The Promise of Digital Biopsy for the Prediction of Tumor Molecular Features and Clinical Outcomes Associated With Immunotherapy

Giuseppe Luigi Banna*, Timothée Olivier, Francesco Rundo, Umberto Malapelle, Filippo Fraggetta, Massimo Libra and Alfredo Addeo

1 Oncology Department, United Lincolnshire Hospital Trust, Lincoln, United Kingdom, 2 Oncology Department, University Hospital Geneva, Geneva, Switzerland, 3 ADG Central R&D - STMicroelectronics of Catania, Catania, Italy, 4 Department of Public Health, University Federico II of Naples, Naples, Italy, 5 Department of Pathology, Cannizzaro Hospital, Catania, Italy, 6 Oncologic, Clinic and General Pathology Section, Department of Biomedical and Biotechnological Sciences, University of Catania, Catania, Italy

Immunotherapy by immune checkpoint inhibitors has emerged as an effective treatment for a slight proportion of patients with aggressive tumors. Currently, some molecular determinants, such as the expression of the programmed cell death ligand-1 (PD-L1) or the tumor mutational burden (TMB) have been used in the clinical practice as predictive biomarkers, although they fail in consistency, applicability, or reliability to precisely identify the responding patients mainly because of their spatial intratumoral heterogeneity. Therefore, new biomarkers for early prediction of patient response to immunotherapy, that could integrate several approaches, are eagerly sought. Novel methods of quantitative image analysis (such as radiomics or pathomics) might offer a comprehensive approach providing spatial and temporal information from macroscopic imaging features potentially predictive of underlying molecular drivers, tumor-immune microenvironment, tumor-related prognosis, and clinical outcome (in terms of response or toxicity) following immunotherapy. Preliminary results from radiomics and pathomics analysis have demonstrated their ability to correlate image features with PD-L1 tumor expression, high CD3 cell infiltration or CD8 cell expression, or to produce an image signature concordant with gene expression. Furthermore, the predictive power of radiomics and pathomics can be improved by combining information from other modalities, such as blood values or molecular features, leading to increase the accuracy of these models. Thus, “digital biopsy,” which could be defined by non-invasive and non-consuming digital techniques provided by radiomics and pathomics, may have the potential to allow for personalized approach for cancer patients treated with immunotherapy.

Keywords: radiomics, pathomics, omics, predictive, immunotherapy, cancer, digital biopsy, prognostic
INTRODUCTION

In the data deluge era, there is a unique opportunity to explore biological processes at multiple scales. Deriving useful information from data, often poorly structured, at large scales, led to the emergence of the so-called “-omics” disciplines (genomic, transcriptomic, proteomic, metabolomic, etc.) (1). Powerful bioinformatic tools allow for high-throughput extraction processes that convert images into data, from which biostatistical analysis, combined with clinical or other “-omics” data, may enhance diagnostic accuracy and find new predictive or prognostic factors (2). Applied to radiological images (most often computed tomography [CT], magnetic resonance [MR] imaging, and positron-emission tomography [PET]), it is called radiomics, which has been the pioneer in the field of images data analysis. Pathomics, that is a more recent discipline, ensues when the same processes are being applied to histopathological images.

In this review, we describe the basic background based on which these new disciplines have emerged and the important steps involved in imaging acquisition to clinical supporting correlations. Selected radiomics and pathomics reports will illustrate achievements in this field, with a focus on immunotherapy. Challenges and future development will be then considered.

BACKGROUND FOR RADIOMICS AND PATHOMICS

The founder hypothesis supporting the use of radiomics and pathomics in medical care is that data derived from images have a correlation with the underlying biological processes. More precisely, data derived from images would give additional information in relation with the underlying biological processes in comparison with the visual interpretation of the image as a picture, which is the traditional way of interpreting images (3).

Radiomics and, at a lesser extent, pathomics, fill the need to assess tumor heterogeneity. The presence, within the tumor, of distinct molecular cell clones, is a hallmark of cancer physiopathology (4). Natural history of cancer, as well as resistance mechanisms acquired through therapeutic selective pressure, manifest spatial and temporal heterogeneity of tumor cells (5, 6). Addressing tumor heterogeneity is one of the major goals of new therapeutic approaches and blood biomarkers may present limitations that could be overcome by radiomics and pathomics. In particular, radiomics represents a promising non-invasive and repeatable tool during the course of the disease.

