

# Radiomics analysis and metastatic patients: can we really continue to sidestep intra-patient tumor heterogeneity ?

**Théophraste Henry**

Molecular Radiotherapy and Therapeutic Innovation, Inserm 1030

**Roger Sun**

Molecular Radiotherapy and Therapeutic Innovation, Inserm 1030

**Marvin Lerousseau**

Molecular Radiotherapy and Therapeutic Innovation, Inserm 1030

**Théo Estienne**

Molecular Radiotherapy and Therapeutic Innovation, Inserm 1030

**Charlotte Robert**

Molecular Radiotherapy and Therapeutic Innovation, Inserm 1030

**Benjamin Besse**

Gustave Roussy

**Caroline Robert**

Gustave Roussy

**Nikos Paragios**

TheraPanacea

**Eric Deutsch** (✉ [eric.deutsch@gustaveroussy.fr](mailto:eric.deutsch@gustaveroussy.fr))

Molecular Radiotherapy and Therapeutic Innovation, Inserm 1030

---

## Article

**Keywords:** Artificial Intelligence, Diagnostic Imaging, Metastases, Machine Learning

**Posted Date:** July 7th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1775619/v1>

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

[Read Full License](#)

---

# Abstract

After promising results for localized cancer disease, radiomics analysis has been applied to the metastatic setting, sidestepping the potential intertumoral inpatient heterogeneity captured by radiomics features. Our aim was to evaluate if radiomics features can capture intertumoral inpatient heterogeneity, and the impact of tumor subsampling on the computed heterogeneity. To this end, We delineated all visible tumors from single acquisition pre-treatment computed tomography of patients with metastatic lung cancer or melanoma, and extracted tumor radiomics features. All robust radiomics indexes exhibited at least 20% of variation with significant heterogeneity both in heavily and oligo metastasized patients, and also at the organ level. None of the segmentation subsampling strategies were able to recover the true tumoral heterogeneity obtained by exhaustive tumor sampling. Further, we discussed potential limitations of predictive algorithms when this kind of heterogeneity is not taken into account.

## Introduction

In the race for accessible, non-invasive and generic biomarkers, radiomics, [1, 2] the extraction and analysis of a high number of features from imaging data, has emerged as a promising approach. From routinely obtained medical images, the radiologist or radiation oncologist delineates the patient's tumor and its digital signature is extracted. Acting as a digital biopsy, radiomics analysis has the potential to unravel what is hidden from the trained eyes of an experienced radiologist: the objective quantification of the image-based heterogeneity of the whole tumor and its significance [3, 4]. In contrast, standard biopsies are invasive, time and money consuming, and can only access a small random fraction of the complete tumor, making tissue-based heterogeneity assessment incomplete at best, misleading at worst [5].

Some promises of radiomics, such as completeness and cost-efficiency, can be fulfilled in its initial scope of application: local or loco-regional cancer, with a limited number of lesions, ideally only the primitive tumor. But, extending the scope to metastatic disease, where the total number of lesions per patient can be greater than one hundred, can be challenging. While spatial heterogeneity at the tumor level is of great interest in the localized cancer setting [6], heterogeneity between patient's lesions is also a quantity to measure in the metastatic setting [7, 8]. Indeed, polyclonality is a key factor of treatment resistance [9, 10], making any biomarker of clonal heterogeneity a game-changer for patient's care. Immune checkpoints inhibitors (ICI) have further emphasized such consideration given the high rate of non-binary response pattern at the patient level [11]. Thus, the following question arises: can radiomics analysis be performed for metastatic disease in the same way as for localized disease? In light of the critical difference between heterogeneity at a specific tumor site, and between different tumor deposits, the logical answer would be to exhaustively segment and analyse all tumors of each patient. Yet, top cited radiomics research in the metastatic setting from 2010 to 2020 has chosen another path (Table 1), mimicking what is usually done in RECIST evaluation [12], or sampling only one lesion per patient, potentially missing an entire level of heterogeneity (Fig. 1).

Table 1

Sampling strategy and rate of the top 5 most cited journal articles as ranked by Scopus between 2010 and 2020 using the following query: TITLE-ABS ( radiomics AND metastatic AND response ) AND PUBYEAR > 2009 AND PUBYEAR < 2021 AND ( LIMIT-TO ( DOCTYPE, "ar" ) ) AND ( LIMIT-TO ( SUBJAREA, "MEDI" ) ).

