nature biomedical engineering

ARTICLES

https://doi.org/10.1038/s41551-020-0578-x

Integrating spatial gene expression and breast tumour morphology via deep learning

Bryan He[1], Ludvig Bergenstråhle[2], Linnea Stenbeck[2], Abubakar Abid[3], Alma Andersson[2], Åke Borg[4], Jonas Maaskola[2]✉, Joakim Lundeberg[2]✉ and James Zou[1,3,5,6]✉

Spatial transcriptomics allows for the measurement of RNA abundance at a high spatial resolution, making it possible to systematically link the morphology of cellular neighbourhoods and spatially localized gene expression. Here, we report the development of a deep learning algorithm for the prediction of local gene expression from haematoxylin-and-eosin-stained histopathology images using a new dataset of 30,612 spatially resolved gene expression data matched to histopathology images from 23 patients with breast cancer. We identified over 100 genes, including known breast cancer biomarkers of intratumoral heterogeneity and the co-localization of tumour growth and immune activation, the expression of which can be predicted from the histopathology images at a resolution of 100 μm. We also show that the algorithm generalizes well to The Cancer Genome Atlas and to other breast cancer gene expression datasets without the need for re-training. Predicting the spatially resolved transcriptome of a tissue directly from tissue images may enable image-based screening for molecular biomarkers with spatial variation.

The spatial organization and heterogeneity of gene expression within a tissue has important biological effects on the properties of the tissue. Regular transcriptome analyses using bulk sequencing or single-cell sequencing do not capture high-resolution spatial heterogeneity. Using these techniques, the rich spatial information about gene expression is lost, and works studying heterogeneity have focused on sampling from multiple regions of tissue rather than heterogeneity within a section of tissue[1].

Recent developments in spatial transcriptomics capture spatial information by using DNA barcodes to distinguish different spots in the tissue[2]. The expression of the tens of cells in each spot can then be recovered from RNA sequencing (RNA-seq). This results in a gene expression vector being measured for hundreds of spots in each section of tissue. Spatial transcriptomics technology is complementary to fluorescence in situ hybridization[3–5] and in situ sequencing techniques[6], in that it can measure more genes albeit at a lower spatial resolution. The corresponding histopathology images of tissue stained with haematoxylin and eosin (H&E) are readily available and can be aligned to the spots. An important question is how to integrate the histology images with the spatial transcriptomics.

Machine learning is becoming a powerful approach for integrative analysis of whole-slide histology images. Older methods use hand-crafted features for extracting information and making predictions[7–9]. More recently, convolutional neural networks (CNNs)[10–13] have been used on histopathology images for predictive tasks, including tumour classification[14–16], mutation prediction[17] and classification of cancer subtypes[18,19]. Although these works have been used successfully to perform challenging tasks using histopathology images, the targets of prediction have been slide-level labels rather than fine-grained labels for small groups of cells.

Here we introduce ST-Net, a deep learning algorithm that combines spatial transcriptomics and histology images to capture high-resolution gene expression heterogeneity. ST-Net was trained on a new spatial transcriptomics dataset of 30,612 spots in 68 breast tissue sections from 23 patients with breast cancer. For held-out test patients, ST-Net can predict the spatial variation in the expression of 102 genes—including several breast cancer biomarkers—at a resolution of around 100 μm (false discovery rate (FDR) of 0.05). As an independent external test, ST-Net accurately predicts spatial expression in the 10x Genomics breast cancer data without any modification or tuning. This suggests that it robustly generalizes to new samples. ST-Net is more accurate than methods that use standard cellular features to predict expression. Moreover, it can predict substantial variation in expression within regions that are all labelled as tumour by clinicians, demonstrating that ST-Net captures intratumor heterogeneity.

Results

We provide a schematic of the workflow of ST-Net as well as summary statistics of our spatial transcriptomics data in Fig. 1. To ensure that our predictions generalized across patients, all results have been reported using leave-one-patient-out cross validation, where we iteratively trained ST-Net on 22 of the patients and made predictions on the remaining held-out patient. In addition, we normalized and log-transformed the count data to reduce confounding due to cell density and to reduce the effect of outliers (Supplementary Fig. 1).

Following the workflow shown in Fig. 1a, we used ST-Net to predict the target gene expression of each spot in the whole-slide images at a ×20 magnification. ST-Net takes a 224 × 224-pixel patch of the image centred on each spot—corresponding to around 150 × 150 μm^2—and predicts the log expression of the 250 target genes. The window used by ST-Net is slightly larger than the size of the spots (100 × 100 μm^2). Using leave-one-patient-out cross validation, we computed the root mean squared error and Pearson's correlation between the predicted and observed log expression of each

[1]Department of Computer Science, Stanford University, Stanford, CA, USA. [2]School of Biotechnology, KTH Royal Institute of Technology, Stockholm, Sweden. [3]Department of Electrical Engineering, Stanford University, Stanford, CA, USA. [4]Division of Oncology and Pathology, Lund University, Lund, Sweden. [5]Department of Biomedical Data Science, Stanford University, Stanford, CA, USA. [6]Chan–Zuckerberg Biohub, San Francisco, CA, USA. ✉e-mail: jonas.maaskola@scilifelab.se; joakim.lundeberg@scilifelab.se; jamesz@stanford.edu

