medRxiv preprint doi: https://doi.org/10.1101/2021.01.18.21249895; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Title: Radiomics-based patient inclusion model improves clinical trial performance

Authors: Michal R. Tomaszewski*[1], Shuxuan Fan*[1,2], Alberto Garcia[1], Jin Qi[1], Youngchul Kim[3], Robert A. Gatenby[4], Matthew B. Schabath[5], William D. Tap[6,7], Denise K. Reinke[8], Rikesh Makanji©[4], Damon R. Reed©[9], Robert J. Gillies[1] ©
*equal contributor
©corresponding

Affiliations: 1. Department of Cancer Physiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA. 2. Department of Radiology, Tianjin Medical University, Tianjin, China 3. Department of Biostatistics, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA. 4. Department of Radiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA. 5. Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA. 6. Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA. 7. Department of Medicine, Weil Cornell Medical College, New York, NY, USA. 8. Sarcoma Alliance for Research through Collaboration, Ann Arbor, MI, USA. 9. Sarcoma Department, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA

Abstract

Background

Quantitative image analytics ("radiomics") is a powerful tool for predicting and prognosing cancer patient outcomes in response to therapy. We hypothesize that radiomic features would be useful as inclusion/exclusion criteria for patient enrichment in clinical trials and aimed to develop the appropriate framework for this analysis.

Methods

This was tested among soft-tissue sarcoma (STS) patients accrued into a randomized clinical trial (SARC021) that evaluated the efficacy of evofosfamide (Evo), a hypoxia activated prodrug, in combination with doxorubicin (Dox). Notably, SARC021 failed to meet its survival objective. We tested whether a radiomic biomarker-driven inclusion/exclusion criterion could have been used to result in a significant treatment benefit of the Evo+Dox combination compared to the standard Dox monotherapy. A total of 166 radiomics features were extracted from 303 patients from the SARC021trial with lung metastases, divided into a training and test set. Univariable and multivariable models were utilized to discriminate OS in the two treatment groups.

Findings

A single radiomics feature, Short Run Emphasis, was the most informative. When combined with histological classification and smoking history, an enriched subset (42%) of patients had longer OS in Evo+Dox vs. Dox groups [p=0.01, Hazard Ratio (HR) =0.57 (0.36-0.90)]. Application of the same model and threshold value in an independent test set confirmed the significant survival difference (p=0.002, HR=0.29 (0.13-0.63)). This process also identifies patients most likely to benefit from doxorubicin alone.

Interpretation

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.01.18.21249895; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The study presents a first of its kind radiomic approach for patient enrichment in clinical trials. In particular, we have shown that had the radiomic model been used for selective patient inclusion into the SARC021 trial, it would have met its primary survival objective for patients with metastatic STS.

Introduction

In the last decade, there has been an explosion in the use of advanced image analysis with machine learning, known as "Radiomics" (1,2). Radiomic analyses of cancer can be used to stage, prognose patient outcome, predict response to specific therapies and, most recently, to inform therapeutic choices (3) with increasing connectivity between image features and tumor biology (4). We hypothesized that radiomic approaches can be used in clinical trials for patient enrichment. We tested this hypothesis in a retrospective analysis of data from the SARC021 phase 3 clinical trial in metastatic soft tissue sarcoma that compared overall survival (OS) in cohorts treated with doxorubicin (Dox) to those treated with Dox + Evofosfamide (Evo), a hypoxia activated pro-drug (NCT01440088). Although Dox+Evo had shown promise for sarcoma control in a phase II study (5), the phase 3 trial failed to meet its threshold of increased survival in the Dox+Evo cohort (6).

Soft tissue sarcomas are a heterogeneous group of malignancies originating in mesenchymal tissue that most often metastasize to the lungs (7). The a historical median OS of 12 months for metastatic soft tissue sarcoma patients has steadily improved to 20.4 months on trials may be attributed to better patient selection along with additional options in second line and beyond therapies (8-10). The shifting survival with Dox monotherapy complicated this particular study (11). Biomarkers that can enrich and accrue a patient cohort that is unlikely to benefit from standard therapy will be useful to identify active agents in first line therapy. In this first of its kind study, we present a novel quantitative imaging framework that can identify patients most and least likely to benefit from trial enrollment. Such radiomics-based biomarkers could be used as companion diagnostics for treatment decision support of approved agents.

Methods

Patient populations

This study was approved by the University of South Florida Institutional Review Board. The analysis includes patients who participated in the TH CR-406/SARC021 multicenter clinical trial of Doxorubicin plus Evofosfamide (Dox+Evo) versus Doxorubicin alone (Dox) in locally advanced, unresectable or metastatic soft-tissue sarcoma. Full trial protocol and results were published by Tap et al. (6). A total of 640 patients were enrolled. The primary endpoint of the trial was OS. CT images obtained prior to treatment were available for analysis in 588 patients.

Patient data and CT images

Patient covariates and CT image were obtained from the Sarcoma Alliance for Research through Collaboration (SARC). Patient data and CT images, obtained prior to treatment, was available for 588 patients. The CT images were uploaded into HealthMyne Quantitative Imaging Decision Support (QIDS) software (https://www.healthmyne.com, Madison, WI), where a radiologist with 10 years (S.F.) identified and segmented all visible lesions. 346 patients were found to have at least one lesion in the lung, the most common metastatic site in the considered cohort (followed by liver lesions, identified in 106 patients), as anticipated (7). Only lung

medRxiv preprint doi: https://doi.org/10.1101/2021.01.18.21249895; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

patients were included in the study to enable comparison of image features between individuals, and hence the use of radiomics. Of these patients, 303 had contrast enhanced CT scans of the lung which could be analyzed, and these were used for quantification. This total cohort of 303 patients used in this study was randomly divided 70:30 into training and test sets using the *sample* function in R version 4.0.2.

Radiomic feature extraction

The images were pre-processed as described in the Supplementary Methods. For each patient, a total of 166 features were calculated for each patient using standardized algorithms from the Image Biomarker Standardization Initiative (IBSI) v5 (12). The radiomic features included statistical, histogram, shape & size, Grey Level Cooccurrence Matrix (GLCM), Grey Level Run Length Matrix (GLRLM), Grey Level Size Zone Matrix (GLSZM) and Neighboring Grey Tone Difference Matrix (NGTDM) features, as well as 16 peritumoral features as described before (13). Laws and Wavelet features were not extracted due to their poor reproducibility (14). As standard in radiomic studies (15), to ensure the radiomic signatures provide additional information compared with tumor volume, the features strongly correlated to volume (Pearson $|r|>0.8$) were excluded from further analysis, while volume itself was included. Spatial stability of the features was assessed, and unstable features excluded. 100 features remained following the exclusion, as detailed in **Supplementary Table 1**.

Feature Selection

The goal of this analysis was to identify the radiomic features and patient covariates differentially associated with OS in the two treatment groups, which was the primary endpoint in the original trial (6). First, univariable Cox proportional hazards regression analysis was used to assess the degree and direction of statistical association of each feature and covariate with post-treatment OS, separately in Dox and Dox+Evo arms. For each arm, features and covariates were considered promising in either of the two scenarios: (i) They showed association ($p<0.05$) with survival in one group AND no association ($p>0.20$) in the other group, or (ii) they showed significant association ($p<0.05$) with survival in both groups in opposite directions (HR >1 in one group and <1 in the other). Multiple comparison correction was not used, as the statistical significance assessment was carried out in the next step.

Correlation between the remaining features was calculated (Pearson's correlation coefficient for continuous and Chi Square independence test statistics for categorical variables). For significantly correlated ($p<0.05$) feature groups, feature with lowest univariable Cox regression p value in the corresponding treatment group was retained as a representative of the group, and others excluded to avoid redundant information. Of the remaining features and covariates, the one with lowest p-value ratio in the two treatment groups (low divided by high) was used in model training.

