Defining the biological and clinical basis of radiomics

Please check the document version of this publication:

Defining the Biological and Clinical Basis of Radiomics

Patrick B.H.J. Großmann

UNIVERSITAIRE
PERS MAASTRICHT

Defining the Biological and Clinical Basis of Radiomics

Towards clinical imaging biomarkers

Dissertation

to obtain the degree of Doctor
at the Maastricht University,
on the authority of the Rector Magnificus, Prof. dr. Rianne M. Letschert,
in accordance with the decision of the Board of Deans,
to be defended in public on
Thursday
March 8th, 2018 at 14:00 hours

by

Patrick Benedict Hans Juan Großmann

Promotor

Prof. dr. P. Lambin

Co-promotors

Dr. ir. H. J.W.L. Aerts

Assessment committee

Prof. dr. ir. A. L.A.J. Dekker   (Chair)
Prof. dr. T. Leiner   (UMC Utrecht, The Netherlands)
Dr. ir. S. Klein   (Erasmus MC, The Netherlands)
Prof. dr. ir. F. Verhaegen

CONTENTS

PART 1

Introduction and Outline

CHAPTER

1

General introduction and outline of the thesis

Patrick B.H.J Großmann

INTRODUCTION

With currently over 14 million new cases and 8.2 million deaths per year worldwide (2012), cancer is one of the leading causes of mortality [1]. The total global burden is expected to increase to over 22 million new cases and 13 million deaths per year, mainly due to a general raise in life expectancy, as well as behavioral factors involving environmental carcinogens [2,3]. The result of this development urges for novel approaches to prevent, diagnose, and treat cancer more effectively. These active research areas already provide successful translation into a holistic cancer management strategy. For example, vaccines against human papillomavirus infection have already contributed in reducing the overall frequency of cervical cancers [4], mammography screening has led to a reduction in breast-cancer related death [5], and immunotherapy has the potential for unprecedented responses for certain patients [6,7].

Herein lies the general challenge in cancer management; cancers across patients are a particularly heterogeneous disease [8], which makes it imperative to select a personalized treatment combination in order to overcome resistances [9]. This paradigm of applying tailored treatment to every individual patient, is commonly referred to as 'precision medicine'. This thesis investigates how novel quantitative radioimaging based diagnostic approaches to phenotype tumors can be used in conjunction with tissue based genotyping of tumors to assess tumors more accurately. We present results of multiple studies, in which we have documented mechanistic connections between radioimaging, gene expression traits, somatic mutations, and clinical outcomes including overall survival, progression-free survival, and pathologic histology.

The core of this thesis are radiogenomic analyses linking radioimaging and molecular biology for precision medicine in cancer. In the following, the thesis objectives are formulated and the structure of the thesis is outlined. Next, we recapitulate the fundamental importance of precision medicine in cancer and the role of radiation oncology in clinical care. Subsequently, we introduce the concept of "radiomics", an automated, quantitative approach to assess the radiographic phenotype of a tumor based on machine learning. Furthermore, we review current radiomics literature with a focus on prognostic and predictive value of radiomic approaches across several cancer types, as well as preliminary results exploring molecular relationships. In addition, we will build on current radiomics literature to motivate the thesis goals of unraveling the mechanistic connections of radiomics, tumor biology, and clinical outcomes.

Thesis Objectives

Radiomics is a novel field that seeks to innovate current cancer management by automatically converting standard medical images of tumors into actionable quantitative data. This data can then be objectively characterized using computational statistics and machine learning to create predictive models of disease progression. Despite the enormous poten-

tial to complement current diagnostic approaches, including genomic profiling, little is known about the biological principles that underlie radiomics.

The goal of this thesis is to elucidate biological mechanisms that drive radiomic imaging phenotypes in solid tumors, information that could be leveraged to predict benefits of targeted drugs (i.e., imaging-biomarkers). To this end, we present integrative analyses that combined quantitative imaging, molecular interrogation, and clinical parameters. These results were obtained on large, independent, and novel cohorts of patients with non-small cell lung cancer and glioblastoma multiforme, which are two of the most aggressive and common cancer types in adults. Moreover, we made all underlying data as well as documented analysis code publicly available wherever possible, to ensure reproducibility of our results and enable further translational research.
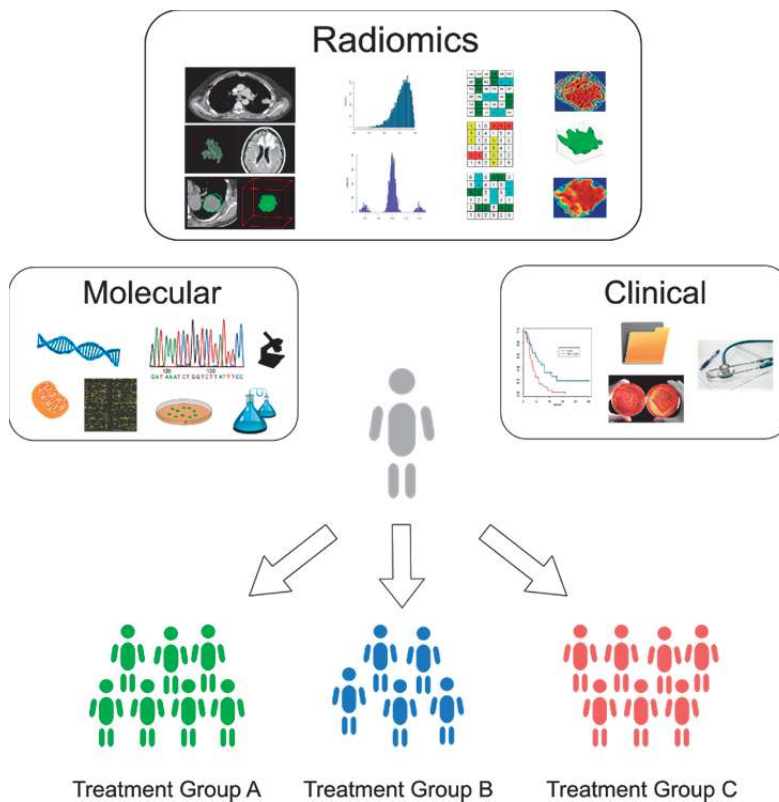
Figure 1: *Integrative Radiomics for Precision Medicine.*
The goal of this thesis is to create holistic views of solid tumors from three different combined angles: imaging, biology, and clinical factors. This is achieved by profiling tumors with radiomics, molecular methods, and clinical follow-up, respectively. The end goal of this holistic view is an advanced understanding of the current state of a tumor to assign more effective therapies to individual patients.

Outline

This doctoral thesis is structured in five parts comprising ten chapters in total. Part 1 is a general introduction to radiomics, its potential impact of precision medicine in cancer, and the challenges of biological reasoning of radiomic based biomarkers. The core of this thesis is structured in Part 2 to 4 that present published studies conducting research on the underlying biology of radiomics, their implications on targeted therapies, and the need for rigorous and technical machine learning methods for radiomic prognostication. Finally, Part 5 discusses those results in the context of current clinical practice and future developments. We briefly summarize the thesis chapters below.

Part 1: Introduction

Chapter 1 provides background information and literature research on precision medicine in cancer, biomarkers, radiation oncology and its potential for cancer diagnostics, the emerging field of quantitative radiomics, and challenges in revealing the biological basis of radiomic based predictions for clinical decision support.

Part 2: Radiomics and its Underlying Biology

Chapter 2 defines the biological basis of radiomics by linking a broad set of radiomic features extracted from CT lung tumor scans to molecular pathways based on global gene expression analyses. Furthermore, we link these results to clinical factors, including tumor stage, histology, and overall survival. Importantly, we validated our results on several levels, including statistical validation with independent datasets and biological experiments. Finally, we shed light on the current academic discussion whether radiomics adds prognostic value to existing predictors based on genomic or clinical data.

Chapter 3 extends the results from Chapter 3 to associations between volumetric phenotypic features of glioblastoma derived from MRI and gene expression data. **Chapter 4** adds novel insight about how such volumetric features are connected to somatic mutations of oncodriver and tumor suppressor genes.

Part 3: Radiomics for Targeted Therapies

Chapter 5 presents an extensive study that investigated prognostic value of radiomics for patients with recurrent glioblastoma treated with bevacizumab, an anti-angiogenic treatment currently approved for recurrent glioblastoma in the US. This study was conducted on data from a clinical phase II trial that contributed to accelerated approval be the US FDA.

Chapter 6 aimed at defining a radiomic response phenotype to gefitinib, an EGFR inhibitor. Similar to Chapter 5, this study investigated whether radiomics could be a valuable tool to predict an EGFR target in patients with non-small cell lung cancer. Here as well, this study was performed on data from a clinical phase II trial.

Part 4: Prognostic Value of Radiomic Machine Learning

Chapter 7 evaluates a diverse host of machine learning algorithms for radiomic prognostication in lung cancer. This study aimed at suggesting optimal machine learning methods for future radiomic studies using overall survival as clinical endpoint.

Chapter 8 provides insights on how pathological response can be predicted in lung cancer by incorporating unsupervised machine learning methods.

Chapter 9 extends those results by incorporating advanced supervised machine learning methods to optimize radiomic predictions. The clinical endpoint here was, for the first time, distant metastasis and included independent validation of the results.

Part 5: Discussion and Future Perspectives

Chapter 10 concludes this doctoral thesis by discussing the overall impact that our radiomic studies have on the general scientific and oncological community. Individual studies that are part of this thesis are linked to each other and embedded into future perspectives to innovate clinical decision making in cancer care.

Precision Medicine and Tumor Heterogeneity

Precision medicine refers to the paradigm of offering patients individualized treatment on the basis of their personal data. In oncology, different layers of patient data are already considered when treating patients, including tissue, family history, demographics, and blood markers [10–12]. More sophisticated approaches that take the biological makeup of a tumor into account are increasingly available in clinical practice [13–16], especially genotyping of tumors to test applicability of compounds that target specific genetic mutations.

One of the major obstacles of individualizing patients with cancer is that cancer is a heterogeneous disease, which limits both diagnostics and treatment. It is well established that the degree of heterogeneity varies temporally [17], intensifying these issues in later stage cancers where diagnostic and treatment are even more decisive. Generally, the scientific and clinical community distinguish between inter- and intra-tumor heterogeneity.

Inter-tumor heterogeneity refers to the heterogeneity of tumors observed across different patients [18]. This entails that gene mutations, expression, and metabolism in tumors may

differ even for cancers of the same tissue. Consequently, the disease phenotype may differ as well, requiring adaptation of treatment. To date, however, treatment is largely based on tissue type, histology of the tumor, and risk factors such as age, gender, or other risk factors [19]; basing treatment on molecular profiling would enable to individualize treatment even further. Cases where molecular profiling has already been successful in treating patients mainly include genetic testing for targetable mutations [20].

Intra-tumor heterogeneity refers to heterogeneity of the population of cancerous cells that a tumor is comprised of [18]. The genetic and genomic landscape of these cells varies substantially explained by darwinian clonal evolution [9]. As a result, genomic profiling of tumors may deliver contradicting views on mutations, expression, and metabolism of a cell mass, depending on the spatial and temporal properties of the sampled tissue [8,9]. This means, for example, that even though a certain number of cancerous cells may contain a specific mutation that is susceptible to a targeted drug, the tumor as a system can become resistant to that drug as soon as cells without the corresponding mutation have outgrown the other susceptible cells. This adaptation of disease phenotype is one of the most common mechanism of acquired drug resistance [21].

While molecular profiling via single-needle biopsy can account for inter-tumor heterogeneity, single-needle biopsy driven assessment is limited by intra-tumor heterogeneity [8] as the spatial location of the tissue sample is crucial. Multiple approaches have been suggested to account for intra-tumor heterogeneity, including combination therapies, single-cell analysis, and assessment of molecular cell-free DNA [22–25]. The ideal scenario of precision medicine in practice would be a holistic view that incorporates multiple of those approaches on a longitudinal scale for immediate intervention. Common goals of those approaches include understanding of the mechanisms of treatment response and combating the development of resistances to expand life expectancies of patients living with cancer.

Biomarkers

Biomarkers are a fundamental tool in precision medicine. Biomarkers are defined as an objective state of a biological process that can be measured reproducibly and which indicates a medical condition of a patient [26,27]; examples range from blood pressure, over genomic assays, to radioimaging based parameters. Biomarkers are developed to predict clinical endpoints based on which treatment options can be designed to improve morbidity and mortality. The underlying assumption is that the same treatment may not show the same efficacy in patients that show different biological indications.

In oncology, a variety of biomarkers have been proposed on the basis of nucleic acids (e.g. genetic mutations or expression), proteins, or antibodies [28]. Such biomarkers can be used for screening, differential diagnosis, prognosis, and prediction of treatment response or recurrence. Most often, these biomarkers have been developed from analyz-

ing alterations to healthy individuals. A biomarker can also be a simultaneous measurement of a group of biological processes; for example, the combined expression of a group of genes, termed gene signature, can be used for prognosis [29].

Biomarkers for cancer subtype identification: An important application of biomarkers in oncology is predicting tumor subtypes to stratify patients into more effective treatment groups. A prominent example of this is inhibition of epidermal-growth-factor-receptor (EGFR) in cases where a sensitizing mutation is present in this gene [30,31]. Drugs that target this mutation, such as Erlotinib, Gefitinib, and Afatinib, have shown dramatic efficacy in lung cancer [32] and similar results have been suggested for other EGFR targets in colon cancer [33]. Without an effective biomarker to assess mutational status of tumors, targeted drugs could not be prescribed.

As noted earlier, applicability of single-needle biopsy based biomarkers is limited in clinical practice due to intra-tumor heterogeneity. Therefore, alternative approaches towards biomarkers are currently being explored to augment molecular profiling of tumors. One of these approaches is tumor phenotyping based on radiographic imaging [34]. Imaging biomarkers aim at assessing tumor phenotype by analyzing images of tumors generated from radiographic scanners [35].

