Radiomics: a new era for tumour management ?

Irène Buvat
Unité Imagerie Moléculaire In Vivo (IMIV)
CEA – Service Hospitalier Frédéric Joliot
Orsay, France
irene.buvat@u-psud.fr
http://www.guillemet.org/irene

Outline

- Tumour management in PET: back to basics
- Quantification in PET for lymphomas: why discrepant results?
- Radiomics: towards a new era for tumour management

Introduction

Basic assumption:

PET reflects relevant quantitative metabolic information

about the disease

Tumour management in PET: back to basics (1)

- 1st point of attention: Metabolic information is sound only if a number of prerequisites are met

- Major prerequisites pertain to:
  - fasting time > 4h
  - delay between injection and acquisition times : ~ 60 min±10 min
  - blood glucose level : < 120 mg/L in non-diabetic patients
  - appropriate attenuation correction (no oral contrast agent should be used)

Tumour management in PET: back to basics (1)

- 1st point of attention: Metabolic information is sound only if a number of prerequisites are met

- Major prerequisites pertain to:
  - fasting time > 4h
  - delay between injection and acquisition times : ~ 60 min±10 min
  - blood glucose level : < 120 mg/L in non-diabetic patients
  - appropriate attenuation correction (no oral contrast agent should be used)

Obvious ? a review from the literature* shows that even recent reports do not always meet these requirements

after CT attenuation correction. CT images were acquired with 130 mAs, 130 kV, and slice width (or 5 min and table feed) of 8 mm per rotation. Intravenous or oral contrast agents were used in all patients, and a standardized breathing protocol was

[redacted] 2012

Biograph PET/CT scanner. Patients were instructed to fast for at least 6 hours and blood glucose level was measured to ensure that it was less than 200 mg/dL before radiotracer injection. Approximately, sixty minutes after

[redacted] 2013

** Only lymphoma-related studies referred to in this talk*

Tumour management in PET: back to basics (2)

- 2nd point of attention: when prerequisites are met, many other factors can introduce differences between PET measurements, including:
  - PET scanner model
  - PET reconstruction algorithm (with or w/o PSF modelling) and associated parameters (post-filtering)
  - Voxel size
  - Measurement methods

Tumour management in PET: back to basics (2)

- 2nd point of attention: when prerequisites are met, many other factors can introduce differences between PET measurements, including:
  - PET scanner model
  - PET reconstruction algorithm (with or w/o PSF modelling) and associated parameters (post-filtering)
  - Voxel size
  - Measurement methods

All associated parameters should be **carefully reported** when using quantitative criteria to interpret images

dence mode, followed by a 1-min transmission scan (^{137}Cs source). Images (144 × 144 matrix; voxel size, 4 × 4 × 4 mm^3) were reconstructed using an iterative ordered-subsets expectation maximization (OSEM) algorithm with attenuation correction. The last 11 patients were scanned on a Gemini PET/CT system

[redacted] 2007

intravenously, and whole-body images were acquired a median of 69 min after injection and were reconstructed using iterative protocols with body weight-normalized SUV computation.

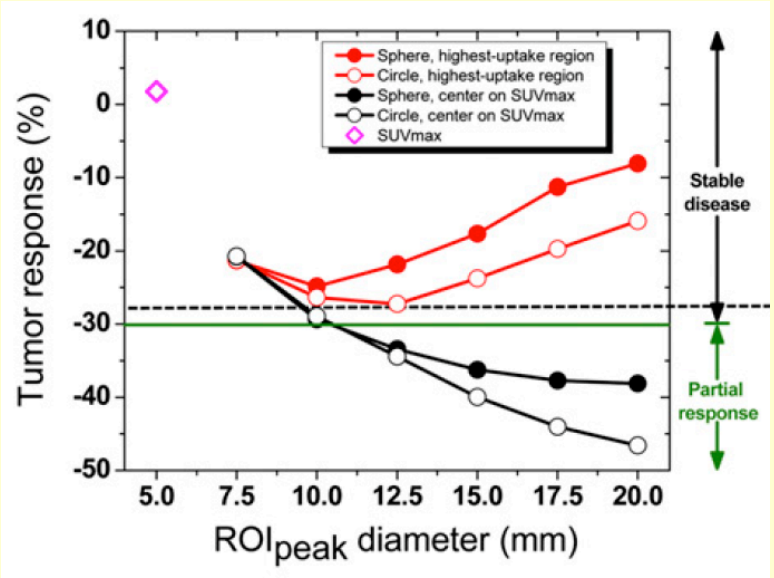
[redacted] 2014

Tumour management in PET: back to basics (3)

- 3rd point of attention: concepts that may not seem ambiguous can actually be equivocal
  - SUVmax
  - SUVpeak
  - Lean body mass

At baseline, the average SUVmax values in the most active tumour were 19.3±9.9, 19.1±9.7 and 18.6±9.3 for the three observers, respectively. After two cycles, these values decreased to 3.9±2.8, 3.8±2.7 and 3.5±2.7, respectively.

Itti et al, EJNMMI 2013



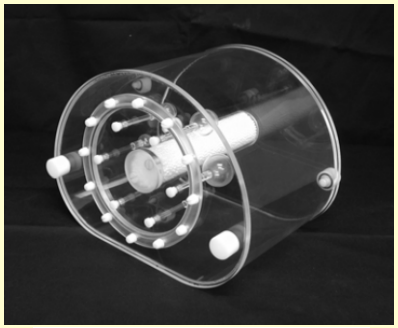
Vanderhoek et al, J Nucl Med 2012

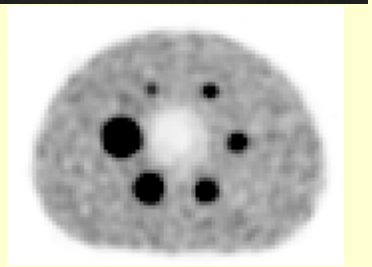
Protocols and definitions should be harmonized*
The resulting image quality should be characterized (QC phantom) and reported (so that data can be interpreted accordingly)*

** On-going efforts in these directions*

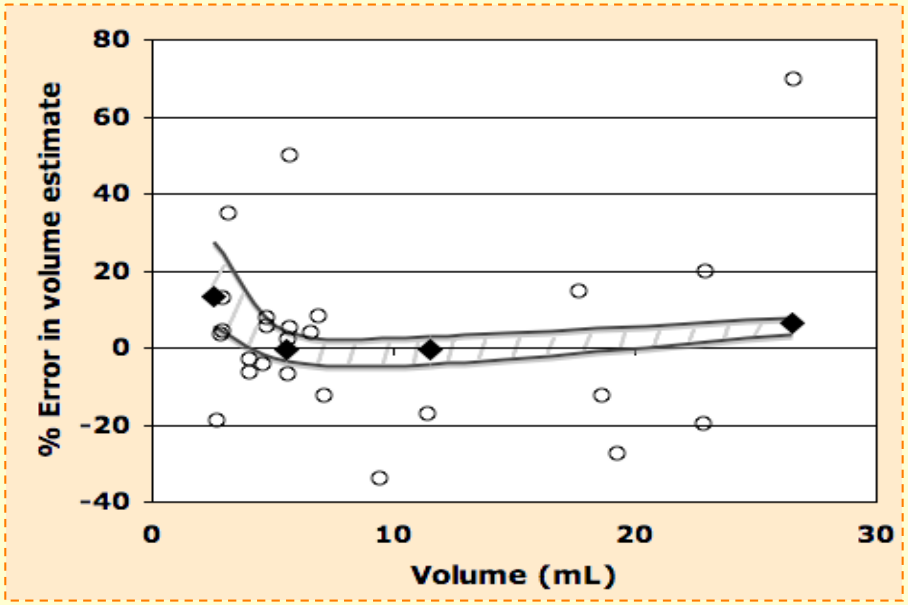
Tumour management in PET: back to basics (4)

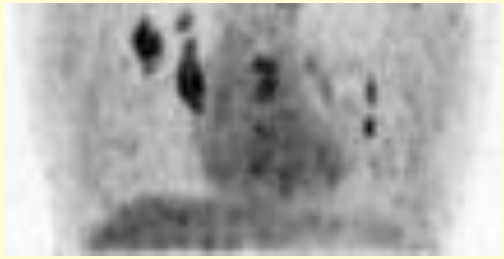
- 4th point of attention: although phantom studies are needed, they do not accurately predict performances that will be acquired *in vivo*





◆ spheres





○ highly realistic simulated data

Parameter tuning and extrapolation of performance from phantom data to patients are subject to errors

Stute et al, IEEE MIC Conf Proceedings 2008

Quantification in PET: usefulness and limitations

- SUVs, MTV or TMTV, TLG (most avid or all lesions)

All these indices have been shown to be useful in some lymphoma-related applications (staging, monitoring)

Yet, confusing statements regarding which index is best, how to measure it, which cut-off to choose, and resulting performance

Example: Baseline PET of DLBCL as a predictor of outcome*

Paper	Index yielding significant results	Usefulness	Cut-off
Chihara 2011	SUVmax	3y PFS	SUVmax=30
Song 2012	TMTV	3y PFS	TMTV=200 mL
Kim 2012	$TLG_{50\%}$	2y PFS	TLG=415.5 (g)
Esfahani 2013	TLG	PFS	TLG=704.77 (g)

Why such discrepancies in conclusions ?