Final model construction

The two final sets of features and covariates predicted to be most informative of the differential response to Dox or Dox+Evo were used to build the corresponding separate multivariable Cox proportional hazards regression models. Risk scores that are log-transformed relative risks of death were calculated using the *predict.coxph* function in R for all patients in the model training cohort and used to determine threshold for patient virtual inclusion and exclusion from the trial. The process of determining the optimum risk score threshold is described in the results section.

medRxiv preprint doi: https://doi.org/10.1101/2021.01.18.21249895; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Risk score values were predicted for all patients in the test cohort from the final multivariable Cox models constructed as above. The threshold values found to result in best separation of the treatment arms as found in the training set were applied to enrich the test cohort, and survival was compared between the treatment arms in the included subset of the test cohort using log-rank test.

For the Dox model, where patients with high risk score values were expected to perform poorly under Dox, and thus be more likely to favor Dox+Evo treatment, a search for the optimal threshold was performed iteratively including sub-cohorts of patients with risk score above 1st, 2nd,3rd etc to 97th percentile of the total training cohort, evaluating survival difference between the treatment arms in terms of Cox regression p-value and hazard ratio each time. Thus, the entire range of possible thresholds was interrogated, to check if such selection can lead to significant treatment arm separation and identify an optimal threshold value.

medRxiv preprint doi: https://doi.org/10.1101/2021.01.18.21249895; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Results

Patients

Clinical covariates included in the analysis are listed in **Table 1**. Description of the clinical features is included in **Supplementary Table 2.** Presence of lung metastasis was associated with significantly poorer overall survival in the entire cohort of 588 patients (p=0.048, HR=1.24 (1.00-1.54)). This was not the case for patients with liver metastases, which was the second most common metastatic site (p=0.41). Among patients with lung metastases, no significant survival difference was observed between treatment groups (p=0.8), similarly to the entire cohort (p=0.45).

No significant difference in OS was observed between the full training and test cohorts (p=0.40, median OS: 17.4 (15.2-20.6) vs. 20.4 (14.0-26.9) training vs test). No significant differences were also seen between Dox and Dox+Evo treatment groups in the training (p=0.67, HR=1.07 (0.78-1.49) median OS: 18.3 (12.6-21.2) vs. 17.2 (15.2-22.1) months Dox vs. Dox+Evo) or test cohorts (p=0.30, HR=1.31 (0.46-1.23) median OS: 23.3 (16.5-31.8) vs. 14.9 (11.1-27.2) months Dox+Evo vs. Dox).

Feature selection

Results of univariable Cox proportional hazards regression, performed separately in the Dox and Dox+Evo treatment groups to identify those radiomic features and clinical covariates differently associated with OS, are shown in **Table 2**. Among clinical covariates, the histological classification of the primary tumor, tumor grade and prior radiotherapy were significantly associated with survival in the Dox and not in the Dox+Evo groups. Following elimination of correlated features, histology and prior radiotherapy remained, and histology was chosen to be included in the model due to lower p value ratio (0.015 vs. 0.029). There were no clinical features significantly related to survival in the Dox+Evo group. A history of smoking was significantly associated with shortened survival (p=0.04, HR=1.62 (1.01-2.61)) in the Dox group yet was insignificantly associated with longer survival in the Dox+Evo group (p=0.17, HR=0.73 (0.46-1.15).

No features were found to be significant in the Dox+Evo and not in the Dox group. Seven uncorrelated radiomic features were found significantly associated with survival in the Dox and not in the Dox+Evo group: Volume Density, Number of Connected 3D Components, Inverse Variance, Correlation, Short Run Emphasis, and Small Zone Emphasis. Of these features, Short Run Emphasis, a measure of heterogeneity, showed the lowest ratio of p value in the Dox to the p value in Dox+Evo groups, and was chosen for training a prediction model of post-treatment survival.

Multivariable model

The three features identified above (histology, non-smoking history, and radiomic Short Run Emphasis) were combined in a multivariable Cox model trained on the Dox cohort of the training data set, producing a highly significant signature (p=0.0001) of survival. Details of the model are shown in **Supplementary Table 3**. No corresponding model was developed in the Dox+Evo group, as no clinical or radiomic features specific to this treatment arm were identified. The Dox model was used to predict risk scores for the entire training set cohort, including the Dox+Evo group, providing a predicted measure of risk of death if the Dox treatment was applied to all patients. Because patients with highest risk scores for Dox monotherapy are expected to benefit the most from the alternative (Dox+Evo) treatment, they should be included in the trial.

medRxiv preprint doi: https://doi.org/10.1101/2021.01.18.21249895; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

On the other hand, patients with a low risk score should be excluded and undergo Dox monotherapy instead. Such patient enrichment strategy for the trial should result in an improved treatment benefit of Dox+Evo in the included patients. To assess this, we performed the log-rank test for difference in survival between Dox vs. Dox+Evo as a function of the risk score threshold for the remaining patients whose score was above that threshold. The schematic of the process is shown in **Figure 1**. As described in Methods, the threshold separating high- from low-risk groups was incrementally increased to identify an optimum that reached a significant difference in OS, while including the largest fraction of patients. The results of this analysis are shown in **Figure 2**, demonstrating that increasingly different OS can be observed for the two treatment groups when patients with low-risk scores are excluded from the analysis (**Figure 2A**).The smaller p-values encountered with increasing thresholds were consistent with decreasing HR <1 (**Figure 2B**), showing increasingly significant treatment benefit of Dox+Evo vs. Dox with more stringent inclusion criteria. A threshold of 1.00 allowed inclusion of 42% of the initial training cohort in the trial and showed a significant advantage of Dox+Evo over Dox [p=0.017, HR=0.57 (0.36-0.90)]. This result was visualized in divergent Kaplan Meier curves for the treatment groups in the included patients and longer survival in the Dox+Evo group (**Figure 2C**, median survival 15.8 (14.2-21.4) vs 9.1 (7.6-13.7) months Dox+Evo vs Dox) and a reverse trend for the excluded patients (**Figure 2D,** median survival 20.9 (14.7-27.2) vs 27.0 (20.1-Not estimable(N.E)) Dox+Evo vs Dox, p=0.036), with Dox treated group showing significantly longer survival, compared to no difference in the whole training cohort (**Figure 2E**). These figures make apparent that the most significant difference ($p<10^{-5}$) between included and excluded groups is their response to Dox monotherapy. Indeed, the difference in survival between included and excluded groups in the Dox+Evo trial arms was insignificant (p=0.48). A model using only clinical features as input was also trained and its performance evaluated as above, showing significant (p=0.0007) advantage in survival of the Dox+Evo vs. Dox, at and above a risk score threshold of 1.23 corresponding to a lower inclusion rate of 32% of initial training cohort (**Supplementary Figure 1**). Although the Short Run Emphasis feature by itself did not significantly discriminate (minimum p-value 0.097 including only patients with SRE<0.84, **Supplementary Figure 2**), it added to the significance of clinical features and thus increased the number of potential patients on trial from 32% to 42%.