Furthermore, traditional medical practice, based on human visual interpretation of images, is known to be inaccurate in up to 20% of cases in radiology and almost the same discrepancy rates are found in pathology reports (2). Despite many explanations accounting for these reporting errors, the result is the high prevalence of diagnosis unreliability, with clinical consequences for patients.

As far as cancer immunotherapy is concerned, immune checkpoint inhibitors (CPIs) have emerged as an effective therapeutic option for patients with aggressive tumors such as lung cancer (7, 8), although a few patients seem to benefit from the long-term benefit from this treatment (9). Aiming at identifying these patients, the expression of programmed cell death ligand-1 (PD-L1) has been widely explored as a predictive biomarker with contrasting results across different tumor subtypes and several methodological issues, mostly related to its variability and spatial intratumoral heterogeneity, that have been undermining its role and use (10). Other predictive biomarkers, such as the tumor mutational burden (TMB), are currently poorly applicable in the clinical practice and, noteworthy, identify a different sensitive population from the one selected by the PD-L1 (11). Thus, there is a need for new biomarkers to integrate into clinical practice in order to early identify patient response (or progression) to CPIs and avoid their potential sever toxicity (12–14).

PROCESS DESCRIPTION AND METHODS

Every “-omics” analysis requires a multistep process. Each stage has its own specificities. Radiomics process has been established as a model for other disciplines in image data analysis (such as pathomics) and essentially consists in the following five steps: image acquisition, identification of the target volumes, segmentation of the volumes, features extraction from the volumes and analysis [see Figure 1; (3, 15)].

After the first step, the identification of the volumes must identify tumor location and determine distinct parts within the tumor. These regions will be called habitats, and present specific biological properties (blood flow, cell density, edema, necrosis). Image data analysis can help to identify such habitats (16) before data extraction. This step is intentionally done before data extraction, thus giving additional data that would not be automatically detected by subsequent data analysis (17).

The next step, the most critical one, is the segmentation. It consists in contouring the volumes of interest. Its importance derives from the fact that all the data extraction process will be generated by each segmented volume, and any error at this point could mislead further interpretation. Given inter-operator variability and the time consuming of manual delineation, semi-automated tools seem to be the most reliable and cost-effective approaches to this step (18).

Next stages, highly technical, allow for high-throughput extraction of quantitative data and their analysis. Data extraction results in image-based “features.” These features are mathematically and bioinformatically derived from images through first-, second-, or higher order statistical processes. Radiomics features could be “texture” feature, “tumor heterogeneity” feature, etc. Quantitative features may be presented based on histograms for each volume of interest.

Analysis of radiomics features, along with clinical data or other “-omics” data try to find correlations with biological processes. The analysis aims to define and validate image-derived features as biomarkers that could have prognostic or predictive values helping thus to support medical decisions.

Different methods could apply to exploit this process, but we will exclusively describe, as an example to understand the
full operation, the bio-inspired system we have been currently investigating within a multi-disciplinary joint lab (engineers, mathematicians, and clinicians) for pathomics and radiomics. The mathematical core is based on recent Machine Learning (ML) approaches. The high capability of the ML systems in addressing complex problems and, in particular, those related to healthcare and medical applications, has already been confirmed (19, 20). As an additional validation, we have also implemented a joined mathematical-ML system for the early discrimination of skin lesions by dermoscopic images with high diagnostic accuracy (21). The bio-inspired system is based on the correlation between the tumor aggressiveness and fractal dimension of the related lesions (22).

Currently, we have been testing this approach within two specific subject areas. The first one in the field of pathomics for lung cancer (reported in Figure 2A), regards specifically the prediction of PD-L1 overexpression (a biomarker predictive of response to immunotherapy in this tumor subtype) by the analysis of histopathological hematoxylin stained images; this could represent a useful guide to pathologists (and physicians). The second one concerns radiomics for urothelial cancer and it is aimed to correlate tumor response to immunotherapy with CT-scans medical images and other blood data (i.e., radiomics).

