vallon-Christersson:** Resources, Reviewing the manuscript. **Keith Humphreys:** Supervision. **Johan Hartman:** Resources, Supervision, Writing – original draft, Writing – review & editing. **Mattias Rantalaainen:** Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: J.H. has obtained speaker's honoraria or advisory board remunerations from Roche, Novartis, AstraZeneca, Eli Lilly and MSD. J.H. has received institutional research grants from Cepheid, Roche and Novartis. M.R. and J.H. are shareholders of Stratipath AB. All remaining authors have declared no conflicts of interest.

Acknowledgements

The authors would like to acknowledge patients, clinicians, and hospital staff participating in the SCAN-B study, the staff at the central SCAN-B laboratory at Division of Oncology, Lund University, the Swedish National Breast Cancer Quality Registry (NKBC), Regional Cancer Center South and the South Swedish Breast Cancer Group (SSBCG).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ejca.2023.112953.

References

- [1] Li Z, Seehawer M, Polyak K. Untangling the web of intratumour heterogeneity. *Nat Cell Biol* 2022;24:1192–201.
- [2] Joseph C, Papadaki A, Althobiti M, Alsaleem M, Aleskandarany MA, Rakha EA. Breast cancer intratumour heterogeneity: current status and clinical implications. *Histopathology* 2018;73:717–31.
- [3] Morris LGT, Riaz N, Desrichard A, Şenbabaoğlu Y, Hakimi AA, Makarov V, et al. Pan-cancer analysis of intratumor heterogeneity as a prognostic determinant of survival. *Oncotarget* 2016;7:10051–63.
- [4] Greaves M, Maley CC. Clonal evolution in cancer. *Nature* 2012;481:306–13.
- [5] Polyak K, Shipitsin M, Campbell-Marrotta LL, Bloushtain-Qimron N, Park SY. Breast tumor heterogeneity: causes and consequences. *Breast Cancer Res* 2009;11:1. <https://doi.org/10.1186/bcr2279>.
- [6] Brouckaert O, Paridaens R, Floris G, Rakha E, Osborne K, Neven P. A critical review why assessment of steroid hormone receptors in breast cancer should be quantitative. *Ann Oncol* 2013;24:47–53.

- [7] Lindström LS, Yau C, Czene K, Thompson CK, Hoadley KA, Van't Veer LJ, et al. Intratumor heterogeneity of the estrogen receptor and the long-term risk of fatal breast cancer. *J Natl Cancer Inst* 2018;110:726–33.
- [8] Filho OM, Viale G, Stein S, Trippa L, Yardley DA, Mayer IA, et al. Impact of HER2 heterogeneity on treatment response of early-stage HER2-positive breast cancer: phase II neoadjuvant clinical trial of T-DM1 combined with pertuzumab. *Cancer Discov* 2021;11:2474–87. <https://doi.org/10.1158/2159-8290.cd-20-1557>.
- [9] Jamal-Hanjani M, Hackshaw A, Ngai Y, Shaw J, Dive C, Quezada S, et al. Tracking genomic cancer evolution for precision medicine: the lung TRACERx study. *PLoS Biol* 2014;12:e1001906.
- [10] Zhang XC, Xu C, Mitchell RM, Zhang B, Zhao D, Li Y, et al. Tumor evolution and intratumor heterogeneity of an oropharyngeal squamous cell carcinoma revealed by whole-genome sequencing. *Neoplasia* 2013;15:1371–8.
- [11] Casasent AK, Schalek A, Gao R, Sei E, Long A, Pangburn W, et al. Multiclonal invasion in breast tumors identified by topographic single cell sequencing. *Cell* 2018;172:205–217.e12.
- [12] Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, et al. Tumour evolution inferred by single-cell sequencing. *Nature* 2011;472:90–4.
- [13] Marx V. Method of the Year: spatially resolved transcriptomics. *Nat Methods* 2021;18:9–14.