Model testing

The multivariable Cox model trained in the above section was used to predict risk scores for all patients in the test cohort. Similar to the training cohort, an increase in minimum risk score threshold for inclusion lead to a monotonic decrease in p value (**Figure 3A**) and HR (**Figure 3B**) for the overall survival comparison between treatment groups. Applying the threshold of 1.00 determined a priori in the training set as the optimum threshold for inclusion showed a significantly better survival in the Dox+Evo vs the Dox treated group [p=0.002, HR=0.29 (0.13-0.63) **Figure 3C**]. As in the training cohort, this was associated with an increased median survival of 22.8 (12.3-N.E) for Dox+Evo vs. 6.3 (3.1-13.7) for Dox. As shown in **Figure 3D** the differences in the two treatment arms for the remaining excluded patients was insignificant for both OS (p=0.72) and median survival 26.0 (18.4-N.E) vs. 27.2 (18.6-N.E), similarly to the starting whole test cohort (**Figure 3E**). The threshold corresponded to inclusion of 38% of the initial test cohort. As in the training set (**Figure 2C,D**), the selection by risk score threshold separated patients who did and did not respond to Dox ($p<10^{-4}$) , whereas it did not discriminate (p=0.54) responses of the Dox+Evo group. The plot of p value vs inclusion threshold (**Figure**

medRxiv preprint doi: https://doi.org/10.1101/2021.01.18.21249895; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

3A) shows a matching profile of improving treatment benefit of the Dox+Evo treatment (because of decreasing effectiveness of Dox) with increasing risk score, further supporting the model and the use of radiomics in patient selection.

Radiomic model interpretation

Given the complexity of the question, directly interpreting the imaging information in the model may be challenging. The analysis of p value vs. feature value threshold (as shown in Figures 3 and 4 for risk scores) for the Short Run Emphasis (SRE) feature, shows that the treatment benefit of Dox+Evo is maximized if only patients with tumors of low SRE are included (minimum p= 0.097, HR=0.62 (0.35-1.09) at SRE<0.845, Supplementary Figure 2). Taking the entire dataset into account, the p value at this threshold reaches p=0.01 signifying the relevance of this radiomic feature in patient selection. The multivariable model developed above also favors low SRE values, as shown in **Figure 4A** for both training and test cohorts. Comparing representative tumors with extreme SRE values reveals visual differences. In line with the model, a patient censored after over 2.5 years showed low SRE in the lung lesion (**Figure 4B**) at baseline; visually associated with regularity and homogeneity of the mass. Conversely, another patient deceased on Dox therapy less than 5 months after enrollment presented a lung lesion with high SRE and highly heterogeneous appearance (**Figure 4C**). While these show extremes, the value of using a quantitative SRE threshold is to identify those patients whose scans may be more equivocal.

Discussion

Herein, we identified a novel radiomic model, based on the combination of pre-treatment CT data and clinical information, that can be used as patient enrichment strategy for a clinical trial. This method predicts patients that would have relatively long OS with Dox monotherapy and thus can be excluded from trials with novel agents. In the SACRC021 trial this would have improved the relative effect evofosfamide. These results were successfully validated in the test set and, if applied, the phase 3 trial would have been successful.

Patient selection for drug trials remains a challenge in clinical trial design. In this study radiomic methods (15) in combination with novel statistical analysis was used to provide and validate a patient inclusion framework based on widely available standard of care imaging data in a retrospective cohort. While radiomics methods have been used to predict patient survival following different treatments (16-19), and correlate to tumor hypoxia(20), this is the first study to derive the prognostic radiomic features and multivariable models required to discriminate between two therapeutic arms and determine the optimal population to most benefit from the novel Evo intervention.

Notably, while there were a number of prognostic features associated with positive outcomes in both groups, there were no features were associated with survival in the Dox+Evo cohort independent of the Dox group. This suggests that the biological factors that govern Evo response may not be related sufficiently strongly to the information available from clinical and imaging data. Conversely, the strong predictive model of Dox response shows significant promise for potentially both clinical care and more optimal patient selection in future trials. The sarcoma community has long sought additional efficacious agents in metastatic soft tissue sarcomas with large trials dedicated to alkylators such as ifosfamide (21), and derivatives like palifosfamide (22) that have shown increased response rates but not overall survival benefit. Excluding patients

medRxiv preprint doi: https://doi.org/10.1101/2021.01.18.21249895; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

who are likely to have survival benefit to Dox monotherapy may increase the effect size and hence power of such trials.

Notably, the failure of the SARC21 trial is at least in part due to a shifting OS for doxorubicin only that is likely multifactorial inclusive of better patient selection, understanding of histologies and patients most likely to benefit from systemic therapy and additionally available subsequent therapies (6,21,22).

An observational trial in soft tissue sarcoma patients bearing lung metastases is planned at our institution to validate the model and understand its biological underpinnings (4), prospectively comparing the model-calculated risk score to overall survival. The prospective data can be used to support this radiomic biomarker as a companion diagnostic or integrated biomarker for patient selection.

There exist some limitations to the presented study. As radiomic patterns and relationships with outcome are known to vary significantly between disease sites, the image quantification in this study was confined to the metastatic lesions in the lung. Lung metastases are both the most common site and leading cause of death in sarcoma (7). The advantage of this analysis is that there was not special radiologic protocol needed for image acquisition and indeed no planning for this radiomic analysis was considered in the trial design. The dataset was collected as part of the trial and the hypothesis and methods applied to this dataset. While the original lack of survival benefit of the novel treatment was conserved in this sub-cohort, in the future the analytical approach proposed should be extended to include other metastatic sites.

In summary, in this work we demonstrate for the first time that machine learning can be used to predict differential survival to distinct treatment regimens. We show that radiomic analysis of CT imaging data can be used in combination with clinical information to develop a first of its kind model capable of identifying soft tissue sarcoma patients likely to benefit from novel combination of Doxorubicin+Evofosfamide *vs*. standard Doxorubicin. Application of the proposed model shows that should patient selection be performed a significant survival benefit could have been observed in an otherwise negative Phase 3 trial. Used prospectively, this approach may in the future improve the chance of determining efficacy of novel therapeutic regimens through better patient selection and guide therapeutic decisions for all metastatic STS through actionable, personalized, image-based, survival prediction.

medRxiv preprint doi: https://doi.org/10.1101/2021.01.18.21249895; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Funding statement: This work was supported by National Cancer Institute (grants U54CA193489, R01CA187532, U01CA143062, P30CA76292) and pilot funds from the Moffitt Cancer Center Department of Radiology. Dr. Fan was partially supported by funds from the Dept. Radiology, Tianjin Medical University, Tianjin, China

Disclosures statement:
RJG has a competing interest as investor with research support and advisory board appointment from HealthMyne, Inc and research support from Helix Biopharma, not related to work. DRR reports personal fees from Threshold, during the conduct of the study; travel reimbursement from Salarius, and serves on advisory board of Epizyme, both outside the submitted work. WDT reports personal fees from Eli Lilly, personal fees from EMD Serono , personal fees from Eisai , personal fees from Mundipharma, personal fees from C4 Therapeutics, personal fees from Daiichi Sankyo , personal fees from Blueprint , personal fees from GlaxoSmithKline , personal fees from Agios Pharmaceuticals , personal fees from NanoCarrier, personal fees from Deciphera, outside the submitted work; In addition, Dr. Tap has a patent Companion Diagnostic for CDK4 inhibitors - 14/854,329 pending to MSKCC/SKI and Scientific Advisory Board - Certis Oncology Solutions, Stock Ownership Co-Founder - Atropos Therapeutics, Stock Ownership. All other authors declare no conflicts of interest.

Authors' contributions:
M.R.T processed the data, designed and implemented the analysis and wrote and edited the manuscript. S.F. selected and annotated the data and performed segmentation. A.G. collected, processed and made the imaging data available for analysis. J.Q. helped select and quality control the imaging data and provided feedback for analysis, Y.K. helped design the analytical model, consulted on statistical approaches and edited the manuscript. R.A.G. helped conceive the study, facilitated obtaining the data and edited the manuscript. M.B.S. Participated in obtaining the data, conception of the study, design of the statistical model, and edited the manuscript. W.D.T was involved in collection, compilation of the data, advised on the trial and helped write the manuscript. D.K.R oversaw collection and compilation of the data and management of the trial. R.M. participated in conception of the study, identification and annotation of the imaging data, and validated the segmentation. D.D.R. helped conceive the study, obtain the data, advise on clinical matters and wrote and edited the manuscript. RJG conceived the study, advised on the analysis and wrote and edited the manuscript
MRT and AG have verified the data, SF and RM have verified the segmentation.

