Supplementary information

Radiomic phenotype features predict pathological response in Non-Small Cell Lung Cancer

Thibaud P. Coroller[1, #] M.Sc., Vishesh Agrawal[1] B.A., Vivek Narayan[1] B.Sc., Ying Hou[1] M.D., Patrick Grossmann[1] M.Sc, Stephanie W. Lee[1] B.A., Raymond H. Mak[1,*] M.D. & Hugo J.W.L. Aerts[1, 2,*] Ph.D.

Departments of Radiation Oncology[1] and Radiology[2], Dana-Farber Cancer Institute, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

[#] Corresponding author

* Equal contribution

Contents:

Supplement I: Radiomic features selection

Supplement II: Univariate analysis results

Supplement I: Radiomic features selection

A total of 1603 Radiomic features [1] that quantify tumor phenotype were extracted for this analysis. Individual description of the features and filters can be found in supplement material from a previous study [2]. Using data mining technique based on features stability and variance, we selected fifteen features for this study. Only these 15 features were included in our analysis to evaluate their power to predict pathological response. All dimension reduction were done on R software [3] version 3.1.3.

Features Stability

Radiomic feature reproducibility was evaluated prior to selection in order to account for variation in image acquisition using the RIDER test / retest dataset [4] images. Contours for the primary tumor were created for twenty-eight patients that received two same-day CT scans with primary tumors contoured on both scans (**Figure S1**) using a CT single click ensemble segmentation algorithm and a full description of the dataset can be found in the supplementary from a previous study from Aerts et al [5].

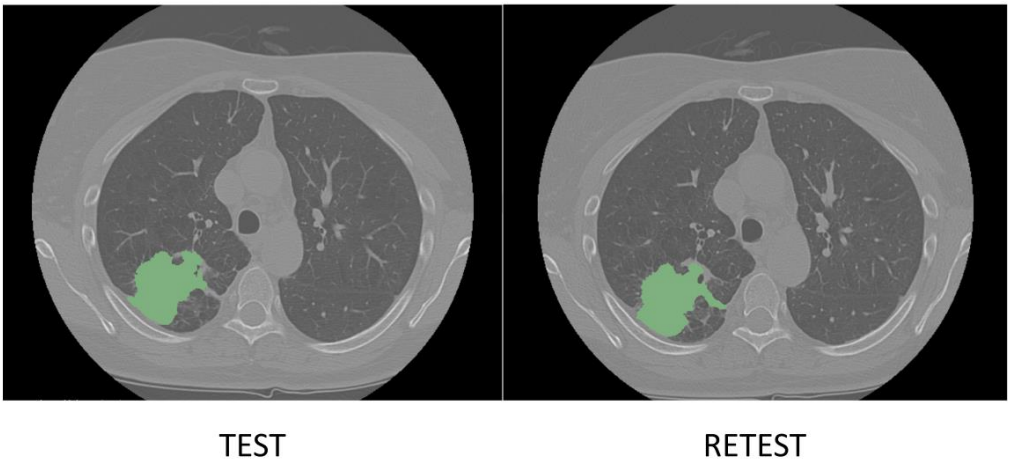

Figure S1: Chest CT images of lung cancer patients on test/retest. Primary lung tumor was contoured (in green) for both images.

From this dataset, we extracted Radiomic features (m=1603) and computed the intraclass correlation coefficient (ICC) [6] was computed using “irr” R package [7] version 0.84. The ICC is a measure of reproducibility with 0 representing non-reproducible features, and 1 representing perfectly reproducible features.

A threshold of ICC > 0.8 was set to consider feature robust [8] to acquisition variation, and were retained based on the hypothesis that non-robust features would be too sensitive to noise to be predictive of clinical outcomes. Under this criteria, a total of 855 Radiomic features were considered as robust (**Figure S2**).