Author	Title	Sampling strategy	Sampling rate (per patient)
Trebeschi, S. et al (2019)	Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers	Measurable lesions defined as any tumor lesion (primary or metastatic) whose entire border could be identified. Lesions with ill-defined borders or poorly visualized were excluded	3 (median)
Antunes, J. et al (2016)	Radiomics analysis on FLT-PET/MRI for characterization of early treatment response in renal cell carcinoma: A proof-of-concept study	Primary tumor lesion	1
Dohan, A. et al (2020)	Early evaluation using a radiomic signature of unresectable hepatic metastases to predict outcome in patients with colorectal cancer treated with FOLFIRI and bevacizumab	Contouring of the dominant liver lesion (DLL)	1
van Helden, E. J. et al (2018)	Radiomics analysis of pre-treatment [18F]FDG PET/CT for patients with metastatic colorectal cancer undergoing palliative systemic treatment	All visually identifiable tumour lesions defined as SUV <sub>peak</sub> higher than two times the SUV <sub>mean</sub> of the blood pool, with a minimum volume of 4.2 mL	4 (mean)
Dercle, L. et al (2020)	Radiomics Response Signature for Identification of Metastatic Colorectal Cancer Sensitive to Therapies Targeting EGFR Pathway	Target liver lesions selected per RECIST 1.1 criteria	4 (median)

The aim of our study was to quantify image-based heterogeneity captured by radiomics features as a function of the number of tumor lesions analyzed, and to evaluate the degree of information loss induced by commonly used subsampling strategies.

## Results

### Participants

Of the four patients manually explored with available CT TAP IV + in cohort L, three were included in the study (patient A, B and C), and one excluded given his low number of lesions. A total of 127 lesions were segmented, each patient exhibiting more than 25 lesions. The most frequent lesion location was the lung

(total: 92–72.4%). The patient with the highest number of lesions had 69 lesions, with three different metastatic sites, while the maximum number of metastatic sites was five.

Regarding cohort M, 22 patients with at least 15 tumor lesions and 18 patients with between 2 and 5 tumor lesions were respectively included in cohorts M-multi and M-oligo, for a total of 475 lesions segmented. The most frequent metastatic location was the lymph node (total 188 – 39.6%). A complete description of the lesion distribution across cohorts, patients and metastatic locations is provided in Table 2.

Table 2

Nature and number of lesions across the three patients included in the study.

	Overall	Cohort A			Cohort B	
		Patient A	Patient B	Patient C	Oligometastatic (18)	Multimetastatic (22)
n_lesions	<b>602</b>	<b>127</b>			<b>475</b>	
		30	28	69	62	413
nature (%)						
bone	27 (4.5)	18 (60.0)	0 (0.0)	0 (0.0)	9 ( 2.2)	0 ( 0.0)
lung	199 (33.1)	1 (3.3)	24 (85.7)	67 (97.1)	96 (23.2)	11 (17.7)
lymphangitis	1 (0.2)	0 (0.0)	0 (0.0)	1 (1.4)	0 ( 0.0)	0 ( 0.0)
lymphnode	187 (31.1)	0 (0.0)	4 (14.3)	0 (0.0)	150 (36.3)	33 (53.2)
peritoneum	31 (5.1)	5 (16.7)	0 (0.0)	0 (0.0)	26 ( 6.3)	0 ( 0.0)
pleural	5 (0.8)	4 (13.3)	0 (0.0)	1 (1.4)	0 ( 0.0)	0 ( 0.0)
soft tissue	41 (6.8)	2 (6.7)	0 (0.0)	0 (0.0)	32 ( 7.7)	7 (11.3)
adrenal gland	8 (1.3)	0 (0.0)	0 (0.0)	0 (0.0)	5 ( 1.2)	3 ( 4.8)
liver	96 (15.9)	0 (0.0)	0 (0.0)	0 (0.0)	91 (22.0)	5 ( 8.1)
Other	7 (1.2)	0 (0.0)	0 (0.0)	0 (0.0)	4 ( 1.0)	3 ( 4.8)

Variation of the radiomics features at the patient and organ level in cohort L

Radiomics features exhibited high variation across lesions for the same patient. Patientwise, absolute coefficient of variation ranged from 1.6% for the Inverse Difference Moment Normalized (IDMN) to 321% for Skewness, considering all radiomics features. Pooling all features at the patient level, the mean coefficient of variation per patient was respectively 144, 117 and 158% for patient A, B and C.

After feature segmentation robustness assessment, only 27 features remained for the correlation analysis. Based on a visual assessment of the clustered correlation heatmap (Fig. 3), eight clusters were identified, and the following robust and uncorrelated features were selected: 90 Percentile, Large Dependence High Gray Level Emphasis, Large Dependence Emphasis (based on lexicographic sorting), Dependence Variance, Informational Measure of Correlation 1, Sphericity, Maximum 2D Diameter Slice (equivalent to RECIST measurement) and Voxel Volume (rather than Mesh Volume due to better computational efficiency). Absolute coefficient of variation ranged from 12% for Sphericity to 687% for Voxel Volume. Pooling this subset of features at the patient level, the mean coefficient of variation per patient was similar to the mean coefficient of variation for all features, respectively equal to 124, 108 and 170% for patient A, B and C.

