Machine Learning: Basi e Sue Applicazioni

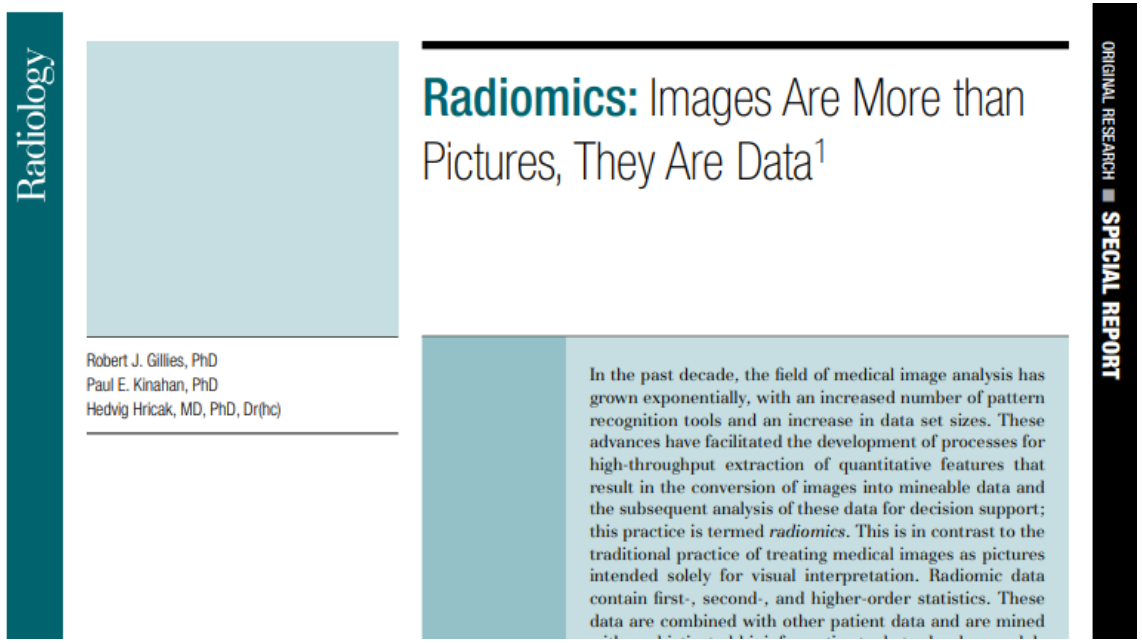
Christian Salvatore
Scuola Universitaria Superiore IUSS Pavia

Radiomic hypotesis

Molecular heterogeneity of cancer lesions is cause of different clinical outcome.

Such heterogeneity can be captured, *in vivo,* on the entire lesion volume, by high-throughput quantitative **radiomics** descriptors from 3D image of cancer lesion.

Different expression level of a signature of radiomic features are able to predict different prognosis or treatment response of patients with similar cancer diagnosis (statistical analysis and predictive models).


Radiology

Robert J. Gillies, PhD
Paul E. Kinahan, PhD
Hedvig Hricak, MD, PhD, Dr(hc)

Radiomics: Images Are More than Pictures, They Are Data[1]

ORIGINAL RESEARCH ■ SPECIAL REPORT

In the past decade, the field of medical image analysis has grown exponentially, with an increased number of pattern recognition tools and an increase in data set sizes. These advances have facilitated the development of processes for high-throughput extraction of quantitative features that result in the conversion of images into mineable data and the subsequent analysis of these data for decision support; this practice is termed *radiomics*. This is in contrast to the traditional practice of treating medical images as pictures intended solely for visual interpretation. Radiomic data contain first-, second-, and higher-order statistics. These data are combined with other patient data and are mined

Radiomics: a new approach for the study of cancer

HHS Public Access
Author manuscript
Eur J Cancer. Author manuscript; available in PMC 2015 August 12.

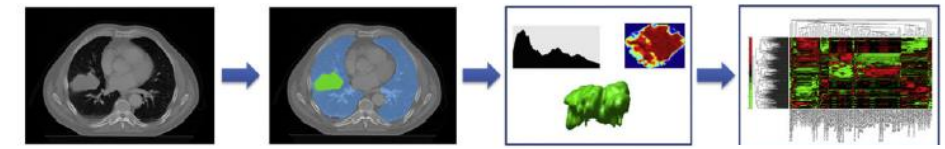
Published in final edited form as:
Eur J Cancer. 2012 March ; 48(4): 441–446. doi:10.1016/j.ejca.2011.11.036.

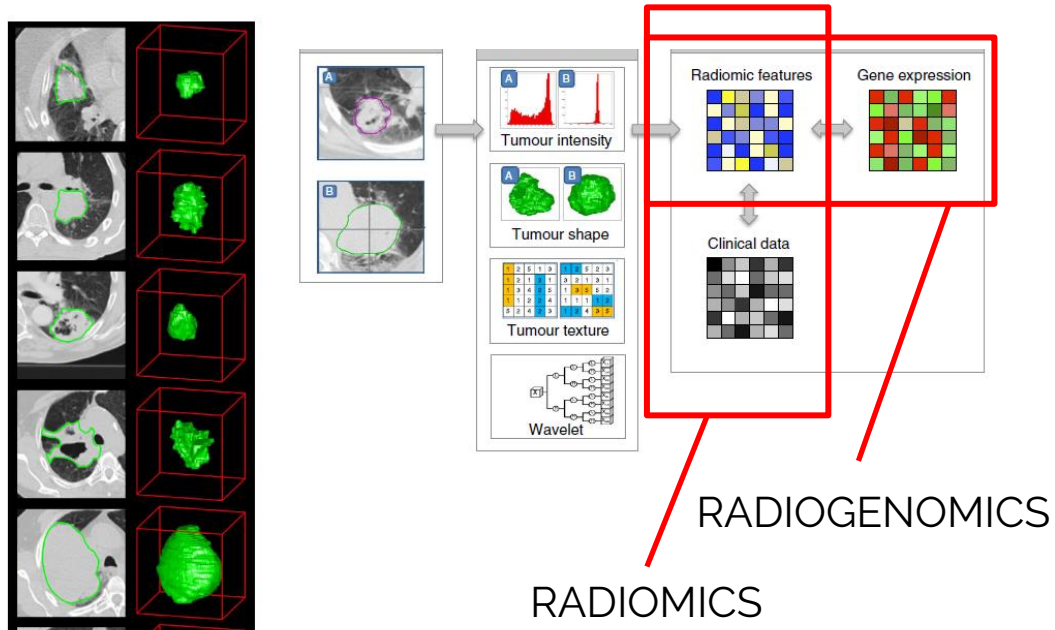
Radiomics: Extracting more information from medical images using advanced feature analysis

Philippe Lambin[a,*,e,f], **Emmanuel Rios-Velazquez**[a,e], **Ralph Leijenaar**[a,e], **Sara Carvalho**[a,e], **Ruud G.P.M. van Stiphout**[a,e], **Patrick Granton**[a,e], **Catharina M.L. Zegers**[a,e], **Robert Gillies**[b,e], **Ronald Boellard**[c,e], **André Dekker**[a,e], and **Hugo J.W.L. Aerts**[a,d,e]

[a]Department of Radiation Oncology (MAASTRO), GROW – School for Oncology and Developmental Biology, Maastricht University Medical Center, Maastricht, The Netherlands [b]H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA [c]U University Medical Center, Department of Nuclear Medicine & PET Research, Amsterdam, The Netherlands [d]Computational Biology and Functional Genomics Laboratory, Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Harvard School of Public Health, USA

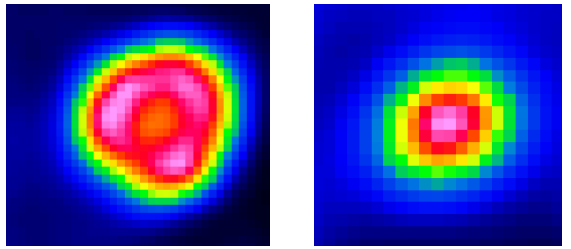
Comprehensive quantification of disease phenotypes by applying a large number of quantitative image features representing lesion heterogeneity and correlating with omics and clinical data

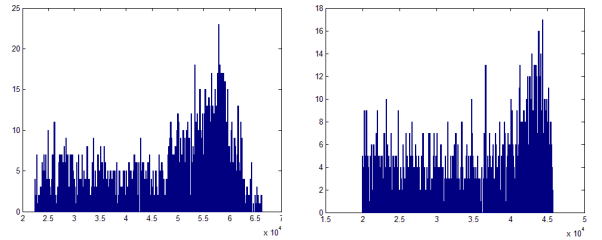


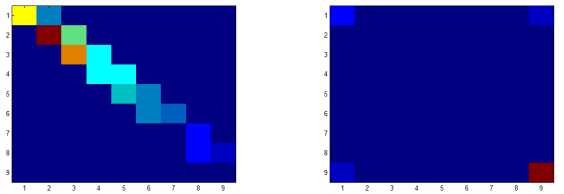


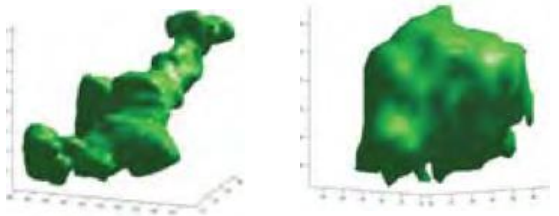
Texture and shape features

Feature	Description		Examples
Texture-First order	Grey level frequency distribution from histogram Analysis	Global	Minimum, mean and maximum intensity Standard deviation Skewness Kurtosis Percentile values Range of intensities
Texture-Second order	From spatial grey level dependence matrices (SGLDM) or co–occurrence matrices *They express how often a pixel of intensity i finds itself within a certain relationship to another pixel of intensity j*	Local	Entropy Energy Contrast Homogeneity Dissimilarity Uniformity Correlation
Texture-Third order	From neighbourhood grey-tone difference matrices (NGTDMs)	Local	Coarseness Contrast Busyness Complexity
	From voxel alignment matrices	Regional	Run-length and emphasis Run-length variability
	From grey level size zone matrices *They reflect regional intensity variations or the distribution of homogeneous regions*	Regional	Zone emphasis Size-zone variability
Shape and Size			Sphericity Compactness Eccentricity Surface Area Spherical Disproportion Surface to Volume ratio Solidity