** Sample of publications, not a comprehensive study*

One reason: variability in the protocols

- Acquisition protocol : prerequisites not always met
- Reconstruction protocol
- Measurement protocol

This variability partly explains variable results and conclusions

Cut-off values depend on the index and protocols

Metastatic colorectal cancer
Interim PET @ day 14 of treatment
Targeting a 95% sensitivity for detecting responding lesions

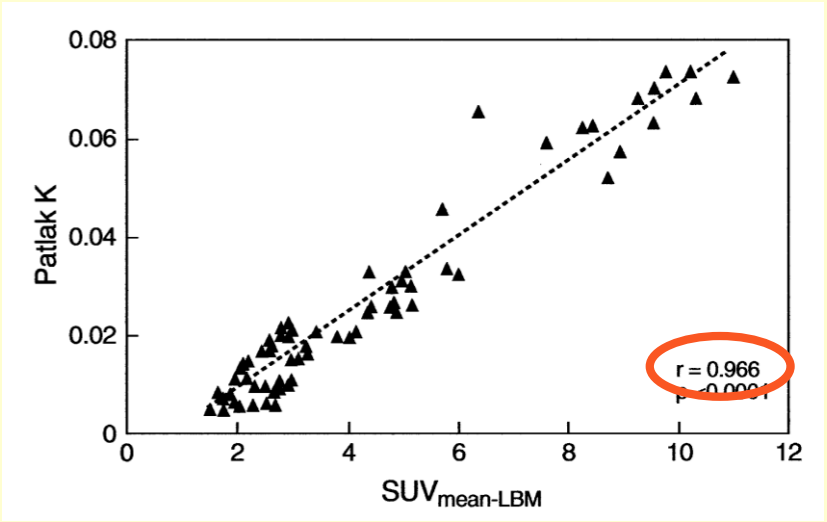
Index	Cut-off	Sensitivity	Specificity
ΔSUVmax	-14%	95%	53%
ΔSUV40%	-22%	95%	64%
ΔSUVmax	-15%	80%	53%
ΔSUV40%	-15%	95%	53%

Buvat et al, EJNMMI 2013

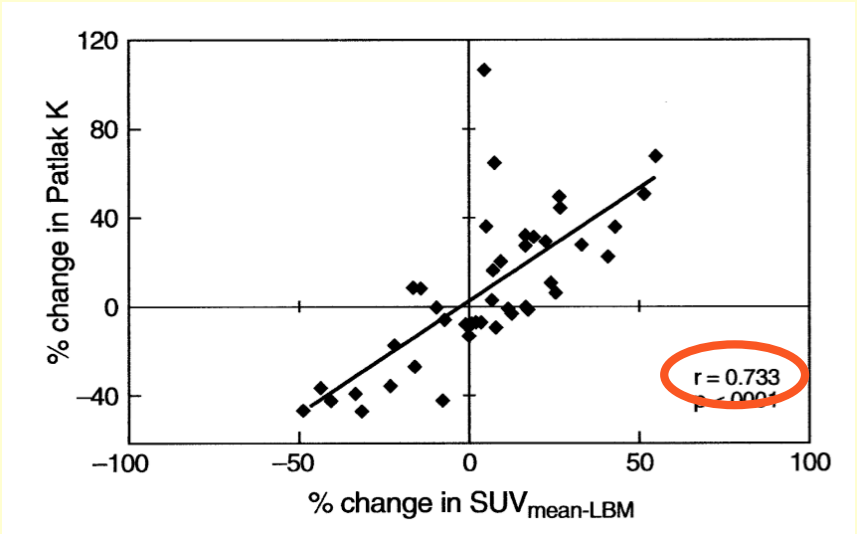
The definition of a cut-off value is meaningful only
if all protocol and processing parameters are set unambiguously

Aggravating factor: variability gets worse when comparing scans*

$$sd_{(1-2)} = \text{sqrt}\ (sd_1^2 + sd_2^2)$$



PET 1: SUV properly estimates K, so SUV is a good biomarker …



PET2 – PET1: … but differences in SUV do not reflect differences in K so well

Freedman et al, Eur J Nucl Med 2003

** Concern for interim scan interpretation*

Other possible reasons

- Differences in patient profiles (IPI, stage) between series

- Differences in end point used to determine the index usefulness and cut-off value (PFS, OS, complete remission, …)

First take home message

- Quantification in PET is tricky
- Not all quantitative results can be trusted due to methodological flaws
- Variability in acquisition / reconstruction / measurement protocols might explain some discordant results

Protocol **S**tandardization
+ careful **Q**uality check
+ more comprehensive **R**eporting *
would certainly improve the consistency of results between centres
Hint to remember: **PQRS strategy**

The same analysis hold true in MR imaging
(ie MR is not easier – possibly worse – than PET in that respect)

And yet it works !

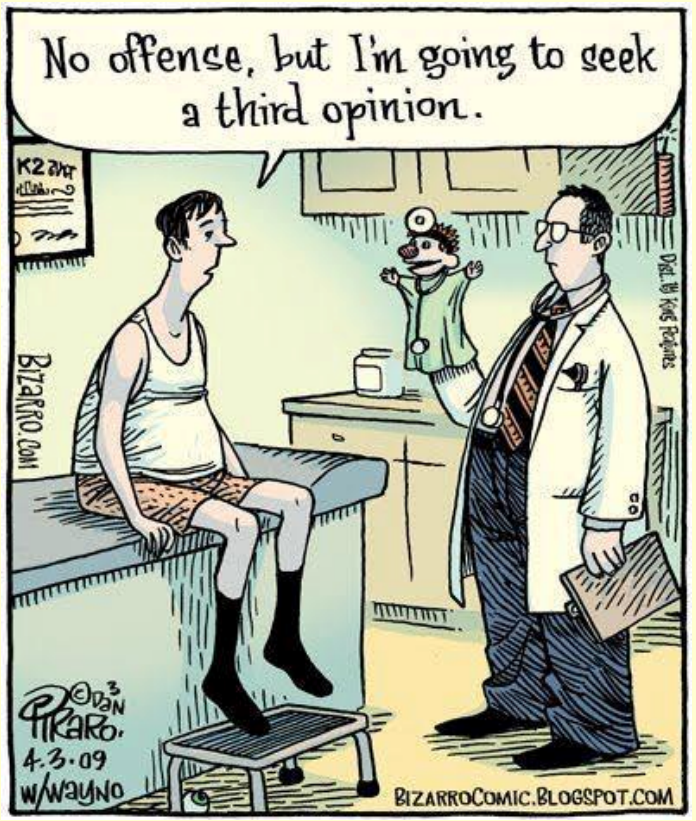
- Despite all limitations mentioned before, there is converging evidence that PET is useful for lymphoma staging and monitoring

- This suggests that the images include very relevant information

- What do we have to gain by using PQRS ?

  - statistical power (fewer patients to get significant results)