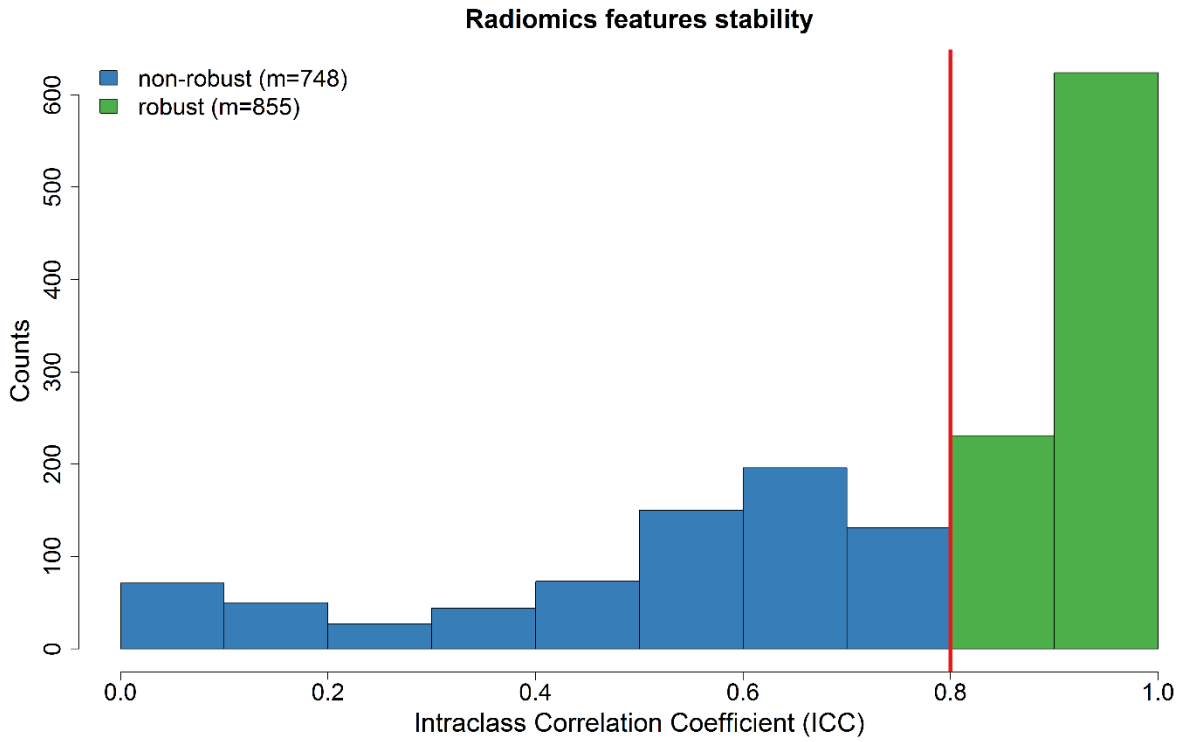

Figure S2: Histogram of intraclass correlation coefficient (ICC) computed for each Radiomic features. A threshold of 0.8 was applied, leaving a total of 855 robust radiomic features.

Feature Selection

After selecting robust features (m=855), we used Principal Component Analysis (PCA) and factor analysis to reduce the resulting high-dimensional dataset to a low-dimensional dataset while retaining most of the variation contained within the data. PCA creates a new principal components space using sing value decomposition of the features also called scores. We selected scores retaining 95% of the variability, then we selected features that correlated at least by 99% to the PCA scores.

PCA analysis was performing using the ”factoMineR” R package [9] version 1.29. After this procedure, a final subset of fifteen features were obtained based on user-set parameters described in **Table S1**.

Table S1: Equation of radiomic features with associated feature group and filter.

Selected Radiomic feature	Radiomic group	Filter associated	Equation
Sphere Disproportionality	Shape	None	$spherical\ disproportion = \frac{Area}{4\pi R^2}$ • $R = is\ the$ radius of a sphere with the same volume as the tumor
Root Mean Square	Statistics	Wavelet HLL	$RMS = \sqrt{\frac{\sum_i^N X(i)^2}{N}}$ • $X = 3\ dimensionnal\ image\ matrix\ intensity$ • $N = number\ of\ voxels$
Range	Statistics	Wavelet LLH Wavelet LHH	$range = \max(X) - \min(X)$