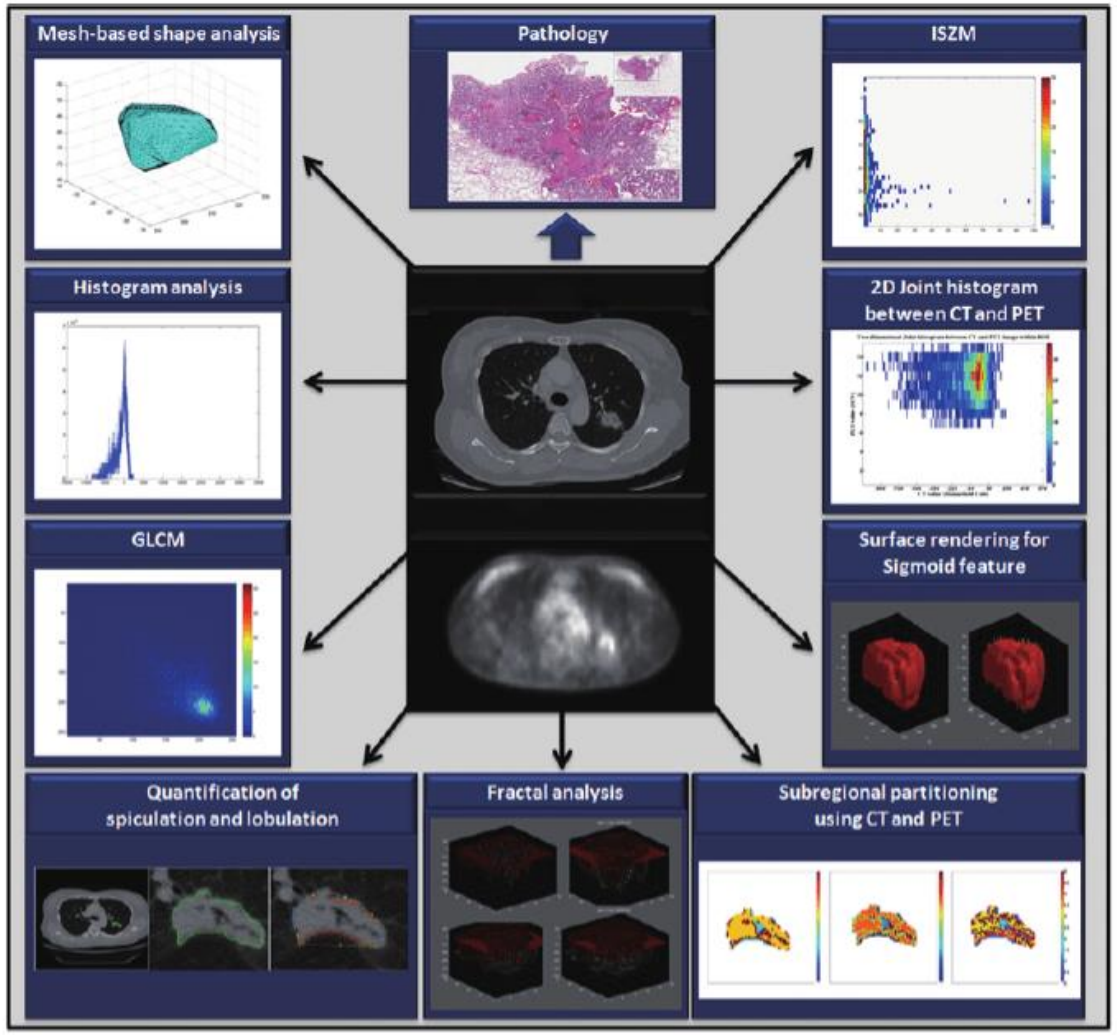








Radiomics

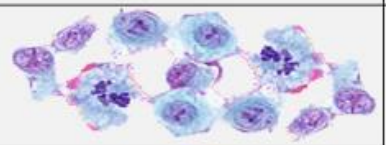
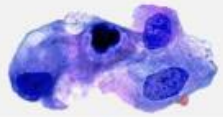


https://www.researchgate.net/figure/Various-radiomic-features-such-as-mesh-based-shape-histogram-gray-level-co-occurrence_fig3_315902486

Radiomics

Morphological features

1. Metabolic Target Volume (MTV)
2. Surface
3. Spherical disproportion
(ratio between measured surface of the lesion and surface of an equivalent-sphere in terms of volume)
4. Sphericity
5. Surface-to-volume ratio

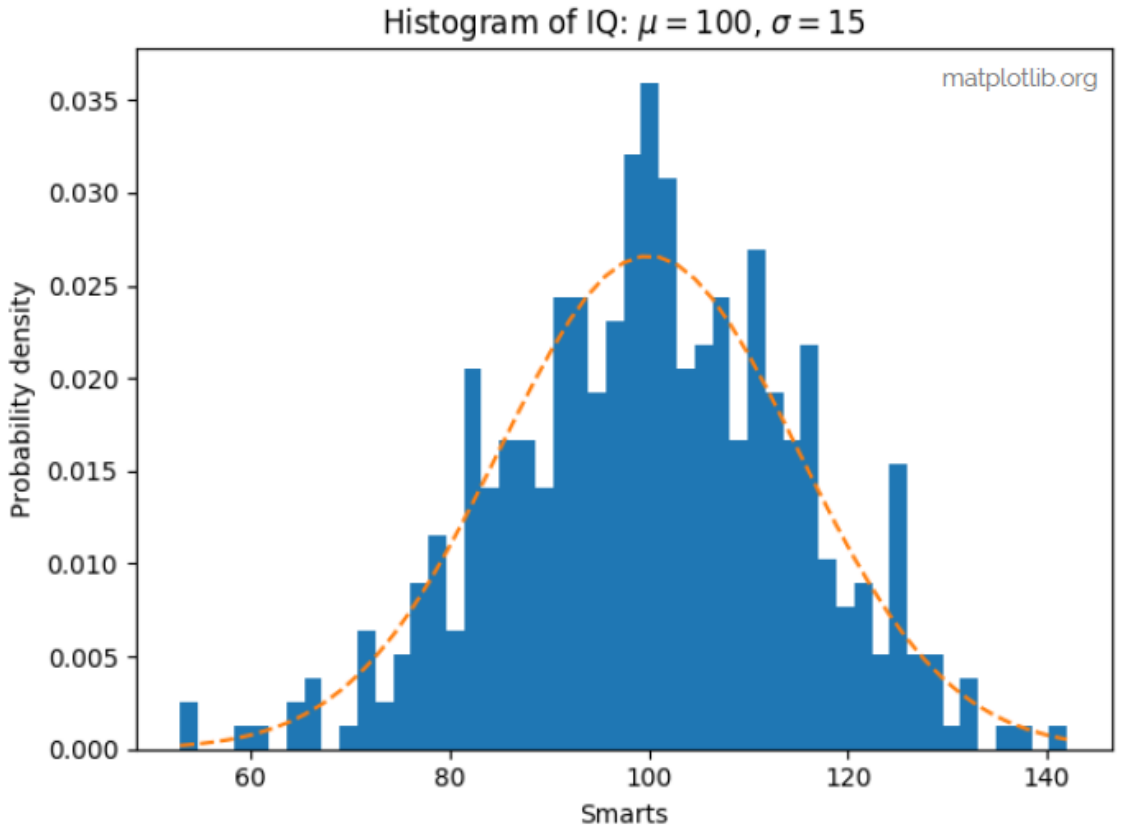
Normal	Cancer	
		Large, variably shaped nuclei
		Many dividing cells; Disorganized arrangement
		Variation in size and shape
		Loss of normal features

http://sphweb.bumc.bu.edu/otlt/MPH-Modules/PH/PH709_Cancer/PH709_Cancer7.html

Radiomics

Histogram-based features

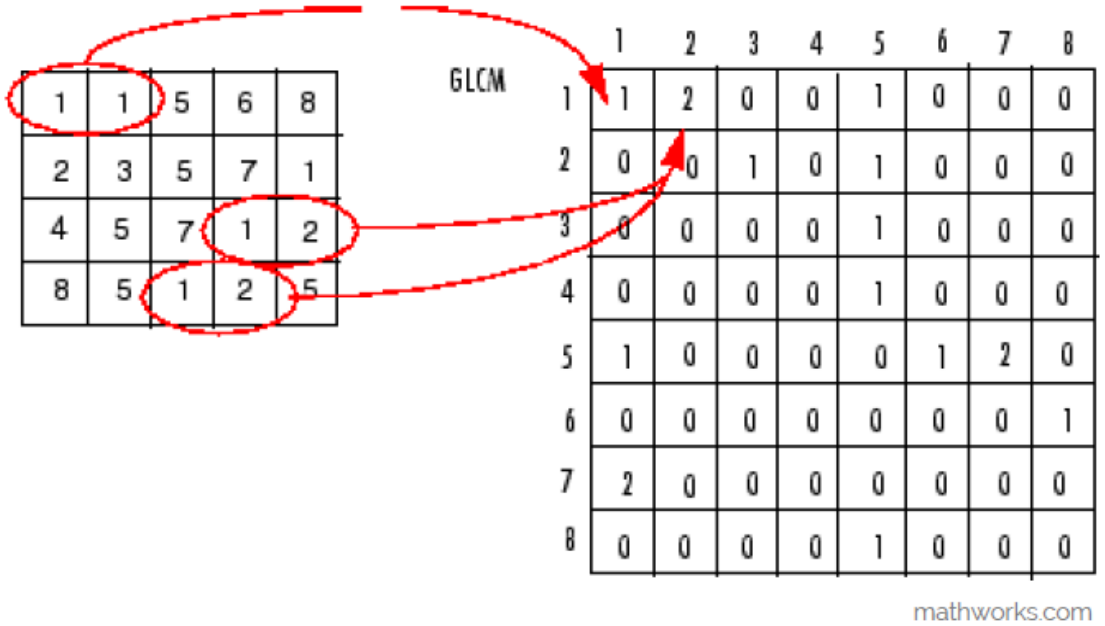
1. Maximum
2. Minimum
3. Mean
4. Median
5. Mean Absolute Deviation (MAD)
6. Root Mean Square (RMS)
7. Energy
8. Entropy
9. Kurtosis
10. Skewness
11. Standard Deviation
12. Uniformity
13. Variance



