BMC Cancer

Open Access

Deep learning-based tumor microenvironment segmentation is predictive of tumor mutations and patient survival in non-small-cell lung cancer

Alicja Rączkowska[1], Iwona Paśnik[2], Michał Kukiełka[1], Marcin Nicoś[3], Magdalena A. Budzinska[4], Tomasz Kucharczyk[3], Justyna Szumiło[2], Paweł Krawczyk[3], Nicola Crosetto[5,6] and Ewa Szczurek[1*]

Abstract

Background Despite the fact that tumor microenvironment (TME) and gene mutations are the main determinants of progression of the deadliest cancer in the world – lung cancer, their interrelations are not well understood. Digital pathology data provides a unique insight into the spatial composition of the TME. Various spatial metrics and machine learning approaches were proposed for prediction of either patient survival or gene mutations from this data. Still, these approaches are limited in the scope of analyzed features and in their explainability, and as such fail to transfer to clinical practice.

Methods Here, we generated 23,199 image patches from 26 hematoxylin-and-eosin (H&E)-stained lung cancer tissue sections and annotated them into 9 different tissue classes. Using this dataset, we trained a deep neural network ARA-CNN. Next, we applied the trained network to segment 467 lung cancer H&E images from The Cancer Genome Atlas (TCGA) database. We used the segmented images to compute human-interpretable features reflecting the heterogeneous composition of the TME, and successfully utilized them to predict patient survival and cancer gene mutations.

Results We achieved per-class AUC ranging from 0.72 to 0.99 for classifying tissue types in lung cancer with ARA-CNN. Machine learning models trained on the proposed human-interpretable features achieved a c-index of 0.723 in the task of survival prediction and AUC up to 73.5% for *PDGFRB* in the task of mutation classification.

Conclusions We presented a framework that accurately predicted survival and gene mutations in lung adeno-carcinoma patients based on human-interpretable features extracted from H&E slides. Our approach can provide important insights for designing novel cancer treatments, by linking the spatial structure of the TME in lung adeno-carcinoma to gene mutations and patient survival. It can also expand our understanding of the effects that the TME has on tumor evolutionary processes. Our approach can be generalized to different cancer types to inform precision medicine strategies.

*Correspondence:
Ewa Szczurek
szczurek@mimuw.edu.pl
Full list of author information is available at the end of the article

Background

In clinical practice it is common to diagnose cancer based on hematoxylin-eosin (H&E) stained slides obtained through biopsy or surgery [1]. Such slides are routinely stored for each patient, so there is an abundance of patient-specific, disease progression-relevant data that until recently has not been utilized at scale in cancer research. This has changed with the advent of digital pathology and the development of machine learning approaches to various predictive tasks based on digitized H&E image data [2, 3].

In addition to tumor cells, H&E images portray the spatial architecture of the tumor microenvironment (TME), including stromal cells, immune cells, and hypoxic/necrotic tissue areas and their reciprocal spatial arrangement. The TME plays an important role in cancer progression and metastasis, and thus it is critical to study its composition extensively [4]. Different tumors, even of the same type, have various genetic profiles resulting from gene mutations [5]. For a given cancer type, survival of individual patients can largely vary [6, 7]. Finally, the TME of different tumors is also different [8]. A burning question in this context is how the structure of the TME relates to patient survival and gene mutations.

This question is particularly relevant for lung cancer. Lung cancer is the deadliest cancer type worldwide, and lung adenocarcinoma (LUAD) is the most often diagnosed subtype of lung cancer [9]. LUAD includes a relatively higher proportion of cases without tobacco exposure, as compared to other lung cancer types. Thus, it has a more balanced molecular background and is more frequently associated with the presence of single somatic driver mutations that may be effectively managed with specific molecularly targeted therapies [10]. Several genes are known markers of response to treatment and survival in LUAD, including *EGFR*, *ALK*, *ROS1*, *BRAF*, *NTRK1–3*, *RET*, *MET*, *KRAS*, and diagnostic panels for targeted gene sequencing for detecting mutations in critical genes are routinely used in clinical practice [11]. H&E images are inspected for LUAD diagnosis in clinical practice [12]. Wide tumor spread, access to vessels, large areas of necrosis visible in H&E images, are associated with poor diagnosis [4], while abundance of immune cells indicates anti-tumor response of the immune system and associates with better survival [13, 14]. The TME plays an important role in LUAD response to immunotherapy. Expression of PD-1 and PD-L1 on cancer or immune cells, as well as tumor mutation burden (TMB), are important biomarkers of immune checkpoint inhibitor efficiency [15, 16]. The interconnections between the spatial TME composition, gene mutations and LUAD patient survival are so far not well understood.

Computational prediction of patient survival from H&E images has been either performed based on spatial metrics or using deep learning approaches. In the former case, spatial metrics are initially used to summarize the spatial arrangement of different tissues and next their correlation with survival is investigated. Alternatively, these spatial metrics are used as features in traditional machine learning algorithms [17–20]. The TME can be very heterogeneous, so it is not obvious how to quantify it and what metrics to use. These metrics include proportion-based [21], clustering-based [22], and methods borrowed from ecology [23], and they have been applied to many different cancer types [18, 22, 24–26]. In general, all these metrics share a common trait, i.e., they incorporate only a limited number of tissue types at once, such as tumor cells and lymphocytes, tumor cells and stroma, etc. This approach cannot comprehensively capture the complexity of the TME. Thus, there is an unmet need for an encompassing spatial metric that would consider many possible TME components at once. In the latter case, deep neural networks are trained to predict patient survival directly from H&E images. Such deep learning-based methods are increasingly used for survival prediction and have been shown to perform comparably to or even better than spatial metric-based approaches [27, 28]. However, one major disadvantage of deep learning methods is the lack of explainability. Due to the complicated structure of these models and number of parameters, it is not easy to surmise which parts of the TME are the most important for patient survival. This creates a need for explainable H&E image-based survival models.

Numerous methods for predicting gene mutations from H&E images were introduced and applied to a spectrum of cancers [29–34], showing that such approaches can reveal links between the TME composition and mutations of selected genes. However, similarly to deep learning-based methods for patient survival, these models take raw image data as input and directly predict the presence of mutations. As such, it is hard to assess what parts of the TME are most predictive of a given mutation. A recent study proved that human-interpretable features extracted from images segmented with deep learning methods can be successfully applied to predict phenotypic expression [35]. This suggests that the same

Rączkowska *et al. BMC Cancer*   (2022) 22:1001

Page 3 of 18

approach can be implemented to predict patient survival and gene mutations, which has not yet been explored.

To address these shortcomings, here we develop a framework for predicting survival and gene mutations in LUAD patients based on H&E images and using human-interpretable features. First, we train a deep learning classifier and apply it to segment the H&E images from 467 tumor LUAD samples into nine tissue classes. Next, we compute two human-interpretable spatial features that describe the composition of the TME in segmented H&E slides. Finally, we use these features in combination with clinical data to predict patient survival, as well as to predict tumor mutations. The predictions generated by our model are human interpretable, i.e., it is possible to pinpoint exactly which component of the TME is associated with a change in survival hazard or with a given mutation. Our framework is readily generalizable to other lung cancer subtypes and can be extended to other tumor types to make predictions of patient survival and cancer mutations based on digital pathology. This work is a step forward to a better understanding of the interplay between the TME, gene mutations and survival of LUAD patients.

Methods
Clinical samples
We obtained the formalin-fixed paraffin embedded (FFPE) tissue samples from 55 primary tumors of lung cancer (35 lung adenocarcinoma, 20 lung squamous cell carcinoma - hereafter denoted LUSC). The material was derived from FFPE surgical resections at the Medical University of Lublin, Poland. At the moment of diagnosis and surgical resection of the primary cancer lesions, none of the patients had received neoadjuvant therapies. We collected clinical and demographic patients' data in a manner that protected their personal information. The study protocol received ethical approval from the Ethics Committee of the Medical University of Lublin, Poland (no KE-0254/235/2016).

Extraction and annotation of the training dataset for ARA-CNN
We extracted the training dataset from the H&E slides sourced from 55 lung cancer patients in total, with 1 slide per patient (Fig. 1a). From these 55, 26 were selected and from them regions of contiguous tissue were annotated using QuPath [36] by an expert pathologist, marking them as one of the following nine classes: *tumor* with neoplastic epithelial cells; *stroma* composed of connective tissue within tumor or extra-tumoral connective tissue; *mixed* where connective tissue was strongly infiltrated with immune cells; *immune* composed of lymphocytes and plasma cells or fragments of pulmonary

lymph nodes; *vessel* composed of smooth muscle layers (veins and arteries) with red blood cells within lumen; *bronchi* composed of cartilage and bronchial mucosa; *necrosis* including necrotic tissue or necrotic debris; *lung* (lung parenchyma); and *background* of the tissue scan (no tissue). In the 26 annotated slides, 13 were from LUAD, 10 from LUSC, 2 from large cell carcinoma and 1 from small-cell lung cancer patients. The TME in the original slides differed between patients, which gave us a diverse set of training examples (Sup. Figure 1 in Additional file 1). Some of the slides were more covered by tumor and necrotic cells or stroma, while in others immune infiltration, vessels or mixed class were dominating. In most of the slides we observed the "normal" lung structures, so bronchi was less common and needed more training data from many sections. All annotated regions were chosen for the purpose of providing the best material for model training. To this end, for a given class, we were annotating tissue that was undoubtedly of that class, and there was enough of that tissue visible in the slides to provide enough annotated patches. For example, for the *vessel* class, we did not consider arterioles, dilated capillaries or venules, as these tissues were too small for the chosen patch scale. Lymphatics were ignored due to them being imperceptible on H&E slides. For the *immune* class, intrapulmonary lymph nodes were included due to their high concentration of well-visible lymphocytes, even though the presence of such lymph nodes is not correlated with the tumor's immune response.

The annotated regions were then traced by a moving window, which cut out non-overlapping square patches of tissue with side size of 87 μm (which corresponded to 172 px). In addition to 87 μm, we also tested the training performance for patches with sizes of 74 μm and 100 μm (see Sup. Table 1–3 in Additional file 1). They resulted in worse performance (mean classification accuracy 84.64% for 74 μm, 84.35% for 100 μm, 85.21% for 87 μm), so we proceeded with using the 87 μm sized ones. This gave us an initial version of the training dataset, which was then improved upon by utilizing human-in-the-loop active learning, as part of the previously proposed accurate, reliable and active (ARA) image classification framework [37] (Fig. 1b; see below). In total, we ended up with 23,199 patches, divided in the following manner: 3311 *tumor* patches, as well as 1511 *stroma*, 716 *mixed*, 1196 *immune*, 1236 *vessel*, 2030 *bronchi*, 4448 *necrosis*, 6031 *lung*, and 2211 *background* patches.