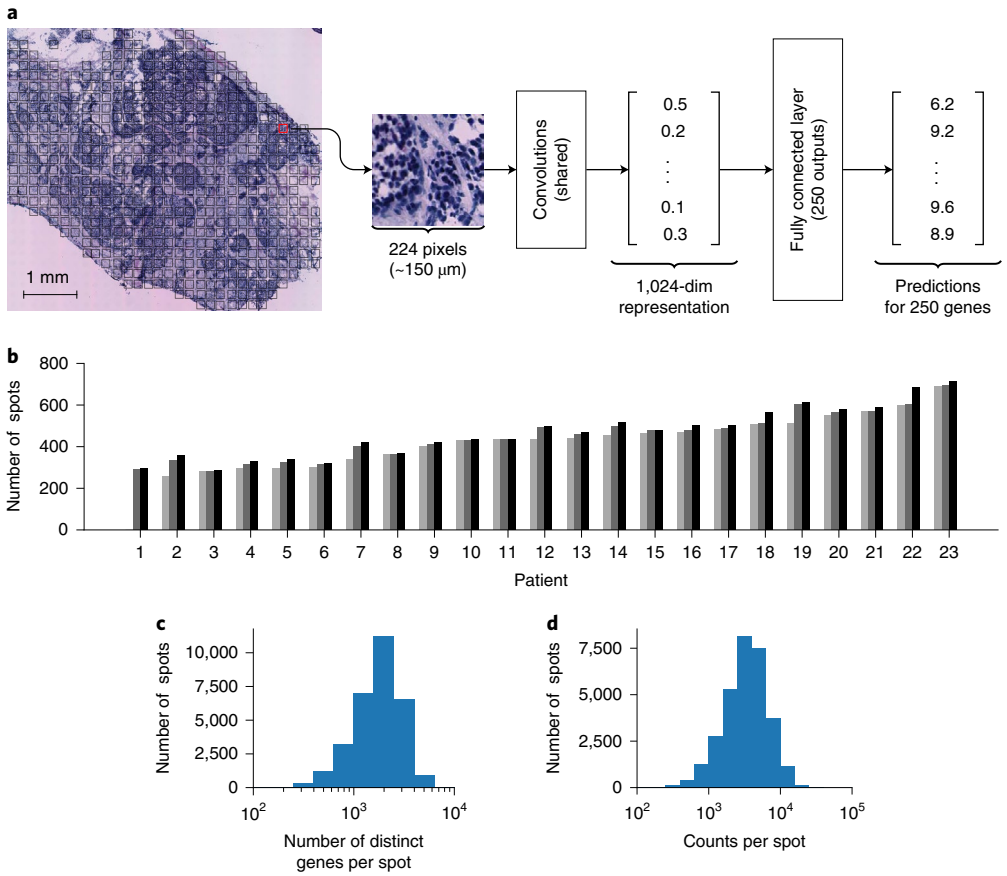

Fig. 1 | ST-Net pipeline and data. a, Prediction pipeline for a whole-slide histopathology image. We begin by taking patches of 224 × 224 pixels centred on the spatial transcriptomics spots. A shared set of 120 convolutional layers is used for all genes, and a fully connected layer with 250 outputs makes the final prediction for the log of the normalized mRNA count. **b**, Number of spots sampled in each section. The sections are grouped by patient (two sections for patient 1 and three sections for all other patients). **c**,**d**, Distribution of the number of genes with non-zero counts (**c**) and the total number of reads captured in each spot (**d**).

gene for each spot in the held-out patient. We found that the predictions have a root mean squared error of 0.31, corresponding to predicting the expressions to within an average factor of 1.4. Moreover, in 102 of the 250 genes, the predicted expressions correlated positively with the experimental measurements in 20 or more of the 23 patients, corresponding to an FDR of 0.05. This result suggests that a substantial number of gene expression levels can be predicted from the tissue morphology.

Pathway analysis using DAVID[20] showed that many of these 102 genes are involved in cancer, pharmacogenomic and immune pathways—accounting for 37, 35 and 31 genes, respectively. In contrast, only 11, 14 and 9 genes of the 148 remaining genes could be attributed to these respective pathways, indicating that the predictive power of ST-Net is enriched in genes relevant to breast cancer. The five genes that could be best predicted by ST-Net (in terms of median correlation across patients) are listed in Table 1. *GNAS*, *FASN*, *DDX5* and *XBP1* are known cancer biomarkers; *ACTG1* encodes γ-actin, which is expressed in cytoskeletons. We provide two representative example sections of tissue to visualize the spatial data and our predictions in Fig. 2a. For each section, we show the raw image, the true expression per spot—measured by sequencing the tissue—and the prediction from ST-Net.

The performance of ST-Net is affected by the inherent experimental noise in spatial transcriptomics and from the limited sequencing depth. To reduce experimental noise, we smoothed the gene expression values of each spot by averaging these with those of its adjacent neighbours. The predictions made by ST-Net demonstrated better agreement with the smoothed experimental measurements, suggesting that the model learns some amount of denoising (Table 1). The total number of patients that were correctly predicted for each gene using the smoothed data is shown in Fig. 2b. In addition, we compared and demonstrated the improved performance of ST-Net against other CNN architectures in Supplementary Table 1 using several evaluation metrics.

External validation on the 10x Genomics data. As an independent external test, we applied ST-Net to a breast cancer sample from the 10x Genomics Spatial Gene Expression dataset. This dataset consists of two sections from an invasive ductal carcinoma breast tissue sample and had deeper sequencing, resulting in a median of 17,531 counts per spot (compared with 3,426 in our original dataset). The sections were scanned at ×20 magnification, and the sequencing of this dataset included measurements for 234 of the 250 genes that ST-Net was trained to predict. The 10x Genomics data were generated independently at a different facility from our data, using a different staining protocol, and ST-Net was applied without any modification (no parameter or hyperparameter was tuned). Therefore, this constitutes a strong test for the reliability of ST-Net.

Despite the data differences, we found that 207 of the 234 genes in 10x could be predicted by ST-Net with a positive correlation. The average Pearson's correlation between the ST-Net prediction and the experimental measurement was 0.33 across all 234 genes. Moreover, the area under the receiver operating characteristics (AUROC) for predicting whether a gene is expressed at high or low levels in a spot