- [14] Berglund E, Maaskola J, Schultz N, Friedrich S, Marklund M, Bergenstråhl J, et al. Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. *Nat Commun* 2018;9:2419.
- [15] Moncada R, Barkley D, Wagner F, Chiodin M, Devlin JC, Baron M, et al. Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat Biotechnol* 2020;38: 333–42. <https://doi.org/10.1038/s41587-019-0392-8>.
- [16] Fuchs TJ, Buhmann JM. Computational pathology: challenges and promises for tissue analysis. *Comput Med Imaging Graph* 2011;35:515–30.
- [17] Acs B, Rantala M, Hartman J. Artificial intelligence as the next step towards precision pathology. *J Intern Med* 2020;288:62–81.
- [18] Schmauch B, Romagnoni A, Pronier E, Saillard C, Maillé P, Calderaro J, et al. A deep learning model to predict RNA-Seq expression of tumours from whole slide images. *Nat Commun* 2020;11:3877.
- [19] Fu Y, Jung AW, Torne RV, Gonzalez S, Vöhringer H, Shmatko A, et al. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nat Cancer* 2020;1:800–10.
- [20] Wang Y, Kartasalo K, Weitz P, Ács B, Valkonen M, Larsson C, et al. Predicting molecular phenotypes from histopathology images: a transcriptome-wide expression–morphology analysis in breast cancer. *Cancer Res* 2021;81:5115–26. <https://doi.org/10.1158/0008-5472.can-21-0482>.
- [21] He B, Bergenstråhl L, Stenbeck L, Abid A, Andersson A, Borg Å, et al. Integrating spatial gene expression and breast tumour morphology via deep learning. *Nat Biomed Eng* 2020;4:827–34.
- [22] Weitz P., Wang Y., Kartasalo K., Egevad L., Lindberg J., Grönberg H., et al. Transcriptome-wide prediction of prostate cancer gene expression from histopathology images using co-expression based convolutional neural networks 2021.
- [23] Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med* 2018;24:1559–67.
- [24] Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology. *Nat Rev Clin Oncol* 2019;16:703–15.
- [25] Shmatko A, Ghaffari Laleh N, Gerstung M, Kather JN. Artificial intelligence in histopathology: enhancing cancer research and clinical oncology. *Nat Cancer* 2022;3:1026–38.
- [26] AbdulJabbar K, Raza SEA, Rosenthal R, Jamal-Hanjani M, Veeriah S, Akarca A, et al. Geospatial immune variability illuminates differential evolution of lung adenocarcinoma. *Nat Med* 2020;26:1054–62.
- [27] Parker JS, Mullins M, Cheang MCU, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* 2009;27:1160–7.
- [28] Wang Y, Acs B, Robertson S, Liu B, Solorzano L, Wählby C, et al. Improved breast cancer histological grading using deep learning. *Ann Oncol* 2022;33:89–98. <https://doi.org/10.1016/j.annonc.2021.09.007>.
- [29] Vallon-Christersson J, Häkinen J, Hegardt C, Saal LH, Larsson C, Ehinger A, et al. Cross comparison and prognostic assessment of breast cancer multigene signatures in a large population-based contemporary clinical series. *Sci Rep* 2019;9:12184. <https://doi.org/10.1038/s41598-019-48570-x>.
- [30] Wang Y, Acs B, Robertson S, Liu B, Solorzano L, Wählby C, et al. Improved breast cancer histological grading using deep learning. *Ann Oncol* 2022;33:89–98.
- [31] Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature* 2000;406:747–52. <https://doi.org/10.1038/35021093>.
- [32] Woolf DK, Li SP, Detre S, Liu A, Gogbashian A, Simcock IC, et al. Assessment of the spatial heterogeneity of breast cancers: associations between computed tomography and immunohistochemistry. *Biomark Cancer* 2019;11:1179299×19851513.
