

# Using a Convolutional Neural Network for Classification of Squamous and Non-Squamous Non-Small Cell Lung Cancer Based on Diagnostic Histopathology HES Images

**Anne Laure Le Page**

Caen University Hospital, Normandy University

**Elise Ballot**

Georges François Leclerc Cancer Center – UNICANCER

**Caroline Truntzer**

Georges François Leclerc Cancer Center – UNICANCER

**Valentin Derangère**

Georges François Leclerc Cancer Center – UNICANCER

**Alis Ilie**

Georges François Leclerc Cancer Center – UNICANCER

**David Rageot**

Georges François Leclerc Cancer Center – UNICANCER

**Frederic Bibeau**

Caen University Hospital, Normandy University

**François Ghiringhelli** (✉ [fghiringhelli@cgfl.fr](mailto:fghiringhelli@cgfl.fr))

François Leclerc - UNICANCER

---

## Research Article

**Keywords:** Lung Cancer, deep learning, artificial intelligence, histology, non-small cell lung cancer.

**Posted Date:** July 6th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-646715/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Scientific Reports on December 1st, 2021.  
See the published version at <https://doi.org/10.1038/s41598-021-03206-x>.

## Abstract

Histological stratification in metastatic non-small cell lung cancer (NSCLC) is essential to properly guide therapy. Morphological evaluation remains the basis for subtyping and is completed by additional immunohistochemistry (IHC) labelling to confirm the diagnosis, which delays molecular analysis and utilises precious sample. Therefore, we tested the capacity of convolutional neural networks (CNNs) to classify NSCLC based on pathologic H&E diagnostic biopsies. The model was estimated with a learning cohort of 132 NSCLC patients and validated on an external validation cohort of 65 NSCLC patients. Based on image patches, a CNN using InceptionV3 architecture was trained and optimized to classify NSCLC between squamous and non-squamous subtypes. Accuracies of 0.99, 0.85, 0.87, 0.85 was reached in the training, validation and test sets and in the external validation cohort. At the patient level, the CNN model showed a capacity to predict the tumour histology with accuracy of 0.73 and 0.78 in the learning and external validation cohorts respectively. Selecting tumour area using virtual Tissue Micro-Array (TMA) improved prediction, with accuracy of 0.79 in both learning and external validation cohorts. This study underlines the capacity of CNN to predict NSCLC subtype with good accuracy and to be applied to small pathologic samples without annotation.

## Introduction

The standard of care in first line treatment of Non-Small Cell Lung Cancer (NSCLC) patients is based on chemoimmunotherapy or tyrosine kinase<sup>1,2</sup>. Treatment is assigned on the basis of specific histologic and genomic characteristics of the patient's tumour<sup>3</sup>. In a first step, NSCLC must be classified into a particular histological type: non-squamous NSCLC versus squamous cell carcinoma. This classification is essential for further molecular examination of the tissue sample to orient patients towards the optimal therapeutic treatment<sup>4</sup>. In case of non-squamous NSCLC, it is mandatory to obtain a list of molecular biomarkers, such as *EGFR* or *BRAF* V600E mutations, or *ALK* and *ROS1* rearrangements<sup>5</sup>. In addition, many emerging biomarkers require histological material for adenocarcinoma NSCLC (Met, NRG1, NTRK) but also for non-squamous NSCLC (PI3KCA, HRAS)<sup>6</sup>. However, while the molecular and histological material needed for treatment determination is increasing, the amount of histologic tumour tissue available is often small. Therefore, strategies that can help to reduce the material required for histological assessment will be helpful.

The development of Artificial Intelligence algorithms, which can be used to automatically classify histological slides, opens new perspectives in virtual and digital pathology. For example, in the setting of lung cancer, automatic analysis of whole-slide images of lung tumour resection has recently been studied to predict survival outcomes<sup>7</sup>, and can be used to predict histological type or mutational status<sup>8</sup>. However, such data are not relevant for clinical use because in most cases, pathologists only have a small tumour biopsy, or cytology fine needle aspiration.

To limit the volume of material required for histological diagnosis, we propose a deep learning convolutional network aimed at predicting the histological classification of non-squamous versus

squamous cell carcinoma. Our analysis was based on tumour biopsy using whole tissue from biopsy or virtual TMA, based on annotation of the tumour zone.