Figures:

medRxiv preprint doi: https://doi.org/10.1101/2021.01.18.21249895; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

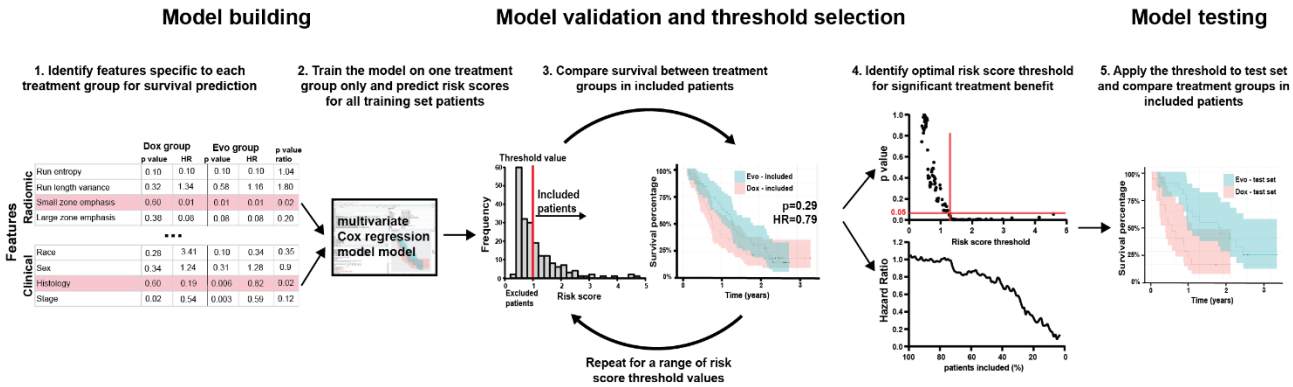

Figure 1 Patient inclusion model. Patient selection into the trial based on Dox group survival was executed according to the following method: firstly (1) radiomic and clinical features associated in training cohort with survival in Dox but not Dox+Evo treatment group were included in a multivariable Cox regression model (2), trained on Dox treated patients. The risk score assigned by the model to each training set patient was then used as a biomarker for inclusion into analysis, iteratively calculating the p value and hazard ratio for survival comparison between treatment arms depending on minimum risk score threshold (3). If available, threshold corresponding to significant (p value<0.05) treatment benefit of Dox+Evo at highest percentage of patients included was chosen (4), and subsequently tested in the test cohort (5), with risk scores assigned by the multivariable Cox model developed in step (2). A corresponding model can also be developed based on Dox+Evo group survival, using a maximum risk score threshold.

medRxiv preprint doi: https://doi.org/10.1101/2021.01.18.21249895; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

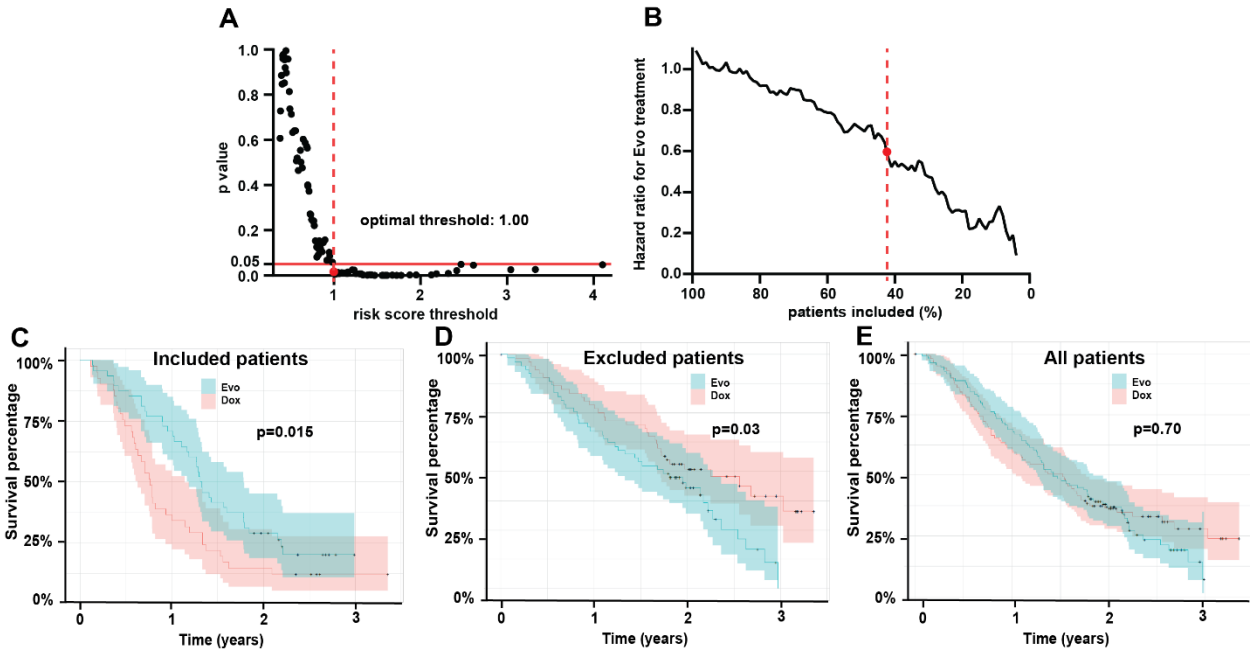

Figure 2 Multivariable Cox model enables selection of patients who benefit from Evofosfamide+Doxorubicin in training cohort. Quantification of the p value of overall survival difference in the training cohort between the Evofosfamide+Doxorubicin (Dox+Evo) vs. Doxorubicin alone (Dox) treatment arms depending on the minimum risk score for patient inclusion, as predicted by the model (A), shows a risk score threshold of 1.00 at which Doxorubicin +Evofosfamide (Dox+Evo) group shows significantly longer OS ($p<0.05$). Exclusion of patients with high risk scores leads to monotonic decrease in the hazard ratio (B), and the 1.00 risk score threshold corresponds to inclusion of 42% of patients in the trial (indicated by red dotted line). The Kaplan-Meier plots by treatment arms show significantly better OS in the included (C) and significantly worse OS in the (D) excluded patients for the Dox+Evo treatment compared to Dox only. In all training set patients (E) no difference between the arms was observed.

medRxiv preprint doi: https://doi.org/10.1101/2021.01.18.21249895; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