The distributions of the absolute coefficient of variation across the seven radiomics classes are depicted in Fig. 4 for both all radiomics features, and the robust and uncorrelated subset. Robust and uncorrelated features still exhibited a significant level of variation. Sphericity and Volume variation showed respectively the minimal and maximal variations for all three patients. All robust radiomic indexes, except for Sphericity, exhibited at least 20% of variation.

Finally, going down to the organ level, lung lesions exhibited heterogeneous features, ranging from 21% for 90 percentile to 419% for tumor Volume, for the same patient.

Analysis of the variation of radiomics features at the patient and organ level in the validation cohort M

To confirm the insights gained from cohort L analysis, We conducted the same analysis on cohort M, using only robust radiomics features. Mean absolute coefficient of variation ranged between 7.2% for Sphericity, to a maximum of 144% for Volume. All features except Sphericity exhibited at least 30% of variation. All radiomics features exhibited variation significantly different from 0 ( $p < 1e-11$ ). These results were consistent across both oligo and multi metastatic sub-cohort, with mean absolute coefficient variation significantly different from 0% for all radiomics features ( $p < 1e-6$  and  $< 1e-5$  in the multi and oligometastatic cohort respectively). Moreover, there was no significant difference in the mean absolute coefficient of variation between oligo and multi metastatic population for four out of eight of the robust radiomics features (Large Dependence Emphasis, Dependence Variance, Sphericity, Maximum 2D Diameter Slice).

Finally, focusing on patients with more than one lesion for either lymph node, liver, or lung, radiomic features were highly unstable, ranging from 6% for sphericity to 106% for Volume. Advanced textural features like Large Dependence High Gray Level Emphasis or Dependence Variance exhibited an absolute variation of 54 and 18% respectively for intra-patient intra-liver lesions.

## Lesion similarity exploration and sampling simulation study

We computed the pairwise dissimilarity between all pairs of lesions for each patient in cohort L. To visualize the dissimilarity between lesions for each patient, we clustered the patient lesions solely based on the dissimilarity matrix previously computed (cf Appendix). An example of distance distribution is visualized in Fig. 5. Then, we summarized this distribution using the two previously defined heterogeneity indexes. The maximal tumoral divergence was respectively 0.94, 0.92 and 0.91 (maximal value: 1.0), and the average tumoral heterogeneity was 0.39, 0.21 and 0.21 (maximal value: 1.0).

Using these heterogeneity indexes, we explored the captured heterogeneity of common subsampling techniques. The case where only one lesion per patient was sampled was trivial: it is impossible to compute pairwise dissimilarity for only one lesion, and this sampling strategy completely failed to capture inter-tumoral intra-patient heterogeneity. For all other sub-sampling strategies, heterogeneity indexes were clearly below the ground truth value obtained by exhaustive tumor sampling. Ground truth values, mean estimated values and their 95% CI obtained by bootstrapping can be found in Table 3.

Table 3

Percentage of heterogeneity recovered by each tumor sampling strategy (mean values and their 95% confidence interval obtained using 500 simulations for each).

	<b><i>Exhaustive ground truth value</i></b>	<b><i>2 lesions at random</i></b>	<b><i>3 lesions at random</i></b>	<b><i>2 lesions per site at random</i></b>	<b><i>3 lesions per site at random</i></b>	<b><i>2 lesions per site at random, excluding bone lesions (RECIST like)</i></b>
<i>average tumoral heterogeneity</i>	<b>0.27</b>	48% [7–93]	67% [30–100]	19% [7–33]	22% [11–33]	19% [4–33]
<i>maximal tumor divergence</i>	<b>0.92</b>	27% [5–55]	55% [22–67]	11% [3–18]	16% [9–24]	10% [3–18]

Using completely random sampling strategies yielded better estimates of the true heterogeneity indexes than sampling by metastatic site or RECIST-like sampling strategies: sampling three lesions at random recovered 67% (95% CI: [30–100]) and 55% (95% CI: [22–67]) of the average tumoral heterogeneity and maximal tumor divergence respectively, whereas sampling three lesions per metastatic site recovered 22% (95% CI: [11–33]) and 16% (95% CI: [9–24]) respectively. When sampling completely at random, increasing the number of sampled lesions improved the heterogeneity index estimate, while empirically this was not the case for sampling by metastatic location. Yet, none of the sub sampling methods were able to recover the true maximal tumor divergence.

Based on the empirical values of maximal tumoral divergence and average tumoral heterogeneity, more than eight lesions per patient would have been necessary to recover at least 75% of the true heterogeneity

captured by the whole lesion distribution analysis (Fig. 6).