Starting from these premises, for pathomics, we have implemented a hyper-filtering pre-processing of histopathological hematoxylin stained images (Figure 2A). Each of the analyzed images has been converted from RGB (red-green-blue) color spaces to luminance (Y) chrominance information (CbCr) spaces with the divided gray-level representation of the histopathologic image. The luminance Y gray-level images have been then pre-processed by the hyper-filtering layer inside the “Pre-processing Block” using an ad-hoc adaptive thresholds-based approach in order to obtain a 1D representation of the source gray-level Y images. From every pre-processed Y images, the system computes the corresponding fractal dimension according to the Hausdorff model allowing to obtain, through an additional computing analysis, a time-series collection of those fractal dimensions (23). These pathomics features, ensued along with histopathologic image-features extracted by the AutoEncoder system (that is designed with one hidden layer of 20 neurons) also included in the “Pre-processing Block” are fed into a regression neural network learned by a classical Scalable Conjugate Gradient (SCG) back-propagation algorithm, with the final classification layer based on the SoftMax approach (21).

For the learning process (training phase), the authors used 70 percent of the histopathologic images while the remaining 30 percent serves for testing and validation. The learning dynamic of the bio-inspired system and an example of the fractal dimension time-series extracted from images are represented in Figure 2B.

For our radiomics project, the system is basically the same as above described (Figure 2A) with the input being the sequence of segmented CT-scan slices in which the lesion is visible along with the possible association of normalized representation of laboratory data (i.e., blood values). Through an innovative patented approach, time-series mapped signals are extracted in the pre-processing layer, starting from an ad-hoc analysis of the morpho-geometric dynamic of the CT-scan lesion in each of the slices. The resulting output (time-series data) feed, as a new input, the regression neural layer and then the SoftMax classifier, which finally provide the binary discrimination of the positive or negative response to the immunotherapy (Figure 2C).

**RADIOMICS AND PATHOMICS APPLICATIONS**

**Diagnosis (Early) and Classification**

Computer-aided diagnosis and detection system (CAD) help for better detection and diagnostic accuracy (24). Radiomics analysis, although sharing some principles with CAD, do not answer only a precise question (detection) but it is a complex process looking for a correlation with biological mechanisms. Magnetic resonance (MR) images from 147 patients with confirmed prostate cancer showed that several MR derived “texture” features were significantly different in benign and malignant prostate tissue and in samples with different Gleason scores (25). Another study confirmed that texture features extracted from MR prostatic images could define with accuracy not only the Gleason score but also score patterns: two patterns of Gleason score 7 (“4 + 3” vs. “3 + 4”) were correctly discriminated with 92% accuracy (26).

Pathomics studies were preceded by computer-aided-system tools, with for instance a fractal analysis set, showing powerful discrimination in grading prostatic cancer (27). In another study, analysis from 39 patients with colorectal lesions finds that analysis of multiscale texture features, extracted through a “3D wavelet transform filter” from histopathological images, were able to correctly distinguish different colorectal cancer grades (28).

In the context of immunotherapy, Tang et al. associated radiomics features with PDL1 expression and CD3 count in two cohorts (training and validation cohort of n = 114 and n = 176, respectively) lung cancer patients (29). Sun et al. developed a radiomic signature for tumor-infiltrating CD8 cells...
FIGURE 2 | Bio-inspired system for radiomics and pathomics. (A) Bio-inspired system for pathomics in lung cancer. (B) The diagram on the left shows an example of fractal dimension time-series extracted from a single histopathological hematoxylin stained image. The one on the right illustrates the learning dynamic of the system during the training session: the lower (blue) curve shows the training dynamic (i.e., the progressive error reduction) while the middle (red) and the upper (green) curves show the testing and validation, respectively. (C) bio-inspired system for radiomics in urothelial cancer. The pre-processing input data used arise from CT-scan images and blood analysis data.
in a retrospective multicohort study on overall 491 patients with advanced solid tumors (30).

**Prognosis**

The prognostic value of radiomics was reported in 108 patients with lung adenocarcinoma (separated in two independent cohorts), radiomics features (including tumor shape complexity and intratumor density variation) were strongly correlated with overall survival (31). Furthermore, Aerts et al. analyzed 440 image-related features extracted from CT images of 1019 patients with lung or head and neck cancer. They could find many radiomic features having a prognosis value and built a prognostic radiomic signature, which was found to be correlated with underlying gene-expression patterns (32).

Pathomics could also yield prognostic information. Pathomics features derived from the analysis of 2186 histopathological images were explored to distinguish short-term and long-term survivors in patients with non-small lung cancer. The survival prediction model was validated on 294 additional images (33).