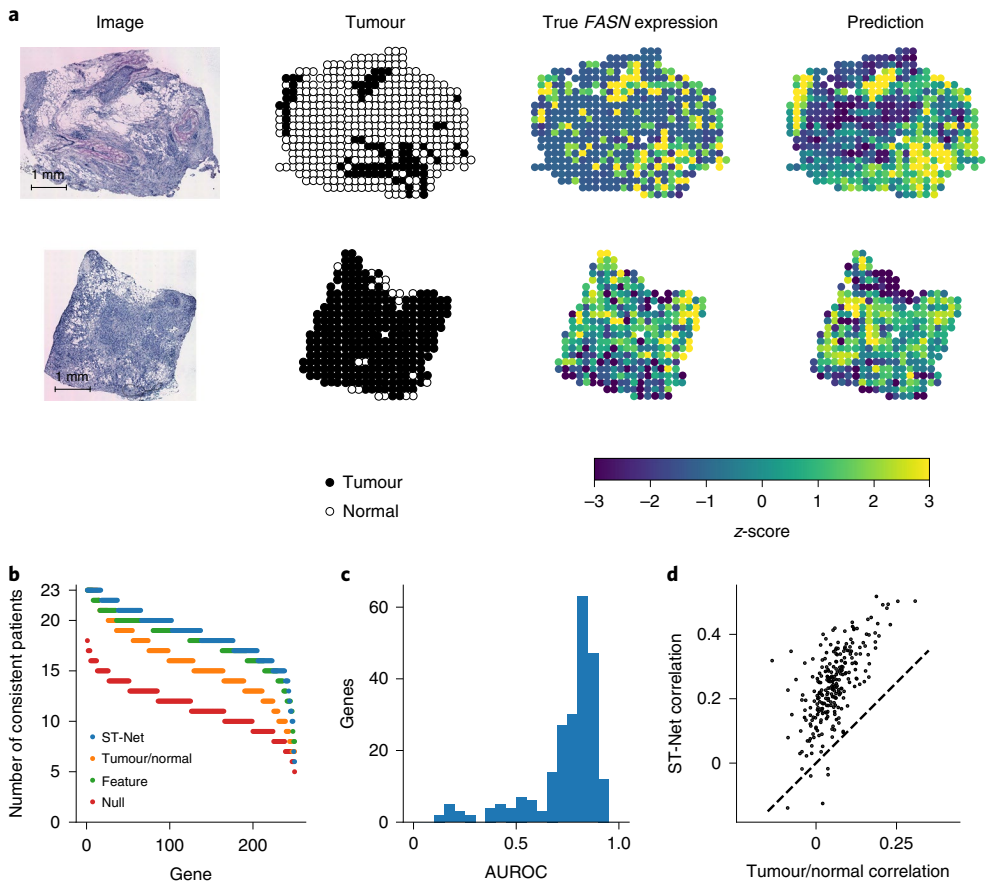

Fig. 2 | Results from the ST-Net predictions. a, Visualization of the true and predicted expression for the tumour biomarker *FASN*. Two separate patients are shown. These two patients are ranked 11 (top) and 12 (bottom) of the 23 patients in the correlation, with correlations of 0.35 and 0.31, respectively. The tumour/normal annotations made by the pathologist is also provided. **b**, The number of patients with a positive Pearson's correlation between the predicted and observed expression of each gene at a resolution of 100 µm. ST-Net is more consistent across patients compared with hand-crafted tissue features, while avoiding the need for feature engineering, and compared to using only the clinical tumour/normal annotations. The permuted null distribution curve is also shown as a baseline. **c**, Histogram of AUROC for predictions on the 10x Genomics Spatial Gene Expression test data. **d**, Comparison of the median Pearson's correlation for the predictions made by ST-Net and the predictions from the tumour/normal labels made by a pathologist. The dashed line indicates the 45-degree diagonal.

was 0.73, averaged across all of the genes. The AUROC distribution across genes is shown in Fig. 2c and the results per slide are shown in Supplementary Fig. 2. The strong performance of ST-Net on this independent validation suggests that the model can reliably generalize to new data and that its performance could improve further with deeper sequencing.

Validation and application to TCGA. We applied ST-Net to breast cancer samples of The Cancer Genome Atlas (TCGA) to further investigate the robustness and clinical applications of the method[21]. There are several notable differences between the TCGA and our breast cancer data—the TCGA samples were gathered and scanned at separate institutions, introducing technical variability to the preparation process, and its sequencing was not performed on the same tissue that was used for imaging—which makes this a substantial challenge for ST-Net. We applied ST-Net directly without any re-training to H&E images of 1,550 TCGA breast tumour samples from 1,093 patients. For each image, ST-Net predicted the spatially resolved expression for 250 genes. As TCGA only has bulk RNA-seq, we first averaged the ST-Net predictions into a pseudo-bulk expression profile for each sample. The ST-Net pseudo-bulk correlated positively with the TCGA RNA-seq for 177 of 249 genes. The correlation was significantly positive for 55 genes (FDR of 0.05), and no gene had a significant negative correlation. Moreover, the ST-Net-inferred pseudo-bulk expression had similar power to distinguish histological subtypes of breast cancer (infiltrating ductal carcinoma and infiltrating lobular carcinoma) as the real TCGA bulk RNA-seq—AUROC of 0.83 (0.78–0.88) and 0.83 (0.78–0.89), respectively—which is slightly higher than directly predicting subtype from the TCGA H&E images (AUROC of 0.81; 0.75–0.87). These analyses suggest that ST-Net can generalize to new clinical data despite the technical differences in the samples and imaging, and it is able to make reliable predictions on TCGA images.

Co-localization of tumour and immune expression. A unique strength of spatial transcriptomics is the ability to characterize how different gene expression activities overlap, which gives insight into cell–cell and molecular interactions. We used DAVID to identify three genes that are primarily markers of tumour growth and three different genes that are primarily markers of immune activation from the 55 genes with significantly positive correlations to the TCGA bulk RNA-seq expression data. We then computed the spatial correlation in the tumour and immune expression profiles in our ST-Net data and in TCGA, where we used spatial expression for these biomarkers that we predicted from the images. In both TCGA and our new data, we found that tumour and immune genes tended

Table 1 | Prediction performance of ST-Net for the top five genes

Gene	No. of consistent patients	Median correlation	Median correlation (smooth)	10x Correlation
GNAS	23	0.34	0.49	0.43
ACTG1	22	0.33	0.50	0.47
FASN	23	0.31	0.50	0.46
DDX5	22	0.30	0.52	0.51
XBP1	21	0.29	0.43	0.54

Four evaluation metrics are shown: (1) number of patients (out of 23) where the ST-Net predictions correlate positively with the experimental measurements across all of the spots in that patient, (2) median Pearson's correlation across patients, (3) median correlation with the smoothed experimental data and (4) correlation with the external 10x Genomics test data.