Energy	Statistics	Wavelet HLL	$energy = \sum_{i}^{N} X(i)^2$
Mean	Statistics	Wavelet HLL	$mean = \frac{1}{N}\sum_{i}^{N} X(i)$
Kurtosis	Statistics	LoG 3D - 5mm	$kurtosis = \frac{\frac{1}{N}\sum_{i=1}^{N}(X(i)-\bar{X})^4}{\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X(i)-\bar{X})^2}\right)^2}$ • $\bar{X} = mean(X)$
Skewness	Statistics	LoG 3D - 4mm	$skewness = \frac{\frac{1}{N}\sum_{i=1}^{N}(X(i)-\bar{X})^3}{\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X(i)-\bar{X})^2}\right)^3}$
Correlation	GLCM	Wavelet LHH	$correlation = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} ijP(i,j)-\mu_i(i)\mu_j(j)}{\sigma_x(i)\sigma_y(j)}$ • (i,j)th element represents the number of times the combination of intensity levels i and j occur in two pixels in the image • $P(i,j)$ be the co-occurrence matrix for an arbitrary δ and α • N_g be the number of discrete intensity levels in the image • μ be the mean of $P(i,j)$ • $p_x(i) = \sum_{j=1}^{N_g} P(i,j)$ be the marginal row probabilities • $p_y(i) = \sum_{i=1}^{N_g} P(i,j)$ be the marginal column probabilities • μ_x be the mean of p_x • μ_y be the mean of p_y • σ_x be the standard deviation of p_x • σ_y be the standard deviation of p_y
Entropy	GLCM	LoG 3D - 5mm	$entropy = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P(i,j)\log_2[P(i,j)]$
Homogeneity 2	GLCM	LoG 2D - 4mm	$homogeneity\ 2 = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{P(i,j)}{1+\lvert i-j\rvert^2}$

Cluster Prominence	GLCM	LoG 3D - 3mm	$cluster\ prominence = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\left[i+j-\mu_x(i)-\mu_y(j)\right]^4 P(i,j)$
Low Intensity Large Area Emphasis	GLSZM	Wavelet LHH	$LILAE = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\left[\frac{p(i,j)j^2}{i^2}\right]}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}p(i,j)}$ • the (i,j)th element describes the frequency of matrices of size j with gray level i • $p(i,j)$ be the (i,j)th entry in the given size matrix p • N_g the number of discrete intensity values in the image • N_r the number of different size matrix • N_p the number of voxels in the image
Large Area Emphasis	GLSZM	LoG 3D - 5mm	$LAE = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}j^2 p(i,j)}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}p(i,j)}$
High Intensity Large Area Emphasis	GLSZM	LoG 3D - 5mm	$HILAE = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}p(i,j)i^2j^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}p(i,j)}$

Label: GLCM = Gray-level Co-occurrence Matrix, GLSZM = Gray-Level Size Zone Matrix, LoG=Laplacian of Gaussian, L=Low, H=High.

Supplement II: Univariate Analysis Results

We are showing the Kaplan-Meier curves for the clinical outcomes (overall survival, distant metastasis and local recurrence). A median split was used for the top 2 features radiomics features and added for comparison to the pathological subgroup.

Two radiomics features were selected for each outcomes based on the two highest concordance index [10] (CI) computed. CI is the probability that among two randomly drawn samples, the sample with the higher risk value has also the higher chance of experiencing an event (e.g., death or development of DM).

None of the radiomics features were significant for the Kaplan Meier (Log-rank test). Pathological response subgrouping was significant for distant metastasis and local recurrence with p-value = 0.013 and 0.036 respectively.