Medical Imaging in Cancer

Medical imaging is crucial to an integral cancer management and is used clinically, as well as pre-clinically [36]. Medical imaging can generate visualizations on multiple scaling levels, ranging from molecular imaging (e.g., cell surface receptors) to anatomical imaging (e.g., organs) [36,37]. In clinical oncology, radiology deals with anatomical imaging techniques to visualize solid and soft tissue within the body of patients, particularly tumors. Importantly, these techniques are non-invasive, meaning that tissue will not be damaged to create a visualization. Non-invasiveness also implies that radiographic interrogation can be applied longitudinally for long-term monitoring without additional acquisition risk to patients, which is a considerable advantage compared to invasive biopsies. Radiation oncologists utilize anatomical imaging to detect, locate, and diagnose a tumor [36]. Furthermore, imaging serves as guidance tool in surgical resection of tumors and is utilized in treatment planning in radiation therapy.

Imaging modalities: Different imaging techniques have been developed; for example, X-Ray, Computed Tomography (CT), Magnetic Resonance Imaging (MRI), or Positron Emission Tomography (PET) [36]. These techniques are used for different purposes, mainly depending on the tissue of the cancer. For example, while CT is the primary imaging modality to visualize lung cancer, MRI is the primary choice for brain cancer. Furthermore, PET provides functional images, which indicate metabolically active areas of tumors [38].

15

Imaging criteria for disease assessment: Several imaging based criteria have been developed to assess characteristics, such as tumor stage or response to treatment, for example the Tumor Nodule Metastasis (TNM) system [39], the Response Evaluation Criteria in Solid Tumors (RECIST) [40], or the Response Assessment in Neuro-Oncology Criteria (RANO) [41,42]. While these criteria are widely implemented in hospitals treating patients with cancer, performance of these metrics is limited due to inherent oversimplifications. For example, treatment response by RECIST has been documented to not correlate with survival outcome in many cases [43], which from a clinical perspective it should. Similarly, reproducibility issues have been reported for TNM staging and response by RANO [44–46]. The greatest challenges with criteria such as RECIST, RANO, or TNM are A) that these require subjective human measurements [46], B) that these criteria incorporate only a limited subset of available information [47], and C) that these criteria provide qualitative assessment only [48]. Hence, there is an urgent need for the development of objective and quantitative imaging metrics that leverage as much available information as possible.

Radiomics

Radiomics is an emerging field aiming at generating actionable insights from standard anatomical imaging in an automated way. Hereby, the goal is to quantify the imaging phenotype of a tumor to inform treatment decisions [49]. As genomics refers to the collectivity of all genes within a cell, radiomics refers to the collectivity of all *quantitative* radiographic features of a subject of interest. Although radiomics can be applied to any condition that can be imaged, radiomics is most developed in oncology. Here, radiomics has the potential to improve human-derived imaging diagnostic and response criteria by providing an objective, comprehensive, and quantitative view of a tumor [37,50]. To achieve this, a large number of quantitative features defining tumor intensity, shape, and texture are automatically extracted from the entire visually segmented tumor volume and subsequently characterized using data mining and machine learning [51]. Human interaction in classification and scoring of tumors is not required, but quality control by a physician is desired.

Radiomics for predictive disease modeling: Multiple studies have documented prognostic and predictive value of radiomics for several clinical endpoints [49,52–57]. Aerts et al. found significant prognostic value of a radiomic CT signature developed for overall survival that validated across multiple independent cohorts of patients with lung and head & neck cancer [49]. Prognostic value across several cancer types has been also observed in a radiomic PET signature [58,59], In colorectal cancer, a seminar study by Huang et al. described a radiomics nomogram to predict lymph node metastases [60]. Further studies suggest that radiomic approaches can distinguish between cancerous and non-cancerous tissue [61,62], which is especially relevant in population screening [5].

These studies have in common that they investigated how radiomic quantities are associated with clinical outcomes to build prognostic and predictive models. The cen-

tral hypothesis of radiomics is that medical images of tumors do not only visualize the tumor burden, but deliver minable data about the tumor type that is currently not used clinically [50]. Furthermore, these data are likely to be complementary to already clinically available data from patient history or biopsies [63]. In this case, radiomics could lead to the development of advanced clinical decision support system that incorporate automated medical image analysis into their workflow.

Radiomic innovation for clinical cancer management: Radiomics has several novel and complementary advantages compared to more traditional approaches, including single-needle biopsy based diagnostics [64]. First, radiomics is medical imaging based and therefore non-invasive, which allows longitudinal monitoring at low additional acquisition risk to patients. Second, radiomics uses objective measurements that are predefined and hence can be reproduced by any physician given the same data. Third, radiomics provides quantitative, not just qualitative, measurements. Fourth, radiomics is comprehensive and not subject to intra-tumor heterogeneity as the entire (visible) tumor is taken into account, not just a sample from it. Finally, radiomics has unprecedented potential to play a central role in clinical cancer management, as imaging is routinely and frequently used in oncology and therefore avoids major additional costs of a novel technology; the innovation of radiomics is not the generation of data, but the translation and analysis of existing and previously unconsidered data.

Radiomics workflow

The radiomic workflow can be grouped into the four steps: 1) image acquisition, 2) tumor segmentation, 3) high-throughput feature extraction, and 4) integrated analysis (Figure 2). All of these steps are active research areas that are hypothesized to ultimately lead to better risk stratification of patients into more beneficial treatment groups.
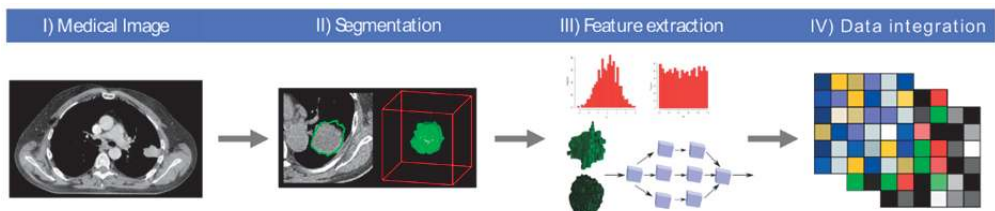
Figure 2: *The Radiomics Workflow*. Radiomics is organized in four steps, starting with the acquisition of medical images of a tumor via radiographic imaging modalities, such as computed tomography (CT) or magnetic resonance imaging (MRI). On the basis of these images, tumors are segmented in 3D, usually along the axial dimension. From the segmented tumor volumes, quantitative radiomic features are automatically extracted in high-throughput, including first-order statistics, shape features, and texture features. Finally, these features are analyzed and can be further integrated with other sources, such as gene expression data and clinical records.

1) Image acquisition: The first step in the radiomics pipeline is the acquisition of images of the region of interest, in this case the tumor. Image acquisition has experienced substantial advances over the last decade, especially in terms of hardware and image reconstruction algorithms [36,37]. In addition, standardized imaging protocols have contributed in harmonizing images across hospitals [37,65]. From those images, segmentations are of the tumors are obtained.

2) Segmentation: Delineation or segmentation of the three dimensional tumor volume is the basis of extracting radiomic tumor imaging features, although historically features were also extracted from two dimensional slides [66–68]. Segmentation of tumors is usually performed on axial slices following the image acquisition of radiographic modalities. While manual segmentation of tumors is the current standard for most tumors and imaging modalities, semi-automated and fully automatic tools are available. For instance, the GrowCut algorithm implemented into the 3D-Slicer platform (www.slicer.org) is a freely available semi-automatic segmentation algorithm [69,70]. An example for fully automatic segmentation of brain lesions is given by the BraTumIA (Brain Tumor Image Analysis) algorithm [71], which has been shown to readily yield accurate segmentations [72,73]. Regardless of whether automated procedures have been employed or not, the final tumor segmentation has to be verified by an experienced radiation oncologist.

3) Feature extraction: Quantitative imaging features are extracted from the segmented tumor images in high-throughput and can broadly be grouped into first-order statistics, tumor shape, and tumor texture [37]. First-order statistics are distributional parameters of the gray-level intensity values of the tumors that include minimum, maximum, mean, variance, dynamic range, kurtosis, skewness, entropy, and energy. Shape features include compactness, sphericity, volume, diameter, and elongation. Texture features are an important group; these features quantify re-occurring gray-level patterns within an image. For example, how often gray-levels of the same discretized values are adjacent to each other, or the number of voxels (i.e., pixels in three dimensions) that a stretch of increasing gray-level values has. In this way, texture features can be a proxy to quantifying intra-tumoral heterogeneity. In fact, texture features seem to be amongst the most promising features in terms of prognosis [49,55,74,75]. First-order statistics and texture features are often also extracted after filtering methods, such as wavelets or Laplace of Gaussian, have been applied to the original image. A comprehensive open source package for radiomic feature extraction is PyRadiomics, available through GitHub at https://github.com/Radiomics/pyradiomics.

4) Integrated analysis: Finally, the crucial part of the radiomics workflow is the analysis of the extracted features. From all features, the most informative are selected with unsupervised and supervised feature selection procedures [63]. A number of current studies also employ a stability dataset for preprocessing to filter for highly reproducible features [49,68,76]. As all features are quantitative, statistical models can be fitted to test for cor-

relations and predictive power for clinical endpoints, including overall and progression-free survival [64]. In addition to clinical records, an important step is integrating the radiomic imaging data with molecular data, such as gene expression and mutation. Hereby, the goal is to gain wider insights of the disease progression by leveraging as many data sources as possible while assuming that these data types deliver complementary information to each other.

Other Type of Image Features

While this thesis deals with automated quantitative radiomic features, other types of features can be extracted from medical imaging [64]. The available feature sets range from qualitative to quantitative features. Semantic features, such as the MRI scoring scheme VASARI (Visually AcceSAble Rembrandt Images, https://wiki.nci.nih.gov/display/CIP/VASARI), provide semantic meaning to their annotation. The advantage of these features is that they are typically developed by an expert radiologist. However, this also implies that human assessment is required and therefore intra- and inter variability and time management may become issues. Furthermore, these features cannot provide quantitative annotations, but yield a qualitative, and sometimes ordinal, perspective. Other types of standardized lexica similar to VASARI have been described and documented [77–81].

The Biological Foundation of Radiomics

A fundamental concern about efforts aiming at clinical translation of radiomics is that the underlying biology of radiomics is poorly understood. As an increasing number of studies document prognostic and predictive value of radiomic features, it is crucial to enrich these associations with their biological rationale to reason about the validity of radiomics based predictive tools. For example, a central hypothesis of radiomics is that molecular heterogeneity of tumors can translate into macroscopic imaging features [82,83]. Furthermore, enhancing radiomics with biological knowledge would allow a more holistic view of the disease state of a tumor and thereby close translational gaps of imaging biomarkers. These goals could be achieved by developing and applying integrative analyses that combine radiomic image analysis, molecular information, and clinical records.

Gene expression analyses: Preliminary studies have investigated associations of radiomic derived features and molecular data. In terms of expression of genes, Segal et al. [84] were among the first to suggest that global gene expression patterns in primary human liver cancer could be decoded by radiomic features derived from CT imaging. The authors hypothesized that a combination of 28 features could reconstruct 78% of expressed genes, revealing cell proliferation and liver functions. A more targeted approach was conducted by Kuo et al. [85] who investigated whether predefined imaging phenotypes correlated with transcriptional responses to doxorubicin in hepatocellular carcinoma. Similarly, Diehn

et al. [86] studied how semantic neuroimaging features in glioblastoma multiforme (GBM) relate to expression of genes, such as EGFR. Associations of radiomic features, gene expression, and patient survival in non-small cell lung cancer (NSCLC) was first well described by Gevaert et al. [66]. More recently, Itakura et al. [75] also suggested three novel radiomic imaging phenotypes in GBM. In a seminal study by Aerts et al. [49], the authors revealed biological pathways that were associated with four feature of a prognostic radiomic signature that validated across multiple datasets [49,87].

Mutations in the genome: In terms of genomic mutations, a number of studies focused on investigating whether radiomics yields predictive value for a small number of widely accepted oncogenetic driver mutations. For example, Liu et al. [88] demonstrated that radiomic features predict the presence of EGFR mutations in an Asian cohort of patients with peripheral NSCLC adenocarcinoma. Similarly, a recent study suggested that radiomic features can predict EGFR mutation status across multiple cohorts [89]. Associations of radiomics and genetic mutations in NSCLC, including EGFR, KRAS, and ALK, have been also described by other work [90–93].

The biological challenge of radiomics: Various further preliminary studies have been published with regard to radiomics and some type of molecular characterization [94–101], but validity of these studies has generally been hampered by the fact that those results are based on small sample sizes, lack independent validation, or focused only on a subset of radiomic features and biological entities, such as genes. Furthermore, the canonical links between radiomics, its underlying biology, and the final clinical patient outcome have not been described with a convincing body of data driven evidence yet. To this end, this thesis aims at uncovering the underlying biology of radiomics and interpreting those results in the context of clinical patient outcome. The end goal of these studies is to point out both phenotypic (on the imaging level) and genotypic (on the molecular level) differences across patients who are treated for solid tumors to gain a holistic view of the disease state of the cancer and enable possibilities of offering more personalized treatments.

REFERENCES

1. Stewart BW, Wild CP. World Cancer Report 2014 (PDF). World Health Organization; 2014.
2. Stewart BW, Bray F, Forman D, Ohgaki H, Straif K, Ullrich A, et al. Cancer prevention as part of precision medicine: "plenty to be done." Carcinogenesis. 2016;37: 2–9.
3. Kitahara CM, Sosa JA. The changing incidence of thyroid cancer. Nat Rev Endocrinol. 2016;12: 646–653.
4. Sangar VC, Ghongane B, Mathur G. Development of Human Papillomavirus (HPV) Vaccines: A Review of Literature and Clinical Update. Rev Recent Clin Trials. 2016;11: 284–289.
5. Myers ER, Moorman P, Gierisch JM, Havrilesky LJ, Grimm LJ, Ghate S, et al. Benefits and Harms of Breast Cancer Screening: A Systematic Review. JAMA. 2015;314: 1615–1634.