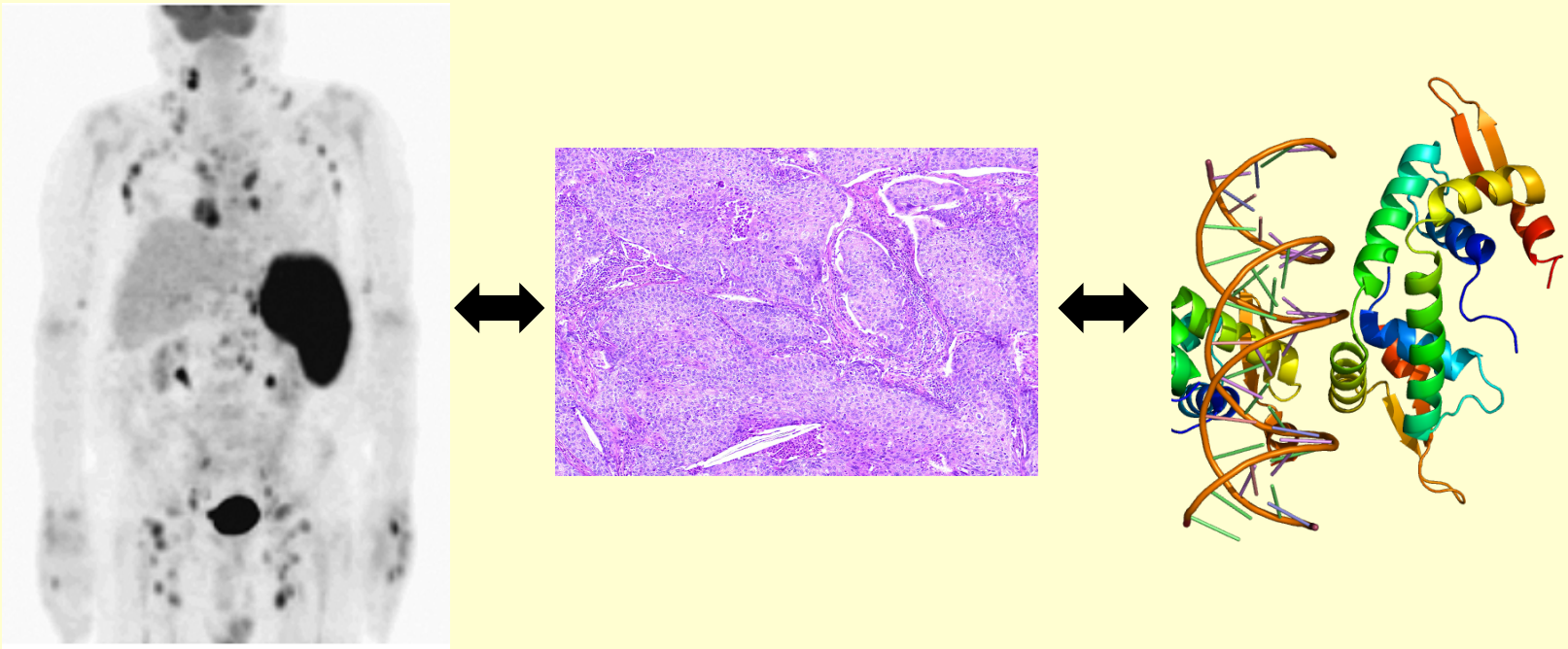
  - credibility towards referring MDs



Beyond SUV, MTV, TLG: Radiomics

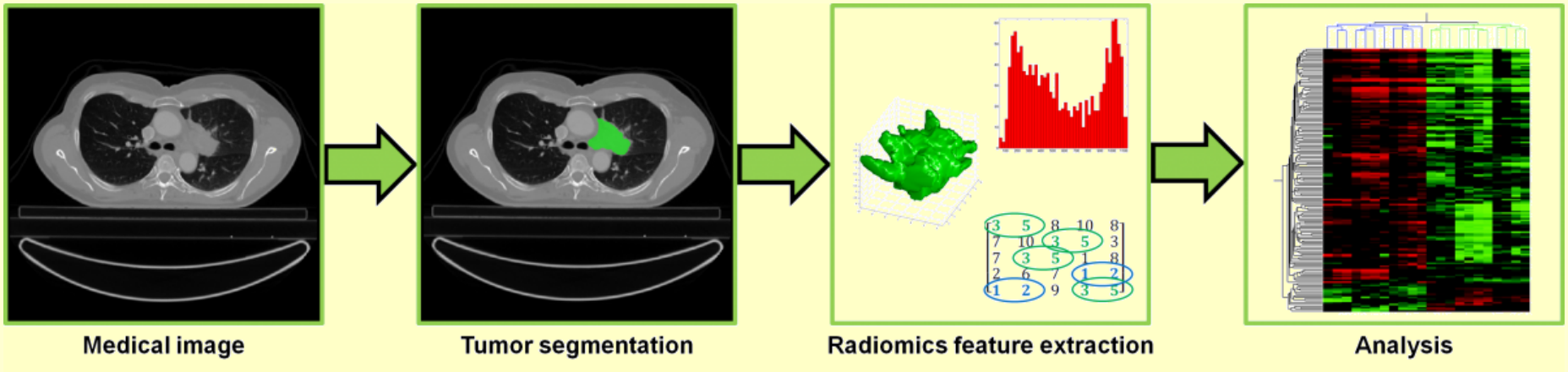
Radiomics: the high-throughput extraction of **large amounts of image features** from radiographic images

Assumption: “advanced image analysis on conventional and novel medical imaging can capture additional information not currently used, and more specifically, that genomic and proteomic patterns can be expressed in terms of macroscopic image-based features”



Lambin et al Eur J Cancer 2012

Radiomics: workflow



Each of these 4 steps has its own challenges:

- Image acquisition: standardization (cf previous slides)

Kumar et al Magn Reson Imaging 2012

Tumour segmentation

- There is no such thing as accurate segmentation of a tumour, no ground truth segmentation in a patient

- What matters is

  - reproducibility: any operator should get the same result from the same image

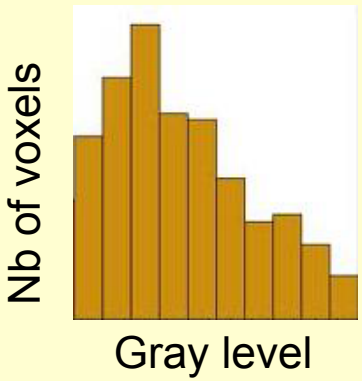
  - understanding how features depend on the segmentation step

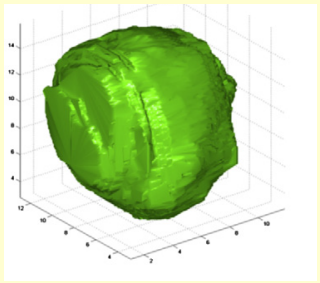
Group	Index	sd_5(%) MCC	NSCLC	BC	Change
1	Homogeneity	3.6	1.9	2.2	Moderate
	Correlation	6.8	4.5	9.1	Large
	Contrast (CM)	7.7	5.1	4.9	Large
	Dissimilarity	4.5	2.7	2.7	Moderate
	Contrast (NGLDM)	9.1	6.1	5.9	Large
2	Energy	6.6	3.0	4.1	Large
	Entropy	1.5	0.7	1.0	Low
	Coarseness	6.9	2.8	3.1	Moderate
3	SRE	0.1	0.1	0.1	Low
	RP	0.2	0.1	0.1	Low
	SZE	1.2	0.9	0.9	Low
	ZP	2.4	1.3	1.4	Moderate
4	LRE	0.5	0.3	0.3	Low
	LZE	7.5	4.4	4.2	Large
	LZHGE	5.4	5.6	5.9	Large
5	LGRE	13.0	10.3	13.5	Large
	SRLGE	12.9	10.3	13.5	Large
	LRLGE	13.1	10.4	13.5	Large
	LGZE	12.3	9.7	13.3	Large
	SZLGE	13.3	10.4	14.5	Large
6	HGRE	4.1	2.7	3.0	Moderate
	SRHGE	4.1	2.6	2.9	Moderate
	LRHGE	3.9	2.8	3.1	Moderate
	HGZE	3.7	2.5	2.9	Moderate
	SZHGE	3.8	2.4	2.8	Moderate
7	GLNUr	10.8	3.6	4.6	Large
	RLNU	9.3	2.3	3.3	Moderate
	GLNUz	8.9	3.0	3.4	Moderate
	ZLNU	6.6	2.4	2.9	Moderate
	MV	9.7	2.4	3.4	Moderate
	TLG	7.7	1.6	2.5	Moderate
8	SUVmax	0.7	0.0	0.0	Low
	SUVmean	2.7	1.1	1.2	Low
	SUVpeak	0.0	0.0	0.0	Low
	SD_{Hist}	8.3	10.4	9.1	Large
	$Entropy_{Hist}$	4.8	5.1	5.5	Large
	$Energy_{Hist}$	10.7	11.5	11.8	Large
9	Skewness	67.9	440.6	158.5	Large
	Kurtosis	15.0	14.5	12.0	Large
10	Busyness	1267.4	119.6	20.1	Large
11	LZLGE	14.8	10.6	13.2	Large

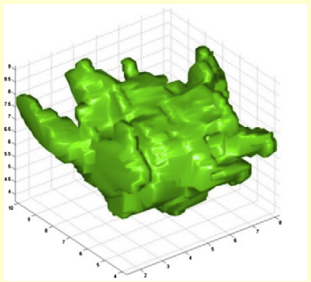
Orlhac et al J Nucl Med 2014

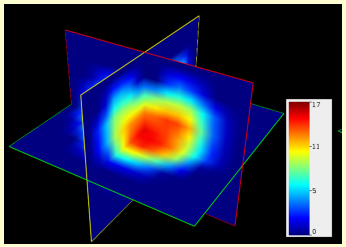
Which features?

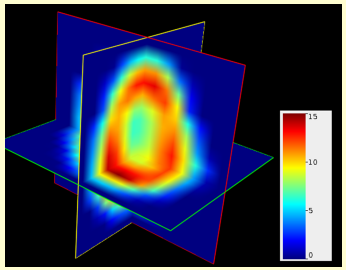
- Tumour intensity histogram indices (mean, standard deviation, skewness, …)
- Shape indices
- Textural indices
- Margin information
- Multi-scale features (wavelet)
- (Whole body biodistribution of lesions)