## Results

### Population description

For the learning set, we included 132 HES slides from Dijon. These samples comprised 66 non-squamous and 66 squamous samples. Samples were obtained from primary lung tumour for all cases. The median tissue area was 11.734 [0.158–111.227] mm<sup>2</sup> and the median tumour tissue area was 0.177 mm<sup>2</sup> [0.002–1.088]. For the validation set, we included HES slides from Caen (n = 65; 45 non squamous and 20 squamous samples).

### A deep learning model for NSCLC subtype prediction using WSI classification

Our objective was to estimate a deep learning model to classify lung carcinoma subtypes using whole HES slides from tumour biopsy, regardless of the percentage of tumour cells contained in the biopsy. As described above, the learning cohort was decomposed into internal training, validation and test sets (Table 1).

Table 1  
Description of learning and external validation cohorts

		Non Squamous cell carcinoma	Squamous cell carcinoma
Learning cohort	Training sample (n = 78)	39 patients (77 640 tiles)	39 patients (62 496 tiles)
	Validation sample (n = 26)	13 patients (23 750 tiles)	13 patients (21 801 tiles)
	Test sample (n = 28)	14 patients (22 911 tiles)	14 patients (20 322 tiles)
External validation cohort	External sample (n = 65)	45 patients (464 022 tiles)	20 patients (106 727 tiles)

Using Inception V3 deep learning architecture, our CNN model was optimized using different approaches. First, we added a threshold for predictions to retain only tiles with high prediction level; in fact, it is expected that WSI include a large number of tiles without tumour cells, which would alter predictions with noisy information. A second strategy used a kernel filter to take into account the spatial environment of the tiles. At the tile level, accuracies from the resulting models underline that the threshold methodology is the best strategy, with values of 0.99, 0.85 and 0.87 respectively in the training, validation and test datasets. Similarly, our model had an accuracy of 0.85 in the external validation cohort, which underlines

the robustness of the model and the absence of overfitting. (Table 2). Supplemental Fig. 1 shows the accuracy and loss across epochs for model estimation.

Table 2  
Accuracy achieved by the different strategies at tile level

	Training	Validation		Test		External validation	
	Overall	Overall	By class	Overall	By class	Overall	By class
<b>WSI</b>							
<i>Thresholds :</i>	0.75	0.64	NS : 0.70	0.58	NS : 0.66	0.68	NS : 0.73
NS = 0.5, S = 0.5			S : 0.57		S : 0.49		S : 0.44
NS = 0.9, S = 0.9	0.99	0.87	NS : 0.88	0.85	NS : 0.84	0.85	NS : 0.89
			S : 0.87		S : 0.86		S : 0.72
<i>Re-estimation using filter kernel</i>	0.84	0.71	NS : 0.79	0.63	NS : 0.71	0.71	NS : 0.77
			S : 0.62		S : 0.53		S : 0.48
<b>TMA</b>							
<i>Thresholds :</i>	0.78	0.69	NS : 0.74	0.65	NS : 0.68	0.66	NS : 0.77
NS = 0.5, S = 0.5			CE : 0.63		S : 0.62		S : 0.49
NS = 0.9, S = 0.9	0.99	0.83	NS : 0.84	0.88	NS : 0.77	0.92	NS : 0.92
			S : 0.82		S : 0.92		S : 0.94
<i>Re-estimation using filter kernel</i>	0.88	0.79	NS : 0.83	0.73	NS : 0.75	0.71	NS : 0.81
			S : 0.76		S : 0.71		S : 0.56

To classify the tumour slide, we pooled tile information using either max pooling or majority voting strategies. The best strategy was majority voting; using this strategy, our model had an accuracy of 0.71 and 0.73 respectively in the learning and external validation cohorts (Table 3).