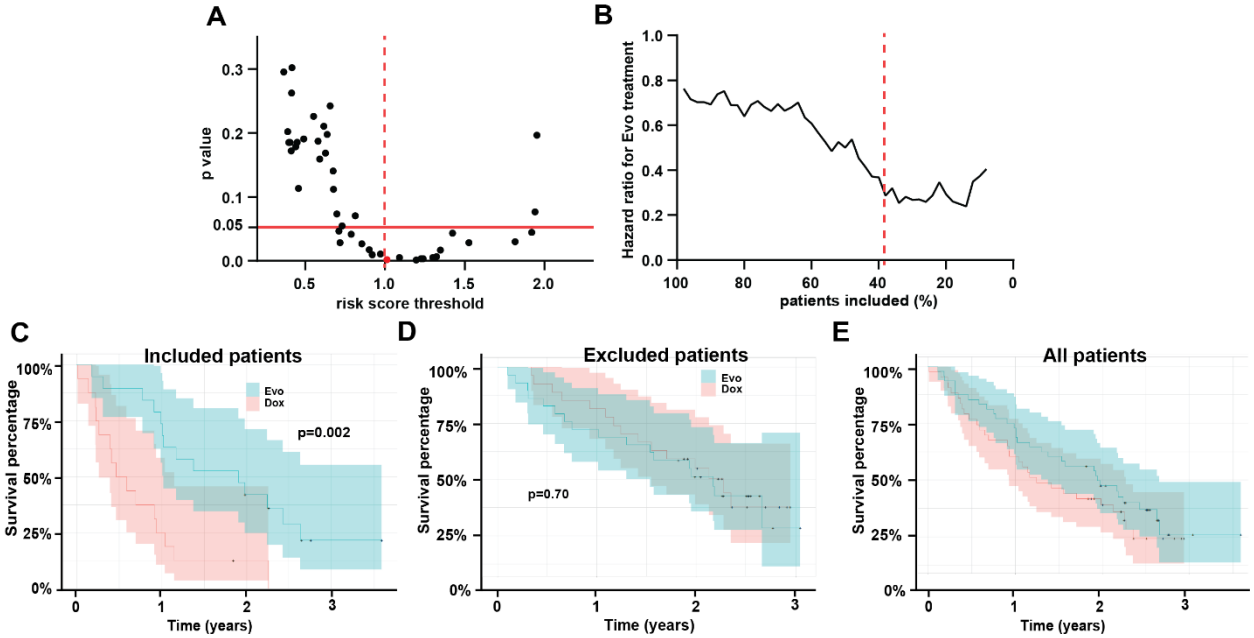

Figure 3. Results in the test cohort confirm the validity of the model. Risk scores predictions in the test cohort based on a multivariable Cox model trained on Dox treated training cohort patients can be used to identify patients who would benefit from Dox+Evo treatment. Graph in (A) shows that increasing the minimum risk score of patients included in the analysis leads to a stronger difference in survival between the treatment groups, as described by the p value of the comparison. For the risk score threshold of 1.00, a highly significant difference is observed (red point and dotted line), which corresponds to a decreased hazard ratio of the combination vs standard therapy (B). These differences are apparent from the Kaplan-Meier curve in the included patients (C) showing significantly longer survival in the Dox+Evo group, while the excluded patients (D), or all test set patients (E) show no difference in survival between treatment groups.

medRxiv preprint doi: https://doi.org/10.1101/2021.01.18.21249895; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

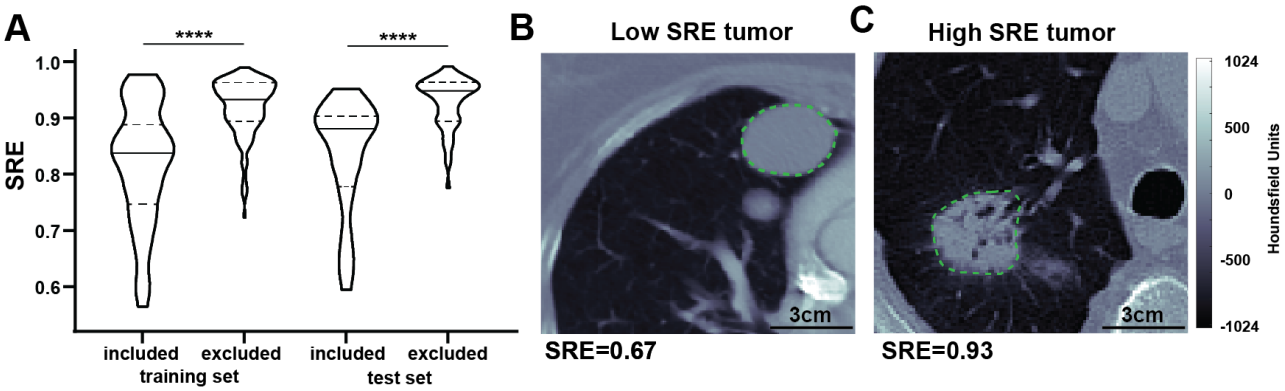

Figure 4. Differences in radiomic features can be apparent visually. The model for selection of patients likely to benefit from Evofosfamide treatment favored low Short Run Emphasis (SRE) radiomic feature for proposed inclusion into the trial. As shown in the violin plot (A), significantly lower SRE is observed in the included vs. excluded patient groups both in training and test cohorts. Qualitatively, a representative tumor with low Short Run Emphasis (SRE, B) appears more regular and homogeneous in a contrast enhanced CT scan than a corresponding tumor with similar volume (15.0 vs 16.5ml respectively), and relatively high SRE (C), which shows higher intratumor heterogeneity. In the violin plot a solid line indicates median while dotted lines indicate 25th and 75th percentile. **** $p<0.0001$

medRxiv preprint doi: https://doi.org/10.1101/2021.01.18.21249895; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Tables

	Training cohort			Test cohort		
	Dox+Evo (n=109)	**Dox (n=103)**	**p value**	**Dox+Evo (n=48)**	**Dox (n=43)**	**p value**
Age (years)	60 (47–73)	55 (33–78)	0.06	60 (44-75)	57 (38-76)	0.82
Sex			1.00			1.00
Female	62 (57%)	59 (57%)		26 (54%)	24 (56%)	
Male	47 (43%)	44 (43%)		22 (46%)	19 (44%)	
Smoking history			0.82			0.40
Never smoker	62 (57%)	56 (54%)		26 (54%)	28 (65%)	
Ever smoker	47 (43%)	47 (46%)		22 (46%)	15 (35%)	
Primary Tumor Site			0.82			0.21
Extremity	35 (32%)	40 (39%)		17 (35%)	20 (47%)	
Head/Neck	7 (6%)	5 (5%)		0	3 (7%)	
Retroperitoneum	17 (16%)	19 (18%)		9 (19%)	4 (9%)	
Visceral	21 (19%)	12 (12%)		9 (19%)	7 (16%)	
Other	29 (27%)	27 (26%)		13 (27%)	9 (21%)	
Metastatic sites number	2 (1–3)	2 (1–3)	0.46	2 (2–3)	2 (1–3.5)	0.66
Stage			0.22			0.43
0	4 (4%)	0		1 (2%)	0	
Stage I	3 (3%)	6 (6%)		2 (4%)	2 (5%)	
Stage II	26 (24%)	21 (20%)		10 (21%)	16 (37%)	
Stage III	45 (41%)	41 (40%)		16 (33%)	12 (28%)	
Stage IV	31 (28%)	35 (34%)		19 (40%)	13 (30%)	
Histology			0.78			0.44
Epitheloid	1 (1%)	3 (3%)		0	0	
Leiomyosarcoma	47 (43%)	41 (40%)		26 (54%)	17 (40%)	
Liposarcoma	7 (6%)	6 (6%)		0	1 (2%)	
Malignant peripheral nerve sheath tumor	4 (4%)	4 (4%)		1 (2%)	4 (9%)	
Myxofibrosarcoma	3 (3%)	4 (4%)		2 (4%)	3 (7%)	
Pleomorphic rhabdomyosarcoma	0	2 (2%)		0	1 (2%)	
Pleomorphic sarcoma/Malignant fibrous histicytoma	17 (16%)	13 (13%)		9 (19%)	7 (16%)	
Other	30 (28%)	30 (29%)		10 (21%)	10 (23%)	

medRxiv preprint doi: https://doi.org/10.1101/2021.01.18.21249895; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