## Discussion

In the era of personalized medicine, radiomics studies have drawn attention, promising to predict treatment response and survival duration, leveraging the hidden “big data” from imaging studies, surpassing trained imaging specialists. But machine learning comes with its own set of strengths and potential pitfalls [13]. Through a series of in-silico experiments, we put in perspective what is achievable and what has been foreseen in previous works. First, using a cohort of 602 contoured lesions across 43 different patients and two different cancer types, we demonstrated that radiomics features vary significantly between tumors from the same patient. To prevent strong correlation between radiomics indexes to lead us to a spurious conclusion, we performed the same analysis on a very small segmentation robust and uncorrelated subset of radiomics features, and found that all of them except one exhibited at least 30% of variation. This level of heterogeneity was corroborated by cohort M, where it was observed not only in heavily metastasized patients, but also in oligo metastasized ones. Finally, tumor heterogeneity was also detected at the organ level: tumors from the same patient, and the same host organ still exhibited heterogeneity that needed to be accounted for. This has both positive and negative implications. On the bright side, radiomics features can indeed capture inter-tumoral heterogeneity, making them promising candidates for imaging based heterogeneity biomarkers. Finding image based heterogeneity between lesions of the same patient is most likely to have a significant impact on patient’s care: in the era of ICI, where pseudoprogression, hyper progression or dissociated progression are all a reality [14], being able to link a tumor-centric signature to a tumor response pattern would be an invaluable insight for patient treatment, e.g by adding local treatment for the predicted non responding tumors. This is a significant paradigm shift, the key unit of interest for response would not only be the patient, but also each of its individual tumor deposits, allowing for an even finer granularity in the treatment response assessment, and treatment options decision.

On the other hand, such levels of variation question the sturdiness of previous radiomics studies carried out on metastatic patients. Indeed, if not all tumor deposits have the same digital signature for the same patient, any predictive machine learning algorithm based on radiomics features would give a different output depending on the tumor considered as input. Thus, the only way to be able to trust such a predictive algorithm would be to define *a priori* completely deterministic and reproducible rules for lesion sampling.

To build on these findings, we crafted two simple heterogeneity indexes that combine robust and uncorrelated radiomics features from an arbitrary number of lesions into a patient-level descriptor. While there is no guarantee that these indexes correlate with common clinical endpoints like progression free or overall survival, they have the advantage of being constructed without modeling prior, in an unsupervised fashion. They are also good metrics to evaluate how much heterogeneity is lost when applying a priori rules for lesion sampling. Our simulations demonstrated that metastatic disease cannot be treated the same way as localized disease. Common subsampling techniques either completely missed inter-tumor

intra-patient heterogeneity (one lesion sampling), or greatly underestimated its extent. On cohort L constituted of completely annotated patients, the empirical value of lesions to annotate for recovering at least 75% of the true heterogeneity would have been 8. This is in stark contrast with what has been done in previously published studies (Table 1).

Our study comes with some limitations. First, we do not provide any conceptual or empirical proof that image-based and histological based heterogeneity are well correlated, nor that image based heterogeneity is predictive of therapy response, progression free survival, or overall survival. There are already multiple works linking radiomics features to histology and/or immunochemistry features [3, 15, 16], While obviously the ultimate goal, we argued that it was necessary to take the time to first carefully consider the implications of the metastatic setting, before heading straight for biological correlation and predictive performance. As such, we found that not all radiomics indexes might be relevant for heterogeneity estimation. For example, sphericity was the least variable robust radiomics index in both cohorts L and M, which is in line with common radiological knowledge: a rounded nodule of variable size is a common radiological sign of metastasis [17]. This feature is most likely to be irrelevant for heterogeneity assessment. Another limitation is the small number of patients included, despite a high number of lesions annotated. Nonetheless, our main goal was first to prove the existence of tumor heterogeneity at the imaging level, and second, roughly estimate its importance. As for the first goal, only one patient would have been sufficient. As for the second goal, we found that all of the 43 patients exhibited significant radiomics index variation. All patient imaging studies were made before immunotherapy treatment, and patients were pre-treated with at least one line of chemotherapy before imaging, which could have increased the observed tumoral heterogeneity. As such, our findings might not be relevant to untreated freshly diagnosed metastatic patients. Moreover, lung cancer and melanoma are known for high tumoral heterogeneity, and radiomics studies for other types of cancer might not be as sensitive to the sampling process [18]. Further work on other cancer types might be needed to confirm our findings. Meanwhile, as a precautionary measure, conclusions drawn from highly heterogeneous cancer types should be transposed to others. Finally, demonstrating that exhaustive tumor sampling is required for accurate heterogeneity estimation is rather cumbersome, as exhaustive tumor segmentation comes with unaffordable trained-radiologist time consumption. As such, two orthogonal future directions can be taken to build upon this work: either searching for better tumor subsampling technique, based on visual analysis of the radiological images; or looking for more efficient ways to perform complete tumor detection and segmentation. AI-driven tools might be a good fit for this job [19, 20].