Table 3  
Accuracy achieved by the different strategies at patient level

	Majority voting				Max pooling			
	Test		External validation		Test		External validation	
	Global	By class	Global	By class	Global	By class	Global	By class
<b>WSI</b>								
<i>Thresholds : NS = 0.5, S = 0.5</i>	0.71	NS : 0.71	0.74	NS : 0.69	0.68	NS : 0.79	0.71	NS : 0.80
		S : 0.71		S : 0.85		S : 0.57		S : 0.50
<i>NS = 0.9, S = 0.9</i>	0.69	NS : 0.69	0.74	NS : 0.69	0.73	NS : 0.77	0.78	NS : 0.76
		S : 0.69		S : 0.85		S : 0.69		S : 0.85
<i>Re-estimation using filter kernel</i>	0.61	NS : 0.57	<b>0.81</b>	NS : 0.77	0.71	NS : 0.79	0.73	NS : 0.80
		S : 0.64		S : 0.90		S : 0.64		S : 0.60
<b>TMA</b>								
<i>Thresholds : NS = 0.5, S = 0.5</i>	<b>0.79</b>	NS : 0.79	<b>0.79</b>	NS : 0.79	<b>0.82</b>	NS : 0.93	0.73	NS : 0.83
		S : 0.79		S : 0.80		S : 0.71		S : 0.50
<i>NS = 0.9, S = 0.9</i>	0.68	NS : 0.67	<b>0.80</b>	NS : 0.79	0.68	NS : 0.67	<b>0.82</b>	NS : 0.81
		S : 0.70		S : 0.83		S : 0.70		S : 0.83
<i>Re-estimation using filter kernel</i>	<b>0.79</b>	NS : 0.79	0.77	NS : 0.79	<b>0.82</b>	NS : 0.86	0.73	NS : 0.79
		S : 0.79		S : 0.75		S : 0.79		S : 0.60

#### Prediction using virtual TMA analysis of WSI.

In order to improve classification and computational time, we created a virtual TMA, using a circle with a radius of 500 micrometres from the centroid of the annotation drawn by the pathologist. The computational time for predictions on TMA was 18 times less than on the entire slide. This strategy also

has a benefit that can translate in clinical routine with quick annotation by a pathologist, who just clicks on the tumour core instead of contouring the whole tumour. Accuracy findings confirmed that using this gating strategy, the threshold methodology was also the best strategy, with model accuracy of 0.99, 0.83 and 0.88 in the training, validation and test datasets at the tile level. Similarly, the model had accuracy of 0.92 in the external validation cohort (Table 2).

We then used the TMA strategy to predict tumour slide classification. Using max pooling, the model had an accuracy of 0.79 in both the learning and external validation data sets (Table 3).

Figures 1 and 2 show two cases of tumour biopsy sections containing respectively squamous and non-squamous tumour, with the correct diagnosis and the predicted WSI diagnosis based on TMA or WSI analysis for each prediction step.

## Discussion

The diagnosis of NSCLC is based on morphological evaluation of tissue specimens. This analysis is the first step before addressing samples for molecular testing and therapy stratification<sup>15</sup>. One issue in the management of metastatic lung cancer is that in most cases, samples are cytological exams or small biopsies. Preservation of the sample in this clinical context for further molecular testing is important. Consequently, even if applying artificial intelligence on such a routine exam may seem irrelevant an experienced pathologist who is well trained in analysis of IHC staining like TTF1 and p40, it clearly assure a sparingly use of biopsy specimen. Our study, like previous reports, showed that the combination of digital pathology and machine learning has the potential to support this decision process in an objective manner<sup>16</sup>. In previous works, the application of deep learning to classify lung histological specimens yielded promising results in lung cancer<sup>17 18 19</sup>. However most of these reports only fostered on surgical samples.

In this study, we analysed whether a CNN-model (InceptionV3 CNN) could be used to differentiate squamous from non-squamous NSCLC, based on the initial tumour biopsy. This study was performed without taking into account the tissue type of the biopsy, or whether the sample was a cytological or histological sample. In this work, we addressed some technical points and show that the whole slide can be used to predict the histological subtype with good accuracy, without prior tumour tissue selection by the pathologist. Surprisingly, adding spatial information using kernel filter did not improve the classification. In contrast, adding quality check with a threshold to select only predictions with a good level of confidence improved the accuracy of the classification. These findings are not unexpected, since WSI include many non-tumour zones.

To improve the prediction, we also used a virtual TMA strategy. Based on the pathologist's hand-drawn tumour annotations, TMA were created by tracing a circle with a radius of 500 micrometres from the centroid of this annotation. This strategy could easily be reproduced by a pathologist, who could click on

the virtual slide to localize the tumour and obtain the prediction for the whole slide using only TMA restricted information.

The limitations of our study include the small sample size, and the small number of extracted image patches in some cases, which may limit the accuracy of the model. Moreover, epithelial lung tumours may be morphologically very different. In particular, the current World Health Organization classification is more complex and separates adenocarcinoma into several different subtypes, such as lepidic, solid, acinar, and papillary. Because of the small learning set, we did not include this information in the model, but using a larger learning set with further non-squamous subtype labelling would undoubtedly improve the capacity of the CNN model to predict histological types with greater accuracy. Further studies are warranted on this point. While the learning set was performed on lung biopsy, the model is validated on either cytological or pathological samples, and also on either lung biopsy or metastatic samples. This heterogeneity in the samples may induce some bias, and may limit the accuracy of the model. However, we chose this heterogeneity to better reflect the clinical reality of lung cancer diagnosis.