- [33] Haralick RM, Shanmugam K, Dinstein I 'hak. Textural features for image classification. *IEEE Trans Syst Man Cybern* 1973;SMC-3:610–21. <https://doi.org/10.1109/tsmc.1973.4309314>.
- [34] Thibault G, Fertil B, Navarro C, Pereira S, Cau P, Levy N, et al. Shape and texture indexes application to cell nuclei classification. *Int J Pattern Recognit Artif Intell* 2013;27:1357002. <https://doi.org/10.1142/s0218001413570024>.
- [35] Galloway MM. Texture analysis using gray level run lengths. *Comput Graph Image Process* 1975;4:172–9. [https://doi.org/10.1016/s0146-664x\(75\)80008-6](https://doi.org/10.1016/s0146-664x(75)80008-6).
- [36] Amadasun M, King R. Textural features corresponding to textual properties. *IEEE Trans Syst Man Cybern* 1989;19:1264–74. <https://doi.org/10.1109/21.44046>.
- [37] Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Ser B (Stat Methodol)* 2005;67:301–20. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>.
- [38] Wang M, Klevebring D, Lindberg J, Czene K, Grönberg H, Rantala M. Determining breast cancer histological grade from RNA-sequencing data. *Breast Cancer Res* 2016;18:1–13.
- [39] Early Breast Cancer Trialists' Collaborative Group (EBCTCG), Darby S, McGale P, Correa C, Taylor C, Arriagada R, et al. Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: meta-analysis of individual patient data for 10,801 women in 17 randomised trials. *Lancet* 2011;378:1707–16.
- [40] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B (Methodol)* 1995;57:289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
- [41] Sokal RR, James Rohlf F. The comparison of dendograms by objective methods. *TAXON* 1962;11:33–40. <https://doi.org/10.2307/1217208>.
- [42] OpenSlide: a vendor-neutral software foundation for digital pathology. *J Pathol Inf* 2013;4:27.
- [43] Abadi M., Agarwal A., Barham P., Brevdo E., Chen Z., Citro C., et al. TensorFlow:large-scale machine learning on heterogeneous systems 2015.
- [44] van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode

- the radiographic phenotype. *Cancer Res* 2017;77:e104–7. <https://doi.org/10.1158/0008-5472.can-17-0339>.
- [45] Pölsterl S. scikit-survival: a library for time-to-event analysis built on top of scikit-learn. *J Mach Learn Res* 2020;21(212):1–6.
- [46] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *JMLR* 2011;12:2825–30.
- [47] Knight V, Palmer G. Applied mathematics with open-source software: operational research problems with Python and R. CRC Press; 2022.
- [48] Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with Python. In: Proceedings of the Python in science conference; 2010. <https://doi.org/10.25080/majora-92bf1922-011>.
- [49] Hunter JD. Matplotlib: a 2D graphics environment. *Comput Sci Eng* 2007;9:90–5. <https://doi.org/10.1109/mcse.2007.55>.
- [50] Waskom M. seaborn: statistical data visualization. *J Open Source Softw* 2021;6:3021. <https://doi.org/10.21105/joss.03021>.
- [51] Wang Y, Kartasalo K, Weitz P, Ács B, Valkonen M, Larsson C, et al. Predicting molecular phenotypes from histopathology images: a transcriptome-wide expression-morphology analysis in breast cancer. *Cancer Res* 2021;81:5115–26.
- [52] Ács B, Fredriksson I, Rönnlund C, Hagerling C, Ehinger A, Kovács A, et al. Variability in breast cancer biomarker assessment and the effect on oncological treatment decisions: a nationwide 5-year population-based study. *Cancers* 2021;13:1166. <https://doi.org/10.3390/cancers13051166>.
- [53] Ács B, Fredriksson I, Rönnlund C, Hagerling C, Ehinger A, Kovács A, et al. Variability in breast cancer biomarker assessment and the effect on oncological treatment decisions: a nationwide 5-year population-based study. *Cancers* 2021;13: 1166. <https://doi.org/10.3390/cancers13051166>.