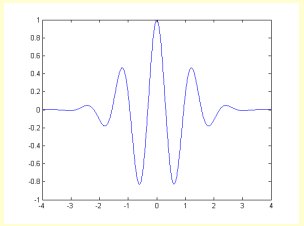


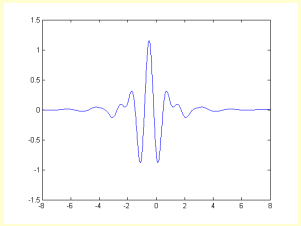


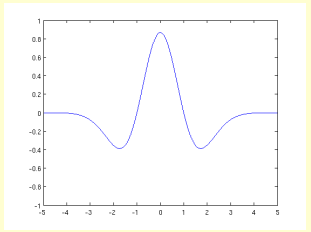


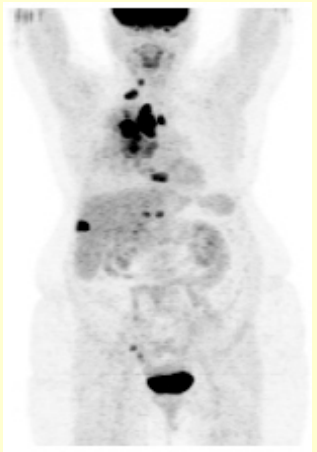






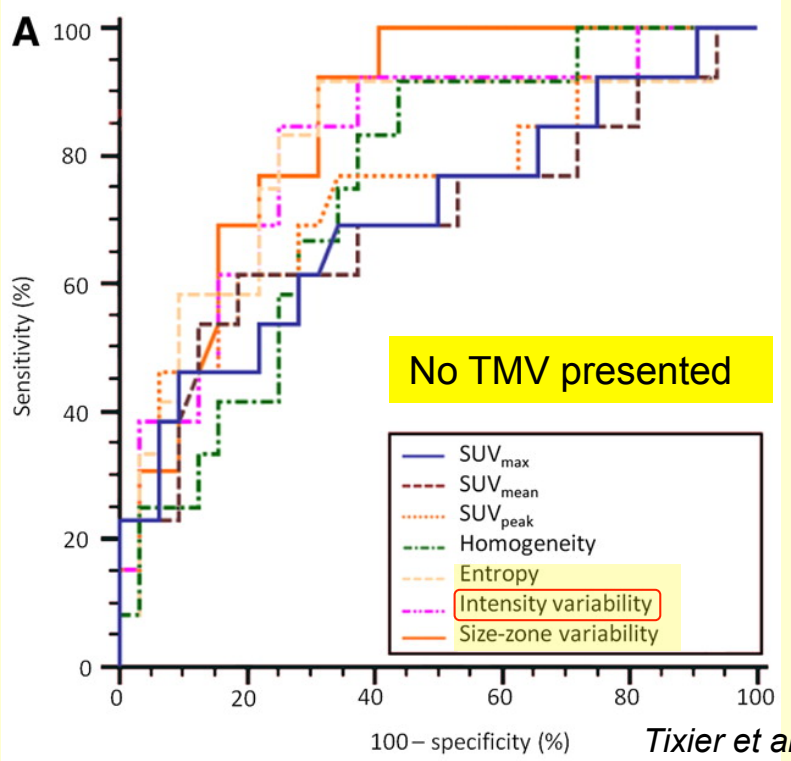




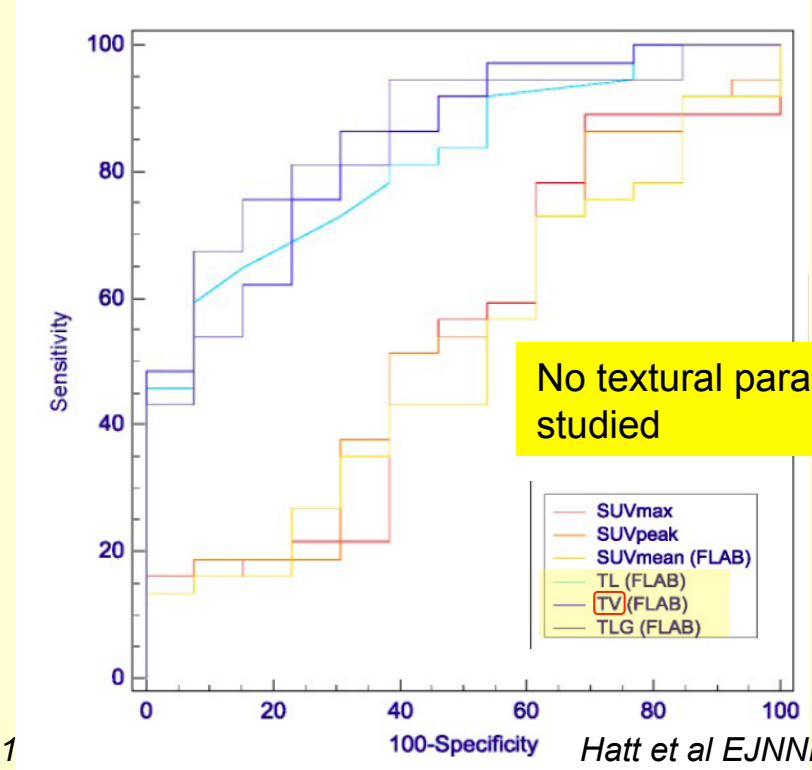


Important: understand the redundancy between features for dimensionality reduction and selection of the best features for subsequent analysis

Poor understanding of features yields misleading conclusion



Tixier et al JNM 2011



Hatt et al EJNNMI 2011

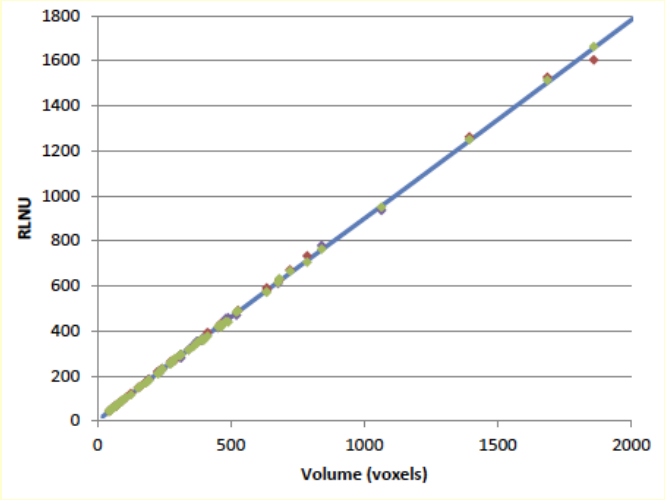
Esophageal cancer treated using radio-chemotherapy
Prediction of treatment response based on the baseline PET

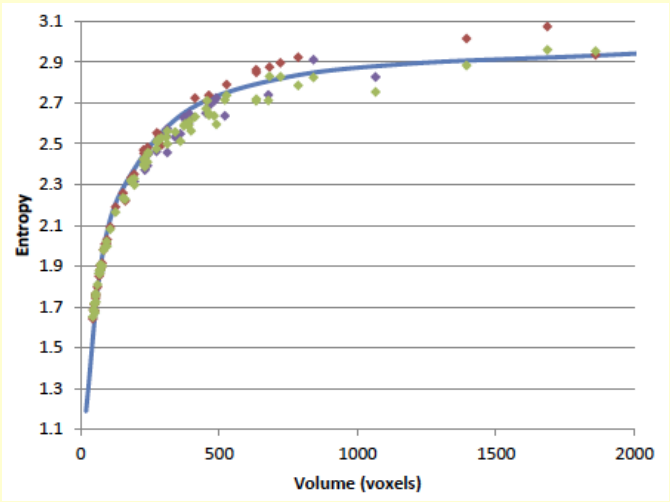
Comparing these two studies suggests that the texture parameters had no added values compared to the metabolic tumour volume!

A multivariate analysis would have been needed

Relationships between textural parameters and TMV

	MV		
	CRC	LUNG	BREAST
Homogeneity	0.62	0.72	0.68
Energy	-0.58	-0.58	-0.46
Correlation	0.67	0.73	0.64
Contrast	-0.56	-0.71	-0.56
Entropy	0.78	0.80	0.64
Dissimilarity	-0.60	-0.73	-0.61
SRE	-0.71	-0.74	-0.77
LRE	0.73	0.57	0.78
LGRE	-0.49	-0.25	-0.43
HGRE	0.32	0.40	-0.04
SRLGE	-0.49	-0.38	-0.43
SRHGE	0.30	0.37	-0.05
LRLGE	-0.48	0.12	-0.41
LRHGE	0.41	0.48	0.04
GLNUr	0.98	0.99	0.97
RLNU	1.00	1.00	1.00
RP	0.99	0.97	0.98
Coarseness	-0.73	-0.78	-0.69
Contrast	-0.56	-0.70	-0.57
Busyness	-0.05	0.05	0.06
SZE	-0.68	-0.80	-0.65
LZE	0.77	0.44	0.78
LGZE	-0.49	-0.57	-0.45
HGZE	0.44	0.42	0.02
SZLGE	-0.47	-0.62	-0.46
SZHGE	0.17	0.15	-0.14
LZLGE	-0.05	0.22	0.21
LZHGE	0.87	0.87	0.78
GLNUz	0.99	0.99	0.99
ZLNU	0.99	0.97	0.90
ZP	0.78	0.54	-0.03





Orlhac et al J Nucl Med 2014

To be kept in mind when working with **textural** parameters

Control quality of features

- The stability of features in **test-retest studies and between observers** should be carefully characterized

- Some results are already available to get an idea of the robustness of each feature (NSCLC tumours)

Index	Test-retest ICC*	Inter-observer ICC*
SUVmax	0.93	1
SUVmean	0.87	0.95
SUVpeak	0.94	1
Volume	0.84	0.98
Entropy	0.90	0.98
GLNU	0.79	0.97
Compactness	0.85	0.98