Texture descriptors
Gray-Level Co-occurrence Matrix (GLCM)*

1. Energy
2. Contrast
3. Entropy
4. Homogeneity
5. Correlation
6. Sum Average
7. Variance
8. Dissimilarity
9. Auto Correlation

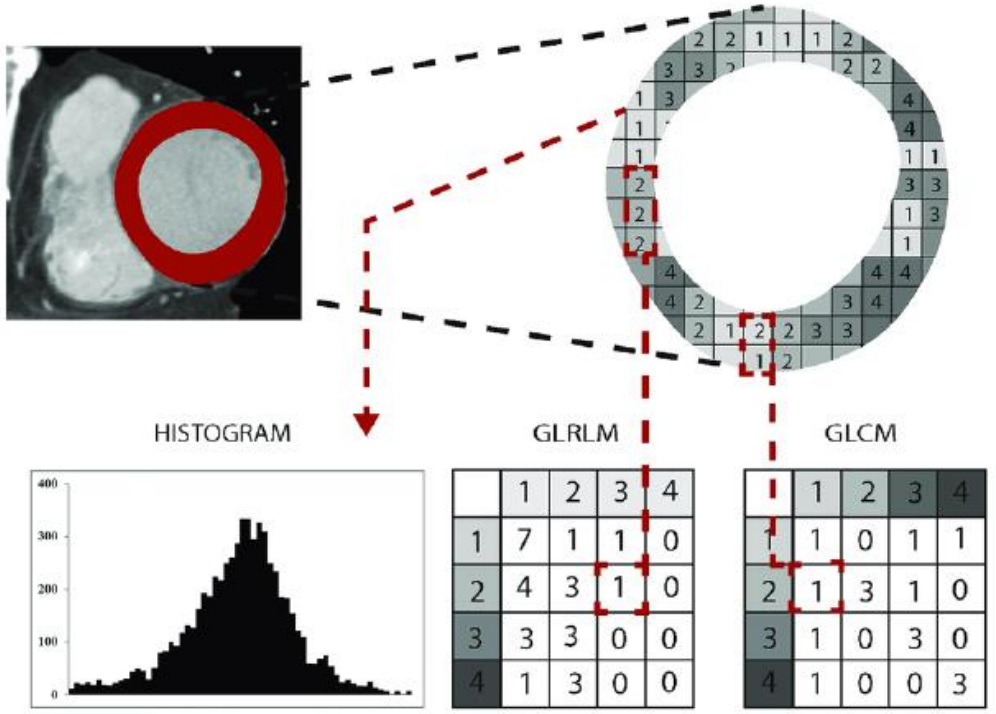
* A Gray Level Co-occurrence Matrix (GLCM) quantifies the number of times the combination of levels X and Y occur in two pixels in the image that are separated by a distance of D pixels along angle A.



Radiomics

Texture descriptors

Gray-Level Run Length Matrix (GLRLM)*



	1	2	3	4
1	7	1	1	0
2	4	3	1	0
3	3	3	0	0
4	1	3	0	0

	1	2	3	4
1	1	0	1	1
2	1	3	1	0
3	1	0	3	0
4	1	0	0	3

https://www.researchgate.net/figure/Principles-of-generating-texture-analysis-features-Principles-of-generating-the_fig2_320821651

*A Gray Level Run Length Matrix (GLRLM) quantifies gray level runs, which are defined as the length in number of pixels, of consecutive pixels that have the same gray level value.

It describes # runs with gray level G and length L that occur in the image along angle A.

Texture descriptors

Gray-Level Size Zone Matrix (GLSZM)

1	2	3	4
1	3	4	4
3	2	2	2
4	1	4	1

Level	*Size zone, s*		
g	1	2	3
1	2	1	0
2	1	0	1
3	0	0	1
4	2	0	1

http://thibault.biz/Research/ThibaultMatrices/GLSZM/GLSZM.html

* A gray level zone is defined as a the number of connected voxels that share the same gray level intensity.

Contrary to GLCM and GLRLM, the GLSZM is rotation independent, with only one matrix calculated for all directions in the ROI

Texture descriptors

Neighbouring Gray Tone Difference Matrix (NGTDM)*

$$\mathbf{I} = \begin{bmatrix} 1 & 2 & 5 & 2 \\ 3 & 5 & 1 & 3 \\ 1 & 3 & 5 & 5 \\ 3 & 1 & 1 & 1 \end{bmatrix}$$

i	n_i	p_i	s_i
1	6	0.375	13.35
2	2	0.125	2.00
3	4	0.25	2.63
4	0	0.00	0.00
5	4	0.25	10.075

https://pyradiomics.readthedocs.io/en/latest/features.html

* A Neighbouring Gray Tone Difference Matrix quantifies the difference between a gray value and the average gray value of its neighbours within distance D

Textures in cancer by PET

Published in final edited form as:
Pattern Recognit. 2009 June 1; 42(6): 1162–1171. doi:10.1016/j.patcog.2008.08.011.

Exploring feature-based approaches in PET images for predicting cancer treatment outcomes

I. El Naqa, Ph.D.[a], P. Grigsby, M.D.[a], A. Apte, M.Sc[a], E. Kidd, M.D.[a], E. Donnelly, M.D.[a], D. Khullar, M.Sc[a], S. Chaudhari, B.Sc[a], D. Yang, Ph.D.[a], M. Schmitt, B.Sc[b], Richard Laforest, Ph.D.[b], W. Thorstad, M.D.[a], and J. O. Deasy, Ph.D.[a]
[1]Department of Radiation Oncology, Washington University School of Medicine St. Louis, MO, USA
[2]Department of Radiology, Washington University School of Medicine, St. Louis, MO, USA

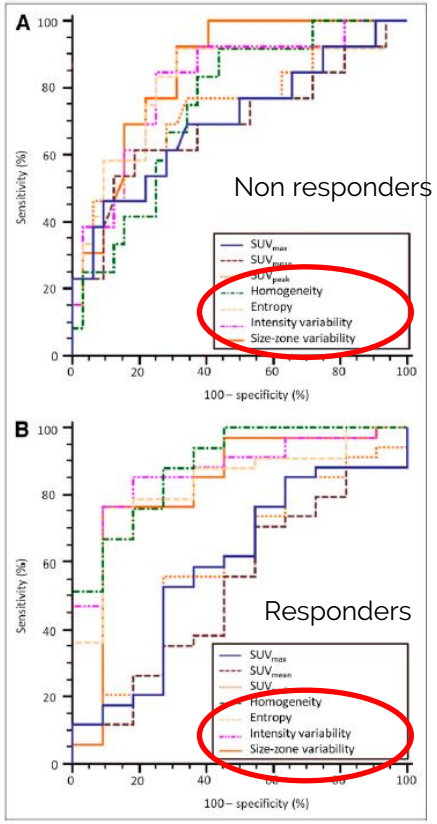


Table 2

Association between different extracted features and overall survival in a cohort of 9 head and neck patients measured by Spearman's rank correlation (rs) and the area under the ROC curve (AUC).

Variable		Spearman (rs)	AUC
Tumor volume		*0.6928*	*0.8750*
SUV Measurements	Maximum	0.3464	0.7000
	Minimum	-0.2642	0.6000
	Mean	0.1732	0.6500
	Standard deviation	0.3464	0.6750
IVH Intensity-volume metrics	I_{10}	0.1732	0.7000
	I_{90}	0.0	0.5000
	I_{10-90}	0.2598	0.6750
	V_{10}	-0.1732	0.5750
	V_{90}	*-0.7794*	*0.9500*
	V_{10-90}	0.0866	0.5000
Texture-based features	Energy	0.0866	0.5000
	Contrast	-0.5196	0.8000
	Local homogeneity	0.5196	0.8250
	Entropy	-0.1732	0.5250
Shape-based features	Eccentricity	0.2598	0.6500
	Euler Number	*0.6166*	*0.8500*
	Solidity	*-0.6088*	*0.8500*
	Extent	*-0.6062*	*0.8500*

J Nucl Med 2011; 52:369–378

Intratumor Heterogeneity Characterized by Textural Features on Baseline ^{18}F-FDG PET Images Predicts Response to Concomitant Radiochemotherapy in Esophageal Cancer

Florent Tixier[1], Catherine Cheze Le Rest[1,2], Mathieu Hatt[1], Nidal Albarghach[1,3], Olivier Pradier[1,3], Jean-Philippe Metges[3,4], Laurent Corcos[4], and Dimitris Visvikis[1]

[1]INSERM, U650, LaTIM, CHU Morvan, Brest, France; [2]Department of Nuclear Medicine, CHU Morvan, Brest, France; [3]Institute of

A study in which we hope not to be cited…

PLOS | ONE

RESEARCH ARTICLE

False Discovery Rates in PET and CT Studies with Texture Features: A Systematic Review

Anastasia Chalkidou*, Michael J. O'Doherty, Paul K. Marsden

Division of Imaging Sciences and Biomedical Engineering, Kings College London 4th Floor, Lambeth Wing, St. Thomas Hospital, SE1 7EH, London, United Kingdom

* anastasia.chalkidou@kcl.ac.uk

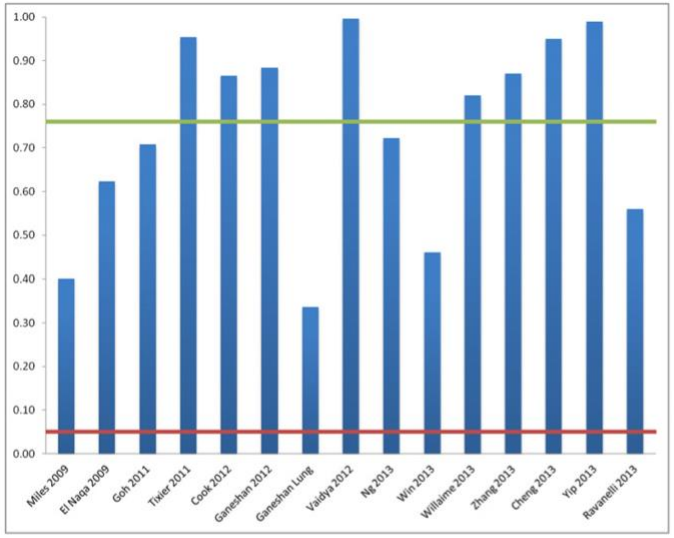


Fig 2. Probability of a false positive result based on number of hypotheses tested per study (blue columns) for all study categories. 5% type-I error probability = red line, average type-I error probability (76%) over all studies = green line (Note—additional inflation of the type-I error probability due to the use of the optimum cut-off approach is not included here).

doi:10.1371/journal.pone.0124165.g002

Key methodological issues

- **Repeatability**, the closeness of the agreement between the results of successive radiomic measurements under the same conditions of measurement

- **Riproducibility**, the closensess of the agreement between the results of radiomic measurement under similar conditions of measurements

- **Significance**, the ability of radiomic in effectively characterizing cancer lesion heterogeneity

Stability

Biological change or radiomics unstability?