- [54] Gorski JJ, James CR, Quinn JE, Stewart GE, Staunton KC, Buckley NE, et al. BRCA1 transcriptionally regulates genes associated with the basal-like phenotype in breast cancer. *Breast Cancer Res Treat* 2010;122:721–31.
- [55] Palacios J, Honrado E, Osorio A, Cazorla A, Sarrió D, Barroso A, et al. Phenotypic characterization of BRCA1 and BRCA2 tumors based in a tissue microarray study with 37 immunohistochemical markers. *Breast Cancer Res Treat* 2005;90:5–14.
- [56] Sridhar S, Rajesh C, Jishnu PV, Jayaram P, Kabekkodu SP. Increased expression of P-cadherin is an indicator of poor prognosis in breast cancer: a systematic review and meta-analysis. *Breast Cancer Res Treat* 2020;179:301–13.
- [57] Chen W-X, Yang L-G, Xu L-Y, Cheng L, Qian Q, Sun L, et al. Bioinformatics analysis revealing prognostic significance of RRM2 gene in breast cancer. *Biosci Rep* 2019;39:BSR20182062. <https://doi.org/10.1042/BSR20182062>.
- [58] Putluri N, Maity S, Kommagani R, Creighton CJ, Putluri V, Chen F, et al. Pathway-centric integrative analysis identifies RRM2 as a prognostic marker in breast cancer associated with poor survival and tamoxifen resistance. *Neoplasia* 2014;16: 390–402.
- [59] Shah KN, Wilson EA, Malla R, Elford HL, Faridi JS. Targeting ribonucleotide reductase M2 and NF-κB activation with didox to circumvent tamoxifen resistance in breast cancer. *Mol Cancer Ther* 2015;14:2411–21.
- [60] Mihály Z, Kormos M, Lánczyk A, Dank M, Budczies J, Szász MA, et al. A meta-analysis of gene expression-based biomarkers predicting outcome after tamoxifen treatment in breast cancer. *Breast Cancer Res Treat* 2013;140:219–32.
- [61] Schettini F, Chic N, Brasó-Maristany F, Paré L, Pascual T, Conte B, et al. Clinical, pathological, and PAM50 gene expression features of HER2-low breast cancer. *NPJ Breast Cancer* 2021;7:1.
- [62] Althobiti M, El-Sharawy KA, Joseph C, Aleskandarany M, Toss MS, Green AR, et al. Oestrogen-regulated protein SLC39A6: a biomarker of good prognosis in luminal breast cancer. *Breast Cancer Res Treat* 2021;189:621–30.
- [63] González de Vega R, Clases D, Fernández-Sánchez ML, Eiró N, González LO, Vizoso FJ, et al. MMP-11 as a biomarker for metastatic breast cancer by immunohistochemical-assisted imaging mass spectrometry. *Anal Bioanal Chem* 2019;411: 639–46.
- [64] Palacios J, Benito N, Pizarro A, Suárez A, Espada J, Cano A, et al. Anomalous expression of P-cadherin in breast carcinoma. Correlation with E-cadherin expression and pathological features. *Am J Pathol* 1995;146:605–12.
- [65] Bao Y, Wang L, Shi L, Yun F, Liu X, Chen Y, et al. Transcriptome profiling revealed multiple genes and ECM-receptor interaction pathways that may be associated with breast cancer. *Cell Mol Biol Lett* 2019;24:1–20.
- [66] Thakkar A, Raj H, Ravishankar, Muthuvelan B, Balakrishnan A, Padigaru M. High expression of three-gene signature improves prediction of relapse-free survival in estrogen receptor-positive and node-positive breast tumors. *Biomark Insights* 2015;10:103–12.