In conclusion, image-based inter-lesions heterogeneity at the patient level can be successfully grasped by radiomics analyses. Further guidelines and AI-driven tools to standardize and alleviate the tumor sampling and segmentation task are required. Future works should account for the variability between tumors in the same patient, and link it to relevant biological and clinical endpoints.

## Methods

### Study design and data collection



Two different cohorts (cohort L and M) were collected. Waiver of informed consent was granted for this study according to French law regulation (MR-004). The study protocol was approved by the Gustave Roussy medical ethics committee (IRB n° 2021-85). Patients with metastatic lung cancer referred to our institution for ICI treatment between 2012 and 2019 were considered for inclusion in cohort L. For each patient, radiological examinations between three months before and one week after the first cure were automatically retrieved from the picture archiving and communication system (PACS) using in-house software. Clinical data for those patients were manually collected from the electronic health records (EHR) generated during patient follow-up. Then, using lexicographic order, consecutive patient's imaging studies were manually explored, and only patients with a single acquisition pre-treatment portal venous contrast enhanced computed tomography of thorax, abdomen and pelvis (CT TAP IV+) were investigated. Patients with less than ten visible lesions were excluded from the cohort. Given the high annotation cost of exhaustive tumor delineation and the exploratory nature of our work, we decided to use a minimum number of lesions rather than a minimum number of patients to manage the segmentation effort. The minimum number of lesions to collect was set to 100 overall for this cohort.

To confirm the findings discovered in cohort L, cohort M was collected from patients with metastatic melanoma referred to our institution treated with anti-PD-1 or anti-PD-L1 (monotherapy or combination) from May 1st, 2012 to February 18, 2019 with similar requirements for imaging and clinical data. Given the annotation cost, not all tumors in a patient were annotated, except for oligometastatic patients (cohort M-Oligo). Lesions in cohort M were annotated by residents, checked and corrected if necessary by an experienced radiation oncologist (R.S). Only patients with oligometastatic disease from two to five lesions (cohort M-oligo), and multi-metastatic patients with at least 15 annotated lesions (cohort M-multi) were included. Cohort M-multi was used to validate the extent of radiomics feature variation at the patient scale, while cohort M-oligo was used to assess if inter-tumoral heterogeneity at the patient's scale was also a reality for oligometastatic disease. Finally, in light of the previous result, we also checked if intra-organ intra-patient tumor heterogeneity was captured by radiomics features.

Patients informed consent was obtained in accordance with their respective trial protocols, and waiver of informed consent was granted for the specific details of this study according to French law regulation (MR-004) and the General Data Protection Regulation (GDPR). The study protocol was approved by the local medical ethics committee.

### Lesion delineation and radiomics features extraction

Lesion detection can be an easy task, with detection speed roughly proportional to the experience of the medical reader. However, lesion segmentation, the act of extensively delineating the tumor contours, does not scale well with practice, and requires a tremendous time and focus from the practitioner.

Accordingly, all visible tumors for each patient in cohort L were delineated, regardless of their size or location, using the 3D Slicer software [21] (<https://www.slicer.org/>). More specifically, lesions that would be deemed non targetable as per RECIST 1.1 were segmented, including bone lesions, and lesions with a

greatest diameter below 1.5cm. The segmentation process was performed by the same senior nuclear radiologist (T.H.) for all patients, in order to remove variability from inter-observer delineation.

Tumor segmentations for patients from cohort M were performed by radiation oncology residents, using Raystation Software version 9B (RaySearch Laboratories, Stockholm, Sweden). Each annotation was validated by a senior radiation oncologist (R.S) and a senior nuclear radiologist (T.H.) .

Imaging data and tumor segmentations were resampled to a common  $1 \times 1 \times 1 \text{ mm}^3$  voxel size using B-spline interpolation, and the following 107 radiomics features were extracted using the pyradiomics [22] package version 3.0 : 18 first order (FO) statistics, 14 shape-based (SB) features, 24 gray level co-occurrence matrix (GLCM) features, 16 gray level run length matrix (GLRLM) features, 16 gray level size zone matrix (GLSZM) features, five neighbouring gray tone difference matrix (NGTDM) features, and 14 gray level dependence matrix (GLDM) features. Before radiomics extraction, image discretization was performed using a fixed bin width of ten, as done in previous studies [3].