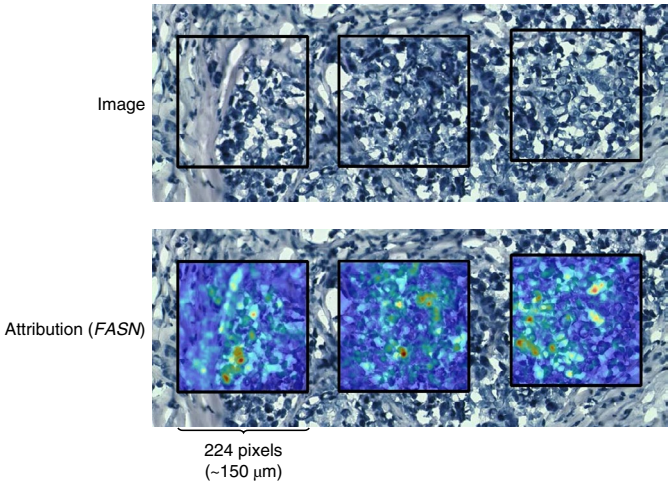

Fig. 3 | Three example patches and regions of interest identified by interpreting ST-Net. The regions highlighted in red are parts of the image that ST-Net believes to be indicative of high *FASN* expression, as captured by integrated gradient attribution. These highlighted regions capture enlarged nuclei.

to co-locate less frequently than expected by chance ($P < 0.001$; a full distribution of the correlations is shown in Supplementary Fig. 3). This result is consistent with the previous finding that 75% of patients with breast cancer have low rates of tumour-infiltrating lymphocytes[22]. ST-Net makes it possible to quantify the spatial co-location of tumour- and immune expression-activities directly from H&E images, which can be a promising biomarker.

ST-Net detects intra-tumour variation. Substantial expression variation was observed within just the tumour or just the normal tissue (Supplementary Figs. 4 and 5) using the tumour/normal annotations of the pathologist. We stratified the ST-Net predictions for tumour and normal tissue separately. ST-Net was still able to predict variation of many genes within both tumour- and normal-only tissues, suggesting that it could identify biomarkers at a potentially higher resolution than the pathology annotations. When predicting on tumour-only tissue, 63 genes were correctly predicted for 20 or more of the 23 patients (FDR of 0.05). When predicting on normal-only tissue, 27 genes were correctly predicted for 18 or more of the 21 patients with normal tissue available (FDR of 0.06). Two of the patients were annotated to have only tumour tissues and were excluded from the last analysis.

GNAS, *FASN*, *AEBP1*, *SPARC* and *BGN* were the top five genes for which ST-Net captured the highest variation in intra-tumour expression. All five genes have been previously identified to be cancer biomarkers. *AEBPT1* is a transcription repressor[23] and, consistent with this, higher *AEBP1* expression correlated with a smaller nucleus size and lower nuclei density in our H&E images ($P < 0.01$). *SPARC* encodes an extracellular matrix protein associated with macrophages[24] and *BGN* expression is known to affect the migration and morphology of tumour cells[25]. The positive predictive performance of ST-Net for these genes suggests that their biological effects are visible in the morphology of tumour tissue.

Comparison with tumour/normal predictions, hand-crafted features and cell-type composition. To provide more context for ST-Net, we compared its prediction performance with that of several other approaches. The first approach we compared with was the use of pathologist labels for each spot as either tumour or normal tissue. Note that these labels were only used for downstream analysis and validation and were not used by ST-Net during training or testing. Certain genes have different expression levels in tumour and normal tissue, so it is expected that the expression of some genes can be predicted using these labels. For each spot, this baseline uses the pathologist label to predict the expression for each gene. If the spot is a tumour, then the baseline predicts the mean expression across tumours in the other patient for that gene, and if the spot is normal, then the baseline predicts the mean expression across normal tissue in the other patient for that gene. We found that the baseline predicted 35 genes with positive correlation for at least 20 of the 23 patients and that ST-Net achieved a higher median correlation for 245 of the 250 genes (Fig. 2d). Two patients had all patches labelled as tumour, so we considered all genes to be correctly predicted for these two patients to evaluate this baseline leniently. This demonstrates that ST-Net captures more fine-grained information than tumour–normal classifications.

Next, we investigated how ST-Net compares with an approach that uses standard cellular features to predict spatial transcriptomics. We used the HistomicsTK library to extract ten commonly used features for each nucleus regarding its size (four features), colour (three features) and its distance to nearby nuclei (three features)[26]. The size features are: area, major-axis length, minor-axis length and the ratio of the major-axis to the minor-axis length. The colour features are the mean red, blue and green channels of the nucleus. The distance features are the maximum, mean and minimum distance to adjacent nuclei in the Delaunay triangulation. Patch-level features were then created from the nuclei features in the patch. For each of the features, the mean, standard deviation, skewness and kurtosis were computed along with a ten-bin histogram. The nuclei count was a particularly informative signal for determining whether a patch contained tumour or normal tissue (Supplementary Fig. 6). We used a random forest over these features to predict the levels of gene expression and found that these features correctly predicted 79 genes with a positive correlation for 20 of the 23 patients. ST-Net achieved a higher median correlation compared with this random forest for 236 of the 250 genes (Supplementary Fig. 7a). This indicates that ST-Net captures more complex signals in the images than these hand-crafted features. The addition of more hand-crafted features might further improve its performance; however, a key advantage of deep learning is that it reduces the need for hand-engineering features.

Finally, we hypothesized that changes in cell-type composition are a key driver of the variation in the transcriptome across different patches. As we did not have the ground-truth estimate of the cell-type composition in each tissue patch, we estimated the composition computationally by applying non-negative matrix factorization to the expression data[27,28]. A CNN model trained to explicitly capture cell-type composition performed reasonably well—it correctly predicted 55 genes with a positive correlation for 20 of the 23 patients. Interestingly, ST-Net still achieved a higher correlation for