6.   Brahmer J, Reckamp KL, Baas P, Crinò L, Eberhardt WEE, Poddubskaya E, et al. Nivolumab versus Docetaxel in Advanced Squamous-Cell Non-Small-Cell Lung Cancer. N Engl J Med. 2015;373: 123–135.

7.   Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma. N Engl J Med. 2015;373: 1270–1271.

8.   Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, et al. Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. N Engl J Med. 2012;366: 883–892.

9.   Gerlinger M, Swanton C. How Darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine. Br J Cancer. 2010;103: 1139–1143.

10.   Sangisetty SL, Miner TJ. Malignant ascites: A review of prognostic factors, pathophysiology and therapeutic measures. World J Gastrointest Surg. 2012;4: 87–95.

11.   Singer PA. Treatment guidelines for patients with thyroid nodules and well-differentiated thyroid cancer. American Thyroid Association. Arch Intern Med. 1996;156: 2165–2172.

12.   Goodison S, Rosser CJ, Urquidi V. Bladder cancer detection and monitoring: assessment of urine- and blood-based marker tests. Mol Diagn Ther. 2013;17: 71–84.

13.   Rouzier R, Pronzato P, Chéreau E, Carlson J, Hunt B, Valentine WJ. Multigene assays and molecular markers in breast cancer: systematic review of health economic analyses. Breast Cancer Res Treat. 2013;139: 621–637.

14.   Sethi S, Ali S, Philip PA, Sarkar FH. Clinical advances in molecular biomarkers for cancer diagnosis and therapy. Int J Mol Sci. 2013;14: 14771–14784.

15.   Wagner PD, Srivastava S. New paradigms in translational science research in cancer biomarkers. Transl Res. 2012;159: 343–353.

16.   Baron JA. Screening for cancer with molecular markers: progress comes with potential problems. Nat Rev Cancer. 2012;12: 368–371.

17.   Jamal-Hanjani M, Hackshaw A, Ngai Y, Shaw J, Dive C, Quezada S, et al. Tracking genomic cancer evolution for precision medicine: the lung TRACERx study. PLoS Biol. 2014;12: e1001906.

18.   Bedard PL, Hansen AR, Ratain MJ, Siu LL. Tumour heterogeneity in the clinic. Nature. 2013;501: 355–364.

19.   Cross D, Burmester JK. The Promise of Molecular Profiling for Cancer Identification and Treatment. Clin Med Res. 2004;2: 147–150.

20.   Harris TJR, McCormick F. The molecular pathology of cancer. Nat Rev Clin Oncol. 2010;7: 251–265.

21.   Sequist LV, Waltman BA, Dias-Santagata D, Digumarthy S, Turke AB, Fidias P, et al. Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. Sci Transl Med. 2011;3: 75ra26.

22.   Alizadeh AA, Aranda V, Bardelli A, Blanpain C, Bock C, Borowski C, et al. Toward understanding and exploiting tumor heterogeneity. Nat Med. 2015;21: 846–853.

23.   Janku F. Tumor heterogeneity in the clinic: is it a real problem? Ther Adv Med Oncol. 2014;6: 43–51.

24.   Hsieh JJ, Manley BJ, Khan N, Gao J, Carlo MI, Cheng EH. Overcome tumor heterogeneity-imposed therapeutic barriers through convergent genomic biomarker discovery: A braided cancer river model of kidney cancer. Semin Cell Dev Biol. 2017;64: 98–106.

25.   Meacham CE, Morrison SJ. Tumour heterogeneity and cancer cell plasticity. Nature. 2013;501: 328–337.

26.   Sakka AP, Whiteside JR. Biomarker Discovery and Medical Diagnostic Imaging. Biomarker Validation. 2015. pp. 59–73.

27.   Strimbu K, Tavel JA. What are biomarkers? Curr Opin HIV AIDS. 2010;5: 463–466.

28.   Henry NL, Hayes DF. Cancer biomarkers. Mol Oncol. 2012;6: 140–146.

29.   Győrffy B, Hatzis C, Sanft T, Hofstatter E, Aktas B, Pusztai L. Multigene prognostic tests in breast cancer: past, present, future. Breast Cancer Res. 2015;17: 11.

30.   Chan BA, Hughes BGM. Targeted therapy for non-small cell lung cancer: current standards and the promise of the future. Transl Lung Cancer Res. 2015;4: 36–54.

31.   Stella GM, Luisetti M, Inghilleri S, Cemmi F, Scabini R, Zorzetto M, et al. Targeting EGFR in non-small-cell lung cancer: Lessons, experiences, strategies. Respir Med. 2012;106: 173–183.

32.   Rosell R. What new therapeutic targets exist for EGFR-mutant NSCLC? Lancet Oncol. 2014;15: 1184–1185.

33. Siena S, Sartore-Bianchi A, Di Nicolantonio F, Balfour J, Bardelli A. Biomarkers predicting clinical outcome of epidermal growth factor receptor-targeted therapy in metastatic colorectal cancer. J Natl Cancer Inst. 2009;101: 1308–1324.

34. Law WP, Phillip Law W, Miles KA. Incorporating prognostic imaging biomarkers into clinical practice. Cancer Imaging. 2013;13: 332–341.

35. O'Connor JPB, Aboagye EO, Adams JE, Aerts HJWL, Barrington SF, Beer AJ, et al. Imaging biomarker roadmap for cancer studies. Nat Rev Clin Oncol. 2016; doi:10.1038/nrclinonc.2016.162

36. Fass L. Imaging and cancer: A review. Mol Oncol. 2008;2: 115–152.

37. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012;48: 441–446.

38. Torigian DA, Huang SS, Houseni M, Alavi A. Functional imaging of cancer with emphasis on molecular techniques. CA Cancer J Clin. 2007;57: 206–224.

39. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. Ann Surg Oncol. 2010;17: 1471–1474.

40. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer. 2009;45: 228–247.

41. Wen PY, Macdonald DR, Reardon DA, Cloughesy TF, Gregory Sorensen A, Galanis E, et al. Updated Response Assessment Criteria for High-Grade Gliomas: Response Assessment in Neuro-Oncology Working Group. J Clin Oncol. American Society of Clinical Oncology; 2016; Available: http://ascopubs.org/doi/full/10.1200/jco.2009.26.3541

42. Chang SM, Wen PY, Vogelbaum MA, Macdonald DR, van den Bent MJ. Response Assessment in Neuro-Oncology (RANO): more than imaging criteria for malignant glioma: Table 1. Neuro-Oncology Practice. 2015;2: 205–209.

43. Villaruz LC, Socinski MA. The clinical viewpoint: definitions, limitations of RECIST, practical considerations of measurement. Clin Cancer Res. 2013;19: 2629–2636.

44. Rami-Porta R, Goldstraw P. Strength and weakness of the new TNM classification for lung cancer. Eur Respir J. 2010;36: 237–239.

45. Huber T, Alber G, Bette S, Kaesmacher J, Boeckh-Behrens T, Gempt J, et al. Progressive disease in glioblastoma: Benefits and limitations of semi-automated volumetry. PLoS One. 2017;12: e0173112.

46. Barnacle AM, McHugh K. Limitations with the response evaluation criteria in solid tumors (RECIST) guidance in disseminated pediatric malignancy. Pediatr Blood Cancer. 2006;46: 127–134.

47. Benson JR. The TNM staging system and breast cancer. Lancet Oncol. 2003;4: 56–60.

48. Toffart A-C, Moro-Sibilot D, Couraud S, Merle P, Perol M, Girard N, et al. Evaluation of RECIST in chemotherapy-treated lung cancer: the Pharmacogenoscan Study. BMC Cancer. 2014;14: 989.

49. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;5: 4006.

50. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology. 2016;278: 563–577.

51. 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: 1234–1248.

52. Grove O, Berglund AE, Schabath MB, Aerts HJWL, Dekker A, Wang H, et al. Quantitative Computed Tomographic Descriptors Associate Tumor Shape Complexity and Intratumor Heterogeneity with Prognosis in Lung Adenocarcinoma. PLoS One. 2015;10: e0118261.

53. Cunliffe A, Armato SG 3rd, Castillo R, Pham N, Guerrero T, Al-Hallaq HA. Lung texture in serial thoracic computed tomography scans: correlation of radiomics-based features with radiation therapy dose and radiation pneumonitis development. Int J Radiat Oncol Biol Phys. 2015;91: 1048–1056.

54. Coroller TP, Grossmann P, Hou Y, Rios Velazquez E, Leijenaar RTH, Hermann G, et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. Radiother Oncol. 2015;114: 345–350.

55. Kickingereder P, Götz M, Muschelli J, Wick A, Neuberger U, Shinohara RT, et al. Large-scale Radiomic Profiling of Recurrent Glioblastoma Identifies an Imaging Predictor for Stratifying Anti-Angiogenic Treatment Response. Clin Cancer Res. 2016; doi:10.1158/1078-0432.CCR-16-0702

56. Chong Y, Kim J-H, Lee HY, Ahn YC, Lee KS, Ahn M-J, et al. Quantitative CT Variables Enabling Response Prediction in Neoadjuvant Therapy with EGFR-TKIs: Are They Different from Those in Neoadjuvant Concurrent Chemoradiotherapy? PLoS One. 2014;9: e88598.

57. Chang K, Zhang B, Guo X, Zong M, Rahman R, Sanchez D, et al. Multimodal imaging patterns predict survival in recurrent glioblastoma patients treated with bevacizumab. Neuro Oncol. 2016;18: 1680–1687.

58. Hatt M, Majdoub M, Vallieres M, Tixier F, Le Rest CC, Groheux D, et al. 18F-FDG PET Uptake Characterization Through Texture Analysis: Investigating the Complementary Nature of Heterogeneity and Functional Tumor Volume in a Multi-Cancer Site Patient Cohort. J Nucl Med. 2014;56: 38–44.

59. Tixier F, Hatt M, Le Rest CC, Le Pogam A, Corcos L, Visvikis D. Reproducibility of Tumor Uptake Heterogeneity Characterization Through Textural Feature Analysis in 18F-FDG PET. J Nucl Med. 2012;53: 693–700.

60. Huang Y-Q, Liang C-H, He L, Tian J, Liang C-S, Chen X, et al. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. J Clin Oncol. 2016;34: 2157–2164.

61. Fehr D, Veeraraghavan H, Wibmer A, Gondo T, Matsumoto K, Vargas HA, et al. Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images. Proc Natl Acad Sci U S A. 2015;112: E6265–73.

62. Wibmer A, Hricak H, Gondo T, Matsumoto K, Veeraraghavan H, Fehr D, et al. Haralick texture analysis of prostate MRI: utility for differentiating non-cancerous prostate from prostate cancer and differentiating prostate cancers with different Gleason scores. Eur Radiol. 2015;25: 2840–2850.

63. Lambin P, van Stiphout RGPM, Starmans MHW, Rios-Velazquez E, Nalbantov G, Aerts HJWL, et al. Predicting outcomes in radiation oncology—multifactorial decision support systems. Nat Rev Clin Oncol. 2013;10: 27–40.

64. Aerts HJWL. The Potential of Radiomic-Based Phenotyping in Precision Medicine: A Review. JAMA Oncol. 2016; doi:10.1001/jamaoncol.2016.2631

65. Kalambo M, Parikh JR. Implementing Standardized Protocols During Geographic Radiology Expansion. J Am Coll Radiol. 2017;14: 84–86.

66. Gevaert O, Xu J, Hoang CD, Leung AN, Xu Y, Quon A, et al. Non–Small Cell Lung Cancer: Identifying Prognostic Imaging Biomarkers by Leveraging Public Gene Expression Microarray Data—Methods and Preliminary Results. Radiology. 2012;264: 387–396.

67. Gevaert O, Mitchell LA, Achrol AS, Xu J, Echegaray S, Steinberg GK, et al. Glioblastoma multiforme: exploratory radiogenomic analysis by using quantitative image features. Radiology. 2014;273: 168–174.

68. Zhao B, Tan Y, Tsai W-Y, Qi J, Xie C, Lu L, et al. Reproducibility of radiomics for deciphering tumor phenotype with imaging. Sci Rep. 2016;6: 23428.

69. Parmar C, Rios Velazquez E, Leijenaar R, Jermoumi M, Carvalho S, Mak RH, et al. Robust Radiomics Feature Quantification Using Semiautomatic Volumetric Segmentation. PLoS One. 2014;9: e102107.

70. Velazquez ER, Parmar C, Jermoumi M, Mak RH, van Baardwijk A, Fennessy FM, et al. Volumetric CT-based segmentation of NSCLC using 3D-Slicer. Sci Rep. 2013;3. doi:10.1038/srep03529

71. Meier R, Knecht U, Loosli T, Bauer S, Slotboom J, Wiest R, et al. Clinical Evaluation of a Fully-automatic Segmentation Method for Longitudinal Brain Tumor Volumetry. Sci Rep. 2016;6: 23376.

72. Rios Velazquez E, Velazquez ER, Meier R, Dunn WD Jr, Alexander B, Wiest R, et al. Fully automatic GBM segmentation in the TCGA-GBM dataset: Prognosis and correlation with VASARI features. Sci Rep. 2015;5. doi:10.1038/srep16822

73. Porz N, Bauer S, Pica A, Schucht P, Beck J, Verma RK, et al. Multi-Modal Glioblastoma Segmentation: Man versus Machine. PLoS One. 2014;9: e96873.

74. Parmar C, Grossmann P, Bussink J, Lambin P, Hugo J W. Machine Learning methods for Quantitative Radiomic Biomarkers. Sci Rep. 2015;5. doi:10.1038/srep13087

75. Itakura H, Achrol AS, Mitchell LA, Loya JJ, Liu T, Westbroek EM, et al. Magnetic resonance image features identify glioblastoma phenotypic subtypes with distinct molecular pathway activities. Sci Transl Med. 2015;7: 303ra138.