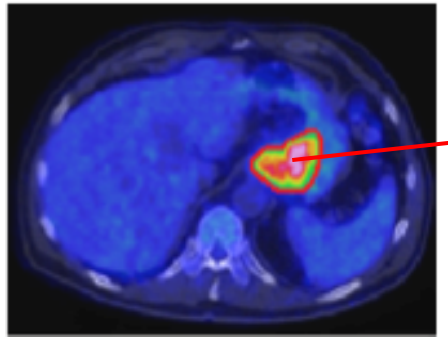
It is necessary that the radiomics features are **repeatable** for the same patient as part of the prognosis and therapeutic monitoring but also **reproducible** when performed across multiple centers and patients.

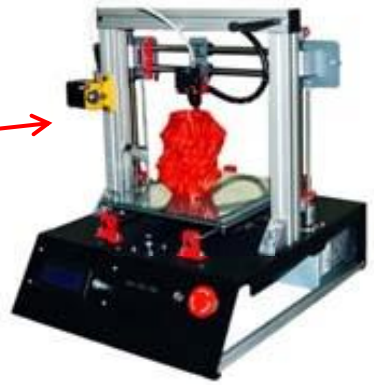
For the SUV and MTV metrics, a cut-off value of ±30% has been accepted for associating the changes to actual metabolic variations (PERCIST).

There is currently no consensus on the tolerated variability of radiomics features for the evaluation of prognosis or response to treatment.

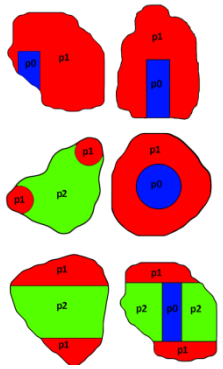
Only radiomic features with high repeatability and reproducibility should be selected as candidate for predictive biomarkers.

Which model to study key radiomics issues?

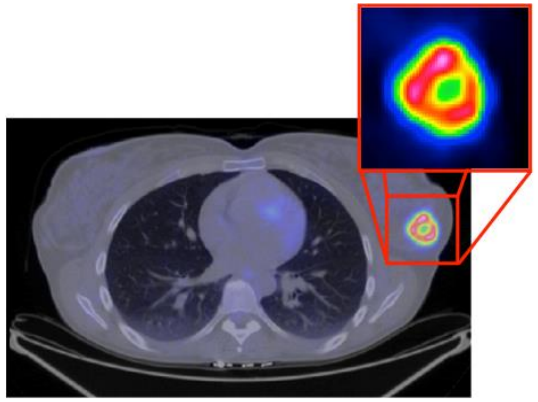




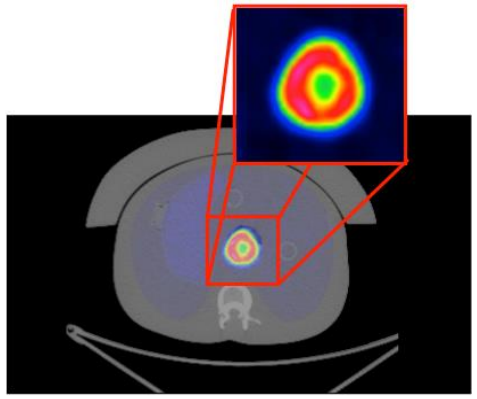
sodium alginate



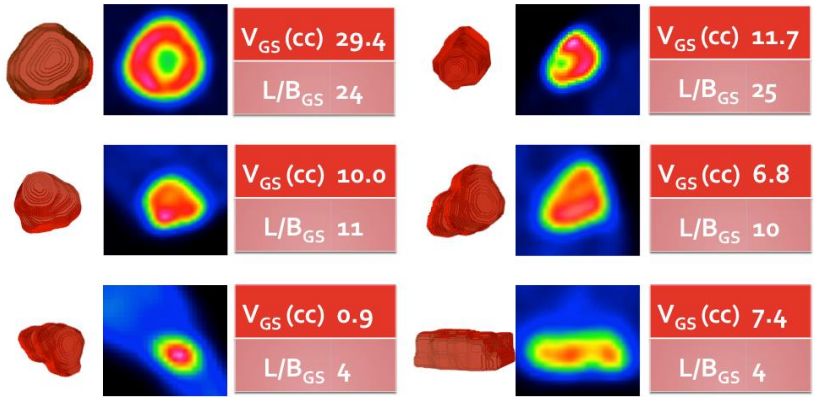
Selected as highlight at IEEE NSS MIC, Strasburgh, November 2016



Real patient



Anthropomorphic phantom with syntehtic lesion



Research Article

A Method for Manufacturing Oncological Phantoms for the Quantification of 18F-FDG PET and DW-MRI Studies

Francesca Gallivanone,[1] Irene Carne,[2] Matteo Interlenghi,[1] Daniela D'Ambrosio,[2] Maurizia Baldi,[3] Daniele Fantinato,[2] and Isabella Castiglioni[1]

[1]*Institute of Molecular Bioimaging and Physiology, National Research Council (IBFM-CNR), Milan, Italy*
[2]*Medical Physics Unit, IRCCS Fondazione S. Maugeri, Pavia, Italy*
[3]*Department of Diagnostic Imaging, IRCCS Fondazione S. Maugeri, Pavia, Italy*

Radiomics repeatability

Hindawi
Contrast Media & Molecular Imaging
Volume 2018, Article ID 5324517, 12 pages
https://doi.org/10.1155/2018/5324517

Research Article

Parameters Influencing PET Imaging Features: A Phantom Study with Irregular and Heterogeneous Synthetic Lesions

Francesca Gallivanone,[1] Matteo Interlenghi,[1] Daniela D'Ambrosio,[2] Giuseppe Trifirò,[3] and Isabella Castiglioni[1]

[1]*Institute of Molecular Bioimaging and Physiology, National Research Council (IBFM-CNR), Milan, Italy*
[2]*Medical Physics Unit, IRCCS Fondazione S. Maugeri, Pavia, Italy*
[3]*Nuclear Medicine Unit, IRCCS Fondazione S. Maugeri, Pavia, Italy*

- **Test-retest** is performed among the distributions of the radiomic values obtained in the subsequent measurements.
- The pairwise Intraclass Correlation coefficient (ICC) is calculated (ICC>0.7 is considered for stability).

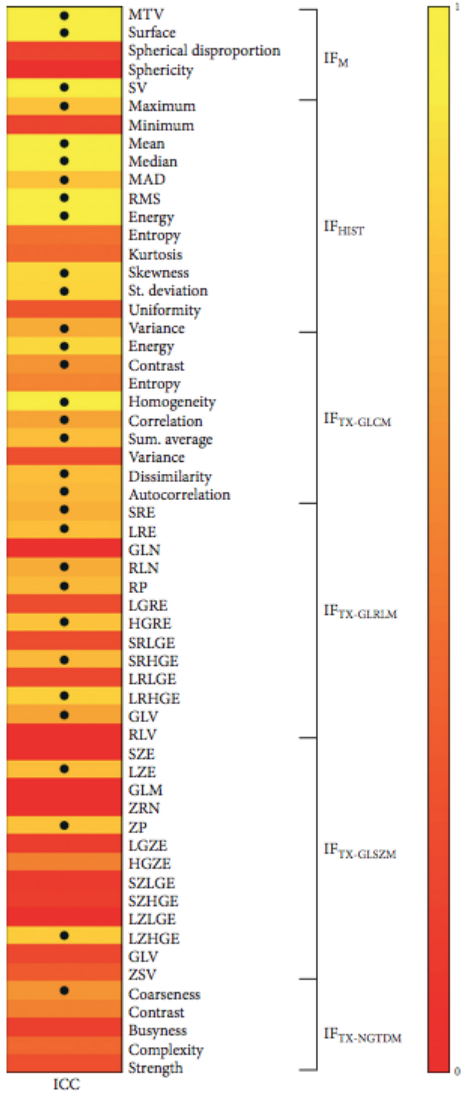


FIGURE 6: Reproducibility of radiomic features on test-retest datasets. ICC results. • indicates ICC ≥ 0.6.

High repeatability

Radiomics reproducibility

- Preparation
- Acquisition
- Reconstruction
- Segmentation
- **Interpolation**
- **Re-segmentation**
- **Discretization**

Preparation and acquisition

- Patient's conditions (e.g. Glycemia)
- Injected dose
- Scan time vs uptake time
- Time per bed position
- Respiratory motion

...

Preparation and acquisition | Uptake time

Lovat et al. 2017 – 54 neurofibromas
a significant **radiomic value change** between two different **uptake times** both for benign and malignant lesions

Preparation and acquisition | Respiratory motion

Vaidya et al. 2012 - 27 lung cancer
Radiomic value change considering or not **respiratory motion** correction by image deconvolution. ***No change in radiotherapy response.***

Yip et al. 2014 - 26 lung cancer/*Oliver et al. 2015* - 23 lung cancer
Radiomic value change considering or not **respiratory motion** correction by gating. ***No results on clinical outcome.***

Grootjans et al. 2016 - 60 lung cancer
Radiomic value change in lower lobes considering or not **respiratory motion** correction by gating. ***No change in prognosis.***

Image reconstruction

- Method (back-projection, iterative –n. it, n. subset…)
- PSF incorporation or not
- TOF incorporation or not
- Matrix size
- Filter
- PVC or not
- Statistical noise

…

Image reconstruction

Galavis et al. 2010 - 20 solid cancer
Radiomic value change with different **reconstruction** settings (method, n iter, matrix size, filter).

Yan et al. 2015 - 20 lung cancer */ Orlhac et al*. 2017 - 54 breast cancer
Radiomic value change with different **reconstruction** settings (method, n iter, matrix size, filter) ±TOF± PSF.
However, matrix size is the more impacting factor.