Pathomics and radiomics studies in glioblastoma patients illustrated how correlations derived from different data scales (neuroimaging, pathologic and genomic) may give a deeper understanding of tumor biology and predict clinical outcomes (34–37).

Regarding immunotherapy, in the above-mentioned study of Tang et al. (29), a radiomic immune pathology-informed model was developed. The model defined four subsets of lung cancer patients significantly associated with overall survival. A group of patients with favorable prognosis was identified, harboring low CT intensity and high heterogeneity (as radiomic features) and low PD-L1 with high CD3 infiltration, indicating a favorable immune activity.

**Outcome Prediction**

To date, fewer works have explored the predictive value of radiomics and pathomics features. MR images-derived texture features from 58 breast cancer patients showed that radiomic features before neoadjuvant chemotherapy could predict response (38).

As far as immunotherapy is concerned, in the study of Sun et al. (30), the radiomic-based biomarker of tumor-infiltrating CD8 cells was validated in 3 independent cohorts and showed predictive value for tumor response to the anti-PD-1 or anti-PD-L1 therapy. Moreover, Colen et al. elaborated a two-feature radiomic model in order to predict immunotherapy-induced pneumonitis characterized by strong internal accuracy (100%) (39).

**FUTURE CHALLENGES OF IMAGE-DERIVED FEATURES**

Some challenges regarding the multistep process of radiomics and pathomics still need to be adequately addressed. Methodologically, quantitative image-derived features biomarkers should undergo a multicenter prospective trial to be validated, as it is for other biomarkers. Technically, each step of image data analysis needs proper benchmarking and reproducibility. Furthermore, curation of big data, time processing and data sharing are other major challenges. In this sense, great efforts have been made by the scientific community to share tools (software, web-based platforms) allowing physicians to explore image data analysis (40–42). The Quantitative Imaging Network, for instance, initiated in 2008 and supported by the National Cancer Institute, is an example of the importance of these new disciplines. Along with the identification of biological biomarkers, assessed by longitudinal repeated tumor samples taken by tissue biopsy and/or liquid biopsy, we postulate that “digital biopsy,” as previously defined, could allow to find potential correlation between biological biomarkers and “radiomics and pathomics biomarkers,” and have the potential to better define prognosis and prediction of response. Interdisciplinarity and integration within “-omics” disciplines and clinicians will certainly be of key importance for greater precision in oncology diagnosis and treatment in the next future.

**DATA AVAILABILITY**

All datasets for this study are included in the manuscript and the supplementary files.

**AUTHOR CONTRIBUTIONS**

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

**FUNDING**

The United Lincolnshire NHS Hospital Trust provided funding for the publication of this paper.

**REFERENCES**

1. Cavallaro S, Paratore S, de Snoo E, Salomone E, Villari L, Buscarino C, et al. Genomic analysis: toward a new approach in breast cancer management. *Crit Rev Oncol Hematol.* (2012) 81:207–23. doi: 10.1016/j.critrevonc.2011.03.006
2. El Naqa I, Ten Haken RK. Can adiomics personalise immunotherapy? *Lancet Oncol.* (2018) 19:1138–9. doi: 10.1016/S1470-2045(18)30429-7
3. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer.* (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036
4. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol.* (2018) 15:81–94. doi: 10.1038/nrclinonc.2017.166
5. Banna GL, Tiseo M. How to develop novel treatments for EGFR-mutant lung cancer. *Future Oncol.* (2015) 11:2375–8. doi: 10.2217/fon.15.172
6. Pilotto S, Rossi A, Vavala T, Follador A, Tiseo M, Galetta D, et al. Outcomes of first-generation EGFR-tkis against non-small-cell lung cancer harboring uncommon EGFR mutations: a post hoc analysis of the BE-POSITIVE study. *Clin Lung Cancer.* (2018) 19:93–104. doi: 10.1016/j.clc.2017.05.016
7. Novello S, Milella M, Tiseo M, Banna G, Cortinovis D, Di Maio M, et al. Maintenance therapy in NSCLC: why? To whom? Which
Banna et al. Radiomics for Immunotherapy

10. Banna GL, Passiglia F, Colonese F, Canova S, Menis J, Addeo A, et al. Immune-checkpoint inhibitors in non-small cell lung cancer: a tool to improve patients’ selection. *Critt Rev Oncol Hematol.* (2018) 129:27–39. doi: 10.1016/j.critrevonc.2018.06.016