76. RIDER Collections - The Cancer Imaging Archive (TCIA) Public Access - Cancer Imaging Archive Wiki [Internet]. [cited 10 Sep 2015]. Available: https://wiki.cancerimagingarchive.net/display/Public/RIDER+Collections

77. Yip SSF, Liu Y, Parmar C, Li Q, Liu S, Qu F, et al. Associations between radiologist-defined semantic and automatically computed radiomic features in non-small cell lung cancer. Sci Rep. 2017;7: 3519.

78. Wang H, Schabath MB, Liu Y, Berglund AE, Bloom GC, Kim J, et al. Semiquantitative Computed Tomography Characteristics for Lung Adenocarcinoma and Their Association With Lung Cancer Survival. Clin Lung Cancer. 2015; doi:10.1016/j.cllc.2015.05.007

79. Opulencia P, Channin DS, Raicu DS, Furst JD. Mapping LIDC, RadLex[TM], and Lung Nodule Image Features. J Digit Imaging. 2010;24: 256–270.

80. Eberl MM, Fox CH, Edge SB, Carter CA, Mahoney MC. BI-RADS Classification for Management of Abnormal Mammograms. J Am Board Fam Med. 2006;19: 161–164.

81. Pinsky PF, Gierada DS, Black W, Munden R, Nath H, Aberle D, et al. Performance of Lung-RADS in the National Lung Screening Trial: a retrospective assessment. Ann Intern Med. 2015;162: 485–491.

82. Jackson A, O'Connor JPB, Parker GJM, Jayson GC. Imaging Tumor Vascular Heterogeneity and Angiogenesis using Dynamic Contrast-Enhanced Magnetic Resonance Imaging. Clin Cancer Res. 2007;13: 3449–3459.

83. Gillies RJ, Anderson AR, Gatenby RA, Morse DL. The biology underlying molecular imaging in oncology: from genome to anatome and back again. Clin Radiol. 2010;65: 517–521.

84. Segal E, Sirlin CB, Ooi C, Adler AS, Gollub J, Chen X, et al. Decoding global gene expression programs in liver cancer by noninvasive imaging. Nat Biotechnol. 2007;25: 675–680.

85. Kuo MD, Gollub J, Sirlin CB, Ooi C, Chen X. Radiogenomic analysis to identify imaging phenotypes associated with drug response gene expression programs in hepatocellular carcinoma. J Vasc Interv Radiol. 2007;18: 821–831.

86. Diehn M, Nardini C, Wang DS, McGovern S, Jayaraman M, Liang Y, et al. Identification of noninvasive imaging surrogates for brain tumor gene-expression modules. Proc Natl Acad Sci U S A. 2008;105: 5213–5218.

87. Leijenaar RTH, Carvalho S, Hoebers FJP, Aerts HJWL, van Elmpt WJC, Huang SH, et al. External validation of a prognostic CT-based radiomic signature in oropharyngeal squamous cell carcinoma. Acta Oncol. 2015;54: 1423–1429.

88. Liu Y, Kim J, Balagurunathan Y, Li Q, Garcia AL, Stringfield O, et al. Radiomic Features Are Associated With EGFR Mutation Status in Lung Adenocarcinomas. Clin Lung Cancer. 2016; doi:10.1016/j.cllc.2016.02.001

89. Rios Velazquez E, Parmar C, Liu Y, Coroller TP, Cruz G, Stringfield O, et al. Somatic Mutations Drive Distinct Imaging Phenotypes in Lung Cancer. Cancer Res. 2017;77: 3922–3930.

90. Rizzo S, Petrella F, Buscarino V, De Maria F, Raimondi S, Barberis M, et al. CT Radiogenomic Characterization of EGFR, K-RAS, and ALK Mutations in Non-Small Cell Lung Cancer. Eur Radiol. 2015; doi:10.1007/s00330-015-3814-0

91. Hong SJ, Kim TJ, Choi YW, Park J-S, Chung J-H, Lee KW. Radiogenomic correlation in lung adenocarcinoma with epidermal growth factor receptor mutations: Imaging features and histological subtypes. Eur Radiol. 2016; doi:10.1007/s00330-015-4196-z

92. Gevaert O, Echegaray S, Khuong A, Hoang CD, Shrager JB, Jensen KC, et al. Predictive radiogenomics modeling of EGFR mutation status in lung cancer. Sci Rep. 2017;7: 41674.

93. Yip SS, Kim J, Coroller T, Parmar C, Rios Velazquez E, Huynh E, et al. Associations between somatic mutations and metabolic imaging phenotypes in non-small cell lung cancer. J Nucl Med. 2016; doi:10.2967/jnumed.116.181826

94. Zhu Y, Li H, Guo W, Drukker K, Lan L, Giger ML, et al. Deciphering Genomic Underpinnings of Quantitative MRI-based Radiomic Phenotypes of Invasive Breast Carcinoma. Sci Rep. 2015;5: 17787.

95. Macyszyn L, Akbari H, Pisapia JM, Da X, Attiah M, Pigrish V, et al. Imaging patterns predict patient survival and molecular subtype in glioblastoma via machine learning techniques. Neuro Oncol. 2016;18: 417–425.

96. Li H, Zhu Y, Burnside ES, Drukker K, Hoadley KA, Fan C, et al. MR Imaging Radiomics Signatures for Predicting the Risk of Breast Cancer Recurrence as Given by Research Versions of MammaPrint, Oncotype DX, and PAM50 Gene Assays. Radiology. 2016; 152110.

97. Jajamovich GH, Valiathan CR, Cristescu R, Somayajula S. Integrative analysis of diffusion-weighted MRI and genomic data to inform treatment of glioblastoma. J Neurooncol. 2016;129: 289–300.

98. Yang D, Rao G, Martinez J, Veeraraghavan A, Rao A. Evaluation of tumor-derived MRI-texture features for discrimination of molecular subtypes and prediction of 12-month survival status in glioblastoma. Med Phys. 2015;42: 6725.

99. Wangaryattawanich P, Hatami M, Wang J, Thomas G, Flanders A, Kirby J, et al. Multicenter imaging outcomes study of The Cancer Genome Atlas glioblastoma patient cohort: imaging predictors of overall and progression-free survival. Neuro Oncol. 2015; doi:10.1093/neuonc/nov117

100. Colen RR, Vangel M, Wang J, Gutman DA, Hwang SN, Wintermark M, et al. Imaging genomic mapping of an invasive MRI phenotype predicts patient outcome and metabolic dysfunction: a TCGA glioma phenotype research group project. BMC Med Genomics. 2014;7: 30.

101. Zinn PO, Majadan B, Sathyan P, Singh SK, Majumder S, Jolesz FA, et al. Radiogenomic Mapping of Edema/Cellular Invasion MRI-Phenotypes in Glioblastoma Multiforme. PLoS One. 2011;6: e25451.

PART 2

The Underlying Biology of Radiomics

CHAPTER

2

Defining the biological basis of radiomic phenotypes in lung cancer

Defining the Biological Basis of Radiomic Phenotypes in Lung Cancer
Patrick Grosmann, Olya Stringfield, Nehme El-Hachem, Marilyn M. Bui, Emmanuel Rios Velazquez, Chintan Parmar, Ralph T.H. Leijenaar, Benjamin Haibe-Kains, Philippe Lambin, Robert J. Gillies, Hugo J.W.L. Aerts

ABSTRACT

Medical imaging can visualize characteristics of human cancer noninvasively. Radiomics is an emerging field that translates these medical images into quantitative data to enable phenotypic profiling of tumors. While radiomics has been associated with several clinical endpoints, the complex relationships of radiomics, clinical factors, and tumor biology are largely unknown. To this end, we analyzed two independent cohorts of respectively 262 North American and 89 European patients with lung cancer, and consistently identified previously undescribed associations between radiomic imaging features, molecular pathways, and clinical factors. In particular, we found a relationship between imaging features, immune response, inflammation, and survival, which was further validated by immunohistochemical staining. Moreover, a number of imaging features showed predictive value for specific pathways; for example, intra-tumor heterogeneity features predicted activity of RNA polymerase transcription (AUC = 0.62, p=0.03) and intensity dispersion was predictive of the autodegration pathway of a ubiquitin ligase (AUC = 0.69, $p<10^{-4}$). Finally, we observed that prognostic biomarkers performed highest when combining radiomic, genetic, and clinical information (CI = 0.73, $p<10^{-9}$) indicating complementary value of these data. In conclusion, we demonstrate that radiomic approaches permit noninvasive assessment of both molecular and clinical characteristics of tumors, and therefore have the potential to advance clinical decision-making by systematically analyzing standard-of-care medical images.

ELIFE DIGEST

Medical imaging covers a wide range of techniques that are used to look inside the body, including X-rays, MRI scans and ultrasound. A process called radiomics uses computer algorithms to process the data collected by these techniques to identify and precisely measure a large number of features that would not otherwise be quantifiable by human experts. By doing so, radiomics can automatically measure the radiographic characteristics of a tumor. For example, radiomics can establish the size, shape and texture of a tumor to help to diagnose cancer and guide its treatment.

Research has suggested that radiomics can predict certain clinical characteristics of cancer, such as how far through the body the cancer has spread, how likely it is to respond to treatment, and how likely a patient is to survive. However, these radiomic characteristics have not yet been precisely linked to the biological processes that drive how cancer develops and spreads.

Cancers develop as a result of genetic changes that activate "molecular pathways" in the cells and trigger processes such as cell division and inflammation. To work out exactly which changes are behind a particular tumor, a sample of the tumor from biopsy or surgery is analyzed using genomics techniques. Linking radiomics features to the molecular processes active in a tumor can generate further information that can complement the molecular data. Images are routinely collected on all cancer patients yet molecular data is not. Hence, in some cases, the images can be used to infer the molecular underpinnings of cancer in individual patients.

Grossmann et al. have now analyzed radiomic, genomic and clinical data collected from approximately 350 patients with lung cancer. The analysis revealed links between biological processes normally detected by genomics – in particular, inflammatory responses – and radiomics features. Furthermore, these features could also be associated with clinical characteristics, such as tumor type and patient survival rates. These results were further validated by using a technique called immunohistochemical staining on tumor tissue obtained by surgery.

Further investigation revealed that certain radiomics features can predict the state of molecular pathways that are key to cancer development (such as the inflammatory response). Furthermore, Grossmann et al. found that combining data from radiomics, genomics and clinical parameters predicts how the cancer will progress better than any of these parameters can predict on their own. These results demonstrate the complementary value of radiomic data to genomic and clinical data.

There are many different algorithms that can be used to process images for radiomics. Before radiomics can be used clinically to assess the biological processes underlying the tumors of patients, a specific algorithm needs to be decided upon and then tested in prospective clinical trials.

INTRODUCTION

'Precision medicine' promotes the molecular characterization of a patient's tumor with genomic approaches, which requires tissue extraction usually obtained via biopsy. A number of examples demonstrate successful translation of genomic information obtained from biopsies into clinical applications (Doroshow and Kummar, 2014), but these approaches also have inherent limitations, such as their invasive nature or sampling artifacts caused by intra-tumor heterogeneity (Sottoriva et al., 2013; Fisher et al., 2013; Gerlinger et al., 2012). These limitations can be addressed by medical imaging that has served as crucial diagnostic tool and treatment guidance in clinical oncology. In contrast to biopsies, medical imaging is usually noninvasive, can be applied longitudinally, and provides information about the entire visible tumor volume. In this way, medical imaging has the potential to characterize phenotypic information of tumors and thus complement molecular interrogation (Choi et al., 2016). As imaging is already used routinely throughout the course of treatment this facilitates ready access to this type of data. Therefore, imaging has the potential to serve as valuable diagnostic tool in clinical decision making by complementing biological interrogation or serving as a surrogate in settings where biospecimen-derived diagnostics is not feasible, such as in longitudinal monitoring.

Radiomics is an emerging field that translates these medical images into mineable data by extracting a large number of quantitative imaging features that objectively define tumor intensity, shape, size, and texture (Gillies et al., 2016; Aerts, 2016; Lambin et al., 2012; Kumar et al., 2012) in a robust and reproducible way (Zhao et al., 2016; Fried et al., 2014; Balagurunathan et al., 2014; Leijenaar et al., 2013). As this approach is applied to existing standard of care images, radiomics can be cost-effectively integrated with genomics or serve as surrogate in cases where biopsies are not feasible (O'Connor et al., 2015). Hence, such strategies can be of value for the development of clinical biomarkers for diagnosis, prognosis, and prediction of response to specific treatments (Choi et al., 2016; Huang et al., 2016a, 2016b; Aerts et al., 2016; Nicolasjilwan et al., 2015; Parmar et al., 2015a, Parmar et al., 2015b; Aerts et al., 2014; Chong et al., 2014; Coroller et al., 2015; Gevaert et al., 2012; Ganeshan et al., 2012; Win et al., 2013; Mattonen et al., 2016; Grossmann et al., 2017). Due to the enormous potential for precision medicine, an increasing number of studies have investigated associations between imaging and tumor biology in different cancer types (Aerts et al., 2014; Gevaert et al., 2012; Diehn et al., 2008; Grossmann et al., 2016; Gutman et al., 2015; Segal et al., 2007; Li et al., 2016; Yoon et al., 2015). However, these studies focused on specific genetic associations, or tended to be underpowered due to a limited number of available samples and lacked validation via independent datasets.

Here, we present a broad radiomic-genomic analysis in independent and large cohorts of patients with lung cancer. We rigorously investigated the mechanistic connections between imaging phenotypes and underlying molecular pathways. Furthermore, we validated key associations via immunohistochemical staining and related these associa-

tions to clinical factors. In addition, we developed and validated radiomic predictors of pathway activation status, and investigated the prognostic value of combining radiomic biomarkers with genetic and clinical data. In this study, we aimed at uncovering whether radiomic approaches have the potential to predict both molecular and clinical characteristics of tumors noninvasively and therefore have the potential to augment clinical decision-making using data extracted from standard of care medical images.