Shiri et al. 2017 - 25 lung, head, neck, liver cancer
Poor reproducibility of radiomic values for different **reconstruction** settings (method, n iter, n subset, matrix size, filter, PSF, TOF, scan time).

Radiomics reproducibility

Hindawi
Contrast Media & Molecular Imaging
Volume 2018, Article ID 5324517, 12 pages
https://doi.org/10.1155/2018/5324517

Research Article

Parameters Influencing PET Imaging Features: A Phantom Study with Irregular and Heterogeneous Synthetic Lesions

Francesca Gallivanone,[1] Matteo Interlenghi,[1] Daniela D'Ambrosio,[2] Giuseppe Trifirò,[3] and Isabella Castiglioni[1]

[1]*Institute of Molecular Bioimaging and Physiology, National Research Council (IBFM-CNR), Milan, Italy*
[2]*Medical Physics Unit, IRCCS Fondazione S. Maugeri, Pavia, Italy*
[3]*Nuclear Medicine Unit, IRCCS Fondazione S. Maugeri, Pavia, Italy*

Coefficient of Variation (COV) can be calculated (COV<0.10 is considered for stability)
but a statistical test is the best choice

High reproducibility

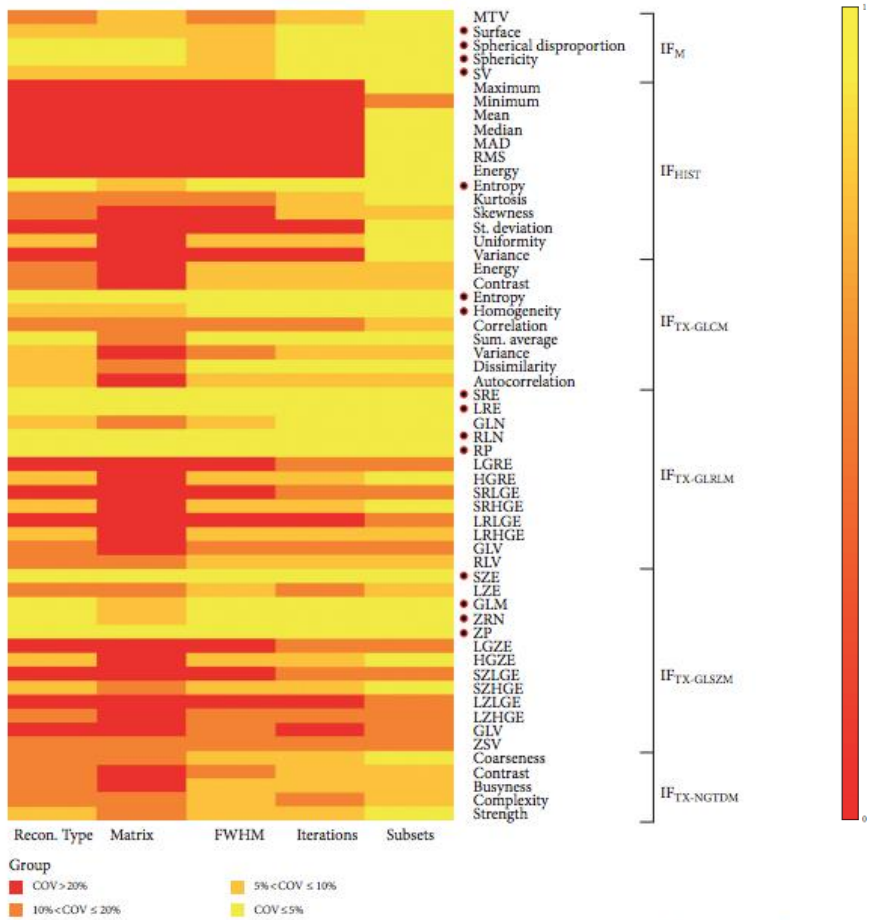
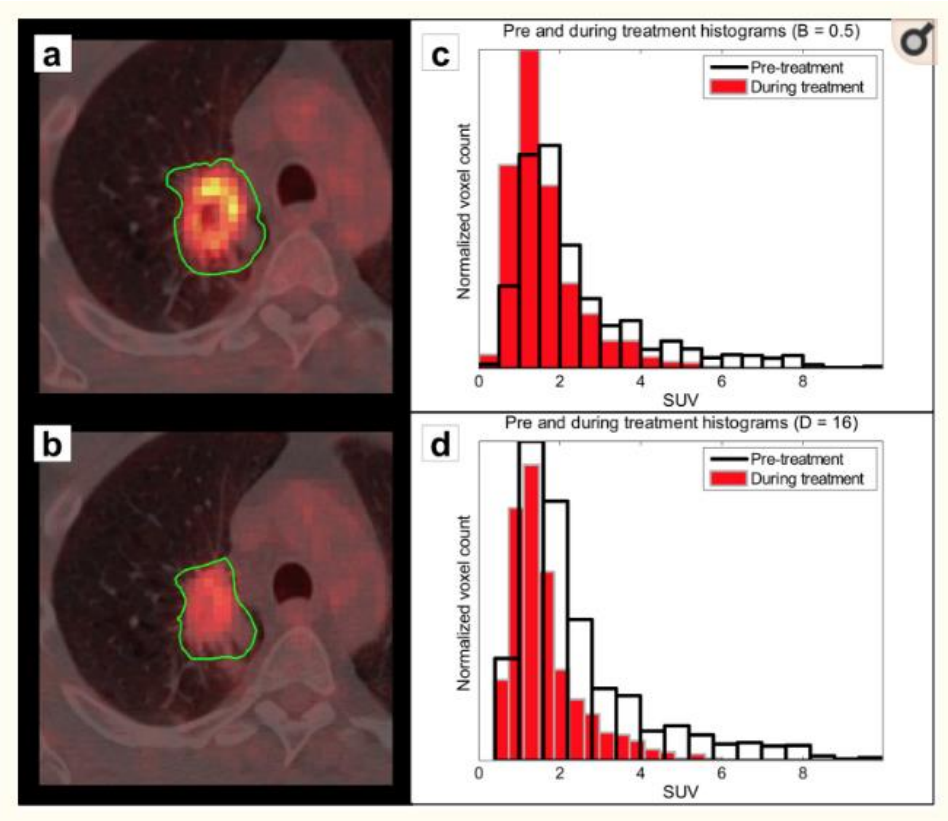


FIGURE 5: Stability of radiomic features on different reconstruction settings. COV results. • indicates COV ≤ 10%

Discretization

Resampling voxels in a limited number of intensity values (bins) (for textural feature calculation).



Fixed bin size

Fixed bin number

Discretization

Leijenaar et al. 2015 - 35 lung cancer
Taxture value is dependent on the method of discretisation
fixed bin size is recommended (constant intensity resolution, more robust, repeatable and less sensitive to segmentation and reconstruction changes)

Lu et al. 2016 - 40 nasopharyngeal carcinoma.
23% of texture features are stable vs **fixed bin size**

Orlhac et al. 2015 - 48 lung cancer & phantom studies /
Desseroit et al. 2017 - 73 lung cancer
fixed bin size is recommended (not requiring MTV of at least 45cc but less intuitive when imaged).

Discretization

Tixier et al.2011 – 41 oesophageal cancer
Textural features are stable and less correlated with MTV for **fixed bin number** (64 bins is recommended since it seems to be sufficient to cover SUV range of lesions with 0.25 increments).

Segmentation

Segmentation of the tumour volume is a crucial step because **all the radiomics features are calculated starting from the segmented volume**.

A variety of methods exists (manual, thresholding, graph-based, region growing, statistical modelling, contour and gradient-based …)

In radiomics **robustness** (e.g. stability vs noise) is more important than accuracy

Segmentation

Hatt et al. 2013- 50 oesophageal cancer
Entropy, homogeneity showed moderate variability for different segmentation.
No change in radiochemotherapy response.

Leijenaar et al. 2013 - 23 lung cancer
Most textural features **are stable** vs4-operator manual contouring .

Orlhac et al. 2014 - 188 colorectal, lung, breast cancer
Entropy and regional textural **are quite stable** for different segmentation methods.

Hatt et al. 2018 ***- 100 lung cancer***
Sphericity, homogeneity and dissimilarity **value changes** depending on the segmentation method
Change in prognosis and prediction of response to treament.

Radiomics significance

Hindawi
Contrast Media & Molecular Imaging
Volume 2018, Article ID 5324517, 12 pages
https://doi.org/10.1155/2018/5324517

Research Article

Parameters Influencing PET Imaging Features: A Phantom Study with Irregular and Heterogeneous Synthetic Lesions

Francesca Gallivanone,[1] Matteo Interlenghi,[1] Daniela D'Ambrosio,[2] Giuseppe Trifirò,[3] and Isabella Castiglioni[1]

[1]*Institute of Molecular Bioimaging and Physiology, National Research Council (IBFM-CNR), Milan, Italy*
[2]*Medical Physics Unit, IRCCS Fondazione S. Maugeri, Pavia, Italy*
[3]*Nuclear Medicine Unit, IRCCS Fondazione S. Maugeri, Pavia, Italy*

Test correlation of radiomic feature with gold-standard heterogeneity H_{GS}

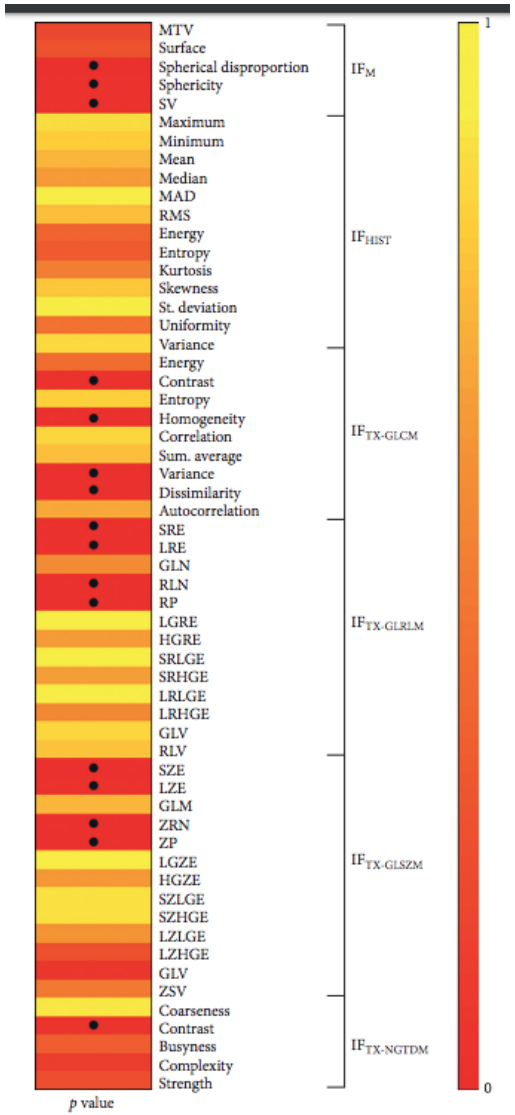


FIGURE 8: Results of correlation analysis between radiomic features and H_{GS} (p value), • indicates p value < 0.05.