## Data Analysis

To quantify the raw captured heterogeneity, the absolute coefficient of variation (CV) of each radiomics index was calculated at the patient level. A sensitivity analysis was then performed using only a subset of all extracted features as advised previously [23]. First, only features robust with respect to the segmentation process were kept: each tumor segmentation was eroded and dilated isotropically by 1mm and radiomics indexes that exhibited a concordance correlation coefficient below 0.8 between the original segmentation and each of the derived ones were dropped. Then, we performed a correlation analysis: radiomics features were clustered using hierarchical clustering, and only one feature per identified cluster was selected, keeping only the most understable one (e.g., tumor volume rather than gray level non-uniformity index), or the first one by lexicographic sorting if the correlated features were all hard to conceptualize.

To confirm the insights gained from the extensively annotated cohort L, the same analysis was performed on cohort M-oligo and M-multi, and then at the patient's organ level for lung, liver and lymph nodes, using only the robust radiomics features. The mean CV coefficient of each radiomics feature was tested for non-nullity using a one sample t-test on the M cohort. Mean CV coefficients of each radiomics feature were compared between cohorts M-oligo and M-multi using the Wilcoxon rank sum exact test. No correction for multiple testing was performed given the exploratory nature of this study. A p-value below 0.05 was deemed significant.

Finally, as a first step towards formal heterogeneity estimation at the patient level, using the extensively annotated cohort L, we proposed two overall heterogeneity measurements with the previously selected features: the maximal tumoral divergence, defined as the maximum pairwise dissimilarity between the patient's tumor deposits, and the average tumoral heterogeneity, defined as the average pairwise dissimilarity (detailed formulation in Appendix). To account for the high dimensionality of the radiomics feature vector of each tumor, we chose the cosine distance as the dissimilarity metric. These

heterogeneity measurements were used to assess the extent of information lost by commonly used tumor sampling strategies. The sampling strategies explored were: two or three lesions randomly selected; random sampling of a maximum of two or three lesions per metastatic location; sampling two lesions at most per metastatic location excluding bone lesions (RECIST 1.1 like sampling strategy). For each sub-sampling strategy, 500 simulations were conducted, and the mean value of each heterogeneity index and its 95% confidence interval (IC 95) was computed. Finally, we performed random lesion sampling of two to a maximum of fifteen lesions per patient, in order to empirically determine the number of lesions required to recover at least 75% of the average tumoral heterogeneity and maximal tumoral divergence.

The complete data pipeline and analysis process is depicted in Fig. 2.

## Declarations

### *Role of the funding source:*

This work was supported by “Fondation pour la Recherche Médicale” [grant number DIC20161236437] and “Fondation ARC pour la recherche contre le cancer” [grant number SIGN'IT20181007805].

### *Ethical approval:*

Patients informed consent was obtained in accordance with their respective trial protocols, and waiver of informed consent was granted for the specific details of this study according to French law regulation (MR-004) and the General Data Protection Regulation (GDPR). The study protocol was approved by the local medical ethics committee.

### *Declaration of Potential Conflicts of Interest:*

R.S. reports research and travel grants from Fondation ARC and Université Paris-Saclay and support from Inserm and Fondation Bettencourt Schueller outside of the submitted work. Ca.R. is a consultant advisor for Amgen, AstraZeneca, BMS, Merck, MSD, Novartis, Pierre Fabre, Roche and Sanofi. E.D. reports grants and personal fees from Roche Genentech, grants from Servier, grants from Astrazeneca, grants and personal fees from Merck Serono, grants from BMS, and grants from MSD, outside the submitted work. All remaining authors have declared no conflicts of interest

### *Data sharing statement:*

In accordance with the General Data Protection Regulation (GDPR), none of the individual participant imaging data can be made available to others. However, the deidentified radiomic features dataset that underlie the results reported in the article will be made available to researchers who provide a methodologically sound proposal beginning 3 months and ending 2 years following article publication. Information regarding submitting proposals and accessing data may be submitted to [eric.deutsch@gustaveroussy.fr](mailto:eric.deutsch@gustaveroussy.fr).

### *Authors' contributions:*

T.H, R.S, Caroline R., Charlotte R., B.B and E.D, contributed to data acquisition. T.H., R.S, E.D, N.P, M.L and T.E contributed to data processing and analysis. T.H and R.S manually inspected all delineated tumors for correctness and had verified the integrity of the radiomics data extracted. T.H wrote the manuscript draft. All authors revisited the manuscript critically and approved the final version. All authors had full access to all the data in the study and accepted responsibility to submit for publication.