RESULTS

To uncover the mechanistic connections between radiomic phenotypes, molecular pathways, and clinical information, we performed an integrated radiomic-genomic analysis of a lung cancer discovery cohort (Dataset1, $n = 262$), and validated our results on an independent validation cohort (Dataset2, $n = 89$). We defined and extracted 636 radiomic features from CT scans (Figure 1A) quantifying tumor intensity, shape, and texture (Figure 1B), detailed in Supplementary file 1. Our approach to integrate radiomic, genomic, and clinical data is outlined in Figure 2 and clinical cohort characteristics are given in Table 1.
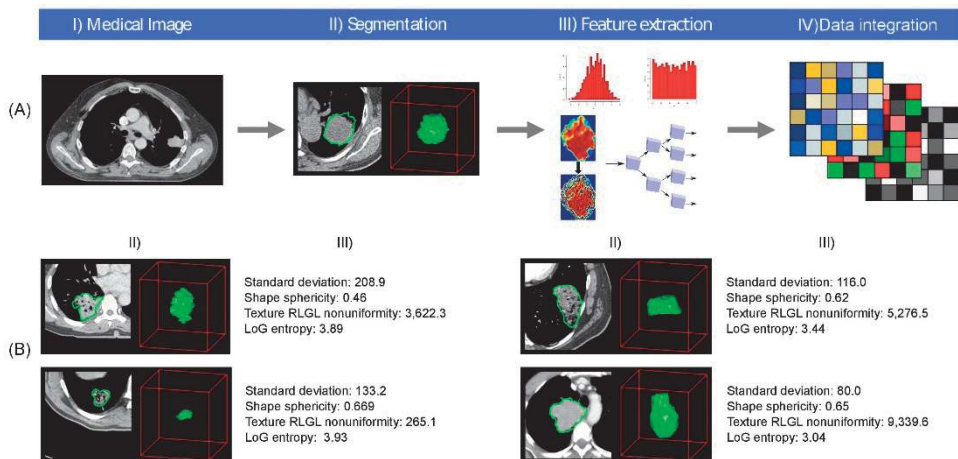
Figure 1

(**A**) Workflow of extracting radiomic features: (I) A lung tumor is scanned in multiple slices. (II) Next, the tumor is delineated in every slice and validated by an experienced physician. This allows creation of a 3D representation of the tumor outlining phenotypic differences of tumors. (III) Radiomic features are extracted from this 3D mask, and (IV) integrated with genomic and clinical data. (**B**) Representative examples of lung cancer tumors. Visual and nonvisual differences in tumor shape and texture between patients can be objectively defined by radiomics features, such as entropy of voxel intensity values ('How heterogeneous is the tumor?') or sphericity of the tumor ('How round is the tumor?').
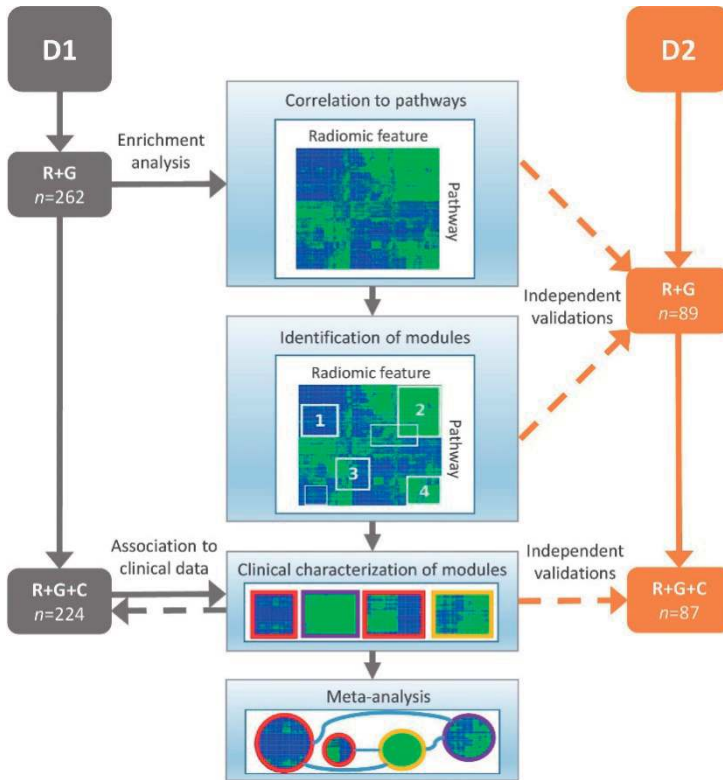
Figure 2

Schema of our strategy to define robust radiomic-pathway-clinical relationships.

Two independent lung cancer cohorts (D1 and D2) with radiomic (R), genomic (G), and clinical (C) data were ana-lyzed. D1 (n = 262) was used as a discovery cohort and D2 (n = 89) was used to validate our findings. A gene set enrichment analysis (GSEA) approach assessed scores for radiomic-pathway associations. These scores were bi-clustered to modules that contain features and pathways with coherent expression patterns. These modules may overlap and vary in size. Clinical association to overall survival (red), pathologic histology (purple), and TNM stage (yellow) was statistically tested in both datasets, and results were combined in a meta-analysis to investigate relationships of modules.

Association modules of radiomic features and molecular pathways

To investigate the main associations of radiomics and underlying molecular pathways, we developed association modules describing radiomic-pathway coherency. Bi-clustering al-lowed simultaneous grouping of coherently expressed features and pathways into a single module, thereby reducing dimensionality. Using this approach, we identified thirteen ra-diomic-pathway modules in Dataset1 that were independently validated in Dataset2 (FDR < 0.05). Figure 3A and Table 2 summarize these modules, while a detailed version of every module is given in Figure 3—source data 1.

Table 1
Proportions of clinical characteristics in Dataset1 and Dataset2, Figure 2. Histology and TNM stage were based on pathology were available.

	Dataset1	Dataset2
Gender		
Male	100 (45%)	59 (68%)
Female	124 (55%)	28 (32%)
Histology		
Adenocarcinoma	129 (58%)	42 (48%)
Squamous	61 (27%)	33 (38%)
Other	34 (15%)	12 (14%)
Stage		
I	123 (55%)	39 (45%)
II	35 (15%)	26 (30%)
III	46 (21%)	12 (14%)
Other	20 (9%)	10 (11%)
Smoking Status		
Current	66 (29%)	NA
Former	141 (63%)	NA
None	17 (8%)	NA
Tumor site		
Primary	224 (100%)	87 (100%)
Endpoints		
Overall survivals	134 (60%)	41 (47%)
Overall deaths	90 (40%)	46 (53%)
Follow up (median months)	32	31
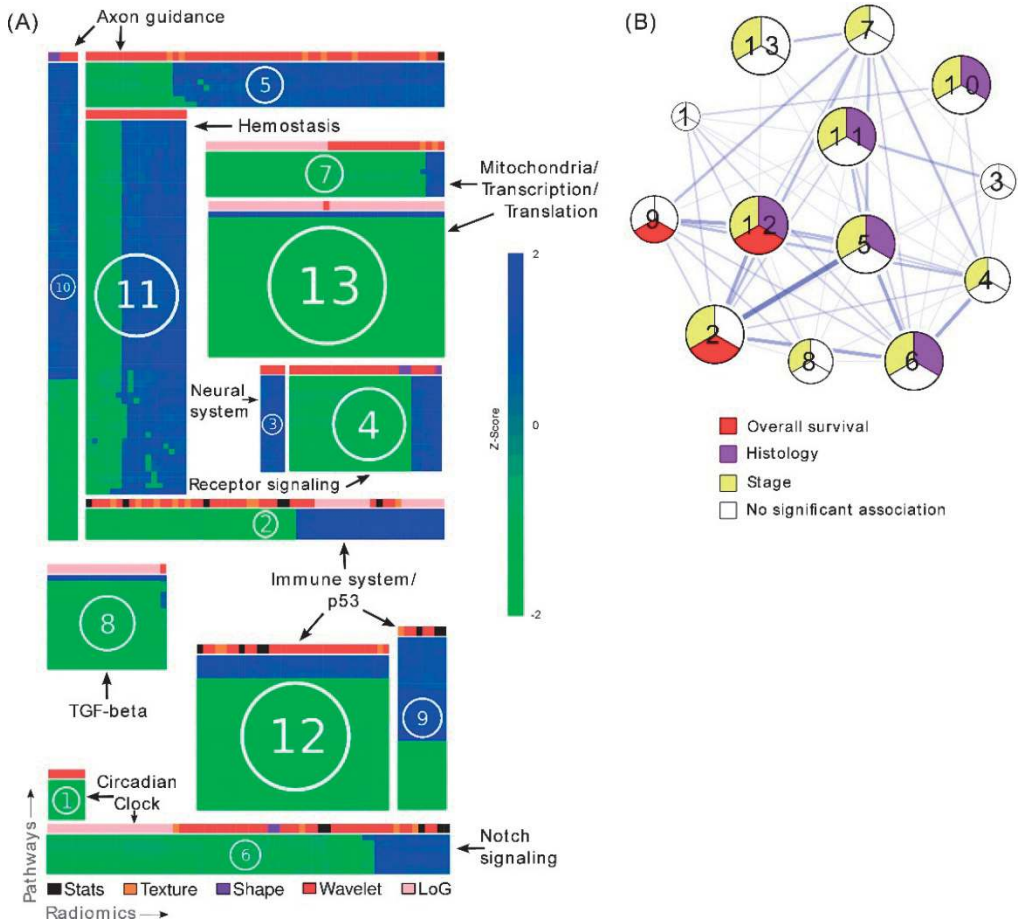
Figure 3

(***A***) *Clustering of significantly validated radiomic-pathway association modules (FDR < 0.05). Normalized enrich-ment scores (NESs) have been biclustered to coherently expressed modules. Every heatmap in this figure corre-sponds to a module (M1 - M13) with radiomic features in columns and pathways in rows. Heatmap sizes are pro-portional to module sizes. Elements are NESs given in Z-scores across features, and are displayed in blue when positive and green when negative. Horizontal color bars above every module indicate radiomic feature groups (black = first order statistics, orange = texture, purple = shape, red = wavelet, and pink = Laplace of Gaussian). Representative molecular pathways are displayed. (***B***) Clinical module network. We investigated if modules were associated with overall survival (red), stage (yellow), histology (purple), or no clinical factor (white). Relationships of modules based on their number of shared radiomic features (thickness of blue lines) are displayed by a net-work. While we found that most modules yield clinical information, overlaps of modules did not indicate relation-ships to similar clinical factors.*

Table 2

Summary of common themes in all of the identified radiomic-pathway association modules. Columns 1–3 display the module name, the number of radiomic features (nr), and pathways (np), respectively. Columns 4–5 hold the radiomic and pathway themes present in each module.

Module	nr	np	Radiomic	Pathway
M1	6	7	Wavelet texture gray-level runs	Lipid and lipoprotein metabolism, Notch signaling, circadian clock
M2	58	5	Wavelet intensity entropy; Laplace of Gaussian intensity standard deviation	Immune system, p53
M3	4	17	Wavelet minimum intensity	Neural system, axon guidance
M4	25	14	Intensity variance and mean; wavelet minimum intensity min	Biological oxidations, signaling by insulin receptor, signaling by GPCR, neuronal system
M5	58	8	Wavelet texture gray-level runs; wavelet intensity range and median; (wavelet) texture information correlation and cluster tendency	Axon guidance and synaptic transmission, lipoprotein metabolism, cell type determination
M6	64	7	Laplace of Gaussian standard deviation; wavelet texture gray-level runs; wavelet texture cluster tendency	Circadian clock, signaling by Notch
M7	39	8	Laplace of Gaussian intensity entropy; wavelet intensity variance; Laplace of Gaussian texture information correlation	Mitochondria, Pol III transcription
M8	20	17	Laplace of Gaussian standard deviation	TCA cycle and electron transport, TGF-beta receptor signaling, response to stress, transcription regulation, protein synthesis,
M9	8	30	Intensity variance; wavelet intensity variance	Immune system, p53, cell cycle regulation checkpoints, cell-cell interaction, circadian clock
M10	5	83	Shape surface (SH); wavelet texture gray-level runs	Axon guidance, neuronal system, (innate) immune system, hemostasis, FGFR signaling, TGF-beta receptor signaling, Notch signaling, circadian clock
M11	17	66	Wavelet intensity range; wavelet texture information correlation	Hemostasis, neural system
M12	32	27	Wavelet texture entropy; intensity variance; wavelet texture cluster tendency	P53, immune system
M13	39	26	Intensity entropy	Gene expression regulation, Pol II/III transcription

In general, we found that distinct radiomic features were associated with distinct biological processes. For example, texture entropy and cluster features, as well as voxel intensity variance features were associated with the immune system, the p53 pathway, and other

pathways involved in cell cycle regulation in modules M2, M9, and M12 (Table 2 and Figure 3A). In another module (M8), we found those features to also be associated with transforming growth factor beta (TGF-β) receptor signaling.

Further examples for radiomic-pathway links included two modules (M13 and M7) that were highly enriched for pathways involved in mitochondrial pathways, transcription, translation, and RNA regulatory mechanisms; with only one exception, all features in the larger module (M13) were voxel intensity entropy features. In addition to this feature type, the smaller module (M7) contained mainly textural variance and information correlation features.

Table 3

Pathway prediction and clinical association. For every module, the independent validation performance of the strongest radiomic based pathway predictors is indicated per module by the area under the curve (AUC) of the receiver operator characteristic. In addition, we highlight whether a module was significantly associated with overall survival (OS), TNM stage (ST), or pathologic histology (HI) (p<0.05).