Training and validation of the ARA-CNN model
The main component of the ARA framework is ARA-CNN, a Convolutional Neural Network (CNN) architecture inspired in part by Microsoft ResNet [38] and DarkNet 19 [39]. It includes standard techniques for
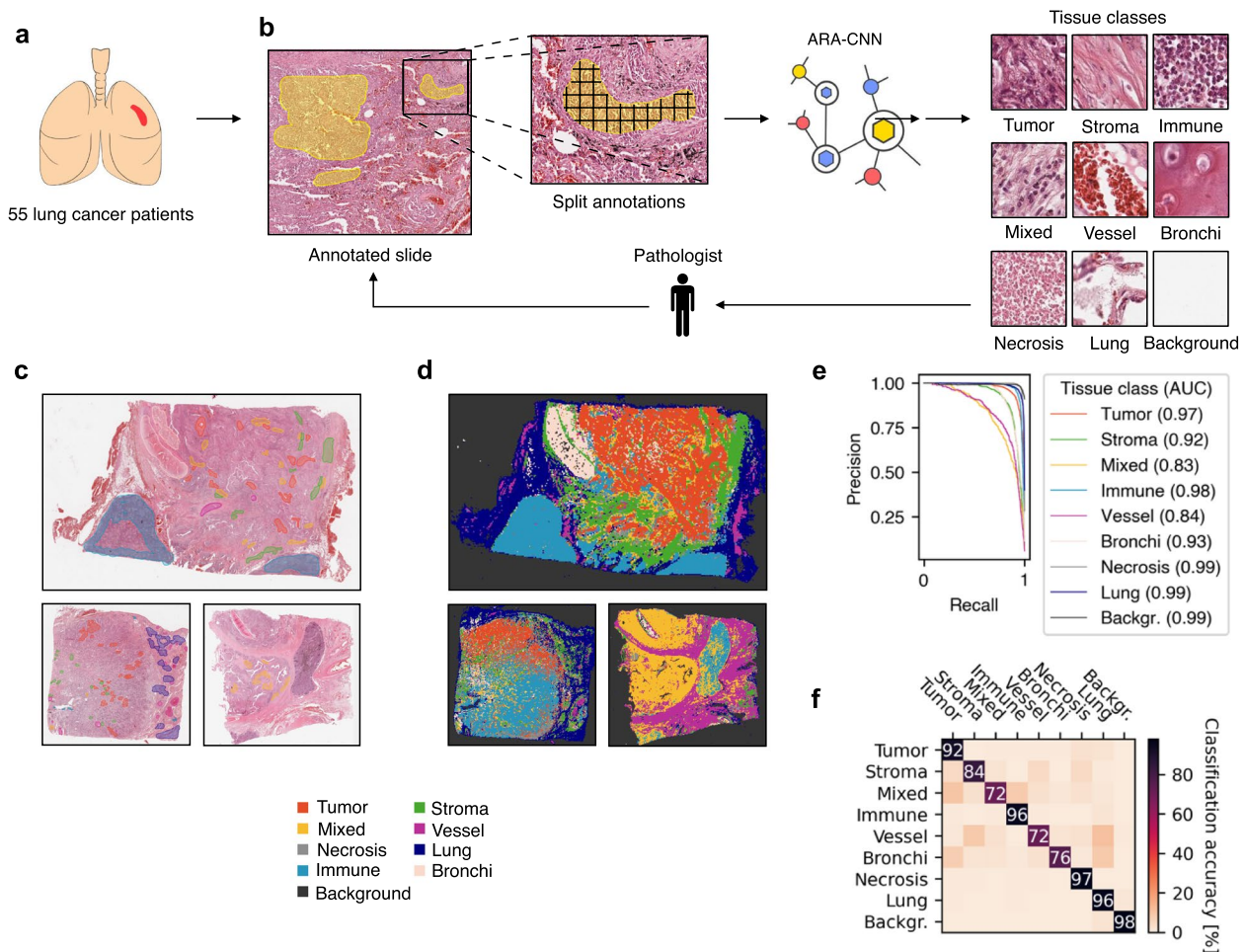
Rączkowska *et al. BMC Cancer*   (2022) 22:1001

Page 4 of 18

Fig. 1 Overview of training ARA-CNN for lung cancer tissue classification. **a** We sourced H&E tissue slides from 55 lung cancer patients. **b** 26 of these slides were annotated by an expert pathologist in an active learning loop with ARA-CNN, which resulted in the *LubLung* dataset and a trained tissue classification model. **c** Example annotations of various tissue regions. **d** Segmentation results from ARA-CNN show that tissue heterogeneity in the TME is captured correctly. **e** Precision-recall curves for each tissue class obtained in a 10-fold cross-validation scheme on the *LubLung* dataset. The mean AUC is 0.94. **f** Confusion matrix for ARA-CNN trained with *LubLung*. Row labels indicate true classes, while column labels describe classes predicted by the model

such models: Batch Normalization [40] (for normalization and to reduce overfitting) and dropout [41] (to reduce overfitting). The latter allowed us to apply variational dropout [42] during testing. Variational dropout is used to estimate uncertainty for every input image, which is returned together with its predicted class. For the overview of the model architecture, refer to Fig. S1 in Rączkowska et al. [37].

The model was trained in three active learning rounds, each one improving upon the previous ones. After the first training process, the distribution of uncertainty for images in each class was measured separately and due to higher median uncertainty, it was concluded that there are three classes in need of more training

examples: *mixed*, *vessel* and *bronchi*. These were passed on to the pathologist, who labeled new regions belonging to these classes. The resulting new training samples were extracted and added to the previous training dataset. After each iteration of the adaptive training procedure, the uncertainty for each class was measured again. After three iterations it was decided that the uncertainty results, with median uncertainty level below 1.5 for each class, were at a satisfactory level (uncertainty was measured using the entropy of the predictions [37]; see Sup. Figure 2 for the obtained entropy levels after the last iteration).

For training and evaluation of the ARA-CNN model on the lung cancer tissue patches, we used stratified 10-fold

Rączkowska *et al. BMC Cancer*   (2022) 22:1001

Page 5 of 18

cross-validation. This was the case during the active learning process as well. In a given iteration, the whole gathered dataset of images was split into a training dataset and a test dataset used for evaluation. For the final evaluation, the full dataset of 23,199 patches was divided into the test dataset containing 2316 patches and the training dataset consisting of 20,883 patches. Each class was split in exactly the same proportion: 10% were sent to the test dataset and 90% to the training dataset.

Additionally, in each training epoch the training data was split into two datasets: the actual training data and a validation dataset. The latter was used for informing the learning rate reducer - we monitored the accuracy on the validation set and if it stopped improving, the learning rate was reduced by a factor of 0.1. This split was in proportion 90 to 10% between actual training data and the validation set, respectively.

Each training process included real-time data augmentation, chosen at random from the following set of transformations: horizontal flip, vertical flip, rotation (up to 90 degrees), zoom (up to 40%), width shift (up to 10%) and height shift (up to 10%).

For parameter optimisation, we used the Adam optimiser [43]. The training time was set to 100 epochs. The training data was passed to the network in batches of 32, while the validation and test data was split into batches of 128 images. The loss function used during training was the categorical cross-entropy. The final ARA-CNN model was trained on the whole dataset of 23,199 patches (which was still split into the training and validation parts in proportion 90 to 10%). This final model was used in further experiments, segmentation of slides from an independent TCGA dataset, predictions of mutations and patient survival (see below).

TCGA data extraction and processing

Independent patient data, including H&E images, mutations, and clinical information was extracted from the TCGA database (as of 2020-05-14) through a REST API provided by TCGA. The database contained 478 LUAD cancer patients with at least one H&E tissue slide per patient. Out of these, frozen tissue slides were filtered out, which left 514 images. We developed a parallelized pipeline that downloaded the slides, ran all necessary calculations and then removed the processed images.

Each processed slide was split into non-overlapping patches with side size of 87 μm, same as for the patches in *LubLung*. As an optimisation step, we filtered the extracted patches and excluded the ones where most of the area was empty. To perform this filtering, we first converted each patch to grayscale using the standard Rec. 601 luma formula. Then, we mapped each pixel to either black (pixel value < 200) or white (pixel value ≥ 200) and counted them. The patch was deemed as relevant if the proportion of black pixels to white pixels was larger than 0.05.

In addition to the slides, clinical and mutation data was also extracted from TCGA. The former was downloaded using the curatedTCGAData R package [44] and it contained data for 518 LUAD patients. We removed data for Asian patients, where the race was determined by the 'race' column present in the TCGA clinical data. These patients are noted to be very distinct when it comes to disease progression [45, 46] and mutational profile. In particular, for Asians *EGFR* is mutated in up to 45% of cases, while for Caucasians it is closer to 10%. These factors and the low number of such cases when compared to the whole dataset contributed to omitting Asian patients in further analysis. This left us with clinical data for 510 patients. Mutation data for 563 LUAD patients was downloaded from the UCSC Xena Browser [47] (dated 07-20-2019), in the form of a *TCGA-LUAD. muse_snv.tsv* file. In addition, it was downsized to include only genes selected as relevant to lung cancer. To this end, we examined a set of genes that are either known to be frequently mutated in lung cancer and are important for patient prognosis and treatment, as characterized by the SureSelect Cancer All-In-One Catalog and Custom Assays [48], or were studied previously by Kather et al. [34]. Specifically, we selected such genes from the Kather et al. study, which showed an AUC higher than 70% or *p-value* < 0.001 in the task of mutation prediction by a deep learning model. From this set of 24 genes, we filtered out such genes for which there were only up to 15 LUAD patients carrying their mutation. This resulted in the following set of 13 selected genes: *ALK, BRAF, DDR2, EGFR, KEAP1, KRAS, MET, PIK3CA, RET, ROS1, STK11, TP53, PDGFRB*.

All datasets were merged together by the TCGA patient identifier. This left an intersection between the image, clinical and mutation datasets, which contained 444 patients and 506 slides.

As a last step, the clinical variables were pre-processed in the following manner. Age was quantified into two groups: 65 years and older, as well as younger than 65 years. Sex was set to 1 for male and 0 for female patients. Pack years were quantified into three groups: non-smoker (0 pack years or smoking history set as 'lifelong non-smoker' or, if information about pack years was missing, smoking history set to 'current reformed smoker for > 15 years'), light smoker (less than 30 pack years or, if information about pack years was missing, smoking history set as 'current reformed smoker for < or = 15 years') and heavy smoker (30

Rączkowska *et al. BMC Cancer*   (2022) 22:1001

Page 6 of 18

or more pack years, or smoking history set to 'current smoker'). Pathologic stage was mapped into three groups as well: early (stage I, Ia, Ib), locally advanced (stage II, IIa, IIb, IIIa) and advanced (stage IIIb, IV).

Due to its limited availability, we did not include treatment information in the main dataset. For the sake of additional analyses, we prepared a smaller dataset, containing 377 patients, with treatment data included (Sup. Methods in Additional File 1).

TCGA H&E patch normalization

Due to the fact that the tissue slides stored in TCGA exhibit high color variation, they needed to be normalized to a common color space, matching that of the training dataset. Three normalization algorithms were considered: Reinhard et al. [49], Macenko et al. [50] and Vahadane et al. [51]. To decide which of these three should be used on the TCGA data, a series of experiments was conducted, in which the *LubLung* training dataset was normalized with each of these algorithms and then ARA-CNN was trained in a cross-validation schema. The results showed that the best classification performance (mean accuracy 81.52% for Macenko et al., 82.39% for Vahadane et al., 85.76% for Reinhard et al.) was achieved for the dataset variant normalized with the Reinhardt et al. algorithm. Consequently, each relevant patch extracted from the TCGA slides was normalized individually with the Reinhardt et al. procedure. The normalization was performed with a region of interest image selected at random from the training dataset. All image patches from the TCGA database were transformed to match the color space of that image.