11. Hellmann MD, Ciuleanu TE, Pluzanski A, Lee JS, Otterson GA, Audigier-Valette C, et al. Nivolumab plus ipilimumab in lung cancer with a high tumor mutational burden. *N Engl J Med.* (2018) 378:2093–104. doi: 10.1056/NEJMoa1801946

12. Paratore S, Banna GL, D’Arrigo M, Saita S, Iemmolo R, Lucenti L, et al. CXCR4 and CXCL12 immunoactivities differentiate primary non-small-cell lung cancer with or without brain metastases. *Cancer Biomark.* (2011) 10:79–89. doi: 10.3233/CBM-2011-0232

13. Banna GL, Anile G, Russo G, Vigneri P, Castaing M, Nicolosi M, et al. Predictive and prognostic value of early disease progression by PET evaluation in advanced non-small cell lung cancer. *Oncology.* (2017) 92:39–47. doi: 10.1159/000448005

14. Addeo A, Banna GL. The crucial role of predicting brain metastases development in non-small cell lung cancer patients. *J Thorac Dis.* (2018) 10(Suppl 26):S3305–S3307. doi: 10.21037/jtd.2018.08.95

15. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: the process and the challenges. *Magn Reson Imaging.* (2012) 30:1234–48. doi: 10.1016/j.mri.2012.06.010

16. Gatenby RA, Grove O, Gillies RJ. Quantitative imaging in cancer evolution and ecology. *Radiology.* (2013) 269:8–15. doi: 10.1148/radiol.13122697

17. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology.* (2016) 278:563–77. doi: 10.1148/radiol.2015151169

18. Parmar C, Rios Velazquez E, Leijenaar R, Jermoumi M, Carvalho S, et al. Integrated morphologic analysis for the identification and characterization of disease subtypes. *J Am Med Inform Assoc.* (2012) 19:317–23. doi: 10.1136/amiajnl-2011-000700

19. Cooper LA, Kong J, Gutman DA, Wang F, Colletteiri SR, Pan TC, et al. An integrative approach for in silico glioma research. *IEEE Trans Biomed Eng.* (2010) 57:2617–21. doi: 10.1109/TBME.2010.2060338

20. Cooper LA, Kong J, Gutman DA, Wang F, Gao J, Appin C, et al. Integrated morphologic analysis for the identification and characterization of disease subtypes. *J Am Med Inform Assoc.* (2012) 19:317–23. doi: 10.1136/amiajnl-2011-000700

21. Liang M, Tang W, Xu DM, Jirapatnakul AC, Reeves AP, Henschke CI, et al. MR imaging predictors of molecular profile and survival: multi-institutional study of the TCGA glioblastoma data set. *Radiology.* (2013) 267:560–9. doi: 10.1148/radiol.13120118

22. Kong J, Cooper LA, Wang F, Gao J, Teodoro G, Scarpace L, et al. Machine-based morphologic analysis of glioblastoma using whole-slice pathology images uncovers clinically relevant molecular correlates. *PLoS ONE.* (2013) 8:e81049. doi: 10.1371/journal.pone.0081049

23. Tsuruyuki CA, Tajima Y, Watanabe K, Iida Y, Hihara M, et al. Prediction of clinical and pathological response to neoadjuvant chemotherapy in patients with locally advanced breast cancer. *NMR Biomed.* (2014) 27:887–96. doi: 10.1002/nbm.3132

24. Colen RR, Fujii T, Bilen MA, Kotrotsou A, Abrol S, Hess KR, et al. Radiomics to predict immunotherapy-induced pneumonitis: proof of concept. *Invest New Drugs.* (2018) 36:601–7. doi: 10.1007/s10637-017-0524-2

25. Eliceiri KW, Berthold MR, Goldberg IG, Ibanez L, Manjunath BS, Martone ME, et al. Biological imaging software tools. *Nat Methods.* (2012) 9:697–710. doi: 10.1038/nmeth.2084

26. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn Reson Imaging.* (2012) 30:1323–41. doi: 10.1016/j.mri.2012.05.001

27. Saltz J, Almeida J, Gao Y, Sharma A, Bremer E, DiPrima T, et al. Towards generation, management, and exploration of combined radiomics and pathomics datasets for cancer research. *AMIA Jt Summits Transl Sci Proc.* (2017) 85–94.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Banna, Olivier, Rundo, Malapelle, Fraggadi, Libra and Addo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.