## References

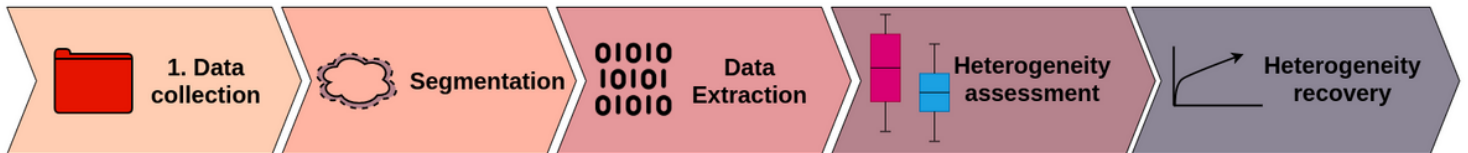
1. Kumar V, Gu Y, Basu S, *et al.* Radiomics: the process and the challenges. *Magn Reson Imaging* 2012; **30**: 1234–48.
2. Lambin P, Rios-Velazquez E, Leijenaar R, *et al.* Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012; **48**: 441–6.
3. Sun R, Limkin EJ, Vakalopoulou M, *et al.* A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD–1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. *Lancet Oncol* 2018; **19**: 1180–91.
4. Parmar C, Leijenaar RTH, Grossmann P, *et al.* Radiomic feature clusters and Prognostic Signatures specific for Lung and Head & Neck cancer. *Sci Rep* 2015; **5**: 11044.
5. Jamal-Hanjani M, Wilson GA, McGranahan N, *et al.* Tracking the Evolution of Non–Small-Cell Lung Cancer. *N Engl J Med* 2017; **376**: 2109–21.
6. Wersto RP, Liblit RL, Deitch D, Koss LG. Variability in dna measurements in multiple tumor samples of human colonic carcinoma.; **67**: 10.
7. Gerlinger M, Rowan AJ, Horswell S, *et al.* Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N Engl J Med* 2012; **366**: 883–92.
8. Naxerova K, Reiter JG, Brachtel E, *et al.* Origins of lymphatic and distant metastases in human colorectal cancer. *Science* 2017; **357**: 55–60.
9. Russo M, Siravegna G, Blaszkowsky LS, *et al.* Tumor Heterogeneity and Lesion-Specific Response to Targeted Therapy in Colorectal Cancer. *Cancer Discov* 2016; **6**: 147–53.
10. McGranahan N, Swanton C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell* 2017; **168**: 613–28.
11. Liu D, Jenkins RW, Sullivan RJ. Mechanisms of Resistance to Immune Checkpoint Blockade. *Am J Clin Dermatol* 2019; **20**: 41–54.
12. Schwartz LH, Litière S, de Vries E, *et al.* RECIST 1.1-Update and clarification: From the RECIST committee. *Eur J Cancer Oxf Engl* 1990 2016; **62**: 132–7.
13. Dercle L, Henry T, Carré A, Paragios N, Deutsch E, Robert C. Reinventing radiation therapy with machine learning and imaging bio-markers (radiomics): State-of-the-art, challenges and perspectives. *Methods* 2021; **188**: 44–60.

14. Borcoman E, Kanjanapan Y, Champiat S, *et al.* Novel patterns of response under immunotherapy. *Ann Oncol* 2019; **30**: 385–96.
15. Mazzaschi G, Milanese G, Pagano P, *et al.* Integrated CT imaging and tissue immune features disclose a radio-immune signature with high prognostic impact on surgically resected NSCLC. *Lung Cancer* 2020; **144**: 30–9.
16. Tang C, Hobbs B, Amer A, *et al.* Development of an Immune-Pathology Informed Radiomics Model for Non-Small Cell Lung Cancer. *Sci Rep* 2018; **8**: 1922.
17. Seo JB, Im J-G, Goo JM, Chung MJ, Kim M-Y. Atypical Pulmonary Metastases: Spectrum of Radiologic Findings. *RadioGraphics* 2001; **21**: 403–17.
18. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer Genome Landscapes. *Science* 2013; **339**: 1546–58.
19. Yan K, Wang X, Lu L, Summers RM. DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. *J Med Imaging* 2018; **5**: 1.
20. Henry T, Carre A, Lerousseau M, *et al.* Brain tumor segmentation with self-ensembled, deeply-supervised 3D U-net neural networks: a BraTS 2020 challenge solution. *ArXiv201101045 Cs Eess* 2020; published online Nov 27. <http://arxiv.org/abs/2011.01045> (accessed Sept 27, 2021).
21. Fedorov A, Beichel R, Kalpathy-Cramer J, *et al.* 3D Slicer as an Image Computing Platform for the Quantitative Imaging Network. *Magn Reson Imaging* 2012; **30**: 1323–41.
22. van Griethuysen JJM, Fedorov A, Parmar C, *et al.* Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* 2017; **77**: e104–7.
23. Lambin P, Leijenaar RTH, Deist TM, *et al.* Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017; **14**: 749–62.