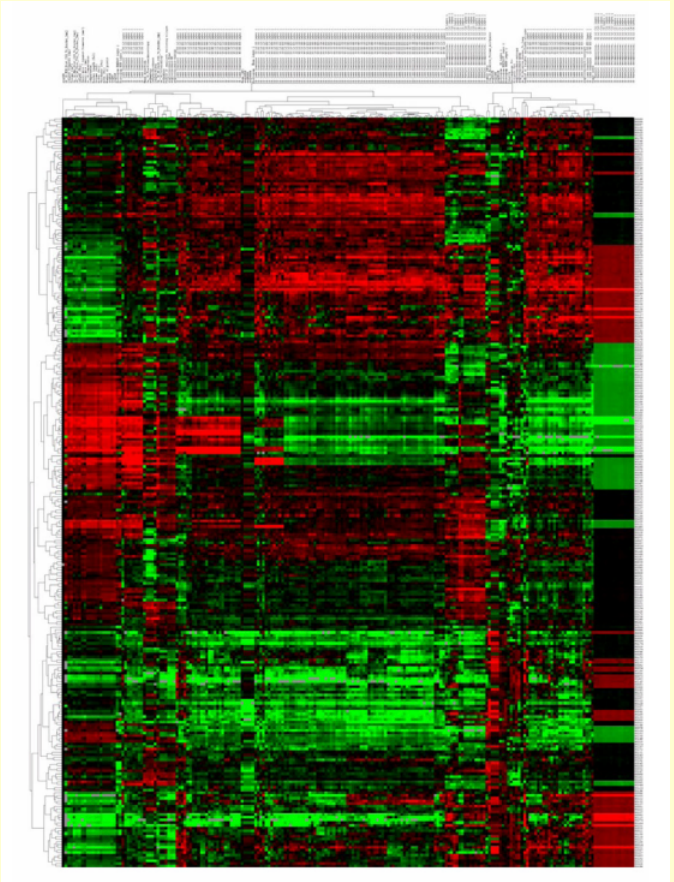
*ICC = Intraclass Correlation Coefficient : 0 = no correlation, 1 = perfect

Leijenaar et al Acta Oncol 2013

Statistical analysis

- Specific issues to be accounted for:
  - Multiple testing ➡ significant findings due to random chance: the False Discovery Rate should be controlled
  - Supervised (building a model to produce an outcome) vs unsupervised approaches (exploring data to identify specific pattern)
  - Sample size issues: learning data cannot be test data, and each has to be big
  - Incorporating non imaging data

Each column : 1 tumour
Each row : 1 feature



Example and preliminary results: in CT (NSCLC and H&N)

- 440 image features related to:
  - Tumour intensity (histogram)
  - Shape
  - Texture
  - Wavelet (=multi-scale features)

+ small data sets (31 pts for test-retest, 21 for multiple delineations)

➡ Characterized the test-retest and inter-operator stability of the features

➡ Selected the 100 most stable features

Aerts et al Nature Communications 2014

Example and preliminary results: in CT (NSCLC and H&N)

- 440 image features related to:
  - Tumour intensity (histogram)
  - Shape
  - Texture
  - Wavelet (=multi-scale features)

+ small data sets (31 pts for test-retest, 21 for multiple delineations)

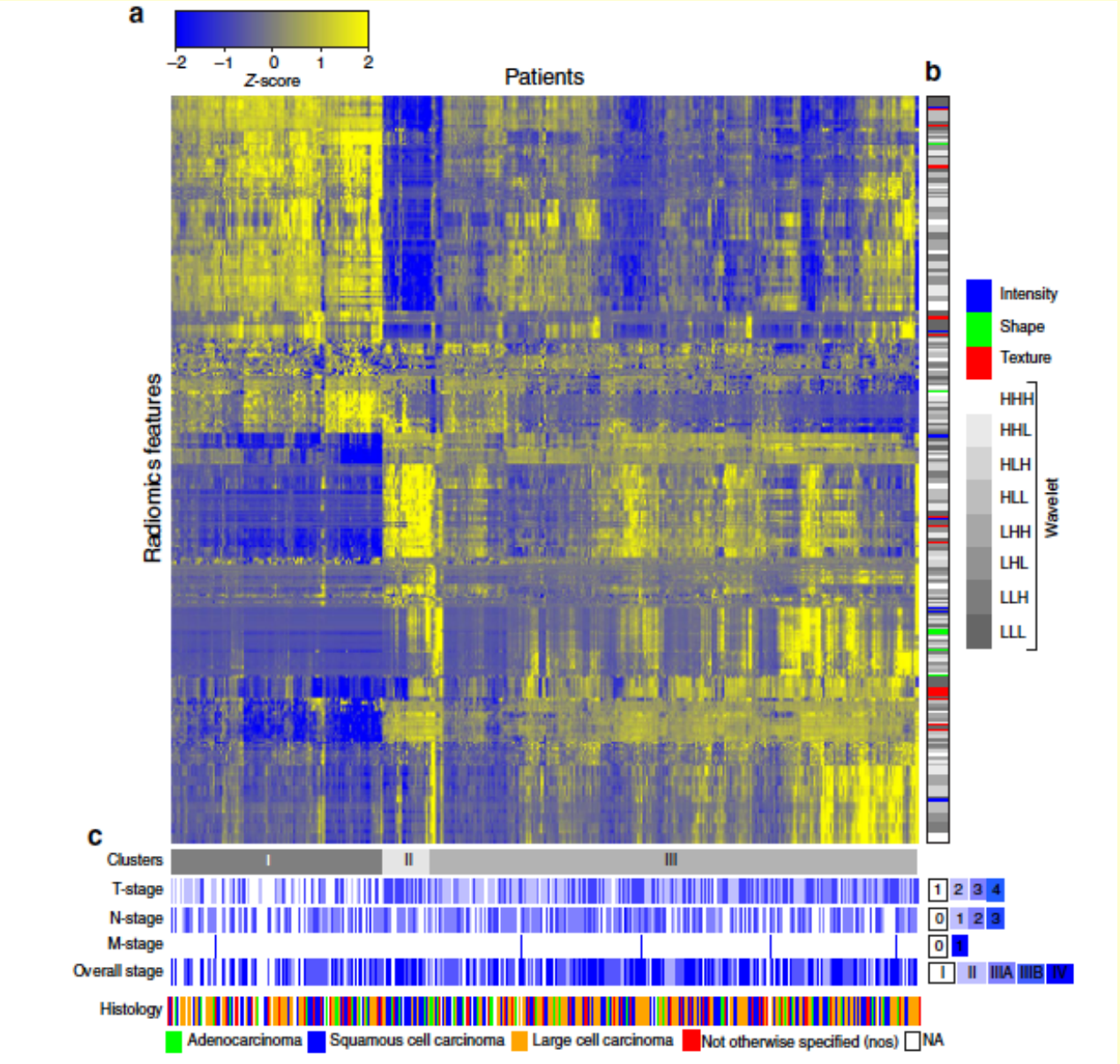
➡ Characterized the test-retest and inter-operator stability of the features
➡ Selected the 100 most stable features

- 422 NSCLC patients :

➡ Identified the 4 best performing features (= radiomic signature) for predicting survival
➡ Determined the weights of a multivariate Cox proportional hazards regression model
➡ Applied that radiomic signature to 3 other cohorts (225 NSCLC, 136 H&N, 95 H&N)