High significance

Radiomics significance

Hindawi
Contrast Media & Molecular Imaging
Volume 2018, Article ID 5324517, 12 pages
https://doi.org/10.1155/2018/5324517

Research Article

Parameters Influencing PET Imaging Features: A Phantom Study with Irregular and Heterogeneous Synthetic Lesions

Francesca Gallivanone[1], **Matteo Interlenghi**[1], **Daniela D'Ambrosio**[2], **Giuseppe Trifirò**[3] **and Isabella Castiglioni**[1]

[1]*Institute of Molecular Bioimaging and Physiology, National Research Council (IBFM-CNR), Milan, Italy*
[2]*Medical Physics Unit, IRCCS Fondazione S. Maugeri, Pavia, Italy*
[3]*Nuclear Medicine Unit, IRCCS Fondazione S. Maugeri, Pavia, Italy*

- test significant differences among each radiomic feature from heterogeneous vs. homogeneous uptake (e.g. Mann-Whitney test)
- measure the ability of radiomic features in discriminating heterogeneous from homogeneous lesions

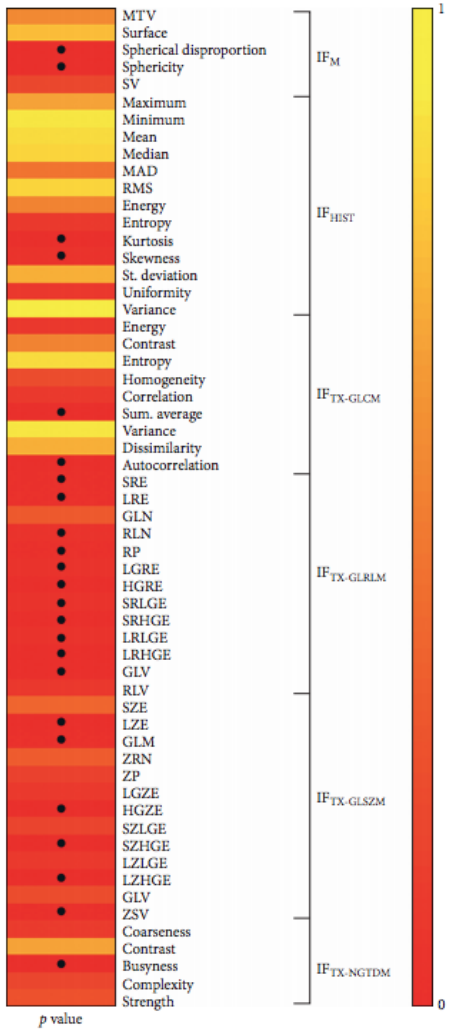


FIGURE 7: Mann–Whitney test results (p value), • indicates p value < 0.05.

High significance

Image Biomarker Standardization Initiative (IBSI)

They are providing:

- image biomarker nomenclature and definitions
- benchmark data sets and values
- reporting guidelines
- consensus-based guidelines for stable radiomic biomarkers

Zwanenburg A, Leger S, Valli`eres M, L¨ock S. Image biomarker standardisation initiative. arXiv preprint arXiv:1612.07003.

Lambin P. Radiomics Digital Phantom, CancerData (2016), DOI:10.17195/candat.2016.08.1

Image Biomarker Standardization Initiative (IBSI)

Some recommandations are delivered:
e.g. re-segmentation and discretization

Imaging intensity units[1]	Re-segmentation range	FBN[2]	FBS[3]
definite	$[a, b]$	✔	✔
	$[a, \infty)$	✔	✔
	none	✔	×
arbitrary	none	✔	×

Table 2.1 — Recommendations for the possible combinations of different imaging intensity definitions, re-segmentation ranges and discretisation algorithms. Checkmarks (✔) represent recommended combinations of re-segmentation range and discretisation algorithm, whereas crossmarks (×) represent non-recommended combinations.
[1] PET and CT are examples of imaging modalities with *definite* intensity units (e.g. SUV and HU, respectively), and raw MRI data of arbitrary intensity units.
[2] *Fixed bin number* (FBN) discretisation uses the actual range of intensities in the analysed ROI (re-segmented or not), and not the re-segmentation range itself (when defined).
[3] *Fixed bin size* (FBS) discretisation uses the lower bound of the re-segmentation range as the minimum set value. When the re-segmentation range is not or cannot be defined (e.g. arbitrary intensity units), the use of the FBS algorithm is not recommended.

They are working on tolerated variability of radiomics features but results are not currently disclosed ....

A possible solution?

JNM The Journal of NUCLEAR MEDICINE

A post-reconstruction harmonization method for multicenter radiomic studies in PET

Fanny Orlhac, Sarah Boughdad, Cathy Philippe, Hugo Stalla-Bourdillon, Christophe Nioche, Laurence Champion, Michaël Soussan, Frédérique Frouin, Vincent Frouin and Irène Buvat

J Nucl Med.
Published online: January 4, 2018.
Doi: 10.2967/jnumed.117.199935

Harmonization method

To pool SUV and textural features measured from different PET protocols, we tested a harmonization method previously described for genomic studies to correct the so-called batch effect. The ComBat harmonization model developed by Johnson et al (*25*) assumes that the value of each feature y measured in VOI j and scanner i can be written as:

$$y_{ij} = \alpha + X_{ij}\beta + \gamma_i + \delta_i\varepsilon_{ij} \qquad \text{Equation 1}$$

where α is the average value for feature y, X is a design matrix for the covariates of interest, β is the vector of regression coefficients corresponding to each covariate, γ_i is the additive effect of scanner i on features supposed to follow a normal distribution, δ_i describes the multiplicative scanner effect supposed to follow an inverse gamma distribution, and ε_{ij} is an error term (normally distributed with a zero mean), as explained in Fortin et al (*30*). ComBat harmonization consists in estimating γ_i and δ_i using Empirical Bayes estimates (noted γ_i^* and δ_i^*) as described in (*25*). The normalized value of feature y for VOI j and scanner i is then obtained as:

$$y_{ij}^{ComBat} = \frac{y_{ij} - \hat{\alpha} - X_{ij}\hat{\beta} - \gamma_i^*}{\delta_i^*} + \hat{\alpha} + X_{ij}\hat{\beta} \qquad \text{Equation 2}$$

where $\hat{\alpha}$ and $\hat{\beta}$ are estimators of parameters α and β respectively. The ComBat harmonization determines a transformation for each feature separately based on the batch (here Department) effect observed on feature values. In the first part of this study, we used ComBat without accounting for any biological covariate (ie X=0), and, in the second part, we used the TN status as the covariate of interest.

For each tissue separately (tumor and liver tissues), we applied ComBat harmonization on all features using the R function called "combat" available at https://github.com/Jfortin1/ComBatHarmonization/.

Results

“Centre effect” on 9 radiomic features from breast cancer patients (63 A vs 74 B)

					After ComBat			
	TN(A) vs TN(B)	non-TN(A) vs non-TN(B)	TN(A+B) vs non-TN(A+B)	TN(B) vs non-TN(A)	TN(A) vs TN(B)	non-TN(A) vs non-TN(B)	TN(A+B) vs non-TN(A+B)	TN(B) vs non-TN(A)
Homogeneity	0.4232	**0.0074**	**0.0014**	0.4635	0.5986	0.8737	**0.0015**	**0.0093**
Entropy	0.5196	0.3906	**0.0031**	0.0875	0.7405	0.9139	**0.0027**	**0.0254**
SRE	0.2995	**0.00044**	**0.0063**	0.9481	0.1294	0.8338	**0.0062**	**0.0061**
LRE	0.2814	**0.0004**	**0.0072**	0.9352	**0.0055**	0.3871	**0.0162**	**0.0004**
LGZE	**0.0405**	**0.0244**	**5.69e-05**	0.3786	0.1102	0.3059	**0.0002**	**0.0003**
HGZE	**0.0494**	**0.0282**	**3.20e-05**	0.2886	0.2814	0.3337	**2.27e-05**	**0.0058**
SUVmax	0.0544	**0.0278**	**7.54e-05**	0.4058	0.5717	0.7943	**4.47e-05**	**0.0072**
SUVmean	**0.0448**	**0.0359**	**3.20e-05**	0.2394	0.4463	0.7747	**3.05e-05**	**0.0052**
SUVpeak	**0.0267**	**0.0306**	**9.75e-05**	0.4736	0.3581	0.7894	**4.99e-05**	**0.0061**

Table 3: P-values of Wilcoxon’s test for all features between TN and non-TN lesions from Departments A and B, before and after ComBat harmonization. Bold values are less than 0.05.

A recommendation

Test radiomic results on many different and independent image data sets!

Radiomics: a new approach for the study of cancer

HHS Public Access
Author manuscript
Eur J Cancer. Author manuscript; available in PMC 2015 August 12.

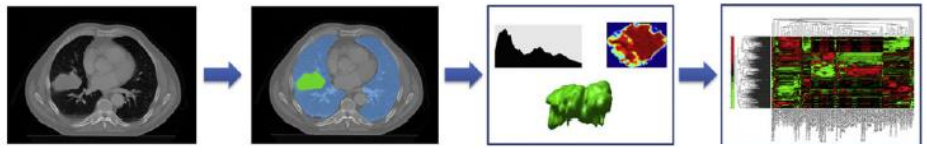
Published in final edited form as:
Eur J Cancer. 2012 March ; 48(4): 441–446. doi:10.1016/j.ejca.2011.11.036.