## Figures

### Figure 1

Radiomic analysis strategy: from one lesion radiomics to metastatic disease data extraction. Created with BioRender.com



**Figure 2**

General view of the data pipeline and the analysis process

**Figure 3**

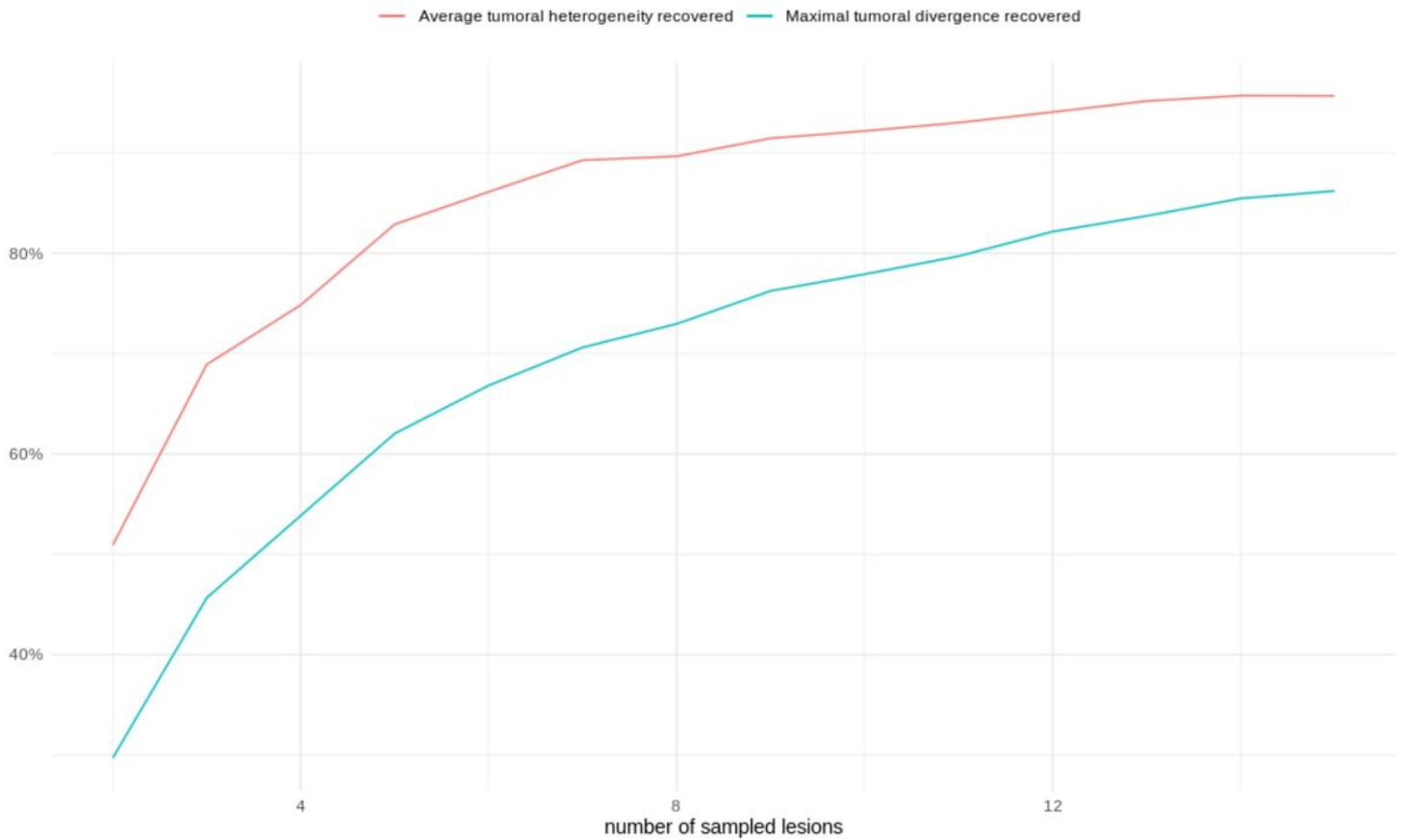
Clustered correlation heatmap of the 27 robust radiomic features. The absolute value of the correlation between features was displayed, for ease of interpretation. 8 clusters are distinguishable.

**Figure 4**

Absolute coefficient of variation considering **A**: all 107 radiomic features ; **B**: eight robust and uncorrelated radiomic features (Sensitivity analysis). Radiomics features are subdivided by patient (one color per patient) and the commonly used radiomic features' classes. *glcm*: Gray Level Co-occurrence Matrix ; *gldm*: Gray Level Dependence Matrix; *glrlm*: Gray Level Run Length Matrix; *glszm*: Gray Level Size Zone; *ngtdm*: Neighbouring Gray Tone Difference Matrix.

**Figure 5**

**A:** Dissimilarity between lesions' pairs for patient A. **B and C:** Volume rendering of the contoured lesions (face + profile).



**Figure 6**

Recovery curves of maximal tumoral divergence and average tumoral heterogeneity for cohort L.

## Supplementary Files

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

- [appendix.docx](#)