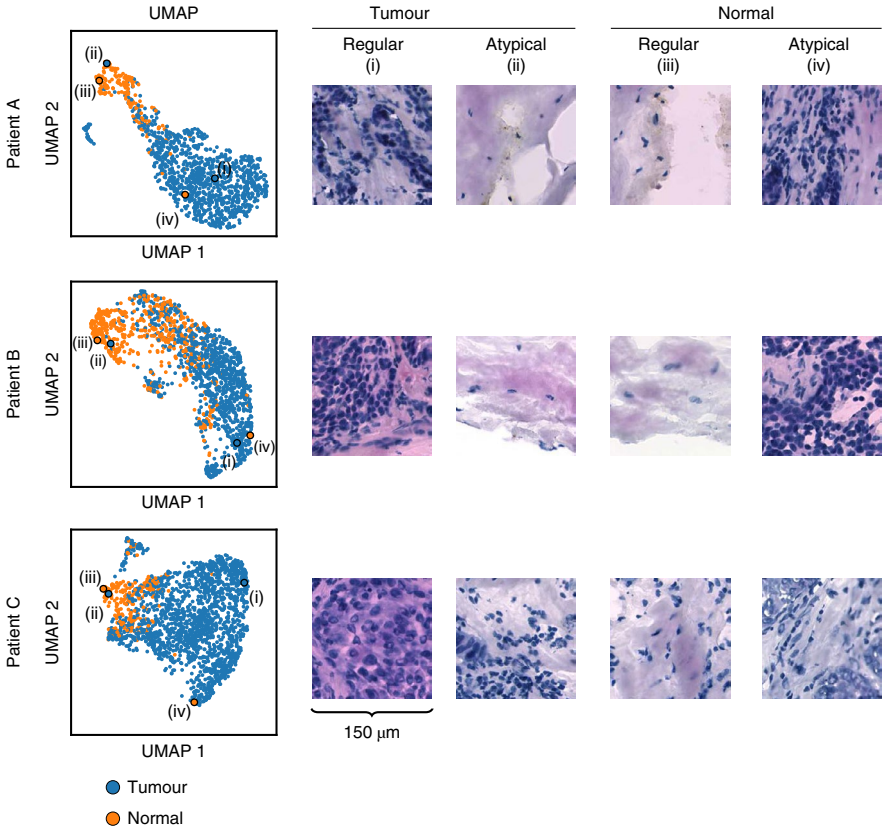

Fig. 4 | UMAP visualization of the ST-Net latent space. Patches from three patients with a representative tumour patch (i), an atypical tumour patch (ii), a representative normal patch (iii) and an atypical normal patch (iv). Each patch corresponds to a point in the final neural network layer of ST-Net.

242 of the 250 genes compared with this composition-based model (Supplementary Fig. 7b), with a mean correlation improvement of 0.03, suggesting that ST-Net captures additional information beyond cell-type composition in the tissue.

Interpretation of predictions. Although ST-Net uses a neural network, we can leverage recent deep learning interpretation methods to understand what it pays attention to in the image when making predictions. In Fig. 3, we show three representative patches and use the integrated-gradients method to identify the image regions causing ST-Net to predict that *FASN* is highly expressed in the patch[29]. These regions are highlighted in red. *FASN* is overexpressed in breast cancer and is a prognostic indicator[30], so we expect tumour-related visual features to also be associated with high *FASN* expression. This interpretation shows that ST-Net learns to associate enlarged nuclei, which are indicative of tumour, with high *FASN* expression. In addition, we considered the relationship between the top predicted genes and specific geometric features to provide more interpretable relationships between morphology and expression in Supplementary Tables 2 and 3. We found that enlarged and less-elongated nuclei generally tended to have higher levels of expression.

Latent space of ST-Net captures semantic similarity between spots. An interesting feature of ST-Net is that it provides a new way to represent and quantify similarities between histology images. We used the final artificial neuron layer of ST-Net as a latent representation of each spot, resulting in a 1,024-dimensional real-valued vector for each spot. We first used Uniform Manifold Approximation and Projection (UMAP) to visualize the latent space of the spots[31] and found that the latent space naturally separates tumour and normal spots, although ST-Net was never trained using the pathologist labels (Fig. 4). In addition, we show a representative tumour patch (a tumour patch whose latent representation is mostly surrounded by other tumour patches), an atypical tumour patch (a tumour patch whose latent representation is mostly surrounded by other normal patches), a representative normal patch and an atypical normal patch. We found that the atypical patches were mostly ambiguous patches or patches near the edge of a tumour and normal tissue (Supplementary Fig. 8), which could suggest potential mis-annotations.

We ran *k*-means clustering on the latent deep features with two clusters as a direct measure of the separation between tumour and normal tissue in the latent space. We used purity—a metric ranging from zero to one representing how close the clusters were to having a single class—to evaluate the extent of separation between the two types of tissue. We found that the clustering resulted in a mean purity of 0.89 across patients (s.d. of 0.07), indicating that each cluster in the latent space consists of either predominantly tumour regions or normal regions.

Next, we used the latent representation of the spots to identify which nearby spots were more similar to the given spot. For each spot, we considered the four adjacent spots and used the distance in the latent space between the spots to rank the similarity of the four neighbours. We used the most similar and most distinct neighbours as predictions and found that the more similar spot in the latent space was also more predictive of the gene expression (Supplementary Fig. 9a). We also found that the nearest neighbours in the latent space were often not simply adjacent spots in the physical layout (Supplementary Fig. 9b), suggesting that the semantic distance between spots was distinct from physical distance.

Discussion

ST-Net links gene expression with visual features in cell morphology. It demonstrates that the combination of spatial transcriptomics

and deep learning can predict gene expression using readily available H&E histopathology images. ST-Net generalizes to held-out patients in our dataset. Moreover, it is accurate on the independent 10x Genomics breast cancer data without any modification. We also find that ST-Net can capture gene expression heterogeneity within tumours. Although this paper focuses on the application to breast tumours, the technical framework of ST-Net can be broadly applied to other tissue types. As additional spatial expression datasets are generated and sample preparation becomes more standardized, we expect the performance of ST-Net to improve further.

In addition to predicting gene expression, ST-Net learns additional structure in the histopathology images. For example, it automatically separates tumour and normal tissue, and identifies patches whose pathology annotation seems to be incongruent with the H&E image. This encourages the use of automated methods such as ST-Net to identify subtle biological signals. We also find it promising that ST-Net can make predictions on TCGA H&E images without any modification, despite the substantial differences in the data. This opens the door to apply ST-Net to the large existing repositories of histology images and infer important molecular biomarkers such as tumour–immune interactions.