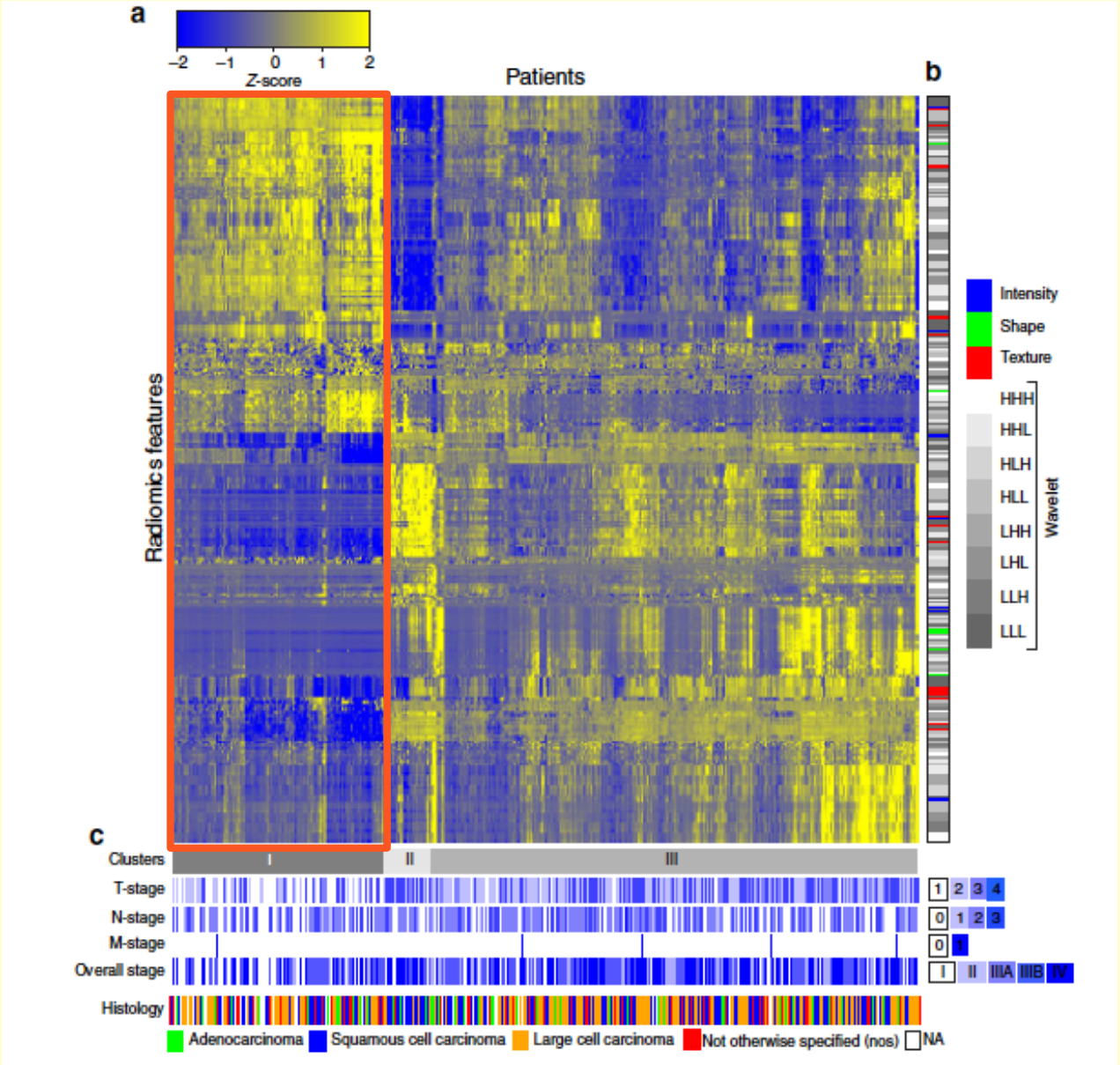
Aerts et al Nature Communications 2014

Results



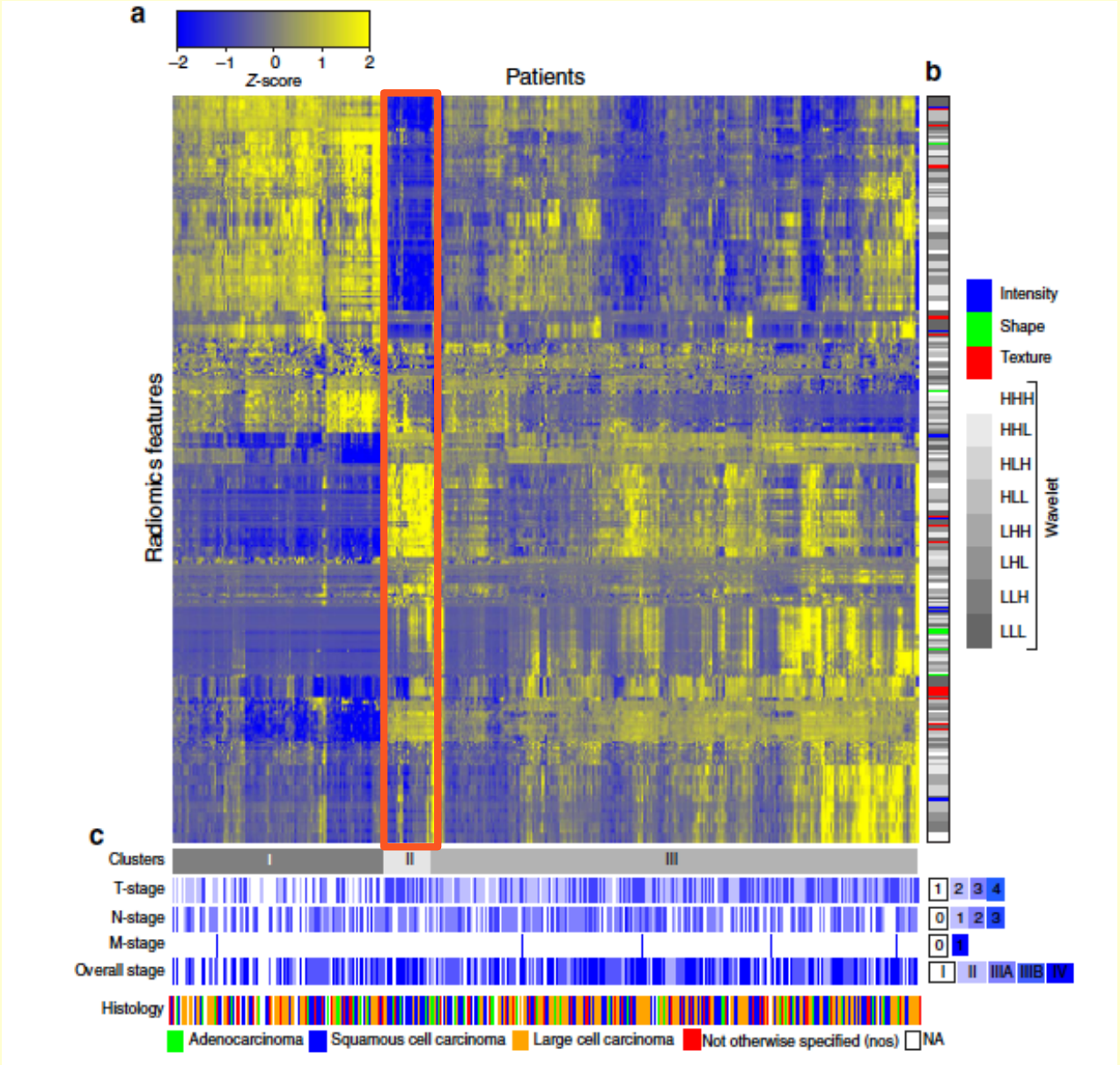
Aerts et al Nature Communications 2014

Results



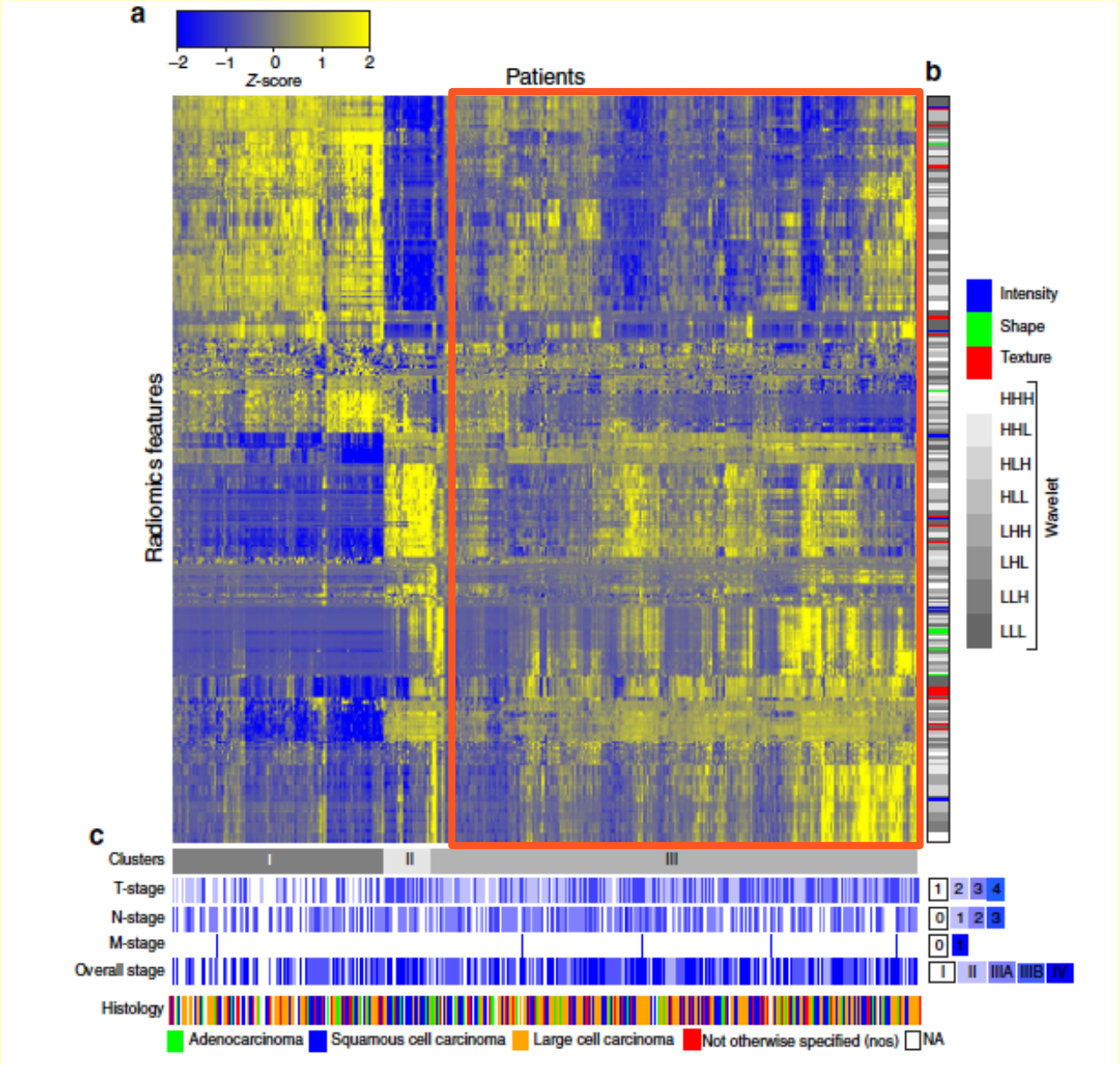
Aerts et al Nature Communications 2014

Results



Aerts et al Nature Communications 2014

Results

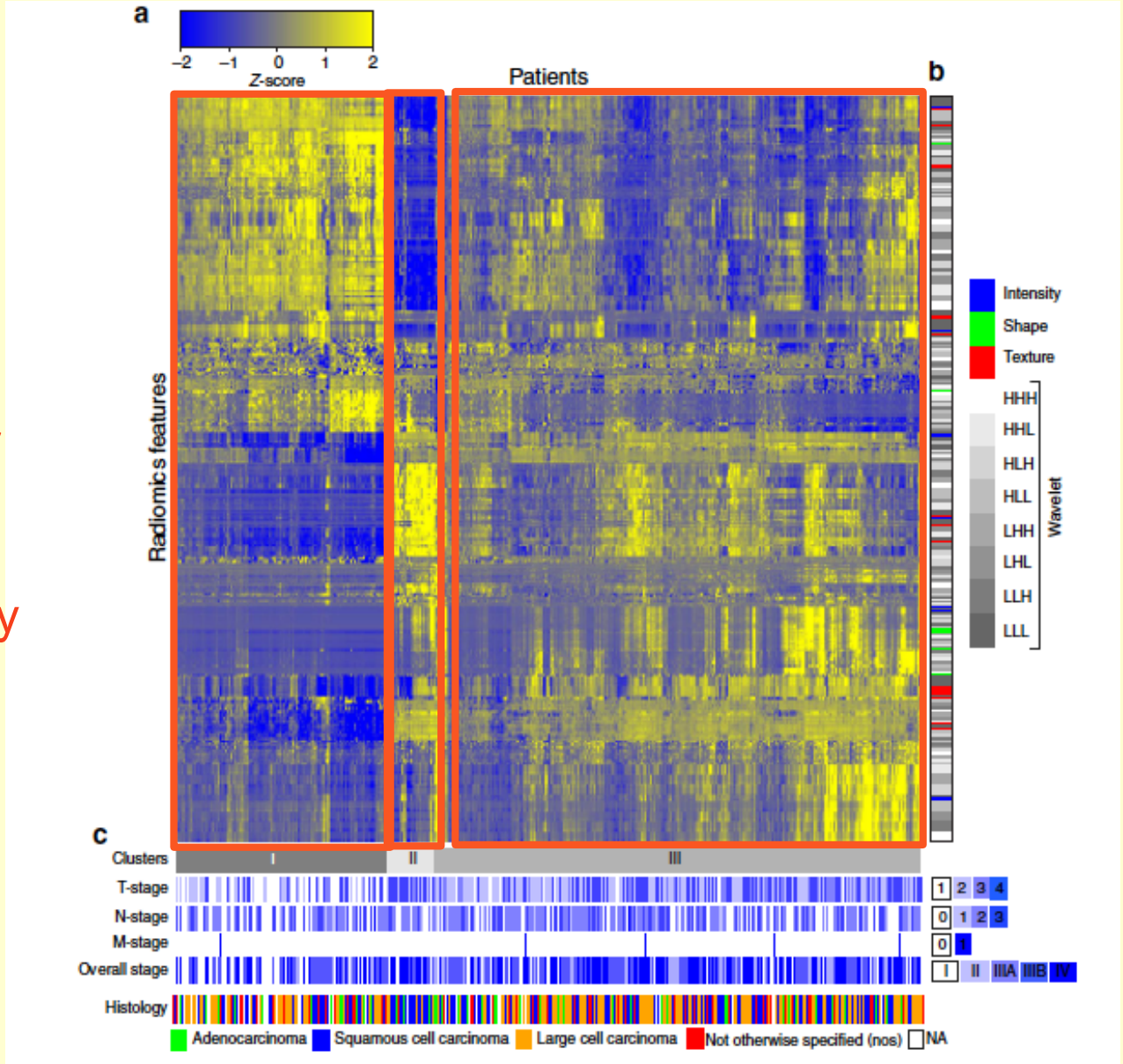


Aerts et al Nature Communications 2014

Results