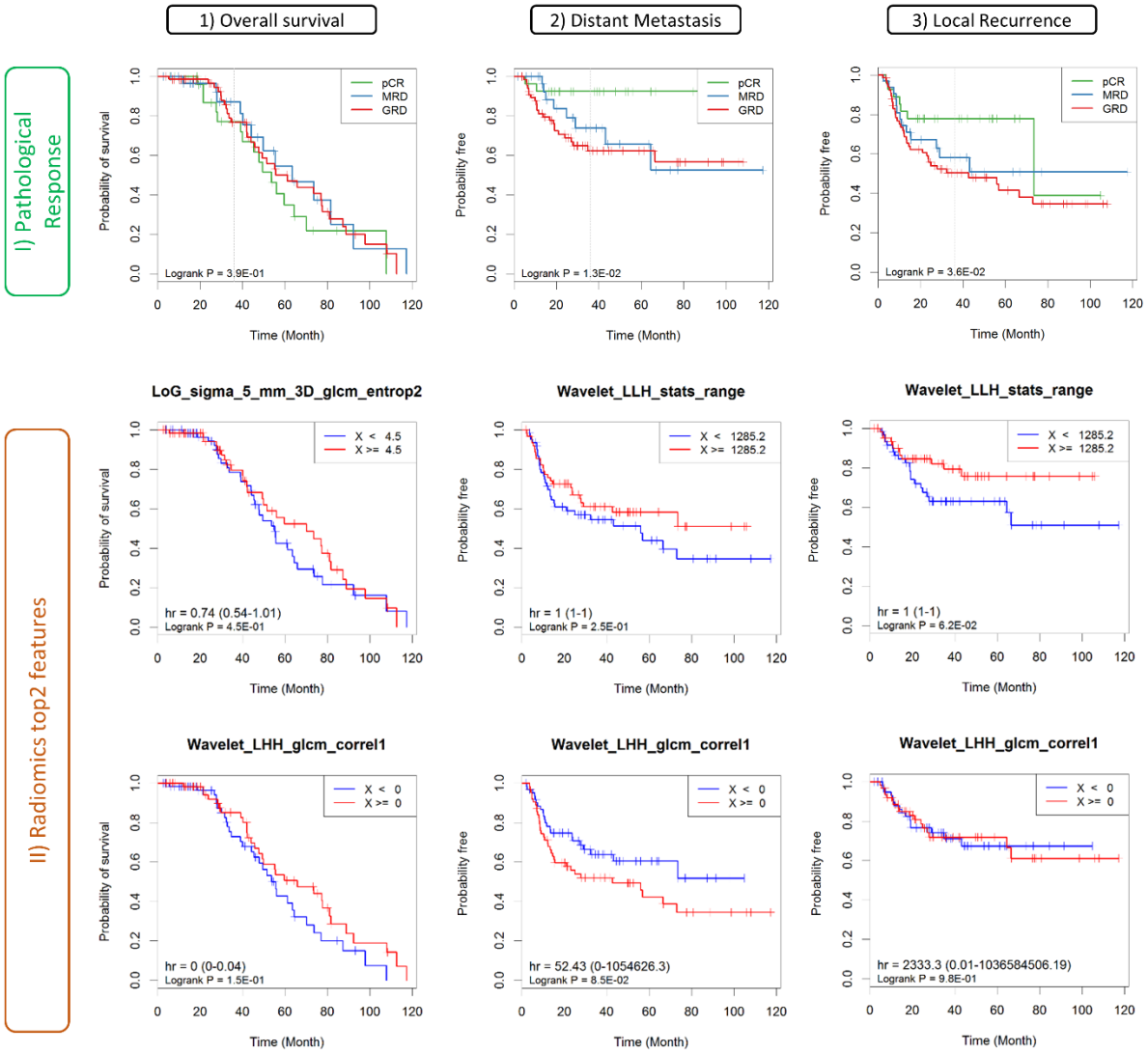

Figure S3: Kaplan-Meier curve for 1) Overall survival, 2) Distant metastasis and 3) Local recurrence. Estimate is reported for each pathological subgroup or using median split for radiomics features. Log-rank test was computed to assess difference across all groups.

In this section we are showing the univariate results for pathological complete response (pCR) in **Table S2** and gross residual disease (GRD) in **Table S3**.

Table S2: Summary of the predictive power of the fifteen selected features for pathological gross residual disease. AUC, AUC flipped, Standard Error (SE), lower/upper boundaries and p-values (Noether test) are reported.

Features	AUC flipped	AUC	SE	lower	upper	Noether p-value
Wavelet LLH stats range	0.533	0.467	0.051	0.368	0.566	0.516
Wavelet LHH glszm lowIntensityLargeAreaEmp	0.536	0.464	0.051	0.364	0.564	0.476
LoG sigma 3 mm 3D glcm clusProm	0.538	0.462	0.050	0.364	0.561	0.454
Wavelet LHH stats range	0.544	0.456	0.050	0.357	0.555	0.380
Wavelet LHH stats energy	0.563	0.437	0.050	0.339	0.536	0.210
LoG sigma 5 mm 3D stats kurtosis	0.568	0.432	0.050	0.335	0.530	0.174
Wavelet LHH glcm correl1	0.583	0.583	0.049	0.486	0.680	0.093
LoG sigma 4 mm 2D glcm homogeneity2	0.589	0.411	0.050	0.312	0.509	0.075
LoG sigma 5 mm 3D glcm entrop2	0.607	0.607	0.049	0.511	0.703	0.030
Wavelet HLL stats rms	0.607	0.607	0.049	0.510	0.704	0.031
LoG sigma 4 mm 3D stats skewness	0.615	0.385	0.049	0.290	0.481	0.019
LoG sigma 5 mm 3D glszm largeAreaEmphasis	0.626	0.374	0.051	0.274	0.474	0.014
Shape spherDisprop	0.627	0.373	0.048	0.279	0.468	0.009
LoG sigma 5 mm 3D glszm highIntensityLarteAreaEmp	0.630	0.370	0.051	0.271	0.470	0.010
Wavelet HLL stats mean	0.661	0.339	0.048	0.245	0.433	0.001
Volume	0.573	0.427	0.050	0.329	0.525	0.145
Maximum Diameter 3D	0.573	0.427	0.050	0.329	0.525	0.144
Maximum Diameter 2D axial	0.590	0.410	0.050	0.12	0.507	0.069

Table S3: Summary of the predictive power of the fifteen selected features for pathological complete response. AUC, AUC flipped, Standard Error (SE), lower/upper boundaries and p-values (Noether test) are reported.