In summary, we trained and optimized an Inception V3 CNN model to classify the two common NSCLC subtypes using routine biopsy or cytological samples. Moreover, we established a virtual TMA strategy to improve predictions. Our results highlight the potential and limitations of CNN image classification models for morphology-based tumour classification.

## Methods

### Study population

The learning cohort comprised 132 NSCLC tumour biopsies (66 non squamous and 66 squamous samples) collected between 2015 and 2018 in the Department of Pathology of the Georges François Leclerc Cancer Center in Dijon, France.

The external validation cohort comprised 65 biopsy samples (45 non squamous and 20 squamous samples) from the University Hospital of Caen, France, using tumours collected between 2017 and 2019.

Only patients from whom informed consent was obtained were included in this retrospective study. The present study was approved by the CNIL (French national commission for data privacy) and the Georges François Leclerc Cancer Center (Dijon, France) local ethics committee, and was performed in accordance with the Helsinki Declaration and European legislation.

### Pathological diagnosis

The pathological diagnosis (adenocarcinoma versus squamous cell carcinoma) was validated for all samples by a pathologist (ALLP). Pathological classification was performed using analysis of morphology on HES stained slides and TTF1 and p40 immunohistological analysis.

### Image processing

Formalin-fixed paraffin-embedded HES stained slides were digitised with a Nanozoomer HT2.0 (Hamamatsu) at  $\times 20$  magnification to generate a whole slide imaging (WSI) file in ndpi format. We partitioned the WSI into non-overlapping 220x220 pixel tiles at 0.5 mm/pixel resolution (equivalent to 20X\_ magnification) using QuPath v.0.2.3 <sup>9</sup>.

In addition, tumour regions of each slide were manually annotated by a pathologist (ALLP). Then, the centroids of each annotation were calculated. A TMA was created based on a circle with a radius of 500 micrometres from the centre of the centroid of the annotation. The same tiling as described above was kept.

## Tile Pre-processing

Tiles were removed if they contained more than 2/3 of white background. The color channel values were normalized by Reinhard normalization to neutralize color differences between slides <sup>10</sup>. This normalization uses a linear transformation to match the mean and standard deviation between slides. The color channel values were scaled to a floating value range of [0, 1].

Training, validation and test sets were generated using respectively 60%, 20% and 20% of tiles. Tiles associated with a given slide were not separated, but associated with one of these sets to prevent overlap of slides between the three sets.

## Deep learning model

We estimated a model based on InceptionV3 <sup>11</sup>. The idea behind the Inception architecture is to use a series of convolutional blocks to both decrease the number of parameters in the network and improve its performance. The main components of a convolutional block are convolutional and pooling layers. To make the algorithm more robust against image variations, and to add a regularisation effect, we applied data augmentation techniques. This included techniques such as randomly flipping the images left-right and up-down with additional random rotations.

The model was trained for one hundred epochs on the augmented training set, starting with an initial learning rate of 0.001, decaying by a factor of 0.9 every five epochs and using the Adam optimisation algorithm <sup>12</sup> with a momentum of 0.9 and epsilon of 1e-7. We used a batch size of 100 tiles.

Due to an unequal number of extracted tiles for each class (unbalanced dataset), we used a weighted loss function allowing direct penalization of false predictions during the training process. Negative and false positives were equally penalized with a 1.5 factor.

The deep learning model was implemented and trained using TensorFlow 2.1.0 and python 3.5. Calculations were performed using HPC resources from DNUM CCUB (Centre de Calcul de l'Université de Bourgogne).

## Patient inference

We then classified each tile and filtered out low-confidence predictions by using thresholding. Thresholds were determined by a grid search over each class, optimizing the correct classification rate<sup>13</sup>.

The CNN can be used directly as a classifier, but it predicts each tile independently and ignores spatial correlations. To take advantage of the neighbourhood of each tile, filter kernel algorithms aimed at extracting spatial information were used; the filter kernel takes advantage of the label distribution of neighbouring patches to re-estimate the output of CNNs. A logistic regression algorithm was used as the strategy for parameter estimation of the filter kernel.