Spatial transcriptomics and single-cell RNA-seq (scRNA-seq) capture complementary insights into RNA expression. Spatial transcriptomics captures the location of transcripts, although not quite at single-cell resolution. Although it does not retain the location information of each cell, scRNA-seq measures the single-cell transcriptome with higher sequencing depth. An interesting direction of future work would be to integrate the spatial data that we analyse here with scRNA-seq from similar samples. For example, the scRNA-seq data could assist with deconvoluting cell types in the spatial data and be used to impute genes that are missed by the spatial probes.

ST-Net demonstrates that spatially resolved expression of key cancer-related genes can be directly inferred from tissue images. Although other machine-learning approaches can predict bulk genomic features from whole histology slides[17], ST-Net captures spatial variation in gene expression at a much higher resolution (at a scale of 100 µm). Our proof-of-concept study finds that the combination of new spatial techniques and deep learning can identify heterogeneity within a patient sample, allowing further study of complex interactions and variations. ST-Net can potentially enable new image-based screening of molecular biomarkers that demonstrate spatial variation, which opens an exciting new direction of clinical applications.

Methods

Spatial transcriptomics. Spatial transcriptomics is a recent technique used to capture the spatial distribution of messenger RNA sequences within tissue sections[2]. DNA-barcoded probes are added on top of flash-frozen histological tissue sections to mark the spatial position of the mRNA sequences. The mRNA sequences are then captured and used to generate sequencing libraries. The DNA barcodes allow the position of individual RNA sequences in the RNA-seq data to be recovered. Details are shown in Supplementary Fig. 10.

We studied a new dataset consisting of 23 patients with breast cancer. For each patient, we obtained three microscope images of slides of H&E-stained tissue and the corresponding spatial transcriptomics data. An example image is shown in Fig. 1a. The patches shown were used by the neural network as inputs and roughly correspond to the receptive field of the spots. For a single section of tissue, spatial transcriptomics measures RNA expression in spots with a diameter of 100 µm arranged in a grid with a centre-to-centre distance of 200 µm. The number of spots in each replicate ranged from 256 to 712, and the distribution across patients is shown in Fig. 1b. Several thousand mRNA sequences were captured in each spot (distribution shown in Fig. 1d). A total of 26,949 distinct mRNA species were observed across the dataset, so each spot was represented as a 26,949-dimensional vector of non-negative integers, where each element represents the number of times a gene was counted.

Sample processing. This study complied with all of the relevant ethical regulations for experiments involving human tissue samples, with informed patient consent and under the ethical permit granted by the Lund University Hospital. Breast cancer biopsies were collected and handled as described previously[2]. The samples were snap frozen and embedded in OCT compound. Cryosections were cut at a thickness of 16 µm. The sections were prepared using standard H&E staining and imaged. To annotate the histological sections from breast cancer, we scored: invasive cancer, cancer in situ, fatty tissue, fibrous tissue and immune cells. The RNA-seq of the barcoded libraries and subsequent analysis from tissue sections were performed using standardized protocols[32].

ST-Net design and training. The introduction of the ImageNet Large-Scale Visual Recognition Challenge[33,34] resulted in the development of numerous CNN architectures[10–13]. CNNs have been used in a variety of medical-imaging applications, including dermatology[35] and retinal disease[36]. CNNs have also been applied to whole-slide histopathology images to identify patches containing tumours[14–16], predict mutations[17] and determine subtypes of cancer[19].

We used CNNs to detect fine-grained spatial heterogeneity in gene expression within the tissue. The whole-slide images are very large (almost $10{,}000 \times 10{,}000$ pixels) and cannot be directly used as inputs to the CNN, so we began by extracting patches of 224×224 pixels from the whole-slide images centred on the spatial transcriptomics spots. The 224×224-pixel patches are the standard input size for the CNNs. In our experiments, we used a DenseNet-121 model with pre-trained ImageNet weights for all convolutional layers. The DenseNet-121 model consists of 120 convolutional layers, followed by a fully connected layer. DenseNets are the most recent CNN architecture with readily available ImageNet weights, and we chose to use the smallest variant of the architecture due to our limited training data. All experiments were implemented using the PyTorch machine learning library[37].

We trained ST-Net to predict the expression of the 250 genes with the highest mean expression. Many of the genes have very low expression (Supplementary Fig. 1b), resulting in a low signal-to-noise ratio, which is difficult to predict. We treated this as a multivariate regression problem, where each gene is a task, and shared weights in all convolutional layers. This results in large savings in computational time and allows the learned representation to be shared across tasks. A single fully connected layer at the end of the network consists of unshared weights, which are used to predict the expression of each gene.

We pre-processed the gene counts with two transformations. First, we normalized the total expression in each spot after adding a pseudo count of one. Regions with higher cell density tend to have higher gene expression, which is orthogonal to the effect related to the morphology that we are interested in. By normalizing the total expression, the cell density cannot be exploited. Second, the counts of a gene are highly skewed (Supplementary Fig. 1a), so we took the log value of the normalized counts to bring the values into a reasonable range. The pseudo count added before normalizing is needed to prevent zeros being used in the logarithm. After these two transformations were applied, we trained ST-Net to minimize the squared error for each of the genes.

To ensure that our predictions generalized to unseen samples, we reported all results using leave-one-out cross validation: we iteratively trained on 22 of the patients and predicted on the remaining patient. Within each cross-validation fold, we initialized the model to use pre-trained ImageNet weights and trained all weights using stochastic gradient descent with a learning rate of 10^{-6} and momentum of 0.9 for up to 50 epochs, with a batch size of 32. To select the number of epochs, we ran an internal four-fold cross validation on the 22 patients. During training time, we augmented the dataset by randomly rotating the image by 0, 90, 180 or 270° and taking the mirror image 50% of the time. During the test time, we averaged the eight symmetries resulting from the rotations and reflections.

To reduce experimental noise from spatial transcriptomics, we performed additional analyses by evaluating the prediction of ST-Net with the smoothed expressions. The smoothed expression of each spot was obtained by taking the average expression of all of the adjacent neighbour spots (that is, average of the three-by-three set of spots, where the spot of interest is in the centre). This simulates the effect of predicting over a larger region that is more deeply sequenced.