	Training cohort			Test cohort		
	Dox+Evo (n=109)	**Dox (n=103)**	**p value**	**Dox+Evo (n=48)**	**Dox (n=43)**	**p value**
Histology Grade			0.40			0.44
Low	3 (3%)	3 (3%)		2 (4%)	2 (5%)	
Intermediate	33 (30%)	32 (31%)		20 (42%)	15 (35%)	
Intermediate/High	1 (1%)	2 (2%)		1 (2%)	3 (7%)	
High	69 (63%)	61 (59%)		22 (46%)	20 (47%)	
Unknown	3 (3%)	5 (5%)		3 (6%)	3 (7%)	
ECOG score			0.50			0.83
0	61 (56%)	61 (59%)		30 (63%)	25 (58%)	
1	48 (44%)	41 (40%)		18 (38%)	18 (42%)	
2	0	1 (1%)		0	0	
Prior radiotherapy			0.64			0.05
Yes	50 (46%)	43 (42%)		15 (31%)	23 (53%)	
No	59 (54%)	60 (58%)		33 (69%)	20 (47%)	
Prior systemic therapy			0.78			0.39
Yes	7 (6%)	11 (11%)		4 (8%)	2 (5%)	
No	102 (94%)	92 (89%)		44 (92%)	41 (95%)	

Table 1. Breakdown of patient characteristics. Numbers are presented for each treatment group in training and test cohort. Data are median (IQR) or n (%). P value by Wilcoxon test (for age) or Chi squared test (all other variables)

medRxiv preprint doi: https://doi.org/10.1101/2021.01.18.21249895; this version posted January 19, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Clinical covariate name	p Evo	HR Evo	HR Evo 95% CI	p Dox	HR Dox	HR Dox 95% CI
Age	0.416	0.993	(0.976-1.01)	0.822	0.998	(0.983-1.014)
Sex	0.349	1.237	(0.793-1.929)	0.314	1.276	(0.794-2.049)
Histology	0.380			0.006		
Leiomyosarcoma		0.192	(0.025-1.46)		0.822	(0.192-3.511)
Liposarcoma		0.179	(0.02-1.587)		0.840	(0.14-5.038)
Malignant peripheral nerve sheath tumor		0.258	(0.028-2.397)		7.058	(1.273-39.141)
Myxofibrosarcoma		0.055	(0.003-0.917)		0.762	(0.106-5.494)
Pleomorphic rhabdomyosarcoma					0.992	(0.09-10.958)
Pleomorphic sarcoma/Malignant fibrous histocytoma		0.125	(0.015-1.017)		2.315	(0.511-10.489)
Other		0.205	(0.027-1.579)		1.491	(0.351-6.33)
Smoking History	0.172	0.729	(0.464-1.148)		1.625	(1.011-2.612)
Primary Tum Site	0.573			0.569		
Head/neck		1.172	(0.484-2.837)		1.300	(0.456-3.705)
Retroperitoneum		0.777	(0.387-1.56)		0.505	(0.195-1.302)
Visceral		0.905	(0.497-1.649)		0.856	(0.445-1.646)
Other		0.636	(0.349-1.161)		0.771	(0.425-1.399)
Tumor Grade	0.443			0.026		
Intermediate Grade		0.992	(0.607-1.622)		0.554	(0.312-0.982)
Intermediate/High		4.368	(0.585-32.594)		3.840	(0.914-16.125)
Low Grade		1.549	(0.479-5.002)		1.268	(0.392-4.1)
Unknown		1.994	(0.615-6.464)		2.039	(0.728-5.709)
Stage	0.024			0.003		
Stage I	0.544	(0.09-3.272)				
Stage II	0.651	(0.189-2.238)			0.592	(0.21-1.666)
Stage III	0.616	(0.187-2.031)			0.486	(0.183-1.288)
Stage IV	1.408	(0.426-4.651)			1.331	(0.512-3.461)
Metastatic Sites #	0.0600	1.135	(0.996-1.294)	0.020	1.154	(0.964-1.38)
ECOG score	0.004	1.919911	(1.234-2.987)	0.003	2.003	(1.262-3.178)
Prior radiotherapy	0.820	1.052591	(0.676-1.639)	0.024	0.565	(0.345-0.927)
Prior systemic therapy	0.527	0.722203	(0.263-1.98)	0.815	1.092	(0.521-2.288)
Radiomic feature name	**p Evo**	**HR Evo**	**HR Evo 95% CI**	**p Dox**	**HR Dox**	**HR Dox 95% CI**
F1.Statistical_Mean	0.290	1.000	(0.999-1)	0.178	1.000	(1-1)
F4.Statistical_SKEW	0.011	0.668	(0.489-0.913)	0.032	0.698	(0.503-0.969)
F5.Statistical_Kurtosis	0.033	1.110	(1.01-1.22)	0.009	1.130	(1.03-1.25)
F6.Statistical_Median	0.177	1.000	(1-1)	0.180	1.000	(1-1)
F9.Statistical_90th_percentile	0.853	1.000	(0.998-1)	0.554	1.000	(0.999-1)
F10.Statistical_Maximum_grey_level	0.557	1.000	(0.999-1)	0.655	1.000	(0.999-1)
F19.Statistical_Root_mean_square	0.211	0.999	(0.997-1)	0.124	0.999	(0.997-1)
F22.Intensity_histogram_skewness	0.010	0.659	(0.48-0.905)	0.036	0.701	(0.503-0.976)
F23.Intensity_histogram_kurtosis	0.022	1.130	(1.02-1.25)	0.010	1.140	(1.03-1.25)
F35.Intensity_histogram_coefficient_of_variance	0.053	0.078	(0.00593-1.03)	0.027	0.031	(0.00148-0.669)
F36.Intensity_histogram_quartile_coefficient_of_dispersion	0.017	0.079	(0.00997-0.632)	0.009	0.041	(0.00367-0.452)
F37.Intensity_histogram.entropy	0.327	0.877	(0.676-1.14)	0.013	0.662	(0.477-0.917)
F38.Intensity_histogram_uniformity	0.243	5.360	(0.319-89.8)	0.007	63.600	(3.05-1330)
F44.Volume_at_intensity_fraction_90	0.860	1.180	(0.194-7.14)	0.540	1.860	(0.255-13.6)
F45.Intensity_at_volume_fraction_10	0.588	0.999	(0.997-1)	0.335	0.999	(0.997-1)
F46.Intensity_at_volume_fraction_90	0.053	1.000	(1-1)	0.084	1.000	(1-1)
F47.Volume_at_intensity_fraction_difference	0.907	1.120	(0.175-7.12)	0.595	0.579	(0.077-4.35)
F48.Intensity_at_volume_fraction_difference	0.089	0.999	(0.997-1)	0.032	0.998	(0.996-1)
F50.Volume_.mm.3.	0.150	1.000	(1-1)	0.005	1.000	(1-1)
F53.Surface_to_volume_ratio_.mm.2.	0.090	0.530	(0.254-1.11)	0.015	0.398	(0.189-0.837)
F54.Compactness_1	0.905	18.700	(2.93e-20-1.2e+22	0.187	0.000	(8.29e-26-78500)
F55.Compactness_2	0.929	0.913	(0.125-6.65)	0.285	0.385	(0.0669-2.22)
F56.Spherical_disproportion	0.564	0.631	(0.132-3.02)	0.080	1.790	(0.933-3.41)
F57.Sphericity	0.837	1.410	(0.0551-35.9)	0.158	0.215	(0.0253-1.82)
F58.Asphericity	0.564	0.631	(0.132-3.02)	0.080	1.790	(0.933-3.41)
F64.Elongation	0.535	1.700	(0.32-9)	0.497	1.750	(0.35-8.72)
F65.Flatness	0.251	2.640	(0.503-13.9)	0.219	2.900	(0.531-15.8)
F66.Volume_density_axis.aligned_bounding_box	0.778	1.410	(0.127-15.7)	0.941	0.913	(0.0817-10.2)
F67.Area_density_axis_aligned_bounding_box	0.939	1.150	(0.0326-40.5)	0.075	7.800	(0.814-74.7)
F68.Volume_density_oriented_bounding_box	0.754	1.480	(0.126-17.4)	0.986	0.978	(0.0867-11)
F69.Area_density_oriented_bounding_box	0.891	1.290	(0.0339-49)	0.064	8.320	(0.882-78.5)
F70.Volume_density_approximate_enclosing_ellipsoid	0.203	5.180	(0.411-65.3)	0.550	0.634	(0.142-2.83)
F71.Area_density_approximate_enclosing_ellipsoid	0.468	1.800	(0.37-8.73)	0.083	2.260	(0.898-5.71)
F72.Volume_density_minimum_volume_enclosing_ellipsoid	0.917	0.914	(0.166-5.01)	0.185	0.263	(0.0367-1.89)