TCGA image data segmentation using ARA-CNN

The normalized patches served as input to ARA-CNN. For each input patch, the model returned a classification probability into each of the nine predefined classes. With these results, each patch was labeled with the class with the highest probability and then the labeled patches were merged back into their full respective slides and colored by the label. This created segmented slides, with clearly visible continuous areas of differing tissue.

The segmented slides were next validated by an expert pathologist, who assessed that 39 slides needed to be excluded from further analysis. There were two reasons for that. The first one involved erroneous classifications returned by ARA-CNN - 21 out of 506 slides contained errors of such nature. The other 18 slides were excluded

due to colored ink markings and other staining errors. After this process, the final dataset contained 467 slides from 411 patients.

Quantification of spatial features for the segmented tumor tissues

The obtained segmented images from TCGA were then processed further in order to extract spatial information in the form of two types of features, which we referred to as *tissue prevalence* (TIP) and *tumor microenvironment composition* (TMEC). TIP is a distribution of tissue classes within the whole tissue area, i.e. excluding the background class. TMEC measures a distribution of tissues that neighbor the tumor tissue within a predefined margin.

The prevalence t_i of tissue i is expressed as:

$$t_i = \frac{n_i}{N},$$

where n_i is the number of patches for tissue i and N is the total number of tissue patches (excluding the background class) and $i \in$ {*TUMOR*, *STROMA*, *MIXED*, *IMMUNE*, *VESSEL*, *BRONCHI*, *NECROSIS*, *LUNG*}. The vector with entries given by t_i makes up the TIP features. The background class was omitted, as it's not relevant to the tissue structure.

The microenvironment composition m_j for tissue j is:

$$m_j = \frac{b_j}{B},$$

where b_j is the number of patches of class j that neighbor the tumor class and B is the total number of all patches neighboring the tumor class (excluding the tumor itself and the background class), with $j \in$ {*STROMA*, *MIXED*, *IMMUNE*, *VESSEL*, *BRONCHI*, *NECROSIS*, *LUNG*}. The TMEC features are organized in a vector with m_j as its entries. The neighbor patches are considered only within a margin around the borders of tumor regions. Each tumor patch is considered separately and up to eight neighbors around it are counted. These patches are summed up to b_j for each class j and to B in total.

Using the microenvironment and prevalence data, we also calculated three spatial metrics that were previously defined in the literature: intra-tumor lymphocyte ratio (ITLR) [21], Simpson diversity index [52], Shannon diversity index [53]. We used a simplified version of these metrics - instead of cell-wise, we calculated them patch-wise. Specifically, these metrics were computed as follows:

Rączkowska *et al. BMC Cancer*   (2022) 22:1001

Page 7 of 18

$$ITLR = \frac{b_{IMMUNE}}{n_{TUMOR}}$$

$$Shannon = -\sum_i t_i \log(t_i)$$

$$Simpson = \sum_i t_i^2,$$

where b_{IMMUNE} is the number of immune patches that neighbor the tumor and t_{TUMOR} is the number of tumor patches in the whole slide.

Multivariate survival modeling using Cox model

The aforementioned predictors were used as input to the Cox proportional hazards model. They were organized into the following basic variants: clinical, clinical + ITLR, clinical + Shannon diversity index, clinical + Simpson diversity index, clinical + TMEC, clinical + TIP, clinical + TMEC + TIP. In addition, variants with mutation data added on top of clinical data were considered. Each variant was trained in a 10-fold cross-validation schema. For categorical variables, the hazard ratio of their basal values was set to 1. For the sex variable, the basal value was 'Female'. For the Stage variable, the basal value was 'Early stage'. For mutation variables (*EGFR*, *STK11* and *TP53*) the basal value was the absence of alteration. Finally, for smoking status, non-smoker was set as basal.

Mutation classification

The processed data from TCGA served as input in the mutation classification task. The predictor variables were the same as in the survival prediction task (minus mutation status). The response variables were binary and were defined by the mutation status for the 13 previously chosen frequently mutated LUAD genes (see TCGA data extraction and processing).

For each classification task, where the class was specified by the presence of mutation of a given gene, the dataset was oversampled so that positive (mutation occurred) and negative (mutation did not occur) subsets of examples were equal in size. Oversampling was done by inserting multiple copies of the positive examples so that their number reached that of the negative ones.

Eight combinations of predictive features were tested: clinical, clinical + ITLR, clinical + Shannon diversity index, clinical + Simpson diversity index, clinical + TMEC, clinical + TIP, clinical + TMEC + TIP, TMEC only. To classify the mutation status for each gene, two distinct machine learning models were trained and compared. The first one was a simple linear model in the form of logistic regression. It was fitted using the Liblinear solver [54], with the L2 (ridge) penalty [55] and up to 2000 iterations. The second one was the Random Forest algorithm [56]. We used the implementation from the sklearn Python library [57] with default parameter values.

All models were trained 100 times with 10-fold cross-validation and the resulting classification accuracy metrics were averaged. Classification performance was evaluated using the AUC metric.

Results

Validation of ARA-CNN

To quantify the classification performance of ARA-CNN, we first inspected how well the final trained model performs in segmenting the whole *LubLung* H&E slides. The segmentation allowed to correctly capture the TME heterogeneity in terms of all trained classes, which was confirmed by an expert pathologist who compared the original H&E slides with the final output of the model (Fig. 1c, d). Next, we used a 10-fold cross-validation procedure on the final set of 23,199 annotated patches obtained in the *LubLung* dataset (Methods). The best performance in a single class versus rest classification was achieved for the background, lung, necrosis, tumor, and immune classes (area under the curve, AUC range: 0.97–0.99) (Fig. 1e). The lowest AUC (0.83) was obtained for the mixed class, which is not surprising given that it is a tissue that is a mix of two other classes (stroma and immune). We then computed a confusion matrix, which confirmed that the best trained classes were background, necrosis, lung, immune and tumor (accuracy range: 92.36%–98.01) (Fig. 1f). In terms of errors, the model most often confused the mixed class with tumor (9.72% of the patches annotated as mixed were classified as tumor) or immune (8.17% of the patches); the vessel class with stroma or lung (8.73 and 10.79% of the patches, respectively); and the bronchi class with tumor or lung (7.30 and 8.53% of the patches, respectively). Given that patches of these classes were also often hard to distinguish by an expert pathologist, we conclude that our trained ARA-CNN model can reliably classify different tissue types in H&E images of LUAD and LUSC tissue sections.

Identification of TME spatial composition features in TCGA slides

We then sought to apply our trained ARA-CNN model to study the spatial architecture of the TME in H&E images from 411 LUAD patients downloaded from the TCGA database (Methods and Additional file 2). Due to the fact that LUAD is more affected by genetic alterations, we focused the further analysis on this particular subtype of lung cancer. We split each image into 87×87 µm patches and then normalized each patch to the same color space as the images in the *LubLung* dataset (Methods).

Rączkowska *et al. BMC Cancer*   (2022) 22:1001

Page 8 of 18

We used each patch as input to our ARA-CNN model, which returned the probabilities of assigning each patch to one of the nine tissue classes. We then segmented each image by assigning the most probable class to each patch (Methods). For each image, we computed two sets of human-interpretable features that reflect the spatial structure of the TME: tissue prevalence (TIP) and tumor microenvironment composition (TMEC) (Methods and Fig. 2a). TIP is represented by a vector of values t_i, computed as the fraction of patches assigned to class i out of all non-background patches in the whole slide image. TMEC is represented by a vector of values m_i, computed as the fraction of patches assigned to class i out of all non-tumor and non-background tissue types in a predefined margin around the tumor tissue.

Across the investigated tissue classes, tumor and lung classes dominated the entire tissue composition, with a median t_{TUMOR} of 0.36 and a median t_{LUNG} of 0.22 (Fig. 2b). The next three most abundant classes in the LUAD slides were mixed, immune and bronchi (with median prevalence of around 0.07). Finally, the least abundant classes were stroma, vessel, and necrosis. The most dominant classes of the tumor microenvironment were lung (median $m_{LUNG}=0.32$), immune (median $m_{IMMUNE}=0.18$) and mixed (median $m_{MIXED}=0.17$). These classes were followed by bronchi (median $m_{BRONCHI}=0.11$). The least abundant in the tumor microenvironment were stroma, vessel and necrosis classes. This indicates that in many patients, the tumor is surrounded by normal lung tissue and is confronted with an immune response. The abundance of all features, however, showed large variability across the analyzed TCGA slides, indicating high heterogeneity of both the entire tissue and the tumor microenvironment composition.

TME features are predictive of patient survival

We then explored if our spatial features can be used to predict patient survival (Fig. 2c), given that the composition of the TME has been previously shown to influence disease aggressiveness and survival in various cancer types [4, 58]. To this end, we first stratified the 411 LUAD patients into two groups based on their TIP and TMEC feature levels (High vs. Low). The stratification was performed using the survminer R package, which selects the cut-off point between high and low values based on the significance to the survival outcome. Specifically, the method implements a test of independence of a response variable and the given feature using maximally selected rank statistics. For each feature, we compared survival between the two groups using the Kaplan-Meier estimator. Six TIP features (vessel $p=0.0016$, immune $p=0.0058$, necrosis $p=0.0001$, stroma $p=0.0352$, bronchi $p=0.0079$ and mixed $p=0.0040$) and five TMEC features (vessel $p=0.0001$, immune $p=0.0045$, necrosis
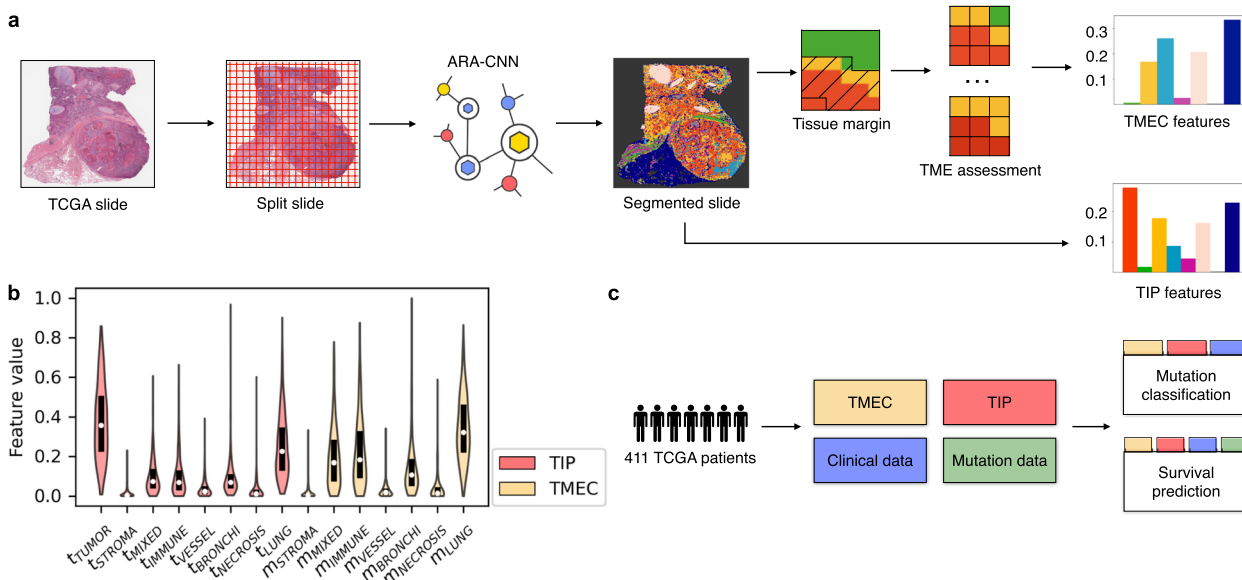