If the label of a tile is the same as the label of the neighbouring tiles, its probability will be increased. Conversely, it will have a lower probability when its label differs from that of its neighbours.

To classify the whole slide, we used two methods<sup>14</sup>. The first, called "majority vote", assigned the most frequent class to the slide. The second, called "max pooling", assigned the class with the highest probability to the slide.

These different strategies were applied on tiles from the whole slide as well as from TMA only in order to focus the results on tumour regions.

## Declarations

**Acknowledgements:** We wish to thank Fiona Ecarnot, PhD (EA3920, University of Franche-Comté, Besançon, France) for English correction and helpful comments.

**Availability of data and materials:** The datasets analyzed during the current study are available from the corresponding author on reasonable request.

**Authors' contributions:** A.-L.L.P., V.D., F.B., F.G., E.B. and C.T. contributed to the design. A.I. and D.R. generated the data. E.B. and C.T. analysed the data. All authors contributed to the writing of the manuscript.

**Competing interests:** No author has any conflicting financial interests.

## References

1. Hanna, N. H. *et al.* Therapy for Stage IV Non-Small-Cell Lung Cancer without Driver Alterations: ASCO and OH (CCO) Joint Guideline Update. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.*, **38**, 1608–1632 (2020).
2. Hanna, N. H. *et al.* Therapy for Stage IV Non-Small-Cell Lung Cancer With Driver Alterations: ASCO and OH (CCO) Joint Guideline Update. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.*, **39**, 1040–1091 (2021).

3. Bernicker, E. H., Miller, R. A. & Cagle, P. T. Biomarkers for Selection of Therapy for Adenocarcinoma of the Lung. *J. Oncol. Pract.*, **13**, 221–227 (2017).
4. Travis, W. D. et al. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. *J. Thorac. Oncol. Off. Publ. Int. Assoc. Study Lung Cancer*, **10**, 1243–1260 (2015).
5. Vanderlaan, P. A. et al. Success and failure rates of tumor genotyping techniques in routine pathological samples with non-small-cell lung cancer. *Lung Cancer Amst. Neth.*, **84**, 39–44 (2014).
6. Emerging Targeted Therapies for the Treatment of Non-small Cell Lung Cancer | SpringerLink. <https://link.springer.com/article/10.1007/s11912-019-0770-x>.
7. Luo, X. et al. Comprehensive Computational Pathological Image Analysis Predicts Lung Cancer Prognosis. *J. Thorac. Oncol. Off. Publ. Int. Assoc. Study Lung Cancer*, **12**, 501–509 (2017).
8. 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).
9. QuPath: Open source software for digital pathology image analysis | Scientific Reports. <https://www.nature.com/articles/s41598-017-17204-5>.
10. Reinhard, E., Ashikhmin, M., Gooch, B. & Shirley, P. Color Transfer between Images. *IEEE Comput. Graph. Appl.*, **21**, 34–41 (2001).
11. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the Inception Architecture for Computer Vision(2016). doi:10.1109/CVPR.2016.308.
12. Kingma, D., Ba, J. & Adam A Method for Stochastic Optimization. *Int. Conf. Learn. Represent.*(2014).
13. Wei, J. W. et al. Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks. *Sci. Rep.*, **9**, 3358 (2019).
14. Hou, L. et al. Patch-Based Convolutional Neural Network for Whole Slide Tissue Image Classification. Proceedings. IEEE Computer Society Conference on Computer Vision and Pattern Recognition vol 2016 (2016).
15. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification - ScienceDirect. <https://www.sciencedirect.com/science/article/pii/S1556086415335711?via%3Dihub>.
16. Cui, M. & Zhang, D. Y. Artificial intelligence and computational pathology. *Lab. Investig. J. Tech. Methods Pathol.*, **101**, 412–422 (2021).
17. Kriegsmann, M. et al. Deep Learning for the Classification of Small-Cell and Non-Small-Cell Lung Cancer. *Cancers*, **12**, 1604 (2020).
18. Gertych, A. et al. Convolutional neural networks can accurately distinguish four histologic growth patterns of lung adenocarcinoma in digital slides. *Sci. Rep.*, **9**, 1483 (2019).
19. Chen, M. et al. Classification and mutation prediction based on histopathology H&E images in liver cancer using deep learning. *Npj Precis. Oncol.*, **4**, 1–7 (2020).