Module	Strongest radiomic based pathway prediction	AUC	OS	ST	HI
M1	Wavelet (HHH) texture (GLCM) correlation → Cholesterol biosynthesis	0.64, *p*=0.014			
M2	Laplace of Gaussian intensity standard deviation → Autodegration of the E3 Ubiquitin ligase COP1	0.69, *p*=8e-4	x	x	
M3	Wavelet minimum intensity → Trafficking of GLUR2 containing AMPA receptors	0.67, *p*=0.003			
M4	Wavelet intensity minimum → Glutathione conjugation	0.68, *p*=9e-4		x	
M5	Texture information correlation → Trafficking of GLUR2 containing AMPA receptors	0.69, *p*=7e-4		x	x
M6	Wavelet texture cluster prominence → Notch1 intracellular domain regulation of transcription	0.66, *p*=0.007		x	x
M7	Laplace of Gaussian intensity entropy → RNA polymerase III transcription	0.62, *p*=0.031		x	
M8	Laplace of Gaussian intensity standard deviation → Pyruvate metabolism and citric acid TCA cycle	0.72, *p*=6e-5		x	
M9	Wavelet intensity variance → Trafficking of GLUR2 containing AMPA receptors	0.64, *p*=0.020	x		
M10	Shape compactness and shape sphericity → TRAF6 mediated NFkB activation	0.66, *p*=0.003		x	x
M11	Wavelet texture cluster tendency → Platelet aggregation plug formation	0.69, *p*=6e-4		x	x
M12	Wavelet texture entropy → G0 and early G1	0.65, *p*=0.007	x	x	x
M13	Laplace of Gaussian intensity entropy → RNA polymerase II transcription initiation and promoter opening	0.68, *p*=0.001		x	

Clinical information contained in modules

For every module, we assessed prognostic association to overall survival (OS) and associations to stage and histology based on the radiomic features of a module (Figure 3B and Table 3). Three modules (M2, M9, and M12) were significantly prognostic for OS (p<0.02), ten modules (M2, M4-8, and M10-13) were significantly associated with stage (p<0.01), and five modules (M5, M6, and M10-12) were significantly associated with histology (p<0.05). The exact p-values of all modules are given in Supplementary file 2.

We examined and summarized the relationships of clinical status, module size, and overlap of modules in a network (Figure 3B and Table 3). We found that smaller modules tended not to be associated with the tested clinical factors. The total number of shared features or pathways was generally low (mean Jaccard index 0.22, range [0.01, 0.59]). Interestingly, certain modules with higher overlap still showed different clinical associations.

Radiomic predictors of pathway status

To test whether radiomic features can predict if a pathway is activated or deleted in individual patients, we fitted univariate models of radiomic features on Dataset1 and selected for every module the strongest predictor in Dataset1 according to the area under the curve (Fawcett, 2006) (AUC) for validation in Dataset2. As shown in Table 3 and Table 3—source data 1, the overall biological and radiomic themes in a module were well represented by these individual predictors. For example, a Laplace of Gaussian intensity standard deviation feature was predictive of the autodegration pathway of the E3 ubiquitin ligase COP1 (AUC = 0.69, $p<10^{-4}$) in module M2, which was also associated with p53. Importantly, COP1 mediates p53 and may interact with autophagy (Rabbani et al., 2014; Kobayashi et al., 2013), which are known drivers of tumorigenesis. Indeed, this module M2 was associated with OS. We found further examples of this radiomic-genomic-clinical link to be important: For example, a texture feature (information correlation) predicted trafficking of GLUR2 containing AMPA receptors (AUC = 0.69, $p<10^{-4}$) in module 5, which was associated with lipoprotein metabolism and stage. Further, two shape features (sphericity and compactness) predicted TRAF6 mediated NFkB activation (AUC = 0.66, p=0.003) in module 10, which was also associated with axon guidance and histology.

Furthermore, we assessed these representative features in terms of their predictive value for driver mutations in the discovery cohort; based on a subset of 60 patients whose tumors were profiled with Sanger sequencing, we estimate that the prevalence of mutated EGFR, KRAS, and TP53 are 15%, 35%, and 20%, respectively. In particular, we found strong performance for mutations in EGFR and KRAS by several features, but only one considerable performance for TP53 (Figure 3—figure supplement 1). Interestingly,

predictive value for EGFR and KRAS were selective in that features had relatively high performance for one gene but not both. Predictive power for smoking history was low to moderate (Figure 3—figure supplement 2).

Immunohistochemical investigation

To further investigate putative connections between radiomics, immune response pathways, and OS we performed immunohistochemical staining of 22 tumors for CD3, a T-cell co-receptor. These tumors were predicted to show relatively high or low immune response by a radiomics feature selected from the three modules (M2, M9, and M12) that were associated with OS. As represented in Figure 4, we found agreement between radiomics and pathology; cases that were pathologically scored to have high CD3 enrichment also expressed significantly higher radiomic values (one-sided Wilcoxon rank sum test, p=0.008). Furthermore, we tested the extent to which radiomic predictors of inflammation can be reproduced immunohistochemically. We built on our previous results suggesting that the radiomic shape feature sphericity predicts NFkB activation (module 10) and analyzed 24 stained tumors that were predicted to have relatively high or low NFkB activity for RelA, the p65 subunit of NFkB (Figure 4—figure supplement 1). Pathological assessment of enrichment for RelA revealed that those cases that indicated high RelA enrichment on average also had higher radiomic feature scores (one-sided Wilcoxon rank sum test, p=0.06).

Prognostic value of radiomic signatures

To build on previously published results, we investigated prognostic value of an existing radiomic signature for survival of lung cancer. We fitted a Cox proportional-hazards model of this signature on Dataset1 and observed significant validation by the concordance-index (CI) on Dataset2 (CI = 0.60, Noether p=0.04). Furthermore, we tested combinations of clinical, genetic, and radiomic data and observed that the combinations of data types tended to result in higher performances than given by the individual data alone (Figure 5). In particular, the performance of a clinical model increased from CI = 0.65 (Noether p=0.001) to CI = 0.73 (p=2×10^{-9}) when adding the radiomic and an existing gene signature (38Hou et al., 2010); this increase was significant at p=0.001 by permutation test. This combined radiomic-genetic-clinical model also performed significantly better than the combined radiomic-clinical model (p=0.007) and the clinical-genetic model (p=0.01). Adding radiomics to clinical data alone did not result in a significant increase (p=0.3). We repeated this analysis with a novel radiomic survival signature and other published gene signatures (Yuan et al., 2004; Chen et al., 2007; Hsu et al., 200939–41), and found that the clinical-genetic-radiomic models consistently yielded the highest performances in nearly all cases (Figure 5—figure supplement 1 and Figure 5—figure supplement 2).
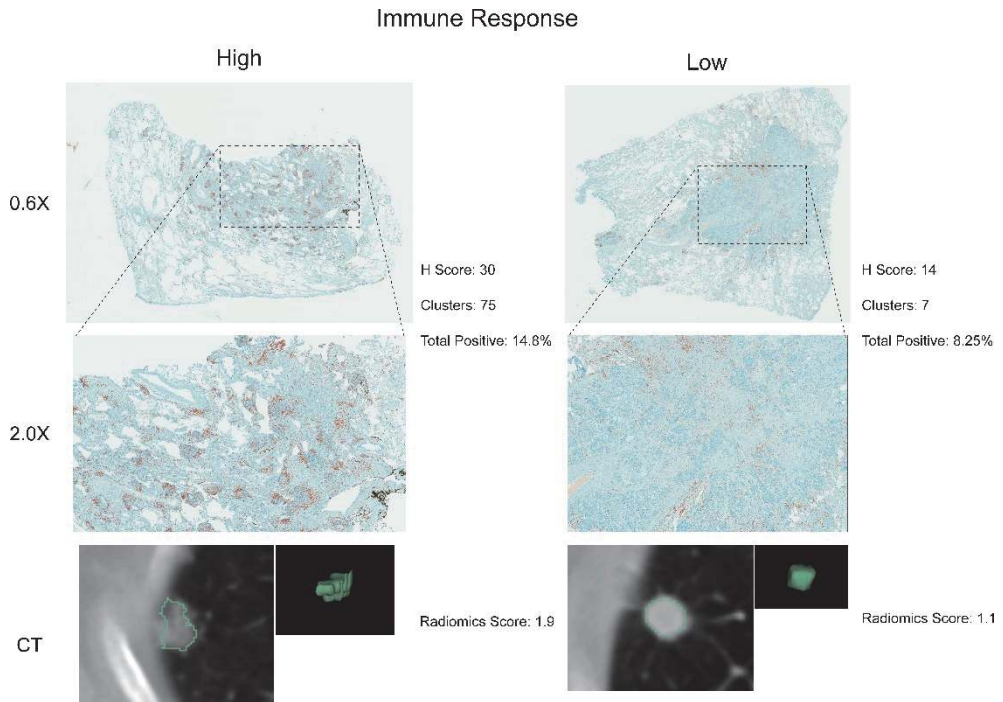
Figure 4
Test for agreement between radiomic and pathological immune response assessment.
Two representative cases are shown where radiomic predictions of immune response were confirmed by im-munohistochemical staining for nuclear CD3 highlighting lymphocytes in brown. Each case is displayed in 0.6X and 2.0X magnification of the tumor slides, and an axial slice of the corresponding diagnostic CT scan and the total tumor volume is given for comparison. Automated quantifications of lymphocytes are displayed in addition to the radiomics score incorporated to classify into high and low responders.

DISCUSSION

Medical imaging plays a crucial role in cancer diagnosis, treatment, and response monitor-ing. Radiomics allows quantification of the radiographic phenotype of a tumor (Kuo and Jamshidi, 2014; Gillies et al., 2010; Rutman and Kuo, 2009), but the underly-ing connections of radiomics to tumor biology and clinical factors have not been elucidat-ed yet. In this study,

Figure 5

Combining prognostic signatures for overall survival.

We tested combinations of clinical, genomic, and radiomic signatures. To a clinical Cox proportional-hazards regression model with stage and histology, we first added a published gene signature and next a published radiomic signature. These models were fitted on Dataset1 and evaluated with the C-index (CI) on Dataset2. An asterisk indicates significance (p<0.05). Combining different data types resulted in increased prognostic performances. By adding radiomic and genomic informati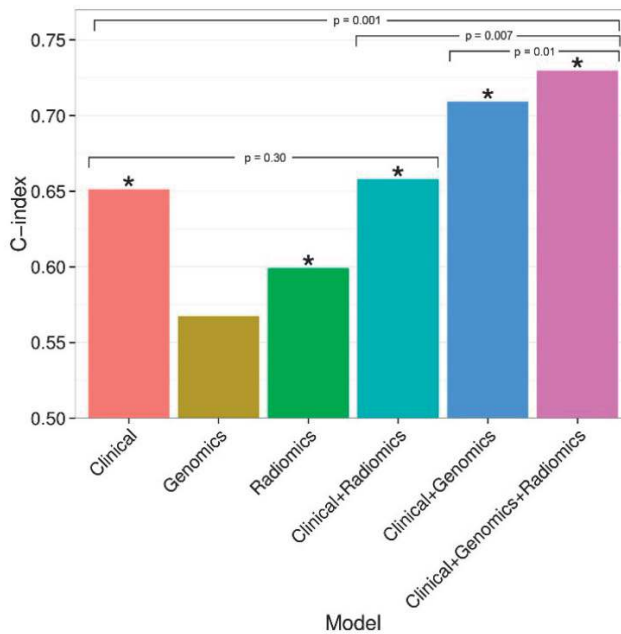on, the initial performance of the clinical model was increased from CI = 0.65 (Noether p=0.001) to CI = 0.73 (p=2×10^{-9}).

we identified novel and consistent associations between radiomic phenotype data, underlying molecular pathways, and clinical factors of patients with lung cancer in a North American cohort, and validated our findings in a European cohort and with immunohistochemical staining. In addition, we presented radiomic predictors for pathway activations, and demonstrated the complementary prognostic value of combining radiomic, genetic, and clinical information.

Preliminary studies have previously investigated associations between imaging features, clinical factors, and molecular data for a number of cancer types as outlined in recent reviews (Gillies et al., 2016; Kuo and Jamshidi, 2014; Gillies et al., 2010; Rutman and Kuo, 2009; Cook et al., 2013). Our analysis builds on these studies in that we performed a rigorous classification of a comprehensive set of radiomic features in terms of underlying molecular pathways on a genome-wide scale and clinical factors in large and independent cohorts. Although the long-term vision is to augment clinical decision making, the current goal of our study is to satisfy the need of the radiomic and oncological community to better understand the underlying biological rationale of radiomic predic-

tions. Furthermore, we are the first to publicly share all study data and analysis code with the growing radiomic and biomedical community to enable further translational research.

We identified and independently validated thirteen radiomic-pathway modules with coherent expression patterns, eleven of which were significantly associated with OS, stage, or histology. By basing these clinical associations exclusively on radiomic features, we could demonstrate that the associated molecular pathways robustly matched radiomic- based hypotheses. For example, based on radiomic features modules M2, M9, and M12 were prognostic and also associated with stage. These modules were highly enriched for immune system, p53, and cell-cycle regulation pathways, biological processes that are widely recognized to play key roles in lung cancer. For example, it has been established that cell cycle regulation is of utmost importance in lung cancer (Baldi et al., 2011). Furthermore, the status of p53 is reported to be a predictor of survival in lung cancer patients (Ahrendt et al., 2003) and a recent review has laid out how p53 can modulate innate immune system responses (Menendez et al., 2013). Radiomic features in these prognostic modules M2, M9, and M12 quantified textural entropy and dispersion image intensity values suggesting associations between textural heterogeneity, cell cycling, and prognosis. Therefore, these results suggest that noninvasive radiomic surrogates may benefit diagnostic methods in assessing cell cycling and immune system states of tumors.

We aimed at confirming our statistical results indicating connections between radiomics, immune response, and survival by immunohistochemical staining of lymphocytes in cases for which a relatively high or low immune response was predicted according to a radiomics score. We generally found high agreement between pathology and radiomics, especially in cases where immune response was predicted to be high. In cases of predicted low responders that showed high pathological immune response, the cause of disagreement may be a heavy distribution of CD3 clusters in the extreme periphery of the tumor with very little staining in the bulk of the tumor. In cases of predicted high responders that showed little to no immune response, this could be due to the lack of normal tissue margin around the edge of the tumor section or a sampling effect. Similarly, we stained tissue for RelA, the p65 subunit of NFkB, to validate radiomic predictions of inflammation. Overall, we found high agreement between pathology and radiomics, although at lower statistical significance. Future studies with whole mount sections stained with multi-plex phenotyping can help determine the relationship between a radiomic and a genetic immune or inflammation signature, and the gold standard.