These 3 clusters presented a significant association with primary tumour stage (T-stage or overall stage)

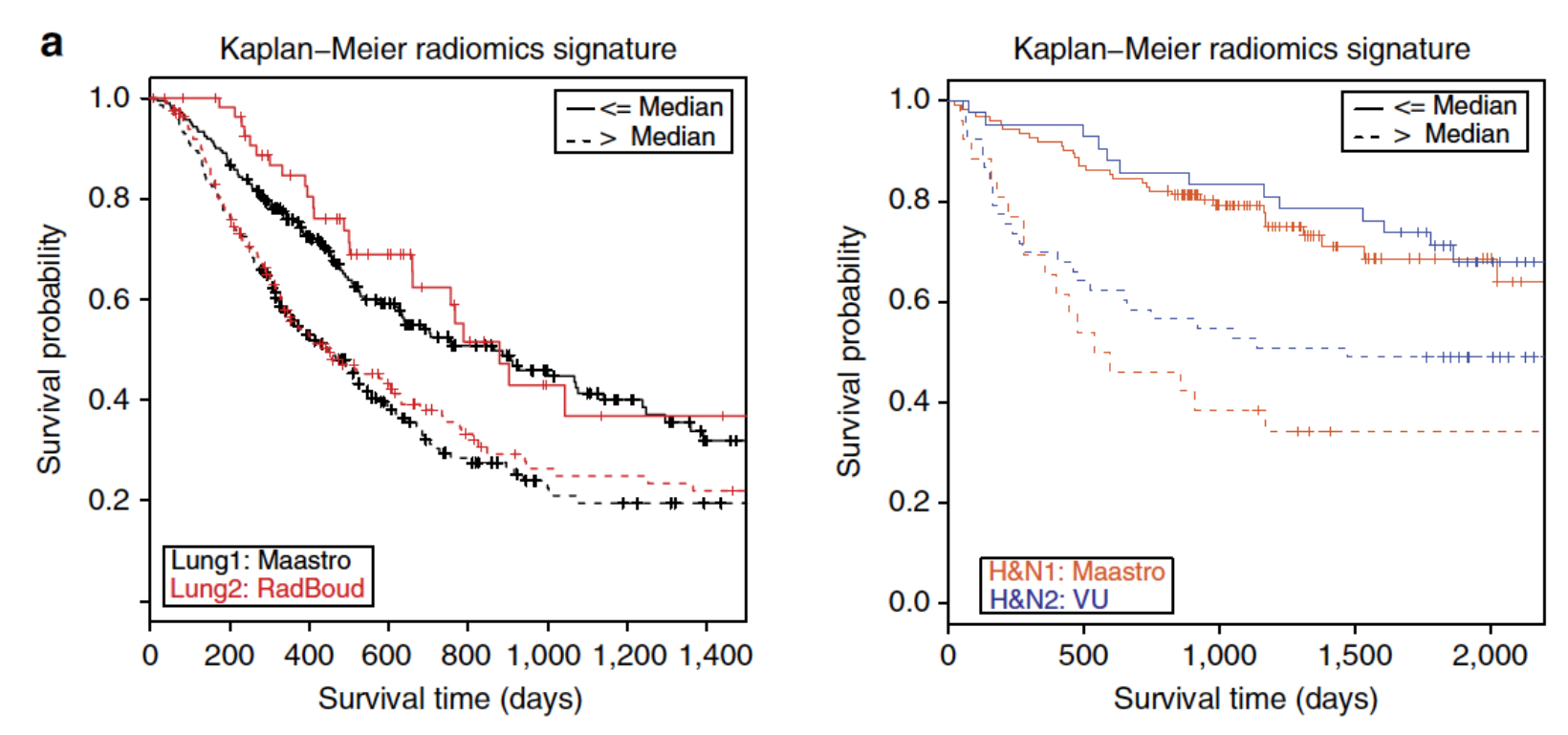
and with histology



Aerts et al Nature Communications 2014

Radiomic signature

- Best discriminating feature in each group
  - Statistics energy (histogram-based)
  - Shape compactness
  - Grey Level Non Uniformity (GLNU)
  - Wavelet Grey Level Non Uniformity