# Figures

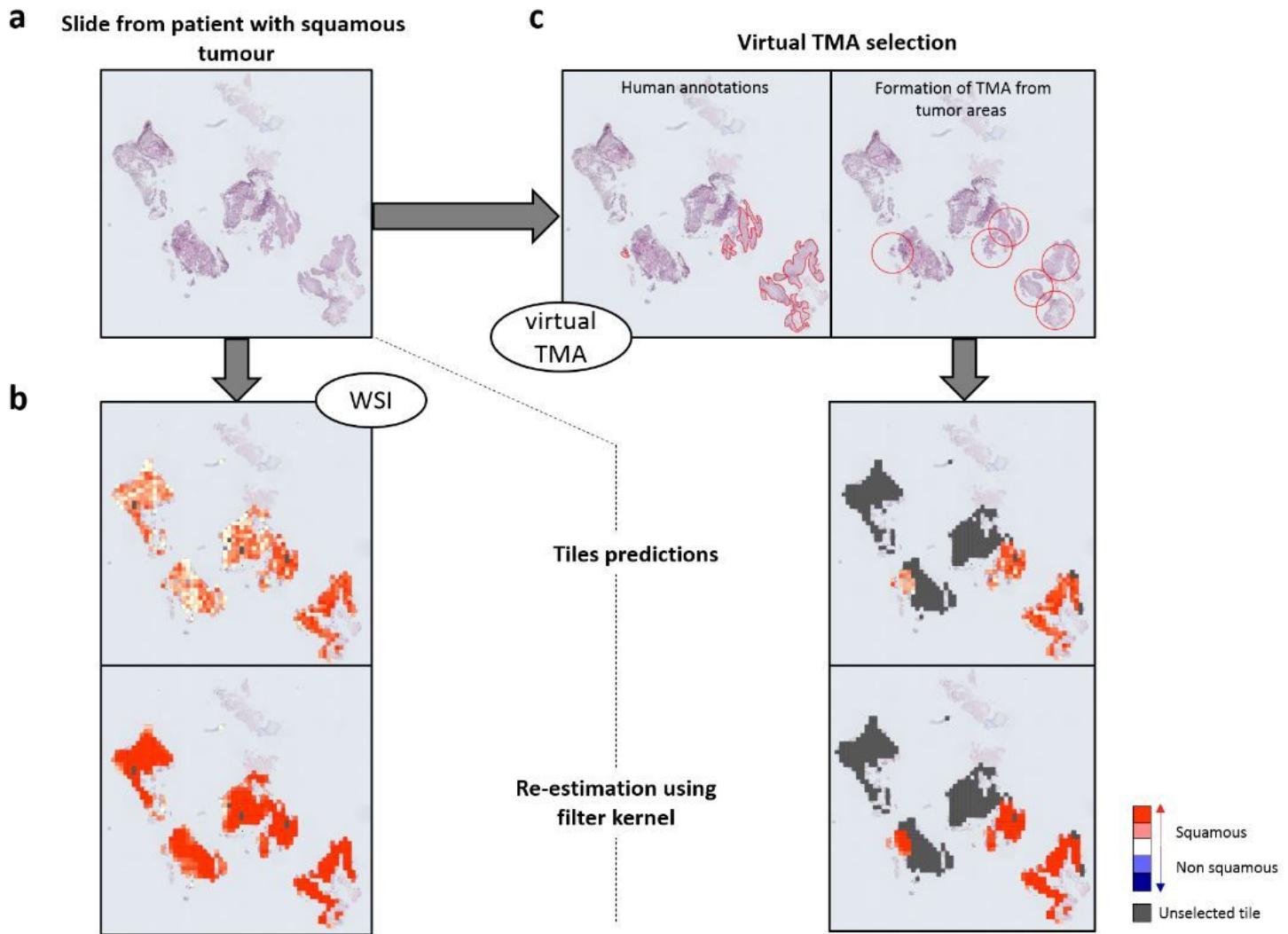
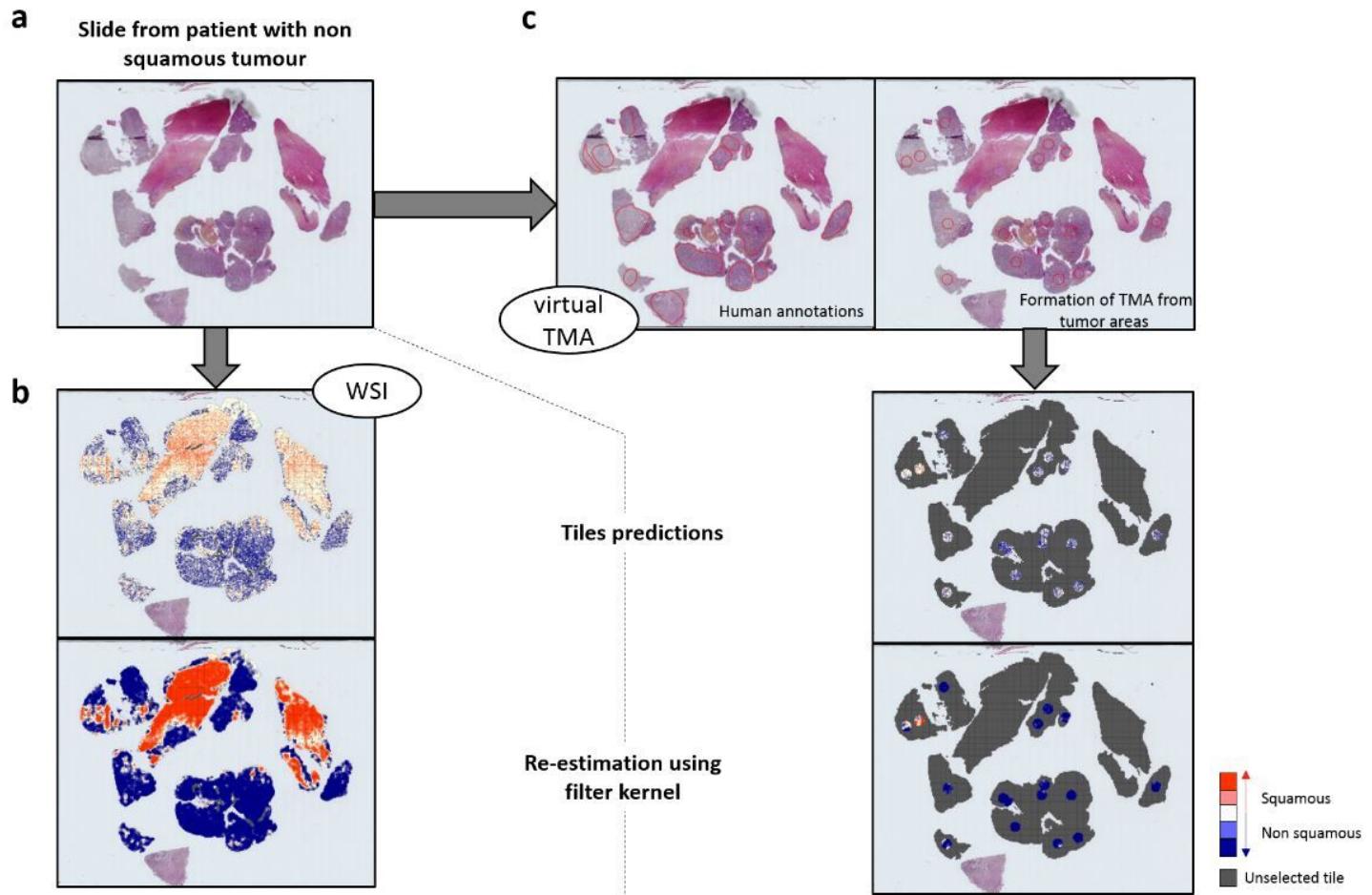


Figure 1

Prediction procedure for a squamous tumour biopsy section (a) Representation of original slide. (b-c) Steps for prediction analysis based on (b) Whole Slide Image (WSI) or (c) virtual TMA strategy. Heatmaps show for each tile probability of being predicted as squamous tumour (red) or non-squamous tumour (blue); the highest probability was kept for coloration; grey was used for tiles non selected by the pathologist or removed by denoising step (tile containing more than 2/3 of white background).



**Figure 2**

Predictions of tumour biopsy section containing non-squamous tumour (a) Representation of original slide. (b-c) Steps for prediction analysis based on (b) Whole Slide Image (WSI) or (c) virtual TMA strategy. Heatmaps show for each tile probability of being predicted as squamous tumour (red) or non-squamous tumour (blue); the highest probability was kept for coloration; grey was used for tiles non selected by the pathologist or removed by denoising step (tile containing more than 2/3 of white background).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFigure1.pptx](#)