Fig. 2 Calculation and utilization of TIP and TMEC features. **a** H&E slides from TCGA were downloaded and split into tissue patches. Each patch was classified with ARA-CNN, producing tissue segmentations. These segmentations were next used to calculate the TIP and TMEC features. **b** Distribution of individual component features in TIP and TMEC. The most often occurring features for TIP were t_{TUMOR} and t_{LUNG}. For TMEC, these were m_{LUNG}, m_{IMMUNE} and m_{MIXED}. **c** Tasks performed with the help of the TIP and TMEC features. In addition to the TIP and TMEC features, clinical and mutation data was also sourced from TCGA. These datasets were combined and served as input in two tasks: survival prediction and gene mutation classification. The results were compared to those obtained using previous spatial metrics instead of TIP and TMEC

Rączkowska *et al. BMC Cancer*   (2022) 22:1001

Page 9 of 18

$p = 0.0009$, stroma $p = 0.0086$, and bronchi $p = 0.0254$) showed statistically significant ($p < 0.05$, log rank test, two-sided) differences in survival between High and Low groups (Fig. 3a-k). Additionally, these same features were found to be significant in the Benjamini-Hochberg procedure (from the total of 15 TIP and TMEC features, significance confirmed if $p_r < c$, where p_r is a *p*-value ranked in ascending order), assuming the False Discovery Rate of 0.1 (Fig. 3a-k). The cut-off values identified for the TIP and TMEC features were chosen to yield the most significant stratification of patients into better and worse surviving groups. Additional Kaplan-Meier analysis, stratifying patients using median values as cut-offs for patient stratification, showed less clear differences in survival between the resulting patient subgroups (Sup. Figure 4 in Additional file 1). We thus consider the identified cut-offs for the features as important findings, since they give additional insights about the corresponding tissues and the way their prevalence or abundance in the proximity of the tumor might affect patient survival. The identified cut-offs are yet to be verified on additional cohorts to validate that they would again yield significant differences in patient survival.

To systematically assess the added value of the TIP or TMEC and to compare them to other predictive features, we trained several versions of a multivariate Cox proportional hazards model of the death hazard for the analyzed LUAD patients and assessed the performance of each model with Harrell's c-index [59]. The versions in question were based on different combinations of input features (Methods). The best performing model yielded a median c-index of 0.723 and included clinical data (age, sex, pathologic stage, and smoking status), *EGFR*, *STK11* and *TP53* gene mutations, as well as TIP features. Inclusion of TMEC instead of TIP features yielded the second best model, with a slightly lower, but still high c-index of 0.709 (Fig. 3l). All other models – including those based on spatial diversity metrics such as Shannon index [53], Simpson index [52] and ITLR [21] – resulted in lower c-index values. These results indicate that the TIP and TMEC features, which respectively reflect the repertoire of different tissues and their proportions across the entire examined tissue and across the TME, are superior to other spatial metrics in predicting patient survival.

Next, we inspected the two best performing Cox models for the association between TIP and TMEC features and the death hazard accounting for the context of other features. A hazard ratio of 1 for a given feature indicates that the feature has no effect on survival, whereas a feature with hazard ratio larger than 1 indicates an increased death hazard and, therefore, a negative impact on survival. According to the best performing model, high abundances of $t_{NECROSIS}$ and t_{VESSEL} features in the

H&E image were associated with increased hazard. Similarly, abundance of $t_{BRONCHI}$ and t_{STROMA} features had a negative effect on survival (Fig. 3m). In contrast, t_{IMMUNE} and t_{MIXED} features were associated with a decreased death hazard and therefore longer survival (Fig. 3m), in line with the established role of the immune system as a barrier against tumor progression [4, 19, 58]. Among mutation features, *TP53* and *STK11* mutations significantly increased ($p < 0.05$, Wald test, two-sided) the death hazard, in agreement with the results of the independent Kaplan-Meier analysis (Sup. Fig. 3 in Additional file 1). The second best model, trained with TMEC instead of TIP features yielded very similar results (Fig. 3n). The impact of clinical features on the hazard agreed with previously published results and our independent Kaplan-Meier analysis (Sup. Results in Additional file 1).

Since the TIP and TMEC features included in the Cox model were binarized into high and low values based on the found cut-offs, we further confirmed the utility of these cut-offs by investigating the results obtained from Cox models where the feature values were used as continuous variables (Sup. Fig. 5 in Additional file 1). For the continuous Cox models, only the features related to necrosis obtain significant *p*-values (with $p < 0.001$ in a two-sided Wald test, both for $t_{NECROSIS}$ and $m_{NECROSIS}$). The hazard ratios of the real-valued features are more difficult to interpret than for the binarized features, as the relation of their values to the values of the hazard ratios is less clear.

Finally, since the treatment information may be an additional, critical predictive factor for patient survival, we repeated the survival analysis presented in Fig. 3 l-n on a smaller dataset of LUAD patients for which the treatment information was available (Sup. Methods in Additional File 1). To this end, we first trained several versions of a multivariate Cox proportional hazards model of the death hazard for that smaller dataset and evaluated the performance of each model with the Harrell's c-index. The best performing model yielded a median c-index of 0.718 and included clinical data (age, sex, pathologic stage, and smoking status), *EGFR*, *STK11* and *TP53* gene mutations, treatment information (targeted therapy, chemotherapy, radiotherapy, combined chemo- and radiotherapy), as well as TIP features. The next best model, with c-index of 0.716, included the same set of features, sans treatment information. Then, replacing TIP features with TMEC features produced two next best models, where including treatment information yielded a c-index of 0.716 and excluding it led to a c-index of 0.707. All other models resulted in lower c-index values. However, in all cases, including treatment information improved the results when compared to feature sets without it (Sup. Fig. 6a). The best obtained c-index of 0.718 for the

dataset with treatment information was lower than the best c-index, 0.723, for the dataset without treatment, presented in Fig. 3l. However, these results should not be compared directly, as the datasets used to generate them differed in size.

Next, we inspected the two best performing Cox models that included treatment information as predictors in terms of inferred hazard ratios (Sup. Fig. 6b,c). The first model, utilizing TIP, yielded results very similar to those presented in Fig. 3m. In terms of the additional treatment features, combined chemo- and radiotherapy had a statistically significant positive impact on survival, while radiotherapy alone had a statistically significant negative impact (in both cases $p < 0.05$, Wald test, two-sided). We speculate that the latter is not an effect of the therapy itself, but rather of the patients' clinical state in general, which might have necessitated the use of this method of treatment [60]. Additionally, as expected, both chemotherapy and targeted molecular therapy were positive survival predictors (Sup. Fig. 5b). The second best model that included treatment information, trained with TMEC instead of TIP features, yielded again very similar results (Sup. Fig. 6c). Interestingly, including treatment information increased statistical significance of spatial image features that were found significant when the treatment was not included in the model, as p for $t_{NECROSIS}$, $m_{NECROSIS}$, m_{VESSEL} decreased when compared to results shown in Fig. 3m,n, although this finding may be a side effect of the reduction in the size of the dataset.

TME features are predictive of disease-relevant mutations
Next, we sought to investigate the association of the human-interpretable spatial composition features of H&E images with mutations in lung cancer genes (Fig. 2c). To this end, we trained classifiers for the mutation status of 13 genes that are frequently mutated in LUAD (Methods). We evaluated eight different feature sets (Methods) with two machine learning algorithms: logistic regression and random forest. Out of all 104 feature set and gene combinations, logistic regression was the better

performing algorithm in 55 cases, while random forest performed better in the remaining 47 cases, indicating that for some genes non-linear relationships between the predictive features may be relevant for prediction of their mutations (Table 1). For 8 out of 13 considered genes (namely, *RET*, *KRAS*, *KEAP1*, *TP53*, *BRAF*, *PDGFRB*, *ROS1*, *STK11*), using the TIP or TMEC features gave the best result. For the remaining 5 genes (*MET*, *ALK*, *DDR2*, *PIK3CA*, *EGFR*), the best AUC was reached for models that utilized one of previously existing spatial metrics as features.

The best AUC (73.5%) was reached for the *PDGFRB* gene mutation by a classifier using clinical data and both TIP and TMEC as features (Table 1). The best model without TIP and TMEC, and with the Simpson metric as a feature, yielded an AUC smaller by 3.4 percentage points (p.p.). This shows that for the *PDGFRB* gene mutation, the full information about tissue distribution, not reduced to a single value using entropy and without focusing on only selected tissues, is highly relevant for its mutation status. The classification performance of the best model using both TIP and TMEC for that gene is only slightly smaller than AUC of 75%, as previously reported for a deep learning model trained on raw H&E images [34], but is less difficult to interpret. For eight other genes (*RET*, *KRAS*, *KEAP1*, *ROS1*, *STK11*, *MET*, *ALK*, *EGFR*), the best AUC ranged between 60 and 70%, while for the four remaining ones (*TP53*, *BRAF*, *DDR2*, *PIK3CA*) the best AUC ranged between 55 and 60%. For some of the genes, the inclusion of TIP or TMEC features resulted in impressive improvements compared to other feature sets. For *RET*, the model trained with clinical data and TMEC outperformed the best model without TIP and TMEC features, but including the Shannon metric, by around 9.1 p.p. Similarly, for *KEAP1* the classification performance increased by 7 p.p. compared to models without TIP or TMEC. These results indicate that, in LUAD, there exists a subset of tumor mutations that correlate with how the TME is structured, and that both TIP and TMEC features are predictive of the presence of these mutations.