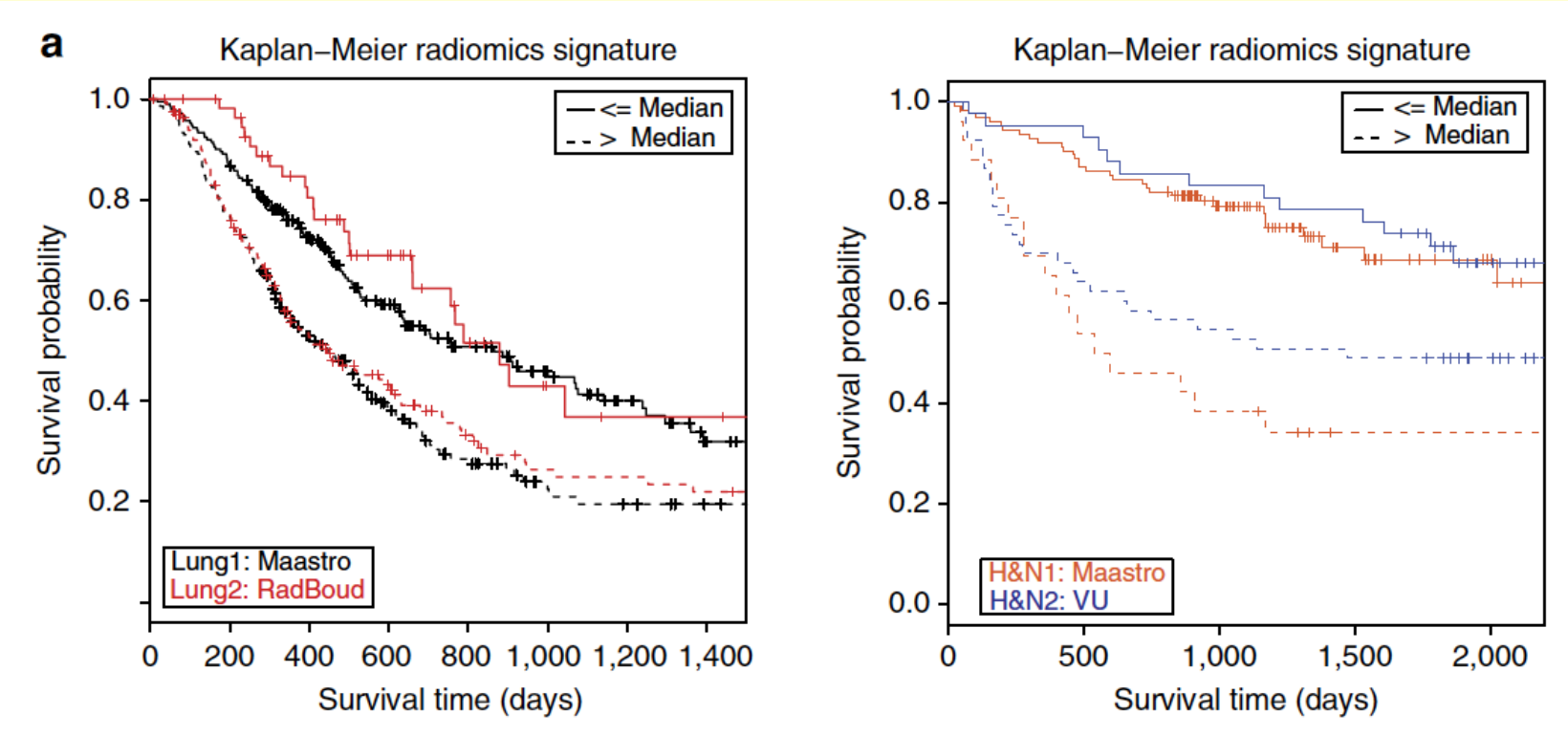


Aerts et al Nature Communications 2014

Radiomic signature

- Best discriminating feature in each group
  - Statistics energy (histogram-based)
  - Shape compactness
  - Grey Level Non Uniformity (GLNU)
  - Wavelet Grey Level Non Uniformity

GLNU highly correlated with the tumor volume.
KM curves for volume only?

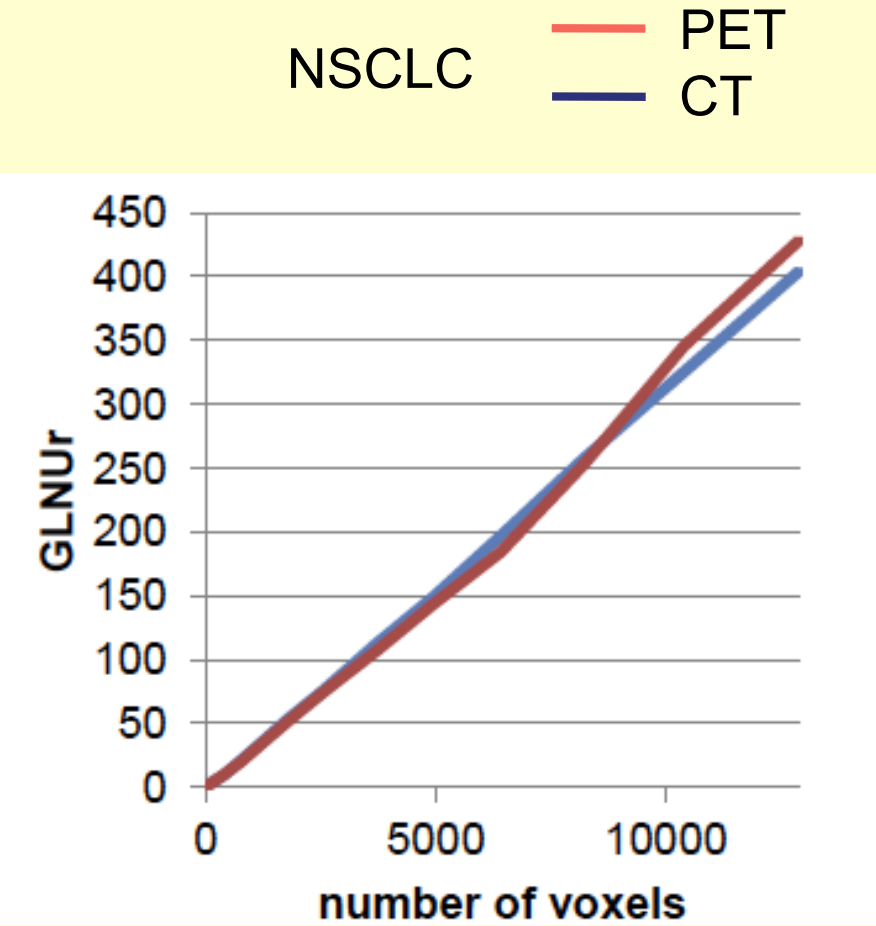


Aerts et al Nature Communications 2014

What is GLNU?

	MV		
	CRC	LUNG	BREAST
Homogeneity	0.62	0.72	0.68
Energy	-0.58	-0.58	-0.46
Correlation	0.67	0.73	0.64
Contrast	-0.56	-0.71	-0.56
Entropy	0.78	0.80	0.64
Dissimilarity	-0.60	-0.73	-0.61
SRE	-0.71	-0.74	-0.77
LRE	0.73	0.57	0.78
LGRE	-0.49	-0.25	-0.43
HGRE	0.32	0.40	-0.04
SRLGE	-0.49	-0.38	-0.43
SRHGE	0.30	0.37	-0.05
LRLGE	-0.48	0.12	-0.41
LRHGE	0.41	0.48	0.04
GLNUr	0.98	0.99	0.97
RLNU	1.00	1.00	1.00
RP	0.99	0.97	0.98
Coarseness	-0.73	-0.78	-0.69
Contrast	-0.56	-0.70	-0.57
Busyness	-0.05	0.05	0.06
SZE	-0.68	-0.80	-0.65
LZE	0.77	0.44	0.78
LGZE	-0.49	-0.57	-0.45
HGZE	0.44	0.42	0.02
SZLGE	-0.47	-0.62	-0.46
SZHGE	0.17	0.15	-0.14
LZLGE	-0.05	0.22	0.21
LZHGE	0.87	0.87	0.78
GLNUz	0.99	0.99	0.99
ZLNU	0.99	0.97	0.90
ZP	0.78	0.54	-0.03

Orlhac et al J Nucl Med 2014



Orlhac et al (submitted)

Comparison of radiomic profile and volume

Prognostic performance as measured using Concordance Indices (0.5 = useless, 1 = perfect prediction)

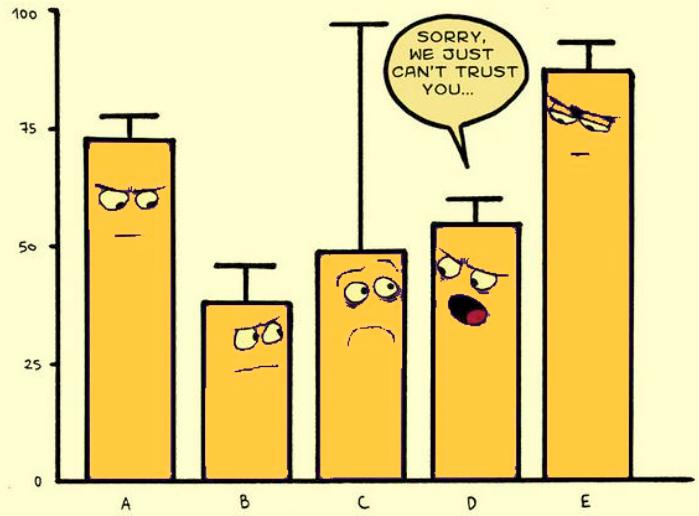
Dataset	TNM	Volume	Radiomics	TNM-Radiomics	Volume-Radiomics
Lung2	0.60	0.63	0.65	0.64	0.65
H&N1	0.69	0.68	0.69	0.70	0.69
H&N2	0.66	0.65	0.69	0.69	0.68

Small but significant added value

Aerts et al Nature Communications 2014

Conclusion

- PET has a role to play in lymphoma management
- Yet, PQRS is needed to make the most of the PET scans
- Large databases that are available could be taken advantage of to explore the Radiomic approach
- Radiomics is actually even more tricky than PET or MR quantification, so studies should be designed and conducted very carefully to avoid mis- or over-interpretation of results



Radiomics: a new era for tumour management ?

Irène Buvat
Unité Imagerie Moléculaire In Vivo (IMIV)
CEA – Service Hospitalier Frédéric Joliot
Orsay, France
irene.buvat@u-psud.fr
Slides available on http://www.guillemet.org/irene/conferences