A variety of textural features were also associated with stage and histology (module M5). Similar associations have been reported by Ganeshan et al. (2010), who suggested that 2D texture features of lung cancer CT scans could predict if tumor stage was II or above. Here, we found that texture features were enriched for axon guidance and lipoprotein metabolism. Furthermore, we observed strong associations between image intensity entropy features and pathways involved in gene expression, transcription regulation, and mitochondrial processes (M13 and M7). Previous research has suggested that imaging can detect consequences of an increase in the hypoxia-inducible

factor as a result of absence of oxygen (Gillies et al., 2010). Hence, if extracting quantitative information about mitochondrial pathways from medical images leads to assessment of hypoxia status of a tumor, this may ultimately aid in clinical decision-making as alternative therapies for hypoxic tumor areas are being developed (Denny, 2010; Bryant et al., 2014). Indeed, previous work has indicated that CT pixel intensities correlate with hypoxia markers such as Glut-1 and pimonidazole (Ganeshan et al., 2013). Those two modules (M1 and M3) that were not associated with any of the tested clinical factors were relatively small modules; these modules suggested radiomic associations to circadian clock and neural system. The impact of these pathways to the clinical factors we tested is not apparent from current lung cancer literature, which could explain why these modules did not show clinical associations.

Our results further suggest that radiomic approaches could have the potential to predict molecular states of pathways. We found that the highest predictors of every module was also a suitable representative of the overall biological and radiomic themes of that module. Amongst these examples of pathways that showed high predictability in terms of radiomics, we found various pathways essential for tumorigenesis such as cell cycle pathways (e.g., G0 and early G1), signaling pathways (e.g., Notch and NfKB), and tumor suppressor pathways (e.g., COP1 autodegration and p53). Furthermore, we tested those radiomic pathway predictors for predictive value of driver mutations. Thereby, the highest performances were found for mutations in EGFR and KRAS, which is in line with current radiomic-genetic literature (Aerts et al., 2016; Gutman et al., 2015; Liu et al., 2016; Rizzo et al., 2016). Interestingly, however, the highest performance for the tumor suppressor and cell cycle regulator TP53 we found was given by a textural entropy feature that also predicted G0 and early G1 (module M12). In addition, features expressed selectivity for predicting mutations, which was suggested previously (Gutman et al., 2015). These results highlight the diagnostic potential, as ready information on pathway and mutation status may permit advanced patient stratification. Previous studies have indicated that gene expression can be predicted by imaging features (Gevaert et al., 2012; Segal et al., 2007; Gevaert et al., 2014). To our knowledge, however, no study has examined and independently validated radiomic models for specific pathways, including biological validation such as immunohistochemical staining.

Finally, we verified a previously described prognostic radiomic signature and observed that the best performance is achieved when combining radiomic, genetic, and clinical data. These results strongly suggest that radiomic data contain complementary prognostic information and are robust, as the published radiomic signature (Aerts et al., 2014) has not been tested on our data before. Notably, these prognostic improvements were relatively stable to substitution of radiomic or gene signatures. A related indication of improved survival predictions by combining imaging features and molecular data has been recently given for glioblastoma, however without validation (Nicolasjilwan et al., 2015). It is worth noting that for the first time we also demonstrate that radiomic prognostication generalizes across cohorts from different continents.

Three research tracks have recently been proposed for clinical translation of such imaging biomarkers (O'Connor et al., 2017), including biological validation, technical validation, and evaluation of cost-effectiveness. Our study conforms with several of these roadmap recommendations by advancing results on a previously proposed radiomic signature (Aerts et al., 2014) with additional biological validation and investigations on how genetic data and clinical factors impact this signature. Fixing a radiomic signature for technical validation and cost-effectiveness verification should be considered in subsequent studies to overcome additional translational gaps. Although the long-term vision would be to augment clinical decision making, the current goal of our study is to contribute in satisfying the need of the radiomic and oncological community to understand the underlying biology of radiomic predictions.

Our study is limited by its retrospective nature. Imaging protocols are not standardized and hence variability in CT acquisition and reconstruction parameters is inherent in clinical practice. However, despite this, no corrections by cohort or scanner type were made in this study illustrating the translational aspect of our results that generalized across institutions. Hence, we expect that the performance of radiomics will further improve, as imaging data are becoming more standardized. In fact, multiple studies have already documented the robustness of radiomic feature extractions in terms of reproducibility and repeatability in test/re-test settings (Fried et al., 2014; Balagurunathan et al., 2014; Leijenaar et al., 2013; Parmar et al., 2015a; Aerts et al., 2014; Grove et al., 2015). Another limitation of this study is that the current cohorts mainly focused on early stage (I - III) tumors, hence generalization of radiomic-genomic associations to late stage tumors should be drawn with precaution only. However, most radiomic applications do focus on early stage tumors as the current radiomic approach requires segmentation of tumors which for late stage tumors remains to be of particular complexity. Furthermore, although our study provides multiple facets of validation, immunohistochemical validation was restricted to considerably smaller sample sizes as compared to our statistical validations due to limited availability of frozen tissue. Prospective protocols can ensure availability of sufficient tissue for additional validation.

Biological material investigated in this study has been acquired by single-needle biopsies, thus the interpretation of our genomic data is limited due to heterogeneity of lung cancer tumors. However, as our results validated in independent data and because known drivers of tumorigenesis were among the main pathways found to be associated with radiomic features, this suggests that these associations have been established in an early evolutionary step in tumorigenesis and are therefore reasonable representatives of the overall tumor. Prospective studies with defined spatial matchings of biopsies and/or single cell analyses could provide deeper insight into whether the strengths of these associations can be further increased. Prospective studies will also be required to assess clinical utility of combining radiomic, genomic, and clinical data into prognostic models.

In conclusion, this study presented novel and consistent associations between radiomics, molecular pathways, and clinical factors. We applied an independent discovery

and validation design on large patient cohorts from different continents with enough variability that allowed confidence in the generalization of our results. Furthermore, we performed biological validation and demonstrated that radiomics predicts molecular pathway status and thus improves the prognostic performances of clinical and gene signatures. The clinical impact of our results is illustrated by the fact that it advances the molecular knowledge of automated radiomic characterization of tumors, information currently not used clinically. This may provide opportunities to improve decision-support at low additional cost as imaging is routinely used in clinical practice as standard of care.

MATERIALS AND METHODS

Discovery and validation data

Data underlying this study is made publically available with this article. We analyzed two cohorts of patients with non-small cell lung cancer (NSCLC), Dataset1 and Dataset2, each consisting of pretreatment diagnostic computed tomography (CT) scans, gene expression profiles, and clinical data. While the larger cohort Dataset1 (North American) is novel and served as a discovery cohort, Dataset2 (European) has been previously published with CT scans and gene expression data (Aerts et al., 2014), and was used for independent validation of our findings. Patients in Dataset1 were treated in the Thoracic Oncology Program at the H. Lee Moffitt Cancer Center, Tampa, Florida, USA; we included patients with diagnosed primary tumors who underwent surgical resection and collected contrast-enhanced CT scans obtained within 60 days of the diagnosis between years 2006 and 2009. Patients in Dataset2 were treated at MAASTRO clinical, Maastricht, NL; we included patients with confirmed primary tumors who received surgery. Further details of Dataset2 are given by Aerts et al. (2014). The majority of CT scans were recorded to be contrast-enhancing (89%% and 71% of patients in Dataset1 and Dataset2, respectively).

      For analyses involving CT scans and gene expression data, 262 and 89 patients were available for Dataset1 and Dataset2, respectively. In addition, clinical data were available for 224 and 87 patients, respectively. Clinical outcomes investigated were overall survival (OS), pathologic TNM stage (combined T, N, and M stages, according to the latest version 7 of the IASLC guideline for lung cancer [Mirsadraee et al., 2012]), and pathologic histology (grouped into adenocarcinoma, squamous carcinoma, and others). Clinical stage and histology were used when pathologic information was not available. Tumors in these cohorts were mainly early stage; in Dataset1 among the 224 clinically annotated cases 26 were stage IIIB or IV and in Dataset2 among the 87 clinically annotated cases 3 cases were stage IV. These late stages have been grouped into 'other' for analysis. Further clinical cohort characteristics are given in Table 1.

For tumors in both cohorts, expression of 60,607 probes was measured on a custom Rosetta/Merck Affymetrix 2.0 microarray chipset (HuRSTA_2a520709.CDF, GEO accession number GPL15048) by the Moffitt Cancer Center. Gene expression of Dataset2 is available also at Gene Expression Omnibus (GEO) through accession number GSE58661. Gene expression values were normalized with the robust multi-array average (RMA) algorithm (Irizarry et al., 2003) implemented in the 'affy' Bioconductor package (Gautier et al., 2004). Probes have been curated by choosing the most variant representative among probes mapping to the same gene identifier (Entrez Gene) resulting in a total of 21,766 unique genes.

Radiomic features

We extracted 636 features grouped into I) tumor intensity (voxel statistics), II) shape, III) texture, IV) wavelet, and V) Laplace of Gaussian features. Group I-IV features have been defined as specified by Aerts et al. (2014). In addition, we added new features to Group III (see GLSZM below). *Group I* features are first-order statistics (e.g. mean, skewness) of all voxel intensity values in the tumor volume mask. *Group II* features describe the shape and size of a tumor (e.g. compactness). *Group III* features quantify texture in tumor images describing clustering of voxels with similar appearance by means of a gray-level co-occurrence matrix (GLCM), a run-length gray-level matrix (RLGL), or a gray-level size-zone matrix (GLSZM). These features quantify how frequent voxels of same gray-level are adjacent to each other (GLCM), how many voxels of the same gray-level appear in a consecutive run (RLGL), or the sizes of flat zones, areas of same gray-level in all directions (GLSZM). *Group IV* features are Group I-III features (except GLSZM) assessed after a wavelet decomposition of the image, which highlights sharp transitions in the intensity frequency spectrum. *Group V* consists of Group I features that have been calculated after applying a Laplace of Gaussian transformation to the image, which highlights edge structures. Detailed description and analytical definitions of the features added to the Aerts et al. (2014) feature set (n = 440) are given in Supplementary file 1. Features were calculated in 3D. For normalization, slice thicknesses of all scans were interpolated to a voxel sizes of $1 \times 1 \times 1$ mm^3.

Pathway analysis

To test if a radiomic feature was associated with a molecular pathway, Spearman's rank correlation coefficient *rho* was calculated for the expression of every gene across all patients and weighted by -log10(*p)*, where *p* is the p-value of *rho*. The resulting gene rank was input to a preranked gene set enrichment analysis (GSEA) algorithm (Subramanian et al., 2005) version 2.0.14 on the C2 collection version 4 of the Molecular Signature Database (MSigDB) (Liberzon et al., 2011). This collection contains the expert-curated set of pathways from the Reactome database (Joshi-Tope et al., 2005). Those 511 out of 674 pathways were considered that contained at least 15 and at most 500 genes. GSEA reports normalized enrichment scores (NESs) for every pathway, which we further analyzed.

Radiomic-pathway association modules

To identify coherently expressed expressed features and pathways, a matrix holding an NESs for every pair of radiomic feature and Reactome pathway was biclustered with the Iterative Signature Algorithm (ISA) using the 'isa2' and 'eisa' packages in R and Bioconductor (Bergmann et al., 2003; Csárdi et al., 2010). As a result, each bicluster contains a set of coherently expressed features and pathways and is referred to as module. Potential module redundancy was limited using the 'isa.unique' function in the 'isa2' package with a maximum correlation threshold of 0.3. To avoid parameter sensitivity with ISA, row and column clustering seed thresholds were set to a liberal sequence of 1.5 to 2.5 by 0.5 to include all potential signals. This procedure yielded 20 putative modules. To validate these modules, we developed and applied a correlation based statistic $r:=mean(C_X)+mean(C_Y)$, where C_X and C_Y are the Spearman rank correlations of all pairs of features and pathways in a module, respectively. The true r was calculated for every module in Dataset1 and validated on Dataset2 with random permutation tests (N = 1000). After correcting for multiple-hypothesis testing with the false-discovery-rate (FDR) (Benjamini and Hochberg, 1995), the validation resulted in 13 significantly enriched modules (FDR < 0.05). In total, the modules captured the associations between 210 radiomic features and 206 pathways.

Module size was defined as $n/N + m/M$, where n and m are the number of features and pathways in a module, respectively, and $N = 636$ and $M = 511$ are the total numbers of features and pathways across all modules, respectively. Overlap of two modules was defined by the Jaccard index (Theodoridis and Koutroumbas, 2008), which is the size of union of features divided by the size of intersection of features of two module. Hereby, same feature names under different transformations were considered equivalent.

Pathway predictions

To test radiomic pathway predictors, we used gene set variation analysis (GSVA) in Bioconductor (Hänzelmann et al., 2013) to calculate pathway enrichment scores per patient. Next, we fitted univariate logistic regression models of every feature to predict the NES sign of pathways (which corresponded to activation or deletion) in Dataset1. We assessed the concordance between the predicted probabilities of the pathway sign and the true sign with the area under the curve (AUC) of the receiver operator characteristic (ROC) (Bradley, 1997). The strongest predictor of each module according to the AUCs in Dataset1 was evaluated on Dataset2 for validation; significance of AUCs was calculated according to Noether for binary outcomes (Pencina and D'Agostino, 2004).

Associations to clinical factors

Associations to OS were assessed by calculating the mean concordance-index (Harrell et al., 1982) of all features in a module univariately using the 'survcomp' package in Bioconductor (Schröder et al., 2011), and by validating this statistic with repeated random permutation tests (N = 1000). Similarly, associations to stage and histology were assessed by the mean of Kruskal-Wallis chi square statistics and permutation tests. As clinical information was not part of the module identification process, a meta-analysis of the results in Dataset1 and Dataset2 was conducted to account for sample size differences and other dataset specific variations. For this, a Fisher Z-transformation (Whitlock, 2005) of the independent p-values in both datasets was employed for every module with weights equal to the respective sample sizes in Dataset1 and Dataset2.