We also studied the effects of several modelling decisions for ST-Net. We found that selecting an appropriate window size and magnification for the patches can affect the accuracy of the model (Supplementary Fig. 11 and Supplementary Tables 4,5). We next found that sharing weights between genes and training for each gene individually had similar performances (Supplementary Fig. 12). However, sharing weights results in large savings in computational time. Finally, we found that using ImageNet pre-trained weights results in substantial improvements.

Comparison to cell-type composition-based model. We modelled the cell types in each patch using a non-negative matrix factorization model. We used ten factors in the model, where each factor modelled a cell type. The variance explained by varying the number of cell types is shown in Supplementary Fig. 7c. The ten factors were sufficient to explain most of the variance. The non-negative matrix factorization was trained with alternating minimization for 200 iterations to minimize the Frobenius norm of the reconstruction.

We trained a DenseNet-121 model to predict the composition of each patch using the same computational pipeline and hyperparameters as ST-Net. For the test patient, we used the trained DenseNet-121 model to predict the cell-type composition of each patch. The non-negative matrix factorization model was then used to map the cell composition to the expected gene expressions.

10x Genomics breast cancer ST-Net analysis. We downloaded the 10x Genomics breast cancer spatial transcriptomics data, which were generated using the Visium Spatial Gene Expression protocol. The data were normalized as per our original breast cancer ST-Net data. We then applied ST-Net to predict the spatial expression level from the 10x histology image and compared the prediction with the 10x experimental measurements. ST-Net was applied to the 10x data without any modification or tuning.

TCGA analysis. The Cancer Genome Atlas consists of image and omics data across cancer types. We used all of the available frozen breast tumour samples (1,550 slides from 1,093 patients) at ×20 magnification. We first processed the whole-slide images into patches containing substantial amounts of tissue. Candidate patches were taken by tiling the slides in non-overlapping windows of 512 × 512 pixels. We then removed the patches that were mostly background. A pixel in a patch was considered background if the mean RGB value was lower than or equal to 220 (that is, close to white), and a patch was removed if more than 50% of the pixels were considered background.

We predicted the spatial expression on TCGA using the trained weights from our dataset with no modifications. To validate our prediction on TCGA, we compared the mean prediction over all patches for each gene against the bulk RNA-seq data. Because no training was performed on TCGA, the Pearson's correlation for each gene was computed over all of the patients. The tumour genes considered for co-localization were *AEBP1*, *BGN* and *P4HB*, and the immune genes considered were *PABPC1*, *TAGLN2* and *HLA-B*.

To distinguish subtypes, we split the TCGA patients into a training and test set, consisting of 70 and 30% of the patients, respectively. We predicted the two main histological subtypes of breast cancer using logistic regression over our pseudo-bulk predictions from the 55 genes with significantly positive correlations to TCGA, logistic regression over real bulk RNA-seq expression from the 249 genes shared between our dataset and TCGA, and fine-tuning the final layer of a Densenet-121 model with pre-trained ImageNet weights. Cross validation using five folds of the training set was used to select the regularization parameter from 10^{-3}, 10^{-2}, 10^{-1}, 1, 10^{1} and 10^{2}.

Statistical analysis. We report several statistical measures of the performance prediction of ST-Net. For each gene and slide, we computed the Pearson's and Spearman's correlations between the expression predicted by ST-Net and the experimentally measured expression. We also measured how accurately ST-Net predicted the spots where the gene was expressed at high (that is, in the top 25% of spots) or low (in the bottom 25% of spots) levels. The AUROC was computed as the standard metric for this prediction task.

Finally, we also report how many patients were consistent with the prediction made by the model. Under the null hypothesis, the predictions of the CNN for a gene on a single patient has a 0.5 probability of having a positive Pearson's correlation. The number of patients with a positive Pearson's correlation would then be distributed as a binomial random variable with 23 trials and a 0.5 probability of success. We used this to compute an uncorrected *P* value. Each of the 250 genes we attempted to predict represented a distinct hypothesis, so we use the Holm–Bonferroni method to correct for multiple hypothesis testing[38]. We used the StatsModels package for computing the FDR values[39]. Comparisons of the tumour and immune spatial correlations were performed using a two-tailed paired *t*-test. All confidence intervals were computed using 1,000 bootstrap samples and correspond to intervals of 95%.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The main data supporting the results in this study are available within the paper and its Supplementary Information. Raw files for the breast cancer samples are available through a Materials transfer agreement with Å.B. (ake.borg@med.lu.se). All images and processed data are available at http://www.spatialtranscriptomicsresearch.org. The 10x Spatial Genomics data can be downloaded from https://wp.10xgenomics.com/spatial-transcriptomics. All data from TGCA are publicly available from the Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov).

Code availability

The code for ST-Net is available at https://github.com/bryanhe/ST-Net.