Radiomic feature name	p Evo	HR Evo	HR Evo 95% CI	p Dox	HR Dox	HR Dox 95% CI
F73.Area_density_minimum_volume_enclosing_ellipsoid	0.801	1.430	(0.0868-23.7)	0.431	2.270	(0.296-17.4)
F74.Volume_density_convex_hull	0.988	0.980	(0.0661-14.5)	0.041	0.148	(0.0238-0.928)
F75.Area_density_convex_hull	0.757	1.940	(0.029-130)	0.077	5.530	(0.831-36.8)
F76.Number_of_connected_3D_components	0.336	0.557	(0.169-1.84)	0.031	1.210	(1.02-1.44)
F77.Asymmetry	0.610	0.488	(0.031-7.69)	0.537	0.426	(0.0286-6.37)
F78.Eccentricity	0.535	0.589	(0.111-3.13)	0.497	0.573	(0.115-2.86)
F79.Orientation	0.855	1.020	(0.81-1.29)	0.297	1.150	(0.882-1.51)
F80.CoM_x_.pxl.	0.719	1.000	(0.999-1)	0.365	1.000	(0.999-1)
F81.CoM_y_.pxl.	0.304	0.999	(0.997-1)	0.741	1.000	(0.998-1)
F82.CoM_z_.pxl.	0.246	1.000	(1-1)	0.349	1.000	(1-1)
F84.CoM_y_.mm.	0.304	0.999	(0.997-1)	0.741	1.000	(0.998-1)
F85.CoM_z_.mm.	0.246	1.000	(1-1)	0.349	1.000	(1-1)
F86.Weighted_CoM_x_.mm.	0.698	1.000	(0.998-1)	0.403	1.000	(0.999-1)
F87.Weighted_CoM_y_.mm.	0.335	0.999	(0.997-1)	0.889	1.000	(0.998-1)
F88.Weighted_CoM_z_.mm.	0.245	1.000	(1-1)	0.355	1.000	(1-1)
F92.avgCoocurrence_Joint_MAX	0.032	11.200	(1.24-102)	0.011	22.400	(2.03-248)
F95.avgCoocurrence_Joint_entropy	0.370	0.933	(0.801-1.09)	0.037	0.822	(0.684-0.988)
F96.avgCoocurrence_Difference_average	0.071	0.886	(0.776-1.01)	0.005	0.767	(0.638-0.922)
F98.avgCoocurrence_Difference_entropy	0.288	0.850	(0.63-1.15)	0.001	0.539	(0.37-0.786)
F102.avgCoocurrence_Angular_second_moment	0.157	29.100	(0.272-3120)	0.017	409.000	(2.89-57900)
F104.avgCoocurrence_Dissimilarity	0.071	0.886	(0.776-1.01)	0.005	0.767	(0.638-0.922)
F105.avgCoocurrence_Inverse_difference	0.116	2.970	(0.765-11.5)	0.001	15.600	(2.99-81.2)
F106.avgCoocurrence_Inverse_difference_normalized	0.046	1130.000	(1.14-1120000)	0.007	2.53E+04	(16.1-39700000)
F107.avgCoocurrence_Inverse_difference_moment	0.131	2.450	(0.765-7.84)	0.001	10.700	(2.61-43.9)
F109.avgCoocurrence_Inverse_variance	0.396	3.230	(0.216-48.2)	0.004	165.000	(4.9-5580)
F110.avgCoocurrence_Correlation	0.328	2.110	(0.473-9.42)	0.048	4.970	(1.02-24.4)
F115.avgCoocurrence_First_measure_of_information_correlation	0.589	1.790	(0.217-14.7)	0.364	2.900	(0.291-28.9)
F116.avgCoocurrence_Second_measure_of_information_correlation	0.790	0.717	(0.0619-8.31)	0.140	0.124	(0.00772-1.98)
F117.avg_3D_SRE_.Short_runs_emphasis.	0.324	0.372	(0.0519-2.66)	0.001	0.015	(0.00125-0.167)
F118.avg_3D_LRE_.Long_runs_emphasis.	0.220	1.050	(0.972-1.13)	0.047	1.090	(1-1.18)
F123.avg_3D_LRLGE_.Long_run_low_grey_level_emphasis.	0.172	0.096	(0.0033-2.77)	0.121	13.200	(0.508-341)
F124.avg_3D_LRHGE_.Long_run_high_grey_level_emphasis.	0.003	1.000	(1-1)	0.175	1.000	(1-1)
F126.avg_3D_GLN_normalize_.Grey_level_non_uniformity_normalised.	0.867	1.430	(0.0214-96.1)	0.007	536.000	(5.64-50900)
F128.avg_3D_RLN_normalize_.Run_length_non_uniformity_normalised.	0.202	0.448	(0.13-1.54)	0.001	0.074	(0.0162-0.339)
F129.avg_3D_RP_.Run_percentage.	0.102	0.322	(0.0827-1.25)	0.001	0.075	(0.0158-0.36)
F131.avg_3D_RLV_.Run_length_variance.	0.100	1.120	(0.979-1.28)	0.105	1.120	(0.977-1.28)
F133.GLSZM_Small_zone_emphasis	0.601	0.533	(0.0503-5.65)	0.009	0.026	(0.00161-0.409)
F134.GLSZM_Large_zone_emphasis	0.380	1.000	(1-1)	0.076	1.000	(1-1)
F135.GLSZM_Low_grey_level_zone_emphasis	0.698	0.426	(0.00571-31.7)	0.085	87.500	(0.539-14200)
F139.GLSZM_Large_zone_low_grey_level_emphasis	0.303	1.000	(1-1)	0.135	1.000	(1-1)
F140.GLSZM_Large_zone_high_grey_level_emphasis	0.106	1.000	(1-1)	0.303	1.000	(1-1)
F141.GLSZM_Grey_level_non_uniformity	0.003	1.000	(1-1)	0.000	1.000	(1-1.01)
F143.GLSZM_Zone_size_non_uniformity	0.002	1.000	(1-1)	0.006	1.000	(1-1)
F144.GLSZM_Zone_size_non_uniformity_normalised	0.457	0.451	(0.0556-3.66)	0.010	0.023	(0.00125-0.405)
F145.GLSZM_Zone_percentage	0.125	0.363	(0.0996-1.32)	0.003	0.086	(0.0169-0.433)
F147.GLSZM_Zone_size_variance	0.711	1.000	(1-1)	0.076	1.000	(1-1)
F148.GLSZM_Zone_size_entropy	0.105	1.420	(0.93-2.16)	0.066	1.440	(0.976-2.13)
F151.NGTDM_Busyness	0.739	0.999	(0.99-1.01)	0.021	1.020	(1-1.04)
F153.NGTDM_Strength	0.126	0.966	(0.923-1.01)	0.029	0.935	(0.88-0.993)
radial.deviation.tumor.mean	0.139	1.010	(0.995-1.03)	0.204	1.010	(0.997-1.02)
radial.deviation.tumor.SD	0.024	1.010	(0.994-1.04)	0.174	1.020	(1-1.04)
radial.gradient.tumor.mean	0.096	0.997	(0.995-1)	0.052	0.998	(0.996-1)
radial.gradient.tumor.SD	0.894	0.997	(0.992-1)	0.133	1.000	(0.997-1)
radial.deviation.tumor.mean..2D.	0.041	1.010	(0.999-1.02)	0.076	1.010	(1-1.02)
radial.deviation.tumor.SD..2D.	0.014	1.010	(0.995-1.03)	0.147	1.020	(1-1.04)
radial.gradient.tumor.mean..2D.	0.086	0.997	(0.995-1)	0.032	0.998	(0.997-1)
radial.gradient.tumor.SD..2D.	0.890	0.997	(0.993-1)	0.101	1.000	(0.997-1)
radial.deviation.border.mean	0.878	0.996	(0.979-1.01)	0.632	1.000	(0.987-1.02)
radial.deviation.border.SD	0.391	1.010	(0.984-1.03)	0.492	1.010	(0.988-1.03)
radial.gradient.border.mean	0.609	0.999	(0.996-1)	0.254	0.999	(0.997-1)
radial.gradient.border.SD	0.905	0.997	(0.993-1)	0.185	1.000	(0.996-1)
radial.deviation.border.mean..2D.	0.529	1.000	(0.987-1.02)	0.827	1.000	(0.991-1.02)
radial.deviation.border.SD..2D.	0.363	1.010	(0.992-1.03)	0.221	1.010	(0.99-1.03)
radial.gradient.border.mean..2D.	0.578	0.998	(0.996-1)	0.177	0.999	(0.997-1)
radial.gradient.border.SD..2D.	0.845	0.998	(0.994-1)	0.286	1.000	(0.996-1)