(See figure on next page.)
Fig. 3 Survival prediction results. **a-k** Kaplan-Meier plots for TIP and TMEC features that result in patient stratification into two groups: with high and low values of the feature. Only features with statistically significant differences in patient survival are shown, as measured using the log rank test and the Benjamini-Hochberg procedure (*p*-values and critical values c in the top right corner, significance confirmed if $p < 0.05$ or $p_r < c$, where p_r is a *p*-value ranked in ascending order). For the latter, we set the False Discovery Rate at 0.1 and included all TIP and TMEC features. The cutoff value ρ (lower left corner) indicates the selected threshold yielding patient strata with high and low values of the feature. The results correlate with previous studies of the relationship between these features and patient survival. **l** c-index scores for Cox models from survival prediction experiments performed with different feature sets. The best results were obtained for models with such feature sets that included TIP and TMEC features. **m** Hazard ratios for the best model that utilized the TIP features. The prevalence of the necrosis tissue class in the whole slide has a statistically significant negative effect on survival. **n** Hazard ratios for the best model that utilized the TMEC features. The presence of the necrosis tissue class and the vessel tissue class in the TME has a statistically significant negative effect on survival
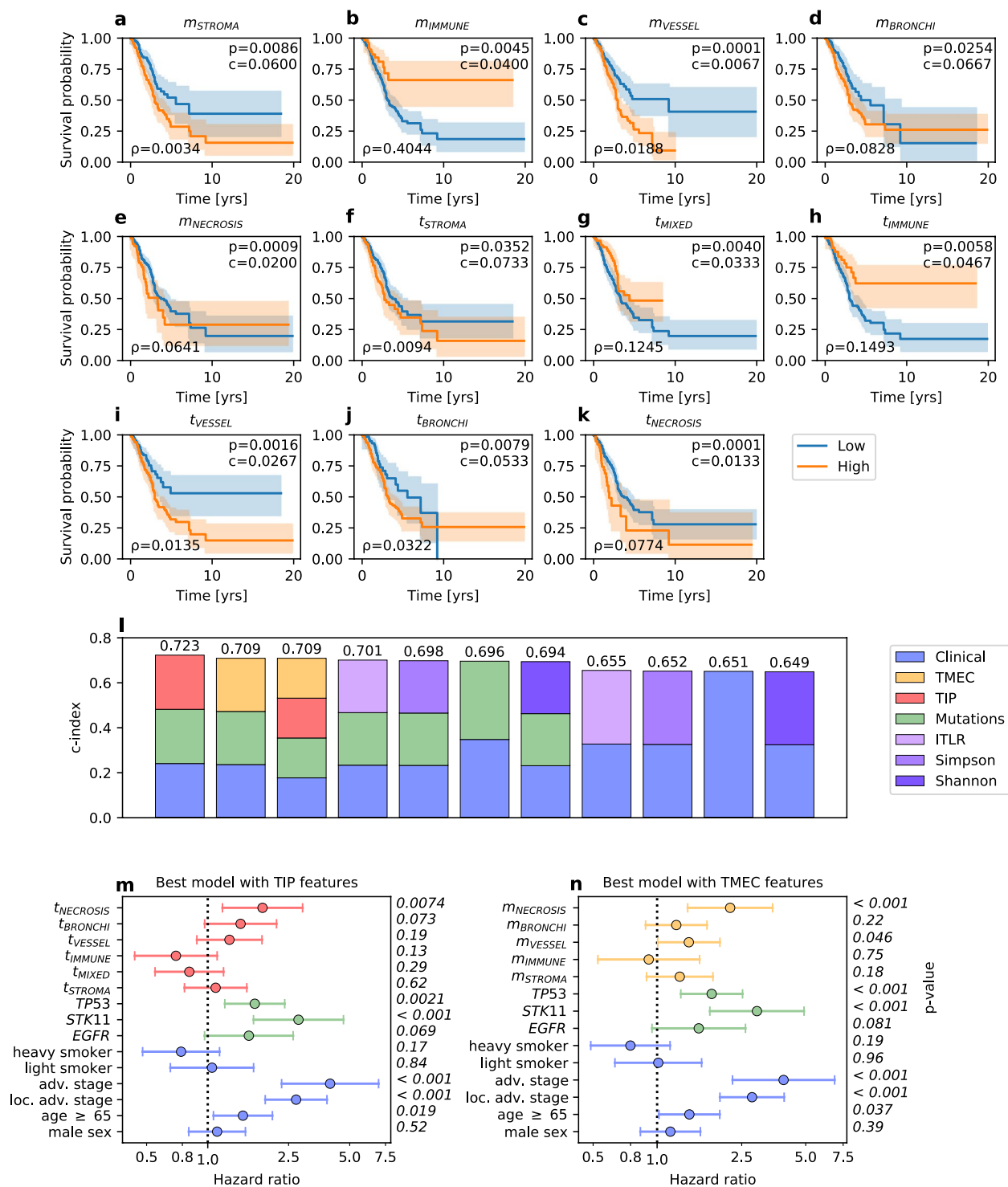
Rączkowska *et al. BMC Cancer*   (2022) 22:1001

Page 11 of 18

Fig. 3 (See legend on previous page.)

We then inspected the two best performing models in the mutation classification task that utilized TIP and TMEC features to find which predictor features were the most important for identifying mutations. Both of the algorithms used – logistic regression and random forest – are easily interpretable because they allow effective

Table 1 Mutation/rearrangement classification AUC scores (given as % of area under the precision-recall curve) for TCGA LUAD patients. The best result for each gene is marked in bold. In cases where the random forest classifier gave the best result, the cells are colored in yellow. Otherwise, if logistic regression gave the best result, the cells are colored in light blue

Gene	Mutation count	Clinical	Clinical + ITLR	Clinical + Shannon	Clinical + Simpson	Clinical + TMEC	Clinical + TIP	Clinical + TMEC + TIP	TMEC
RET	25	53.70 ± 2.21	50.76 ± 2.39	58.03 ± 2.13	57.76 ± 2.09	**67.14 ± 2.75**	59.46 ± 2.13	65.32 ± 1.95	64.39 ± 2.73
KRAS	96	61.12 ± 0.88	61.36 ± 0.84	60.06 ± 1.04	60.51 ± 0.99	**62.37 ± 1.50**	62.07 ± 1.66	60.98 ± 1.71	55.58 ± 1.54
KEAP1	96	51.35 ± 1.12	53.98 ± 1.60	51.37 ± 1.17	50.87 ± 1.21	**61.07 ± 1.60**	60.33 ± 1.85	60.05 ± 1.58	56.85 ± 1.48
TP53	200	55.28 ± 0.74	54.38 ± 0.75	55.18 ± 0.78	55.02 ± 0.79	**58.59 ± 1.30**	57.87 ± 1.26	57.78 ± 1.52	57.14 ± 1.48
BRAF	30	55.38 ± 1.74	55.14 ± 2.49	56.59 ± 1.76	56.87 ± 1.74	55.26 ± 2.66	**57.10 ± 1.95**	56.23 ± 2.14	51.35 ± 2.88
PDGFRB	15	67.85 ± 2.51	69.75 ± 2.79	69.91 ± 3.23	70.06 ± 2.97	72.03 ± 2.04	72.78 ± 2.48	**73.50 ± 2.02**	52.13 ± 3.83
ROS1	18	58.29 ± 3.58	60.74 ± 2.76	51.27 ± 3.20	49.97 ± 3.20	59.44 ± 4.69	60.07 ± 2.55	**61.45 ± 4.54**	59.58 ± 4.08
STK11	44	53.98 ± 1.76	53.91 ± 1.67	53.56 ± 1.83	61.39 ± 2.22	58.41 ± 1.51	54.89 ± 2.64	59.22 ± 2.30	**62.89 ± 2.51**
MET	19	61.22 ± 2.31	**69.20 ± 2.31**	63.14 ± 2.39	64.19 ± 2.35	66.58 ± 3.70	68.10 ± 2.45	63.68 ± 2.76	63.36 ± 4.28
ALK	28	53.55 ± 2.23	48.42 ± 2.49	**62.64 ± 3.05**	59.51 ± 2.23	50.32 ± 2.88	61.53 ± 1.70	58.71 ± 1.95	49.41 ± 2.89
DDR2	18	57.17 ± 3.81	56.89 ± 4.21	**57.38 ± 3.80**	56.69 ± 3.82	55.63 ± 4.30	54.65 ± 3.96	52.56 ± 4.37	51.78 ± 5.13
PIK3CA	23	49.14 ± 1.78	51.30 ± 2.67	51.07 ± 2.11	**58.35 ± 2.78**	53.90 ± 3.27	56.96 ± 3.46	58.01 ± 3.18	56.15 ± 3.22
EGFR	49	**65.09 ± 1.50**	61.48 ± 2.49	63.32 ± 1.25	64.90 ± 1.97	59.01 ± 1.60	59.69 ± 1.56	57.93 ± 1.60	49.05 ± 1.97

identification of the most important features. First, we analyzed the logistic regression classifier of *PDGFRB* mutations with clinical, TMEC and TIP features (Fig. 4a). The most important features positively correlated with *PDGFRB* mutation were sex, m_{MIXED} – corresponding to the proportion of the mixed tissue in the tumor microenvironment – and t_{TUMOR} – corresponding to the fraction of the entire slide occupied by the tumor. The positive correlation of the male sex with *PDGFRB* mutation is not well explored. We hypothesize that there could be a relationship between tobacco carcinogens, to which male patients are more exposed, and the TME composition, which is corroborated by our results. On the other hand, the most negatively correlated (i.e., decreasing the chance of mutation) features were non-smoker status, t_{IMMUNE}, and $m_{BRONCHI}$. Next, we inspected the random forest classifier of *RET* mutations, which included clinical and TMEC features in its feature set (Fig. 4b). The latter proved to be of larger importance than the former ones. Indeed, *RET* mutations were found to be most associated with the prevalence of different tissues in the tumor microenvironment, with bronchi and vessels identified as the most impactful tissues, followed by mixed, stroma, lung, immune and necrosis. This observation might be explained by the fact that, in LUAD, *RET* mutations

mainly consist of rearrangements between *RET* gene and its common fusion partners such as *KIF5B*, *CCDC6*, *CUX1*, *TRIM33*, *NCOA4*, *KIAA1468* and *KIAA1217* genes.

In addition to feature importance, we also inspected the distributions of the values of the TIP and TMEC spatial composition features for patients with and without mutations of the *PDGFRB* and *RET* genes. For both of them, we selected the four most important TIP or TMEC features and assessed their value distributions separately for mutated and non-mutated cases. For *PDGFRB*, these features were: m_{MIXED} and $m_{BRONCHI}$ (TMEC features), as well as t_{VESSEL} and t_{IMMUNE} (TIP features) (Fig. 4c). We detected a statistically significant difference between the value distributions (two-sided Wilcoxon test *p-value* < 0.05) for t_{VESSEL}. For *RET*, the four most important features were TMEC features $m_{BRONCHI}$, m_{MIXED}, m_{VESSEL} and m_{STROMA} with m_{MIXED} and $m_{BRONCHI}$ features having a statistically significant difference in value distributions between mutated and non-mutated tumors (Fig. 4d). These results indicate that the spatial composition features TIP and TMEC are different between tumors with and without *PDGFRB* and *RET* mutations, and their importance for the classification of mutations of these genes is not incidental.

(See figure on next page.)