Received: 10 October 2019; Accepted: 23 May 2020;
Published online: 22 June 2020

References

1. Gerlinger, M. et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
2. Ståhl, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
3. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
4. Eng, C. H. L. et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **568**, 235–239 (2019).
5. Liu, R. et al. Modeling spatial correlation of transcripts with application to developing pancreas. *Sci. Rep.* **9**, 5592 (2019).
6. Lee, J. H. et al. Highly multiplexed subcellular RNA sequencing in situ. *Science* **343**, 1360–1363 (2014).
7. Carpenter, A. E. et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**, R100 (2006).
8. Kamentsky, L. et al. Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software. *Bioinformatics* **27**, 1179–1180 (2011).
9. Yu, K. H. et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat. Commun.* **7**, 12474 (2016).
10. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* 770–778 (2016).
11. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. Densely connected convolutional networks. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* 4700–4708 (2017).
12. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. In *Proc. Int. Conf. on Learning Representations* (2015).
13. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the inception architecture for computer vision. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* (2016).
14. Litjens, G. et al. 1399 H&E-stained sentinel lymph node sections of breast cancer patients: the CAMELYON dataset. *GigaScience* **7**, giy065 (2018).
15. Liu, Y. et al. Detecting cancer metastases on gigapixel pathology images. Preprint at https://arXiv.org/abs/1703.02442 (2017).
16. Wang, D., Khosla, A., Gargeya, R., Irshad, H. & Beck, A. H. Deep learning for identifying metastatic breast cancer. Preprint at https://arXiv.org/abs/1606.05718 (2016).
17. Coudray, N. et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat. Med.* **24**, 1559–1567 (2018).
18. Khosravi, P., Kazemi, E., Imielinski, M., Elemento, O. & Hajirasouliha, I. Deep convolutional neural networks enable discrimination of heterogeneous digital pathology images. *EBioMedicine* **27**, 317–328 (2018).
19. Yu, K. H. et al. Classifying non-small cell lung cancer histopathology types and transcriptomic subtypes using convolutional neural networks. *J. Am. Assoc. Med. Inform. Assoc.* **27**, 757–769 (2019).
20. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2008).
21. The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
22. Kurozumi, S. et al. Prognostic significance of tumour-infiltrating lymphocytes for oestrogen receptor-negative breast cancer without lymph node metastasis. *Oncol. Lett.* **17**, 2647–2656 (2019).
23. Ladha, J. et al. Identification of genomic targets of transcription factor AEBP1 and its role in survival of glioma cells. *Mol. Cancer Res.* **10**, 1039–1051 (2012).
24. Sangaletti, S. et al. Macrophage-derived SPARC bridges tumor cell–extracellular matrix interactions toward metastasis. *Cancer Res.* **68**, 9050–9059 (2008).
25. Yamamoto, K. et al. Biglycan is a specific marker and an autocrine angiogenic factor of tumour endothelial cells. *Br. J. Cancer* **106**, 1214–1223 (2012).
26. Cheng, J. et al. Integrative analysis of histopathological images and genomic data predicts clear cell renal cell carcinoma prognosis. *Cancer Res.* **77**, e91–e100 (2017).
27. Ge, R. & Zou, J. Intersecting faces: non-negative matrix factorization with new guarantees. In *Proc. of the 32nd Int. Conf. on Machine Learning* (2015).
28. Rahmani, E. et al. Sparse PCA corrects for cell type heterogeneity in epigenome-wide association studies. *Nat. Methods* **13**, 443–445 (2016).
29. Sundararajan, M., Taly, A. & Yan, Q. Axiomatic attribution for deep networks. In *Proc. 34th Int. Conf. on Machine Learning* (2017).
30. Hunt, D. A. et al. mRNA stability and overexpression of fatty acid synthase in human breast cancer cell lines. *Anticancer Res.* **27**, 27–34 (2007).
31. McInnes, L., Healy, J. & Melville, J. UMAP: uniform manifold approximation and projection. *J. Open Source Software* **3**, 891 (2018).
32. Salmén, F. et al. Barcoded solid-phase RNA capture for spatial transcriptomics profiling in mammalian tissue sections. *Nat. Protoc.* **13**, 2501–2534 (2018).

33. Deng, J. et al. Imagenet: a largescale hierarchical image database. In *IEEE Conf. on Computer Vision and Pattern Recognition* 248–255 (2009).
34. Russakovsky, O. et al. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115**, 211–252 (2015).
35. Esteva, A. et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115–118 (2017).
36. De Fauw, J. et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat. Med.* **24**, 1342–1350 (2018).
37. Paszke, A. et al. PyTorch: an imperative style, high-performance deep learning library. In *Adv. Neural Inf. Process. Syst.* (2017).
38. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6**, 65–70 (1979).
39. Seabold, S. & Perktold, J. Statsmodels: econometric and statistical modeling with Python. In *Proc. 9th Python in Science Conf.* (2010).

Acknowledgements

J.Z. is supported by the NSF (grant no. CCF 1763191), NIH (grant nos. R21 MD012867-01, P30AG059307 and U01MH098953) and grants from the Silicon Valley Foundation and Chan–Zuckerberg Initiative. J.L. is supported by the Swedish Foundation for Strategic Research, Swedish Cancer Society and Swedish Research Council.

Author contributions

All of the authors contributed to the project planning and writing of the manuscript. B.H. and L.B. performed analysis. L.S., Å.B. and J.L. generated data. J.M., J.L. and J.Z. supervised the project.

Competing interests

J.L. is an author on patent nos. PCT/EP2012/056823 (WO2012/140224), PCT/EP2013/071645 (WO2014/060483) and PCT/EP2016/057355 applied for by Spatial Transcriptomics AB/10x Genomics Inc. covering the described technology.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41551-020-0578-x.

Correspondence and requests for materials should be addressed to J.M., J.L. or J.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): James Zou

Last updated by author(s): May 17, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Protocols for RNA sequencing of barcoded libraries, and subsequent analysis from tissue sections was done as described in detail in previous publications by using open-source Python scripts.
Data analysis	ST-Net and all of the analysis in this paper were conducted in Python v3. All of the code is open-source, and available at https://github.com/bryanhe/ST-Net.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The main data supporting the results in this study are available within the paper and its Supplementary Information. Raw files for the breast-cancer samples are available through a Materials transfer agreement with Åke Borg (ake.borg@med.lu.se). All images and processed data are available at http://www.spatialtranscriptomicsresearch.org. The 10x Spatial Genomics data can be downloaded from https://wp.10xgenomics.com/spatial-transcriptomics. All data from the Cancer Genome Atlas is publicly available from the Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	ST-Net was trained on spatial-transcriptomics data collected from 23 breast cancer patients. Even though the number of patients is relatively small, each patient generated a large amount of imaging and RNA-sequencing data. All together, we analyzed 30,612 sets of spatially resolved RNA-seq profiles with matched histology images.
Data exclusions	In order to produce the most reliable findings, we focused our analyses on 250 genes that have the highest average expression across all of the patients.
Replication	We collected 3 sections from each of the 23 patients. We further validated the prediction performance of the algorithm on a new 10x Genomics breast cancer dataset.
Randomization	We used leave-one-patient-out cross validation. The algorithm was trained on 22 patients and tested on the hold-out patient; then the results are averaged across all 23 possible hold-out patients.
Blinding	Because of our cross-validation approach, blinding was not necessary.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	Twenty three female breast cancer patients agreed with their tumor biopsy samples being used in this study.
Recruitment	The participants were breast cancer patients treated at the Lund University Hospital.
Ethics oversight	Ethical oversight was provided by the Lund University Hospital. This study complied with all relevant ethical regulations involving human tissue.

Note that full information on the approval of the study protocol must also be provided in the manuscript.