Table 2. Association with survival in radiomic features and clinical covariates by treatment arm. Univariable Cox regression model was applied separately in the Doxorubicin + Evofosfamide (Dox+Evo) and Doxorubicin only (Dox) treatment arms to calculate the p value ('p Evo' and 'p Dox' respectively) and hazard ratio ('HR Evo' and 'HR Dox') for the relationship of each feature and covariate with overall survival. P values below 0.05 are highlighted in red, while these above 0.20 are highlighted in green. Table also available as .xls.

medRxiv preprint doi: https://doi.org/10.1101/2021.01.18.21249895; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

References

1. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J*, et al.* Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol **2017**;14:749-62
2. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology **2016**;278:563-77
3. Mu W, Jiang L, Zhang J, Shi Y, Gray JE, Tunali I*, et al.* Non-invasive decision support for NSCLC treatment using PET/CT radiomics. Nat Commun **2020**;11:5228
4. Tomaszewski MR, Gillies RJ. The biological meaning of radiomic features. Radiology **2020**;in press
5. Chawla SP, Cranmer LD, Van Tine BA, Reed DR, Okuno SH, Butrynski JE*, et al.* Phase II study of the safety and antitumor activity of the hypoxia-activated prodrug TH-302 in combination with doxorubicin in patients with advanced soft tissue sarcoma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology **2014**;32
6. Tap WD, Papai Z, Van Tine BA, Attia S, Ganjoo KN, Jones RL*, et al.* Doxorubicin plus evofosfamide versus doxorubicin alone in locally advanced, unresectable or metastatic soft-tissue sarcoma (TH CR-406/SARC021): an international, multicentre, open-label, randomised phase 3 trial. The Lancet Oncology **2017**;18
7. Billingsley KG, Burt ME, Jara E, Ginsberg RJ, Woodruff JM, Leung DH*, et al.* Pulmonary metastases from soft tissue sarcoma: analysis of patterns of diseases and postmetastasis survival. Annals of surgery **1999**;229
8. Tap WD, Wagner AJ, Schoffski P, Martin-Broto J, Krarup-Hansen A, Ganjoo KN*, et al.* Effect of Doxorubicin Plus Olaratumab vs Doxorubicin Plus Placebo on Survival in Patients With Advanced Soft Tissue Sarcomas: The ANNOUNCE Randomized Clinical Trial. JAMA **2020**;323:1266-76
9. Seddon B, Strauss SJ, Whelan J, Leahy M, Woll PJ, Cowie F*, et al.* Gemcitabine and docetaxel versus doxorubicin as first-line treatment in previously untreated advanced unresectable or metastatic soft-tissue sarcomas (GeDDiS): a randomised controlled phase 3 trial. Lancet Oncol **2017**;18:1397-410
10. Lorigan P, Verweij J, Papai Z, Rodenhuis S, Le Cesne A, Leahy MG*, et al.* Phase III trial of two investigational schedules of ifosfamide compared with standard-dose doxorubicin in advanced or metastatic soft tissue sarcoma: a European Organisation for Research and Treatment of Cancer Soft Tissue and Bone Sarcoma Group Study. Journal of clinical oncology : official journal of the American Society of Clinical Oncology **2007**;25
11. Lindner LH. Hypoxia-activated prodrug: an appealing preclinical concept yet lost in clinical translation. The Lancet Oncology **2017**;18
12. Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative. **2016**
13. Tunali I, Stringfield O, Guvenis A, Wang H, Liu Y, Balagurunathan Y*, et al.* Radial gradient and radial deviation radiomic features from pre-surgical CT scans are associated with survival among lung adenocarcinoma patients. Oncotarget **2017**;8:96013-26
14. Tunali I, Hall LO, Napel S, Cherezov D, Guvenis A, Gillies RJ*, et al.* Stability and reproducibility of computed tomography radiomic features extracted from peritumoral regions of lung cancer lesions. Med Phys **2019**;46:5075-85
15. Yip SS, Aerts HJ. Applications and limitations of radiomics. Physics in medicine and biology **2016**;61
16. Chetan MR, Gleeson FV. Radiomics in predicting treatment response in non-small-cell lung cancer: current status, challenges and future perspectives. European radiology **2020**

medRxiv preprint doi: https://doi.org/10.1101/2021.01.18.21249895; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

17. Agrawal V, Coroller TP, Hou Y, Lee SW, Romano JL, Baldini EH*, et al.* Radiologic-pathologic correlation of response to chemoradiation in resectable locally advanced NSCLC. Lung cancer (Amsterdam, Netherlands) **2016**;102
18. Horvat N, Veeraraghavan H, Khan M, Blazic I, Zheng J, Capanu M*, et al.* MR Imaging of Rectal Cancer: Radiomics Analysis to Assess Treatment Response after Neoadjuvant Therapy. Radiology **2018**;287
19. Kickingereder P, Götz M, Muschelli J, Wick A, Neuberger U, Shinohara RT*, et al.* Large-scale Radiomic Profiling of Recurrent Glioblastoma Identifies an Imaging Predictor for Stratifying Anti-Angiogenic Treatment Response. Clinical cancer research : an official journal of the American Association for Cancer Research **2016**;22
20. Ganeshan B, Goh V, Mandeville HC, Ng QS, Hoskin PJ, Miles KA. Non-small cell lung cancer: histopathologic correlates for texture parameters at CT. Radiology **2013**;266:326-36
21. Judson I, Verweij J, Gelderblom H, Hartmann JT, Schöffski P, Blay JY*, et al.* Doxorubicin alone versus intensified doxorubicin plus ifosfamide for first-line treatment of advanced or metastatic soft-tissue sarcoma: a randomised controlled phase 3 trial. The Lancet Oncology **2014**;15
22. Ryan CW, Merimsky O, Agulnik M, Blay JY, Schuetze SM, Van Tine BA*, et al.* PICASSO III: A Phase III, Placebo-Controlled Study of Doxorubicin With or Without Palifosfamide in Patients With Metastatic Soft Tissue Sarcoma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology **2016**;34