Fig. 4 Feature importance for the two best performing mutation classification models that utilized TIP and TMEC features. **a** Feature importance for the *PDGFRB* gene mutation classifier (logistic regression). Here, feature importance is measured by the value of its regression coefficient. **b** Feature importance for the *RET* gene mutation classifier (random forest). Here, the importance is measured by the reduction of the Gini index obtained when the feature is added to the tree, averaged across the trees in the random forest model. **c** Distribution of feature values for four of the most important TIP or TMEC features, as presented in (**a**), divided between patients with the mutated and non-mutated *PDGFRB* gene. **d** Distribution of feature values for four of the most important TMEC features, as presented in (**b**), divided between patients with the mutated and non-mutated *RET* gene
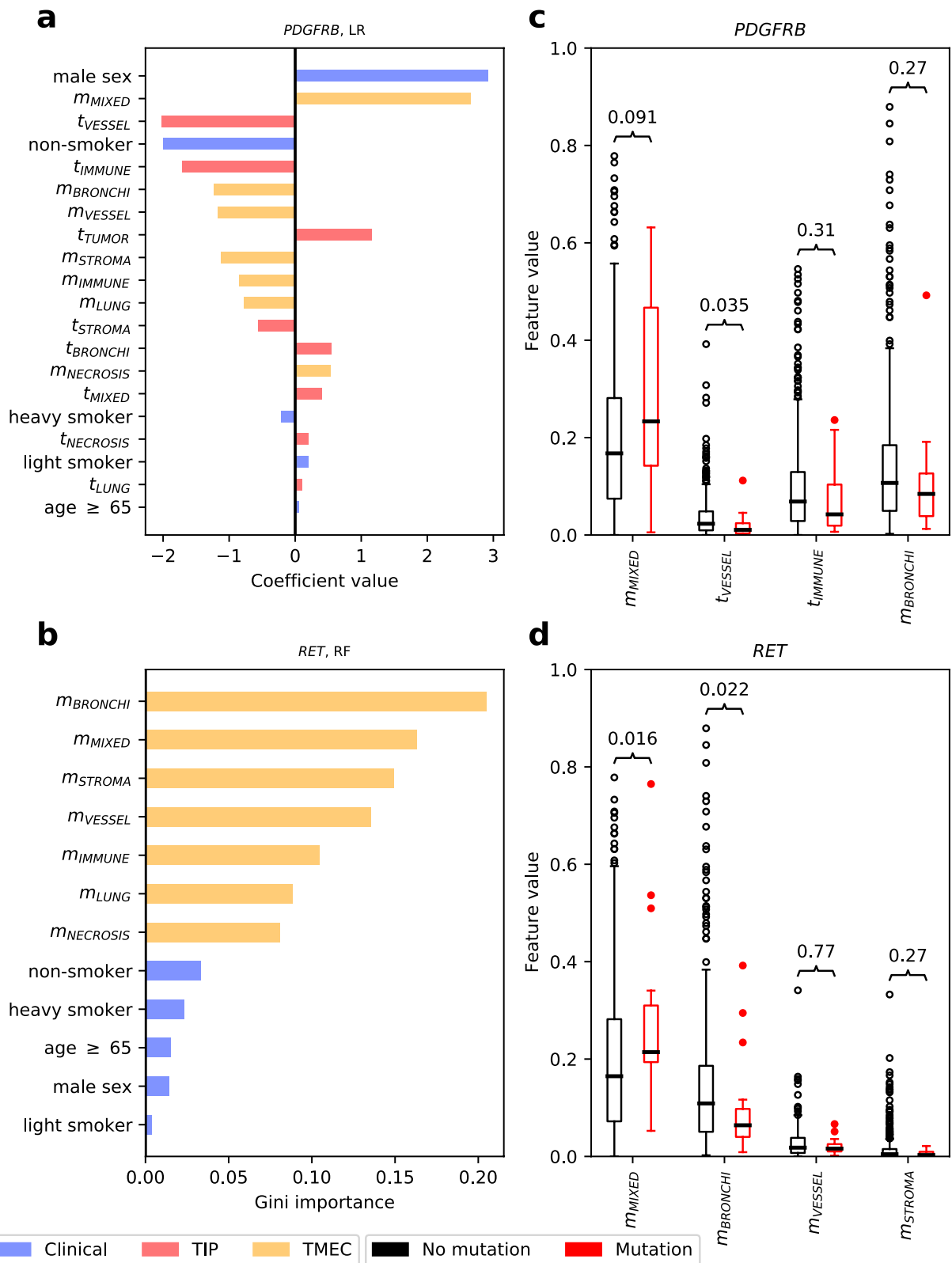
Fig. 4 (See legend on previous page.)

Rączkowska *et al. BMC Cancer*   (2022) 22:1001

Page 14 of 18

Discussion

We have developed a novel H&E image classification model, ARA-CNN, and a training dataset of annotated tissue patches from LUAD and LUSC H&E images, *LubLung*. Both considerably expand the current ability to analyze the TME automatically and quantitatively in lung cancer samples, which in turn has important implications for patient stratification and precision treatment. TIP and TMEC features, which we have introduced in this work, provide a novel way of capturing the composition and spatial structure of the TME, and are predictive of both overall survival and clinically relevant mutations. Spatial statistics of H&E images in the form of metrics that quantify colocalization of cell or tissue types, have been previously shown to be predictive of patient survival [21]. However, these metrics are computed based on a limited number of features, such as counts of tumor and immune cells. Other approaches that try to link the structure of tumor tissue and TME with either gene mutations or patient survival are end-to-end deep learning models and work as 'black boxes' [29, 30, 34, 61–63]. Instead, our approach allows explicit interpretability, as it decouples H&E slide inference from downstream tasks (e.g., mutation classification and survival analysis). The TIP and TMEC features are per se human interpretable, so it is possible to precisely pinpoint which tissue types are the most important. Our approach requires the initial tissue classification to be as accurate as possible. We ensured this to be the case by using ARA-CNN, which performs excellently in classifying nine tissue classes present in lung cancer H&E images. To foster further research in predictive spatial statistics based on a rich repertoire of segmented lung cancer tissues, in addition to *LubLung* we also share the segmented TCGA images as a separate dataset, named *SegLungTCGA*.

Our analysis revealed that patient stratification based on TIP and TMEC features yields significant differences in patient survival between the strata. Moreover, the most predictive survival models included TIP and TMEC features. These findings are supported by previous clinical studies. It has been shown that blood vessel invasion is a major prognostic factor in lung cancer survival [14, 64]. Similarly, there have been studies which proved that tumor necrosis is a significant risk factor for survival in lung cancer [65]. However, the complexity of the entire lung microenvironment plays a key role in the development of primary lung carcinomas and offers a resource of targets for personalized therapy development. Targeting the angiogenesis and immune cells has elucidated the prognostic and pathophysiological roles of other components of the TME in lung cancer [13, 66]. In the end, the combination of the clinical and genetic information with the TME landscape may play a pivotal role in predicting the type and duration of response to personalized therapies.

We found eight genes relevant to lung cancer (*PDGFRB*, *RET*, *KRAS*, *KEAP1*, *ROS1*, *STK11*, *MET* and *ALK*), for which integrating clinical data with our TME features clearly improves the ability to predict mutations in these genes. We speculate that mutations of these genes may alter cellular interactions, and hence the spatial arrangement of the TME visible in H&E images. For *RET*, *ROS1* and *ALK* genes, mutations mainly consist of chromosomal rearrangements which produce chimeric proteins that might affect the cellular organization within the TME [67–69]. Likewise, loss of *STK11/LKB1* overlapping with oncogenic *KRAS* mutations is associated with increased neutrophil recruitment, and decreased T-cells infiltration in lung cancer tumors [70]. Moreover, *STK11* mutations often coexist with *KEAP1* mutations that relate to cellular resistance to oxidative stress [71], and co-occurrence of *KEAP1* mutations and *PTEN* inactivation is an indicator of an immunologically "cold" tumor [72]. We speculate that each of these mutations might slightly affect the cellular morphology in H&E images in a way that is not apparent to the human eye, but can be captured by deep-learning algorithms.

Our findings concern mutations of clinically relevant genes, and as such may have clinical implications. For example, both *RET* and *PDGFRB* are clinically relevant LUAD cancer genes. *RET* has proto-oncogene properties and its fusions, which occur in 1–2% of LUAD [73], are associated with a high risk of brain metastasis [74]. However, last clinical trials indicated that they may be effectively targeted by *RET* tyrosine kinase inhibitors such as pralsetinib, selpercatinib [73]. *PDGFRB* is a member of the PDGF/PDGFR axis that is recognized as a key regulator of mesenchymal cell activity in TME [75], and several new agents (linifanib, motesanib, olaratumab) that block the PDGFR signaling are being tested in LUAD [76]. In breast, colon, pancreas and prostate cancers, the high stromal expression of the PDGFRβ protein has been associated with poor prognosis [76], however its prognostic relevance in tumors of epithelial origin is inconclusive [75]. It was only confirmed that a relative expression of PDGFRs is a strong and independent predictor of longer survival for surgical stages of lung cancer (I-IIIA) [76].

There is a difference in genetic landscape between centrally and peripherally located NSCLC that is affected by exposure to various environmental carcinogens [77]. For instance, exposure to tobacco carcinogens leads to centrally located NSCLC tumors with higher accumulation of alterations in suppressor genes [78]. On the other hand, oncogenic alterations such as EGFR-activating mutations are prone to occur in the peripheral location that is more common to LUAD [79, 80]. Our results

Rączkowska *et al. BMC Cancer*   (2022) 22:1001

Page 15 of 18

indicate that central locations of NSCLC (high neighborhood of bronchi) decrease the probability to detect RET alterations (Fig. 4b,d). This may suggest that RET alterations preferably occur in peripherally located LUADs, however it needs to be confirmed experimentally.

The analysis presented here shows that there is a correspondence between the spatial structure in H&E images for LUAD and both gene mutations and patient survival. Not every mutation is expected to have an effect on tissue prevalence or tumor neighborhood structure, so it is not surprising that for some of the analyzed genes the mutation classification performance did not exceed an AUC of 0.6. In contrast, it is striking that there are genes for which adding tissue composition data to the clinical information improves classification results. Finally, it is also surprising that our TIP and TMEC features, as well as other metrics of TME spatial organization, such as ITLR, can give good results in terms of both mutation classification and survival analysis. In summary, despite having several limitations, discussed in Additional File 1, our approach successfully identifies novel image features that are important for patient survival and mutations.

Conclusions

In this paper, we presented a framework that accurately predicted survival and gene mutations in LUAD patients based on human-interpretable features extracted from H&E slides. Our approach can provide important insights for designing novel cancer treatments, by linking the spatial structure of the tumor microenvironment in LUAD to gene mutations and patient survival. It can also expand our understanding of the effects that the tumor microenvironment has on tumor evolutionary processes. The presented framework is generalisable, so it can be extended to other tumor types. We therefore envision that, in the future, our quantitative approach will become incorporated in routine diagnostics for LUAD and other cancer types.

Abbreviations

TME	Tumor microenvironment
LUAD	Lung adenocarcinoma
LUSC	Lung squamous cell carcinoma
TIP	Tissue prevalence
TMEC	Tumor microenvironment composition
ITLR	Intra-tumor lymphocyte ratio
ARA	Accurate, reliable, active
CNN	Convolutional neural network
TCGA	The Cancer Genome Atlas
H&E	Hematoxylin and eosin

Supplementary Information

Acknowledgements
Not applicable.

Authors' contributions

AR performed all experiments, analyzed the results and prepared visualizations. IP was responsible for H&E image annotation and segmentation validation. AR, IP, MN and ES coordinated the active learning process. MK contributed to the work on survival analysis. MB provided code for extracting mutation data from TCGA. MN and PK helped with the interpretation of clinical data and provided feedback on the results. JS provided the access for the archival tissue material used for the model training. MN collected the material and prepared the tissue sections. TK performed the H&E staining and scanned all the slides used for the model training. ES and NC supervised the research. AR and ES wrote the manuscript. AR, ES, MN and NC conceptualized the project. All authors reviewed the manuscript. The author(s) read and approved the final manuscript.