- [67] Papadakis ES, Reeves T, Robson NH, Maishman T, Packham G, Cutress RI. BAG-1 as a biomarker in early breast cancer prognosis: a systematic review with meta-analyses. *Br J Cancer* 2017;116:1585–94.
- [68] Ali S, Coombes RC. Endocrine-responsive breast cancer and strategies for combating resistance. *Nat Rev Cancer* 2002;2:101–12.
- [69] Toy W, Shen Y, Won H, Green B, Sakr RA, Will M, et al. ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. *Nat Genet* 2013;45:1439–45.
- [70] Kim C, Tang G, Pogue-Geile KL, Costantino JP, Baehner FL, Baker J, et al. Estrogen receptor (ESR1) mRNA expression and benefit from tamoxifen in the treatment and prevention of estrogen receptor-positive breast cancer. *J Clin Oncol* 2011;29:4160–7. <https://doi.org/10.1200/JCO.2010.32.9615>.
- [71] Cheng L-C, Chao Y-J, Overman MJ, Wang C-Y, Phan NN, Chen Y-L, et al. Increased expression of secreted frizzled related protein 1 (SFRP1) predicts ampullary adenocarcinoma recurrence. *Sci Rep* 2020;10:13255.
- [72] Bernemann C, Hülsewig C, Ruckert C, Schäfer S, Blümel L, Hempel G, et al. Influence of secreted frizzled receptor protein 1 (SFRP1) on neoadjuvant chemotherapy in triple negative breast cancer does not rely on WNT signaling. *Mol Cancer* 2014;13:1–12. <https://doi.org/10.1186/1476-4598-13-174>.
- [73] Klopocki E, Kristiansen G, Wild PJ, Klaman I, Castanos-Velez E, Singer G, et al. Loss of SFRP1 is associated with breast cancer progression and poor prognosis in early stage tumors. *Int J Oncol* 2004;25:641–9.
- [74] Sollini M, Cozzi L, Ninatti G, Antunovic L, Cavinato L, Chiti A, et al. PET/CT radiomics in breast cancer: mind the step. *Methods* 2021;188:122–32.
- [75] Ashraf AB, Daye D, Gavenonis S, Mies C, Feldman M, Rosen M, et al. Identification of intrinsic imaging phenotypes for breast cancer tumors: preliminary associations with gene expression profiles. *Radiology* 2014;272:374–84.
- [76] Chang R-F, Chen H-H, Chang Y-C, Huang C-S, Chen J-H, Lo C-M. Quantification of breast tumor heterogeneity for ER status, HER2 status, and TN molecular subtype evaluation on DCE-MRI. *Magn Reson Imaging* 2016;34:809–19.
- [77] Kather JN, Weis C-A, Bianconi F, Melchers SM, Schad LR, Gaiser T, et al. Multi-class texture analysis in colorectal cancer histology. *Sci Rep* 2016;6:27988.
- [78] Linder N, Konsti J, Turkki R, Rahtu E, Lundin M, Nordling S, et al. Identification of tumor epithelium and stroma in tissue microarrays using texture analysis. *Diagn Pathol* 2012;7:1–11.

- [79] Valkonen M, Ruusuvuori P, Kartasalo K, Nykter M, Visakorpi T, Latonen L. Analysis of spatial heterogeneity in normal epithelium and preneoplastic alterations in mouse prostate tumor models. *Sci Rep* 2017;7:44831.
- [80] Yuan Y. Modelling the spatial heterogeneity and molecular correlates of lymphocytic infiltration in triple-negative breast cancer. *J R Soc Interface* 2015;12:20141153<https://doi.org/10.1098/rsif.2014.1153>.
- [81] Natrajan R, Sailem H, Mardakheh FK, Arias Garcia M, Tape CJ, Dowsett M, et al. Microenvironmental heterogeneity parallels breast cancer progression: a histology-genomic integration analysis. *PLoS Med* 2016;13:e1001961.