Features	AUC flipped	AUC	SE	lower	upper	Noether p-value
LoG sigma 5 mm 3D stats kurtosis	0.516	0.484	0.062	0.363	0.606	0.797
Wavelet LHH stats range	0.517	0.517	0.059	0.402	0.632	0.767
Wavelet LHH glszm lowIntensityLargeAreaEmp	0.519	0.519	0.057	0.407	0.630	0.744
LoG sigma 4 mm 2D glcm homogeneity2	0.520	0.520	0.060	0.402	0.638	0.740
Wavelet LHH stats energy	0.526	0.526	0.060	0.409	0.642	0.668
LoG sigma 5 mm 3D glcm entrop2	0.526	0.474	0.061	0.354	0.594	0.671
Wavelet LLH stats range	0.544	0.544	0.058	0.430	0.658	0.448
Wavelet HLL stats rms	0.546	0.454	0.059	0.338	0.570	0.439
LoG sigma 4 mm 3D stats skewness	0.559	0.559	0.057	0.446	0.671	0.306
Wavelet LHH glcm correl1	0.579	0.421	0.065	0.293	0.549	0.227
LoG sigma 5 mm 3D glszm highIntensityLarteAreaEmp	0.581	0.581	0.068	0.448	0.715	0.233
LoG sigma 5 mm 3D glszm largeAreaEmphasis	0.590	0.590	0.066	0.461	0.720	0.171
Shape spherDisprop	0.609	0.609	0.056	0.500	0.717	0.051
LoG sigma 3 mm 3D glcm clusProm	0.611	0.611	0.057	0.498	0.723	0.053
Wavelet HLL stats mean	0.632	0.632	0.051	0.531	0.732	0.010
Volume	0.508	0.508	0.059	0.392	0.624	0.888
Maximum Diameter 3D	0.518	0.518	0.055	0.409	0.627	0.745
Maximum Diameter 2D axial	0.551	0.551	0.058	0.437	0.665	0.384

References

1. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, Zegers CML, Gillies R, Boellard R, Dekker A, Aerts HJWL: **Radiomics: Extracting more information from medical images using advanced feature analysis**. *Eur J Cancer* 2012, **48**:441–446.

2. Coroller TP, Grossmann P, Hou Y, Rios Velazquez E, Leijenaar RTH, Hermann G, Lambin P, Haibe-Kains B, Mak RH, Aerts HJWL: **CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma**. *Radiother Oncol* 2015, **114**:345–350.

3. *R Core Team (2013). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-Project.org/.* .

4. Zhao B, James LP, Moskowitz CS, Guo P, Ginsberg MS, Lefkowitz RA, Qin Y, Riely GJ, Kris MG, Schwartz LH: **Evaluating Variability in Tumor Measurements from Same-day Repeat CT Scans of Patients with Non–Small Cell Lung Cancer 1**. *Radiology* 2009, **252**:263–272.

5. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Cavalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D, Hoebers F, Rietbergen MM, Leemans CR, Dekker A, Quackenbush J, Gillies RJ, Lambin P: **Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach**. *Nat Commun* 2014, **5**.

6. Barkto JJ: **The Intraclass Correlation Coefficient as a Measure of Reliability**. *Psychol Rep* 1966, **19**:3–11.

7. Gamer M, Lemon J, Fellows I, Puspendra S: *Irr : Various Coefficients of Interrater Reliability and Agreement*. 2012.

8. Parmar C, Rios Velazquez E, Leijenaar R, Jermoumi M, Carvalho S, Mak RH, Mitra S, Shankar BU, Kikinis R, Haibe-Kains B, Lambin P, Aerts HJWL: **Robust Radiomics Feature Quantification Using Semiautomatic Volumetric Segmentation**. *PLoS ONE* 2014, **9**:e102107.

9. Husson F, Josse J, Le S, Mazet J: *FactoMineR: Multivariate Exploratory Data Analysis and Data Mining*. 2015.

10. Harrell FE, Califf RM, Pryor DB, Lee KL, Rosati RA: **Evaluating the yield of medical tests**. *JAMA J Am Med Assoc* 1982, **247**:2543–2546.