Authors' information

AR is a PhD candidate at the Faculty of Mathematics, Informatics and Mechanics at the University of Warsaw. They graduated with a Master of Science in Engineering degree in Informatics from the Nicolaus Copernicus University in Toruń, Poland. They are now developing machine learning algorithms for biological data analysis, with a focus on deep learning for cancer imaging. Their research interests include machine learning, computer vision and high-performance computing.
IP is a pathologist at the Department of Clinical Pathomorphology and PhD graduate of the Medical University of Lublin, Poland (2011). In daily practice, she focuses on cytology, pulmonary and oncologic pathology.
MK graduated with a Master's degree in Bioinformatics and Biology Systems from the Faculty of Mathematics, Informatics and Mechanics at the University of Warsaw. His interest is in developing biomedical-related pipelines with strong focus on NLP and NGS applications. Currently, he works as a Bioinformatics Data Scientist in the area of precision medicine.
MN is a Master graduate (2012) and PhD graduate of the Medical University of Lublin, Poland (2016). In that period he studied the signatures of the "druggable" mutations in brain metastases of lung cancer. Then, in 2018–2020, he completed a postdoctoral fellowship in Nicola Crosetto Laboratory at the Karolinska Institutet and Science for Life Laboratory in Stockholm, Sweden where he gained experience in next-generation sequencing. Currently, he is an Assistant professor at the Department of Pneumonology, Oncology and Allergology, Medical University of Lublin, Poland leading two scientific projects. His main research interest is studying intratumor heterogeneity.
MB is a PhD graduate at the Sydney Medical School, The University of Sydney, Australia where she focused on deciphering pre-cancerous genomic alterations leading to development of liver cancer. She obtained her Master's degree in Biotechnology with a major in Bioinformatics from the Wroclaw University of Science and Technology. Currently, she is a Senior Bioinformatician at Ardigen working in an area of cancer biology.
TK is a biotechnologist and laboratory diagnostician with PhD regarding genetic determinants in lung cancer chemotherapy. He is a Master graduate of the Maria Sklodowska-Curie University in Lublin, Poland (2007) and PhD graduate of the Medical University in Lublin, Poland (2014). He is currently working at the Laboratory of Immunology and Genetics, at the Pulmonology Department of Medical University in Lublin, focusing his work on immunohistochemical analyses in lung cancer patients' treatment qualification.

Rączkowska *et al. BMC Cancer*   (2022) 22:1001

Page 16 of 18

JS is a professor at the Medical University of Lublin, heading the Chair and Department of Clinical Pathomorphology. She specializes in pathomorphology, especially hematopathology. Her research interests focus on digestive tract pathology, reproductive toxicology, congenital malformations and hematopathology.

PK is a professor at Medical University of Lublin, heading the Laboratory of Immunology and Genetics at the Department of Pneumonology, Oncology and Allergology. He specializes in internal medicine and laboratory diagnostics, with a focus on medical genetics and medical immunology. He is a member of the Scientific Council of the Medical University of Lublin and the National Institute of Oncology in Warsaw, and the Bioethical Committee of the Medical University of Lublin. His main interests focus on molecular and immunological predictive factors in the qualification of cancer patients to molecularly targeted therapies and immunotherapy.

NC received his M.D. from the University of Pavia, Italy in 2003, and his specialization in Medical Oncology from the University of Turin, Italy in 2007. From 2004 until 2010 he worked in the Dikic lab at the Goethe University in Frankfurt, Germany and in 2011 he obtained a PhD in molecular biology/biotechnology from the University of Pavia, Italy. From 2011 till 2014 he was a postdoc in the van Oudenaarden lab at MIT in Cambridge, MA. Since 2015, he has been an Assistant Professor and Group Leader at Karolinska Institutet and Science for Life Laboratory in Stockholm, Sweden. His main research interest is in chromatin organization and genome fragility, with a special focus on cancer.

ES is an assistant professor at the Faculty of Mathematics, Informatics and Mechanics at the University of Warsaw. She holds two Master degrees, one from the Uppsala University, Sweden and one from the University of Warsaw, Poland. She finished PhD studies at the Max Planck Institute for Molecular Genetics in Berlin, Germany and conducted postdoctoral research at ETH Zurich, Switzerland. She now leads a research group focusing on machine learning and molecular biology, with most applications in computational oncology. Her group works mainly with probabilistic graphical models and deep learning.

Availability of data and materials
The *LubLung* dataset is available publically at github.com/animgoeth/LubLung. The *SegLungTCGA* dataset is available at github.com/animgoeth/SegLungTCGA.

The ARA-CNN weights trained on *LubLung* are available at github.com/animgoeth/ARA-CNN, in the 'pretrained' directory.

The dataset of TCGA data used in this study plus both TIP and TMEC features computed from TCGA slides is included in this published article and its supplementary information files.

Declarations

Ethics approval and consent to participate
The study was performed in accordance with the Declaration of Helsinki. The study protocol received ethical approval from the Ethics Committee of the Medical University of Lublin, Poland (no. KE-0254/235/2016) and oral informed consent was collected prior to storing anonymized tissue samples. The analysis described here was retrospective. It was conducted with archived formalin-fixed paraffin embedded FFPE samples, devoid of any personalized patient data.

Consent for publication
Not applicable.

Competing interests
Other projects in the lab of ES are co-funded by Merck Healthcare KGaA.

Author details
[1]Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Banacha 2, 02-097 Warsaw, Poland. [2]Department of Clinical Pathomorphology, Medical University of Lublin, Jaczewskiego 8b, 20-090 Lublin, Poland. [3]Department of Pneumology, Oncology and Allergology, Medical University of Lublin, Jaczewskiego 8, 20-090 Lublin, Poland. [4]Ardigen, Podole 76, 30-394 Cracow, Poland. [5]Division of Genome Biology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Tomtebodavägen 23a, 17165 Solna, Sweden. [6]Science for Life Laboratory, Tomtebodavägen 23a, 17165 Solna, Sweden.

References
1. Fox H. Is H&E morphology coming to an end? J Clin Pathol. 2000;53(1):38–40.
2. Komura D, Ishikawa S. Machine Learning Methods for Histopathological Image Analysis. Comput Struct Biotechnol J. 2018;16:34–42.
3. Djuric U, Zadeh G, Aldape K, Diamandis P. Precision histology: how deep learning is poised to revitalize histomorphology for personalized cancer care. Npj Precis Oncol. 2017;1(1):1–5.
4. Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. Nat Med. 2018;24(5):541–50.
5. Griffiths AJ, Miller JH, Suzuki DT, Lewontin RC, Gelbart WM, Griffiths AJ, et al. An Introduction to Genetic Analysis. 7th ed. W. H. Freeman; 2000.
6. Abedi S, Janbabaei G, Afshari M, Moosazadeh M, Rashidi Alashti M, Hedayatizadeh-Omran A, et al. Estimating the Survival of Patients With Lung Cancer: What Is the Best Statistical Model? J Prev Med Pub Health. 2019;52(2):140–4.
7. Hassan MRA, Suan MAM, Soelar SA, Mohammed NS, Ismail I, Ahmad F. Survival Analysis and Prognostic Factors for Colorectal Cancer Patients in Malaysia. Asian Pac J Cancer Prev APJCP. 2016;17(7):3575–81.
8. Anderson NM, Simon MC. The tumor microenvironment. Curr Biol. 2020;30(16):R921–5.
9. Bade BC, Dela Cruz CS. Lung Cancer 2020: Epidemiology, Etiology, and Prevention. Clin Chest Med. 2020;41(1):1–24.
10. Ruiz-Cordero R, Devine WP. Targeted Therapy and Checkpoint Immunotherapy in Lung Cancer. Surg Pathol Clin. 2020;13(1):17–33.
11. Imyanitov EN, Iyevleva AG, Levchenko EV. Molecular testing and targeted therapy for non-small cell lung cancer: Current status and perspectives. Crit Rev Oncol Hematol. 2021;157:103194.
12. Duma N, Santana-Davila R, Molina JR. Non–Small Cell Lung Cancer: Epidemiology, Screening, Diagnosis, and Treatment. Mayo Clin Proc. 2019;94(8):1623–1640.
13. Altorki NK, Markowitz GJ, Gao D, Port JL, Saxena A, Stiles B, et al. The lung microenvironment: an important regulator of tumour growth and metastasis. Nat Rev Cancer. 2019;19(1):9–31.
14. Zhang C, Liu Y, Guo S, Zhang J. Different biomarkers in non-small cell lung cancer in blood vessel invasion. Cell Biochem Biophys. 2014;70(2):777–84.
15. Chan TA, Yarchoan M, Jaffee E, Swanton C, Quezada SA, Stenzinger A, et al. Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. Ann Oncol. 2019;30(1):44–56.
16. Rehman JA, Han G, Carvajal-Hausdorf DE, Wasserman BE, Pelekanou V, Mani NL, et al. Quantitative and pathologist-read comparison of the heterogeneity of programmed death-ligand 1 (PD-L1) expression in non-small cell lung cancer. Mod Pathol. 2017;30(3):340–9.
17. Huang YK, Wang M, Sun Y, Di Costanzo N, Mitchell C, Achuthan A, et al. Macrophage spatial heterogeneity in gastric cancer defined by multiplex immunohistochemistry. Nat Commun. 2019;10(1):3928.
18. Heindl A, Sestak I, Naidoo K, Cuzick J, Dowsett M, Yuan Y. Relevance of Spatial Heterogeneity of Immune Infiltration for Predicting Risk of Recurrence After Endocrine Therapy of ER+ Breast Cancer. JNCI J Natl Cancer Inst. 2018;110(2):166–75.
19. Rudolf J, Büttner-Herold M, Erlenbach-Wünsch K, Posselt R, Jessberger J, Haderlein M, et al. Regulatory T cells and cytotoxic T cells close to the epithelial–stromal interface are associated with a favorable prognosis. OncoImmunology. 2020;9(1):1746149.

Rączkowska *et al. BMC Cancer*   (2022) 22:1001

Page 17 of 18

20. Valous NA, Moraleda RR, Jäger D, Zörnig I, Halama N. Interrogating the microenvironmental landscape of tumors with computational image analysis approaches. Semin Immunol. 2020;48:101411.

21. Yuan Y. Modelling the spatial heterogeneity and molecular correlates of lymphocytic infiltration in triple-negative breast cancer. J R Soc Interface. 2015;12(103):20141153.

22. Saltz J. Spatial Organization and Molecular Correlation of Tumor-Infiltrating Lymphocytes Using Deep Learning on Pathology Images. 2018;21.

23. Yuan Y. Spatial Heterogeneity in the Tumor Microenvironment. Cold Spring Harb Perspect Med. 2016;6(8):a026583.

24. Heindl A, Lan C, Rodrigues DN, Koelble K, Yuan Y. Similarity and diversity of the tumor microenvironment in multiple metastases: critical implications for overall and progression-free survival of high-grade serous ovarian cancer. Oncotarget. 2016;7(44):71123–35.

25. Corredor G, Wang X, Zhou Y, Lu C, Fu P, Syrigos K, et al. Spatial Architecture and Arrangement of Tumor-Infiltrating Lymphocytes for Predicting Likelihood of Recurrence in Early-Stage Non–Small Cell Lung Cancer. Clin Cancer Res. 2019;25(5):1526–1534.

26. Xi KX, Wen YS, Zhu CM, Yu XY, Qin RQ, Zhang XW, et al. Tumor-stroma ratio (TSR) in non-small cell lung cancer (NSCLC) patients after lung resection is a prognostic factor for survival. J Thorac Dis. 2017;9(10):4017–4026.

27. Mobadersany P, Yousefi S, Amgad M, Gutman DA, Barnholtz-Sloan JS, Vega JEV, et al. Predicting cancer outcomes from histology and genomics using convolutional networks. Proc Natl Acad Sci U S A. 2018;115(13):E2970–9.