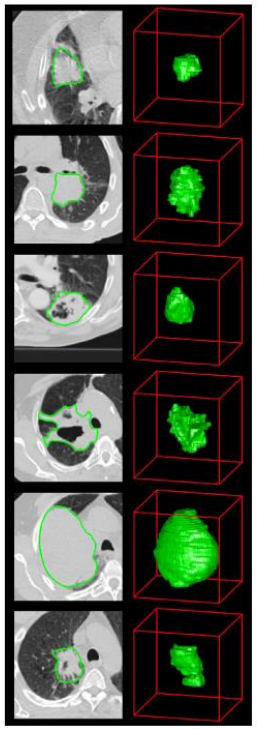
Radiomics: Extracting more information from medical images using advanced feature analysis

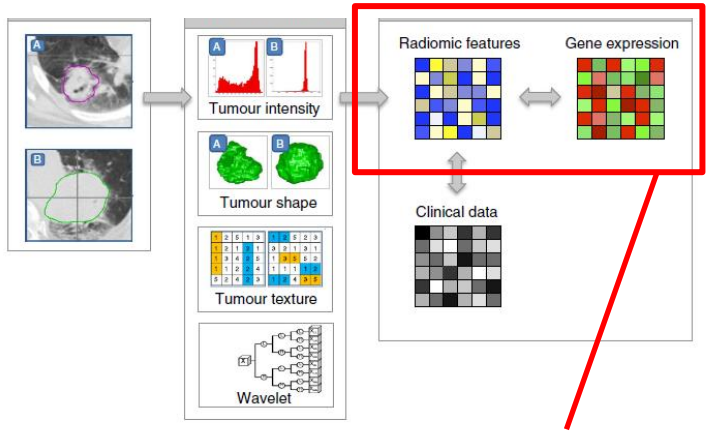
Philippe Lambin[a,*,e,f], **Emmanuel Rios-Velazquez**[a,e], **Ralph Leijenaar**[a,e], **Sara Carvalho**[a,e], **Ruud G.P.M. van Stiphout**[a,e], **Patrick Granton**[a,e], **Catharina M.L. Zegers**[a,e], **Robert Gillies**[b,e], **Ronald Boellard**[c,e], **André Dekker**[a,e], and **Hugo J.W.L. Aerts**[a,d,e]

[a]Department of Radiation Oncology (MAASTRO), GROW – School for Oncology and Developmental Biology, Maastricht University Medical Center, Maastricht, The Netherlands [b]H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA [c]U University Medical Center, Department of Nuclear Medicine & PET Research, Amsterdam, The Netherlands [d]Computational Biology and Functional Genomics Laboratory, Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Harvard School of Public Health, USA

Comprehensive quantification of disease phenotypes by applying a large number of quantitative image features representing lesion heterogeneity and correlating with omics and clinical data

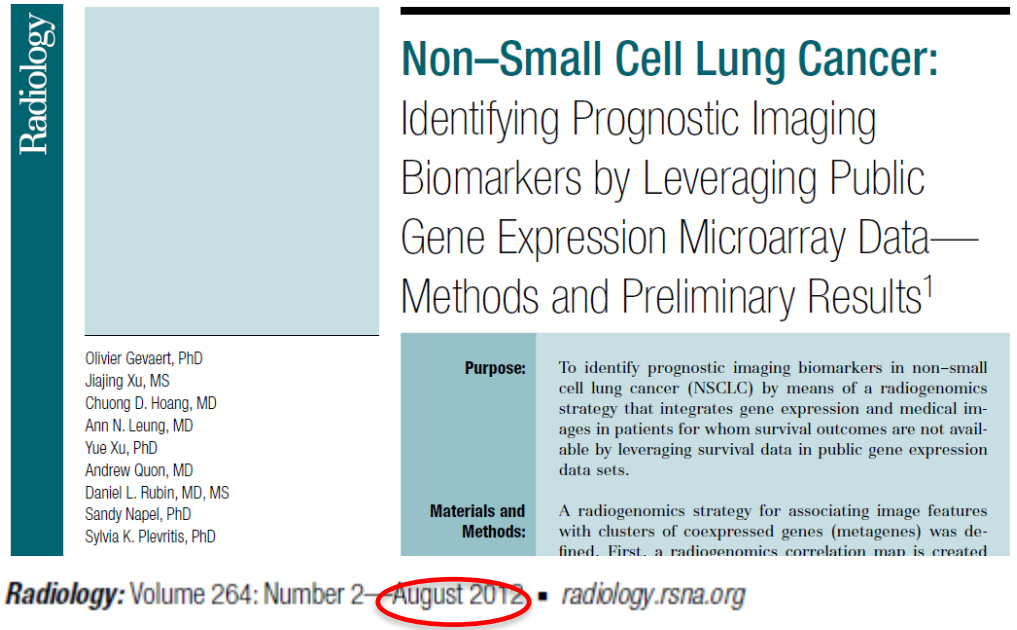






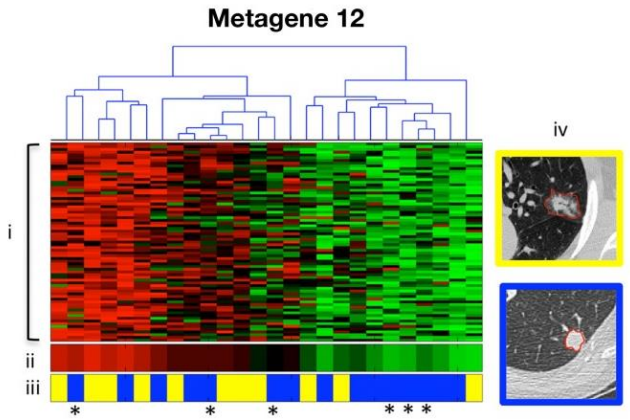
RADIOGENOMICS

CT radiogenomics for cancer

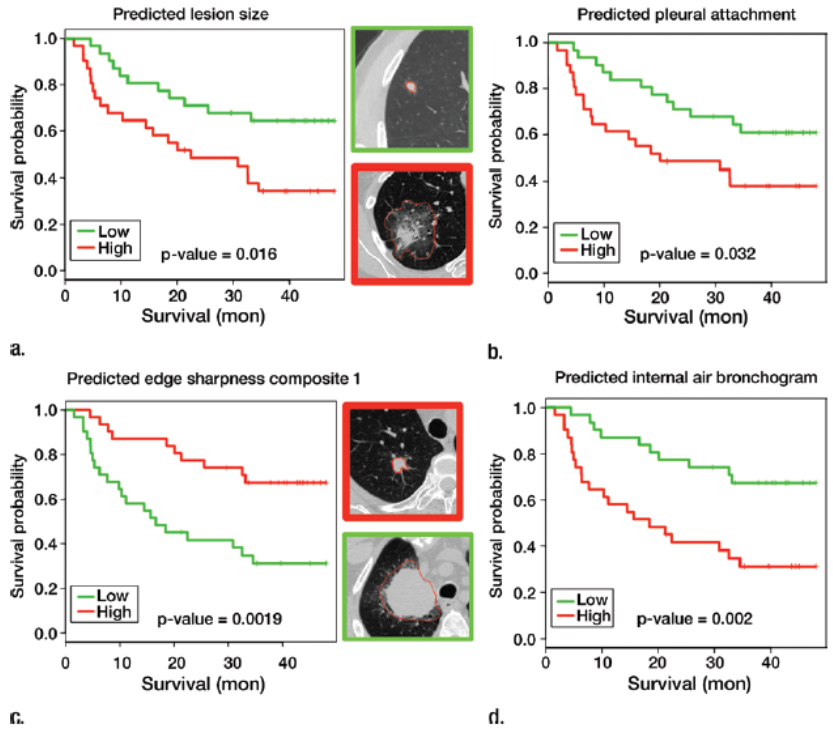


Non-Small Cell lung cancer

153 CT features



4 CT image features



An approach to radiogenomics: system medicine

Published online 23 December 2015

Nucleic Acids Research, 2016, Vol. 44, No. 8 e71
doi: 10.1093/nar/gkv1507

TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data

Antonio Colaprico[1,2,†], Tiago C. Silva[3,4,†], Catharina Olsen[1,2], Luciano Garofano[5,6], Claudia Cava[7], Davide Garolini[8], Thais S. Sabedot[3,4], Tathiane M. Malta[3,4], Stefano M. Pagnotta[5,9], Isabella Castiglioni[7], Michele Ceccarelli[10], Gianluca Bontempi[1,2,*] and Houtan Noushmehr[3,4,*]

[1]Interuniversity Institute of Bioinformatics in Brussels (IB)[2], Brussels, Belgium, [2]Machine Learning Group (MLG), Department d'Informatique, Université libre de Bruxelles (ULB), Brussels, Belgium, [3]Department of Genetics Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, São Paulo, Brazil, [4]Center for Integrative Systems Biology - CISBi, NAP/USP, Ribeirão Preto, São Paulo, Brazil, [5]Department of Science and Technology, University of Sannio, Benevento, Italy, [6]Unlimited Software srl, Naples, Italy, [7]Institute of Molecular Bioimaging and Physiology of the National Research Council (IBFM-CNR), Milan, Italy, [8]Physics for Complex Systems, Department of Physics, University of Turin, Italy, [9]Bioinformatics Laboratory, BIOGEM, Ariano Irpino, Avellino, Italy and [10]Qatar Computing Research Institute (QCRI), HBKU, Doha, Qatar

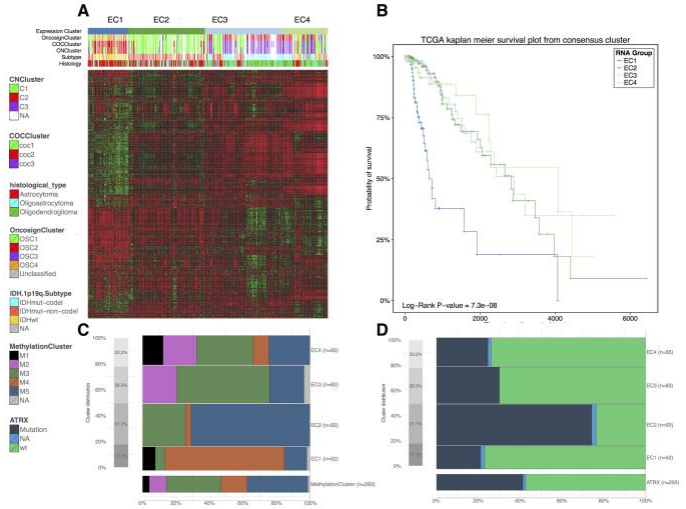


Figure 4. Case study n.2 Integrative (or Downstream) analysis of gene expression and clinical data from LGG disease with unsupervised clustering and crossing expression clusters with clinical and molecular information. (**A**) Heatmap of 1187 more variables genes clustered with tree $k = 4$ in EC1, EC2, EC3, EC4. (**B**) Kaplan Meier survivals plot for EC clusters. (**C** and **D**) Distribution of the DNA Methylation clusters and ATRX mutation within the EC clusters.