28. Yousefi S, Amrollahi F, Amgad M, Dong C, Lewis JE, Song C, et al. Predicting clinical outcomes from large scale cancer genomic profiles with deep survival models. Sci Rep. 2017;7(1):11707.

29. Chen M, Zhang B, Topatana W, Cao J, Zhu H, Juengpanich S, et al. Classification and mutation prediction based on histopathology H&E images in liver cancer using deep learning. Npj Precis Oncol. 2020;4(1):1–7.

30. Liao H, Long Y, Han R, Wang W, Xu L, Liao M, et al. Deep learning-based classification and mutation prediction from histopathological images of hepatocellular carcinoma. Clin Transl Med. 2020;10(2). [cited 2021 May 6] Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7403820/

31. Chen X, Lin X, Shen Q, Qian X. Combined Spiral Transformation and Model-Driven Multi-Modal Deep Learning Scheme for Automatic Prediction of TP53 Mutation in Pancreatic Cancer. IEEE Trans Med Imaging. 2021;40(2):735–47.

32. Wu Z, Huang X, Huang S, Ding X, Wang L. Direct Prediction of BRAFV600E Mutation from Histopathological Images in Papillary Thyroid Carcinoma with a Deep Learning Workflow. In: 2020 4th International Conference on Computer Science and Artificial Intelligence. New York: Association for Computing Machinery; 2020 [cited 2021 May 5]. p. 146–51. (CSAI 2020). Available from: https://doi.org/10.1145/3445815.3445840.

33. Schaumberg AJ, Rubin MA, Fuchs TJ. H&E-stained Whole Slide Image Deep Learning Predicts SPOP Mutation State in Prostate Cancer. Pathology; 2016 [cited 2021 May 17]. Available from: http://biorxiv.org/lookup/doi/10.1101/064279

34. Kather JN, Heij LR, Grabsch HI, Loeffler C, Echle A, Muti HS, et al. Pan-cancer image-based detection of clinically actionable genetic alterations. Nat Cancer. 2020;1(8):789–99.

35. Diao JA, Wang JK, Chui WF, Mountain V, Gullapally SC, Srinivasan R, et al. Human-interpretable image features derived from densely mapped cancer pathology slides predict diverse molecular phenotypes. Nat Commun. 2021;(1):1613.

36. Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, et al. QuPath: Open source software for digital pathology image analysis. Sci Rep. 2017;7(1):16878.

37. Rączkowska A, Możejko M, Zambonelli J, Szczurek E. ARA: accurate, reliable and active histopathological image classification framework with Bayesian deep learning. Sci Rep. 2019;9(1):14347.

38. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. ArXiv151203385 Cs. 2015 [cited 2021 Mar 21]; Available from: http://arxiv.org/abs/1512.03385

39. Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. p. 6517–6525.

40. Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In: International Conference on Machine Learning. PMLR; 2015 [cited 2021 Mar 21]. p. 448–56. Available from: http://proceedings.mlr.press/v37/ioffe15.html

41. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. J Mach Learn Res. 2014;15(56):1929–58.

42. Gal Y, Ghahramani Z. Bayesian Convolutional Neural Networks with Bernoulli Approximate Variational Inference. ArXiv150602158 Cs Stat. 2016 [cited 2021 Mar 21]; Available from: http://arxiv.org/abs/1506.02158

43. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. ArXiv14126980 Cs. 2017 [cited 2021 Mar 21]; Available from: http://arxiv.org/abs/1412.6980

44. Ramos M. curatedTCGAData: Curated Data From The Cancer Genome Atlas (TCGA) as MultiAssayExperiment Objects. Bioconductor; 2020 [cited 2021 Mar 21]. Available from: https://bioconductor.org/packages/curatedTCGAData

45. Zhou F, Zhou C. Lung cancer in never smokers—the East Asian experience. Transl Lung Cancer Res. 2018;7(4):450–63.

46. Chen J, Yang H, Teo ASM, Amer LB, Sherbaf FG, Tan CQ, et al. Genomic landscape of lung adenocarcinoma in East Asians. Nat Genet. 2020;52(2):177–86.

47. Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the Xena platform. Nat Biotechnol. 2020;38(6):675–8.

48. SureSelect Cancer All-In-One Catalog and Custom Assays. Available from: https://www.agilent.com/en/product/next-generation-sequencing/hybridization-based-next-generation-sequencing-ngs/cancer-all-in-one-assays/sureselect-cancer-all-in-one-lung-assay-520074

49. Reinhard E, Adhikhmin M, Gooch B, Shirley P. Color transfer between images. IEEE Comput Graph Appl. 2001;21(5):34–41.

50. Macenko M, Niethammer M, Marron JS, Borland D, Woosley JT, Xiaojun Guan, et al. A method for normalizing histology slides for quantitative analysis. In: 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. 2009. p. 1107–1110.

51. Vahadane A, Peng T, Sethi A, Albarqouni S, Wang L, Baust M, et al. Structure-Preserving Color Normalization and Sparse Stain Separation for Histological Images. IEEE Trans Med Imaging. 2016;35(8):1962–71.

52. Simpson EH. Measurement of Diversity. Nature. 1949;163(4148):688.

53. Shannon CE. A mathematical theory of communication. ACM SIGMOBILE Mob Comput Commun Rev. 2001;5(1):3–55.

54. Fan RE, Chang KW, Hsieh CJ, Wang XR, Lin CJ. LIBLINEAR: A Library for Large Linear Classification. J Mach Learn Res. 2008;9(61):1871–4.

55. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. New York: Springer; 2009 [cited 2022 Aug 16]. (Springer Series in Statistics). Available from: http://link.springer.com/10.1007/978-0-387-84858-7

56. Breiman L. Random Forests. Mach Learn. 2001;45(1):5–32.

57. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011;12(85):2825–30.

58. Oliver AJ, Lau PKH, Unsworth AS, Loi S, Darcy PK, Kershaw MH, et al. Tissue-Dependent Tumor Microenvironments and Their Impact on Immunotherapy Responses. Front Immunol. 2018;31(9):70.

59. Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the Yield of Medical Tests. JAMA. 1982;247(18):2543–6.

60. Rodríguez De Dios N, Navarro-Martin A, Cigarral C, Chicas-Sett R, García R, Garcia V, et al. GOECP/SEOR radiotherapy guidelines for non-small-cell lung cancer. World J Clin Oncol. 2022;13(4):237–66.

61. 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(10):1559–67.

62. Bian C, Wang Y, Lu Z, An Y, Wang H, Kong L, et al. ImmunoAIzer: A Deep Learning-Based Computational Framework to Characterize Cell Distribution and Gene Mutation in Tumor Microenvironment. Cancers. 2021;13(7):1659.

63. Courtiol P, Maussion C, Moarii M, Pronier E, Pilcer S, Sefta M, et al. Deep learning-based classification of mesothelioma improves prediction of patient outcome. Nat Med. 2019;25(10):1519–25.

64. Kessler R, Gasser B, Massard G, Roeslin N, Meyer P, Wihlm JM, et al. Blood vessel invasion is a major prognostic factor in resected non-small cell lung cancer. Ann Thorac Surg. 1996;62(5):1489–93.

Rączkowska *et al. BMC Cancer*   (2022) 22:1001

Page 18 of 18

65. Park SY, Lee HS, Jang HJ, Lee GK, Chung KY, Zo JI. Tumor Necrosis as a Prognostic Factor for Stage IA Non-Small Cell Lung Cancer. Ann Thorac Surg. 2011;91(6):1668–73.
66. Chen Z, Fillmore CM, Hammerman PS, Kim CF, Wong KK. Non-small-cell lung cancers: a heterogeneous set of diseases. Nat Rev Cancer. 2014;14(8):535–46.
67. Sholl LM, Sun H, Butaney M, Zhang C, Lee C, Jänne PA, et al. ROS1 immunohistochemistry for detection of ROS1-rearranged lung adenocarcinomas. Am J Surg Pathol. 2013;37(9):1441–9.
68. Wang R, Hu H, Pan Y, Li Y, Ye T, Li C, et al. RET fusions define a unique molecular and clinicopathologic subtype of non-small-cell lung cancer. J Clin Oncol Off J Am Soc Clin Oncol. 2012;30(35):4352–9.
69. Nishino M, Klepeis VE, Yeap BY, Bergethon K, Morales-Oyarvide V, Dias-Santagata D, et al. Histologic and cytomorphologic features of ALK-rearranged lung adenocarcinomas. Mod Pathol Off J U S Can Acad Pathol Inc. 2012;25(11):1462–72.
70. Skoulidis F, Goldberg ME, Greenawalt DM, Hellmann MD, Awad MM, Gainor JF, et al. STK11/LKB1 Mutations and PD-1 Inhibitor Resistance in KRAS-Mutant Lung Adenocarcinoma. Cancer Discov. 2018;8(7):822–35.
71. Papillon-Cavanagh S, Doshi P, Dobrin R, Szustakowski J, Walsh AM. STK11 and KEAP1 mutations as prognostic biomarkers in an observational real-world lung adenocarcinoma cohort. ESMO Open. 2020;5(2):e000706.
72. Błach J, Wojas-Krawczyk K, Nicoś M, Krawczyk P. Failure of Immunotherapy—The Molecular and Immunological Origin of Immunotherapy Resistance in Lung Cancer. Int J Mol Sci. 2021;22(16):9030.
73. Drusbosky LM, Rodriguez E, Dawar R, Ikpeazu CV. Therapeutic strategies in RET gene rearranged non-small cell lung cancer. J Hematol OncolJ Hematol Oncol. 2021;14(1):50.
74. Drilon A, Lin JJ, Filleron T, Ni A, Milia J, Bergagnini I, et al. Frequency of Brain Metastases and Multikinase Inhibitor Outcomes in Patients With RET–Rearranged Lung Cancers. J Thorac Oncol. 2018;13(10):1595–601.
75. Paulsson J, Ehnman M, Östman A. PDGF receptors in tumor biology: prognostic and predictive potential. Future Oncol Lond Engl. 2014;10(9):1695–708.
76. Kilvaer TK, Rakaee M, Hellevik T, Vik J, Petris LD, Donnem T, et al. Differential prognostic impact of platelet-derived growth factor receptor expression in NSCLC. Sci Rep. 2019;9(1):10163.
77. Moon Y, Lee KY, Sung SW, Park JK. Differing histopathology and prognosis in pulmonary adenocarcinoma at central and peripheral locations. J Thorac Dis. 2016 Jan;8(1):169–77.
78. Wang Z, Li M, Huang Y, Ma L, Zhu H, Kong L, et al. Clinical and radiological characteristics of central pulmonary adenocarcinoma: a comparison with central squamous cell carcinoma and small cell lung cancer and the impact on treatment response. OncoTargets Ther. 2018;11:2509–17.
79. Putora PM, Szentesi K, Glatzer M, Rodriguez R, Müller J, Baty F, et al. SUV-max and Tumour Location in PET-CT Predict Oncogene Status in Lung Cancer. Oncol Res Treat. 2016;39(11):681–6.
80. Xie X, Li X, Tang W, Xie P, Tan X. Primary tumor location in lung cancer: the evaluation and administration. Chin Med J (Engl). 2021;135(2):127–36.

Publisher's Note