Theranostics 2015, Vol. 5, Issue 10 1122

IVYSPRING INTERNATIONAL PUBLISHER

Theranostics

2015; 5(10): 1122-1143. doi: 10.7150/thno.11543

Review

MicroRNAs: New Biomarkers for Diagnosis, Prognosis, Therapy Prediction and Therapeutic Tools for Breast Cancer

Gloria Bertoli, Claudia Cava, and Isabella Castiglioni✉

Institute of Molecular Bioimaging and Physiology (IBFM), National Research Council (CNR), Milan, Italy.

✉ Corresponding author: Institute of Molecular Bioimaging and Physiology of the National Research Council, IBFM-CNR, Via F.Cervi 93- 20090 Segrate (Mi), Italy. Email: isabella.castiglioni@ibfm.cnr.it

Received: 2015.01.09; Accepted: 2015.06.17; Published: 2015.07.13

International Journal of *Molecular Sciences*

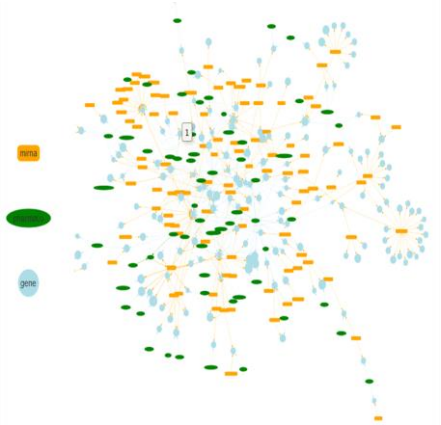
MDPI

Article

SpidermiR: An R/Bioconductor Package for Integrative Analysis with miRNA Data

Claudia Cava [1,*], Antonio Colaprico [2,3], Gloria Bertoli [1], Alex Graudenzi [1], Tiago C. Silva [4], Catharina Olsen [2,3], Houtan Noushmehr [4,5], Gianluca Bontempi [2,3], Giancarlo Mauri [6,7] and Isabella Castiglioni [1,*]

1 Institute of Molecular Bioimaging and Physiology National Research Council (IBFM-CNR), Segrate (Mi) 20090, Italy; gloria.bertoli@ibfm.cnr.it (G.B.); alex.graudenzi@unimib.it (A.G.)
2 Interuniversity Institute of Bioinformatics in Brussels (IB)2, Brussels 1050, Belgium; antonio.colaprico@ulb.ac.be (A.C.); colsen@ulb.ac.be (C.O.); gbonte@ulb.ac.be (G.B.)
3 Machine Learning Group (MLG), Department d'Informatique, Universite libre de Bruxelles (ULB), Brussels 1050, Belgium
4 Department of Genetics Ribeirao Preto Medical School, University of Sao Paulo, Ribeirao Preto, Sao Paulo 14049-900, Brazil; tiagochst@gmail.com (T.C.S.); houtana@gmail.com (H.N.)
5 Department of Neurosurgery, Henry Ford Hospital, Detroit, MI 48202, USA
6 Department of Informatics, Systems and Communication, University of Milan-Bicocca, Milan 20125, Italy; mauri@disco.unimib.it
7 SYSBIO Centre of Systems Biology (SYSBIO), Milan 20126, Italy
* Correspondence: claudia.cava@ibfm.cnr.it (C.C.); isabella.castiglioni@ibfm.cnr.it (I.C.); Tel.: +39-02-21717552 (C.C. & I.C.)



Breast-cancer system biology

The Author(s) *BMC Bioinformatics* 2016, **17**(Suppl 12):348
DOI 10.1186/s12859-016-1196-1

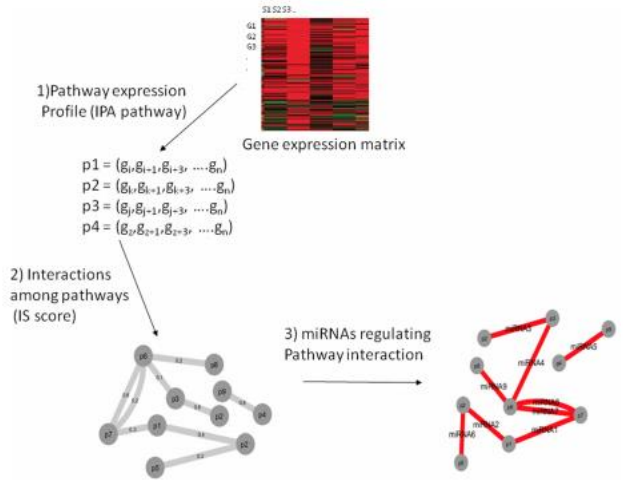
BMC Bioinformatics

RESEARCH Open Access

How interacting pathways are regulated by miRNAs in breast cancer subtypes

Claudia Cava[1], Antonio Colaprico[2,3], Gloria Bertoli[1], Gianluca Bontempi[2,3], Giancarlo Mauri[4] and Isabella Castiglioni[1*]

From Twelfth Annual Meeting of the Italian Society of Bioinformatics (BITS)
Milan, Italy. 3-5 June 2015



[Frontiers In Bioscience, Landmark, 22, 1697-1712, June 1, 2017]

Pathway-based classification of breast cancer subtypes

Alex Graudenzi[1,2], Claudia Cava[1], Gloria Bertoli[1], Bastian Fromm[3], Kjersti Flatmark[3,4,5], Giancarlo Mauri[2,6], Isabella Castiglioni[1]

[1]Institute of Molecular Bioimaging and Physiology of the Italian National Research Council (IBFM-CNR), Milan, Italy, [2]Department of Informatics, Systems and Communication, University of Milan-Bicocca, Milan, Italy, [3]Department of Tumor Biology, Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway, [4]Department of Gastroenterological Surgery, Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway, [5]Institute of Clinical Medicine, University of Oslo, Oslo, Norway, [6]SYSBIO Centre of Systems Biology (SYSBIO), 20126 Milan, Italy

Table 1. Unique pathways enriched of differentially expressed genes for each breast cancer subtype: 17 pathways for luminal A, 5 for luminal B, 19 for basal and 16 for HER2

LUMINAL A	LUMINAL B	BASAL	HER2
REACTOME degradation of the extracellular matrix	KEGG arrhythmogenic right ventricular cardiomyopathy arvc	REACTOME mrna splicing	REACTOME activation of the mrna upon binding of the cap binding complex and eifs and subsequent binding to 43s
BIOCARTA intrinsic pathway	REACTOME crmps in sema3a signaling	REACTOME activation of the pre replicative complex	REACTOME unfolded protein response
REACTOME abc family proteins mediated transport	BIOCARTA cellcycle pathway	KEGG spliceosome	REACTOME developmental biology
REACTOME ethanol oxidation	KEGG ribosome	KEGG apoptosis	KEGG ether lipid metabolism
BIOCARTA ami pathway	REACTOME cell junction organization	REACTOME activation of atr in response to replication stress	REACTOME axon guidance
REACTOME metabolism of carbohydrates		KEGG dna replication	REACTOME nucleotide like purinergic receptors
REACTOME glycerophospholipid biosynthesis		REACTOME interferon alpha beta signaling	REACTOME cgmp effects
REACTOME platelet activation signaling and aggregation		BIOCARTA ranms pathway	REACTOME fgfr ligand binding and activation
KEGG chemokine signaling pathway		KEGG type i diabetes mellitus	REACTOME phospholipase c mediated cascade
BIOCARTA eryth pathway		BIOCARTA g2 pathway	REACTOME glycolysis
BIOCARTA longevity pathway		KEGG bladder cancer	KEGG histidine metabolism
REACTOME triglyceride biosynthesis		REACTOME s phase	KEGG natural killer cell mediated cytotoxicity
REACTOME transmembrane transport of small molecules		REACTOME g1 s transition	REACTOME asparagine n linked glycosylation
BIOCARTA cftr pathway		REACTOME amino acid synthesis and interconversion transamination	REACTOME p2y receptors
REACTOME o linked glycosylation of mucins		REACTOME metabolism of amino acids and derivatives	REACTOME keratan sulfate keratin metabolism
REACTOME transport of glucose and other sugars bile salts and organic acids metal ions and amine compounds		KEGG arginine and proline metabolism	KEGG melanoma
REACTOME factors involved in megakaryocyte development and platelet production		KEGG glycolysis gluconeogenesis	
		BIOCARTA mcm pathway	
		REACTOME extension of telomeres	

Conclusions

Radiomic features have been shown to be sensitive to many factors, i.e. preparation & acquisition, reconstruction, segmentation and new ones, more specific of radiomic (e.g discretization).

Factors not only influence the values of radiomics but their extent is highly variable with different results.

These instability generate fluctuations that should not be misinterpreted as being of biological meaning.

Conclusions

Some solutions are coming and collecting from research groups involved in radiomic harmonization initiatives (e.g. IBSI)

Until clear recommendations on how to harmonize data are defined, you should select only highly repeatable and reproducible radiomic features from your clinical imaging studies and validate in independent studies to select candidates radiomic biomarkers for prognosis and prediction.

However, it is currently not possible to formally exclude any radiomics feature from future investigations solely based upon their low repeatability and riproducibility.

Conclusions

Advanced image processing such as radiomics combined with machine learning can develop models based on imaging signatures for predicting phenotype subtype prognosis and response to therapy

They are opening new role to in vivo medical imaging in predictive personalized medicine

Some radiomic methodological issues (e.g. lesion segmentation, feature harmonization and stability) need robust solutions and validations prior to be traslated in clinicial studies

Radiomic predicting models can be improved by
liquid epigenomics for integrated phenotype models

christian.salvatore@iusspavia.it

https://christiansalvatore.github.io/machinelearning-iuss/