We tested additive prognostic effects of integrating radiomic, gene expression, and clinical data by combining in a Cox proportional-hazards model the predictions of (I) a clinical Cox model with stage and histology, (II) an NSCLC OS gene signature, and (III) an NSCLC OS radiomic signature. We tested five published gene signatures (Hou et al., 2010; Yuan et al., 2004; Chen et al., 2007; Hsu et al., 2009) without inclusion of clinical and radiomic data and retained the strongest performing signature by Hou et al. (2010) to challenge potential performance increases. To test for generalizability of radiomics, we tested a published radiomic signature by Aerts et al. (2014) and a novel signature developed in the current study. We developed this novel radiomic signature using a supervised feature selection algorithm followed by a stepwise Cox regression approach on Dataset1: First, we employed the minimum-redundancy maximum-relevance (mRMR) algorithm implemented in the 'mRMRe' R package (De Jay et al., 2013) on all radiomic features with respect to OS to select a non-redundant, highly informative ranked set of complementary features. Next, we trained Cox models incrementally, adding features starting by the highest ranked feature. We performed repeated random cross-validation (N = 1,000) to measure the performance of each model, and retained the model with the highest mean CI. Finally, these fitted models were tested on Dataset2 for validation.

All statistical analyses were carried out using the R software (R Development Core Team, 2013) version 3.1.0 on a Linux operating system. Details of version numbers of utilized packages are available in Supplementary file 2.

Immunohistochemical staining for CD3

We selected 25 cases each that were predicted to have high and low immune response by using the value of the radiomic feature in the prognostic modules M2, M9, and M12 that showed the highest absolute correlation to the mean expression of genes in the CTLA4 inhibitory pathway that is supported to be associated with immune activity (Postow et al., 2015; Pardoll, 2012; Wolchok and Saenger, 2008). In total, 22 cases were available with enough tumor tissue and sufficient staining quality. Tumor cross section slides were stained using a Ventana Discovery XT automated system (Ventana Medical Systems, Tucson, AZ) as per manufacturer's protocol with recommended reagents. Briefly, slides were deparaffinized with EZ Prep solution (Ventana) and a heat-induced antigen retrieval method was used under mild cell conditioning using CC1 antigen retrieval buffer (Ventana). A rabbit primary antibody for CD3, (790–4341, Ventana) was used at supplied concentration and incubated for 16 min. Next a Ventana OmniMap Anti-Rabbit Secondary Antibody was applied to the samples for 16 min and the Ventana ChromoMap kit was used as the detection system. Slides were then counterstained with Hematoxylin and dehydrated. Finally, the slides were cover slipped as per normal laboratory protocol.

Immunohistochemical staining for RelA

We selected 25 cases each that were predicted to have high and low NFkB activity. The same procedure as for the CD3 staining was applied, with the exception that a standard cell conditioning was used with CC2 antigen retrieval buffer (Ventana). Furthermore, a rabbit polyclonal primary antibody for RelA (NFkB p65), (Spring Biosciences E2750) was used at 1:600 dilution* and incubated for 32 min. In total, 24 cases were available with enough tumor tissue and sufficient staining quality.

Evaluation a immunostained slide

The lymphocytes are highlighted by brown nuclear staining of CD3. The staining pattern was analyzed by a board-certified pathologist (MB) and scored into low and high enrichment. The percentage and intensity (weak 1+, moderate 2+ and intense 3+) of staining were recorded as well as the number and size of clustering of CD3 positive cells. The pathologist also chose the appropriate area from each sample for image analysis. We observed that the tissue section that has a complete cross section of the tumor with a complete rim of adjacent benign lung parenchyma is most ideal for image analysis. This is because the lymphocytic infiltration is commonly present at the periphery of the tumor. In

addition to this assessment by a pathologist, a computational system was implemented for automatic evaluation (Supplementary file 3).

ADDITIONAL FILES

Supplementary file 1
Radiomic feature definition and further description towards meaning of feature groups.
https://doi.org/10.7554/eLife.23421.020
Download elife-23421-supp1-v2.pdf

Source code 1
Analysis code.
Source code used to analyse data and generate figures.
https://doi.org/10.7554/eLife.23421.021
Download elife-23421-code1-v2.zip

Supplementary file 2
Exact p-values of modules and list of used R packages and their versions used for analysis.
https://doi.org/10.7554/eLife.23421.022
Download elife-23421-supp2-v2.pdf

Supplementary file 3
Methods for automated pathological call assessment.
https://doi.org/10.7554/eLife.23421.023
Download elife-23421-supp3-v2.pdf

ACKNOWLEDGEMENTS

REFERENCES

Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, Cavalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D, Hoebers F, Rietbergen MM, Leemans CR, Dekker A, Quackenbush J, Gillies RJ, Lambin P. 2014. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature Communications* **5**:4006.

Aerts HJ. 2016. The potential of radiomic-based phenotyping in precision medicine: a review. *JAMA Oncology* **2**:1636.

Aerts HJ, Grossmann P, Tan Y, Oxnard GG, Rizvi N, Schwartz LH, Zhao B. 2016. Defining a radiomic response phenotype: a pilot study using targeted therapy in NSCLC. *Scientific Reports* **6**:33860.

Ahrendt SA, Hu Y, Buta M, McDermott MP, Benoit N, Yang SC, Wu L, Sidransky D. 2003. p53 mutations and survival in stage I non-small-cell lung Cancer: results of a prospective study. *JNCI Journal of the National Cancer Institute* **95**:961–970.

Balagurunathan Y, Gu Y, Wang H, Kumar V, Grove O, Hawkins S, Kim J, Goldgof DB, Hall LO, Gatenby RA, Gillies RJ. 2014. Reproducibility and prognosis of quantitative features extracted from CT Images. *Translational Oncology* **7**:72–87.

Baldi A, De Luca A, Esposito V, Campioni M, Spugnini EP, Citro G. 2011. Tumor suppressors and cell-cycle proteins in lung Cancer. *Pathology Research International* **2011**:1–12.

Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: a Practical and Powerful Approach to multiple testing. *Journal of the Royal Statistical Society. Series B, Statistical Methodology* **57**:289–300.

Bergmann S, Ihmels J, Barkai N. 2003. Iterative signature algorithm for the analysis of large-scale gene expression data. *Physical Review E* **67**:031902.

Bradley AP. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition* **30**:1145–1159.

Bryant JL, Meredith SL, Williams KJ, White A. 2014. Targeting hypoxia in the treatment of small cell lung Cancer. *Lung Cancer* **86**:126–132.

Chen HY, Yu SL, Chen CH, Chang GC, Chen CY, Yuan A, Cheng CL, Wang CH, Terng HJ, Kao SF, Chan WK, Li HN, Liu CC, Singh S, Chen WJ, Chen JJ, Yang PC. 2007. A five-gene signature and clinical outcome in non-small-cell lung Cancer. *New England Journal of Medicine* **356**:11–20.

Choi ER, Lee HY, Jeong JY, Choi YL, Kim J, Bae J, Lee KS, Shim YM. 2016. Quantitative image variables reflect the intratumoral pathologic heterogeneity of lung adenocarcinoma. *Oncotarget* **7**:67302–67313.

Chong Y, Kim JH, Lee HY, Ahn YC, Lee KS, Ahn MJ, Kim J, Shim YM, Han J, Choi YL. 2014. Quantitative CT variables enabling response prediction in neoadjuvant therapy with EGFR-TKIs: are they different from those in neoadjuvant concurrent chemoradiotherapy? *PLoS One* **9**:e88598.

Cook GJ, Yip C, Siddique M, Goh V, Chicklore S, Roy A, Marsden P, Ahmad S, Landau D. 2013. Are pretreatment 18F-FDG PET tumor textural features in non-small cell lung Cancer associated with response and survival after chemoradiotherapy? *Journal of Nuclear Medicine* **54**:19–26.

Coroller TP, Grossmann P, Hou Y, Rios Velazquez E, Leijenaar RT, Hermann G, Lambin P, Haibe-Kains B, Mak RH, Aerts HJ. 2015. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. *Radiotherapy and Oncology* **114**:345–350.

Csárdi G, Kutalik Z, Bergmann S. 2010. Modular analysis of gene expression data with R. *Bioinformatics* **26**:1376–1377.

De Jay N, Papillon-Cavanagh S, Olsen C, El-Hachem N, Bontempi G, Haibe-Kains B, Jay ND. 2013. mRMRe: an R package for parallelized mRMR ensemble feature selection. *Bioinformatics* **29**:2365–2368.

Denny WA. 2010. Hypoxia-activated prodrugs in Cancer therapy: progress to the clinic. *Future Oncology* **6**:419–428.

Diehn M, Nardini C, Wang DS, McGovern S, Jayaraman M, Liang Y, Aldape K, Cha S, Kuo MD. 2008. Identification of noninvasive imaging surrogates for brain tumor gene-expression modules. *PNAS* **105**:5213–5218.

Doroshow JH, Kummar S. 2014. Translational research in oncology--10 years of progress and future prospects. *Nature Reviews Clinical Oncology* **11**:649–662.

Fawcett T. 2006. An introduction to ROC analysis. *Pattern Recognition Letters* **27**:861–874.

Fisher R, Pusztai L, Swanton C. 2013. Cancer heterogeneity: implications for targeted therapeutics. *British Journal of Cancer* **108**:479–485.

Fried DV, Tucker SL, Zhou S, Liao Z, Mawlawi O, Ibbott G, Court LE. 2014. Prognostic value and reproducibility of pretreatment CT texture features in stage III non-small cell lung Cancer. *International Journal of Radiation Oncology*Biology*Physics* **90**:834–842.

Ganeshan B, Abaleke S, Young RC, Chatwin CR, Miles KA. 2010. Texture analysis of non-small cell lung Cancer on unenhanced computed tomography: initial evidence for a relationship with tumour glucose metabolism and stage. *Cancer Imaging* **10**:137–143.

Ganeshan B, Panayiotou E, Burnand K, Dizdarevic S, Miles K. 2012. Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: a potential marker of survival. *European Radiology* **22**:796–802.

Ganeshan B, Goh V, Mandeville HC, Ng QS, Hoskin PJ, Miles KA. 2013. Non-small cell lung Cancer: histopathologic correlates for texture parameters at CT. *Radiology* **266**:326–336.

Gautier L, Cope L, Bolstad BM, Irizarry RA. 2004. Affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**:307–315.

Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A, Tarpey P, Varela I, Phillimore B, Begum S, McDonald NQ, Butler A, Jones D, Raine K, Latimer C, Santos CR, Nohadani M, Eklund AC, Spencer-Dene B, Clark G, Pickering L, Stamp G, Gore M, Szallasi Z, Downward J, Futreal PA, Swanton C. 2012. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *New England Journal of Medicine* **366**:883–892.

Gevaert O, Xu J, Hoang CD, Leung AN, Xu Y, Quon A, Rubin DL, Napel S, Plevritis SK. 2012. Non-small cell lung Cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data--methods and preliminary results. *Radiology* **264**:387–396.

Gevaert O, Mitchell LA, Achrol AS, Xu J, Echegaray S, Steinberg GK, Cheshier SH, Napel S, Zaharchuk G, Plevritis SK. 2014. Glioblastoma multiforme: exploratory radiogenomic analysis by using quantitative image features. *Radiology* **273**:168–174.

Gillies RJ, Anderson AR, Gatenby RA, Morse DL. 2010. The biology underlying molecular imaging in oncology: from genome to anatome and back again. *Clinical Radiology* **65**:517–521.

Gillies RJ, Kinahan PE, Hricak H. 2016. Radiomics: images are more than Pictures, they are data. *Radiology* **278**:563–577.

Grossmann P, Gutman DA, Dunn WD, Holder CA, Aerts HJ. 2016. Imaging-genomics reveals driving pathways of MRI derived volumetric tumor phenotype features in glioblastoma. *BMC Cancer* **16**:611.

Grossmann P, Narayan V, Chang K, Rahman R, Abrey L, Reardon DA, Schwartz LH, Wen PY, Alexander BM, Huang R, Aerts H. 2017. Quantitative Imaging biomarkers for risk stratification of patients with recurrent glioblastoma treated with Bevacizumab. *Neuro-Oncology*.

Grove O, Berglund AE, Schabath MB, Aerts HJ, Dekker A, Wang H, Velazquez ER, Lambin P, Gu Y, Balagurunathan Y, Eikman E, Gatenby RA, Eschrich S, Gillies RJ. 2015. Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma. *PLoS One* **10**:e0118261.

Gutman DA, Dunn WD, Grossmann P, Cooper LA, Holder CA, Ligon KL, Alexander BM, Aerts HJ. 2015. Somatic mutations associated with MRI-derived volumetric features in glioblastoma. *Neuroradiology* **57**:1227–1237.

Harrell FE, Califf RM, Pryor DB, Lee KL, Rosati RA. 1982. Evaluating the yield of medical tests. *JAMA: The Journal of the American Medical Association* **247**:2543–2546.

Hou J, Aerts J, den Hamer B, van Ijcken W, den Bakker M, Riegman P, van der Leest C, van der Spek P, Foekens JA, Hoogsteden HC, Grosveld F, Philipsen S. 2010. Gene expression-based classification of non-small cell lung carcinomas and survival prediction. *PLoS One* **5**:e10312.

Hsu YC, Yuan S, Chen HY, Yu SL, Liu CH, Hsu PY, Wu G, Lin CH, Chang GC, Li KC, Yang PC. 2009. A four-gene signature from NCI-60 cell line for survival prediction in non-small cell lung Cancer. *Clinical Cancer Research* **15**:7309–7315.

Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, Ma ZL, Liu ZY. 2016a. Development and Validation of a Radiomics Nomogram for Preoperative prediction of Lymph Node Metastasis in Colorectal Cancer. *Journal of Clinical Oncology* **34**:2157–2164.

Huang Y, Liu Z, He L, Chen X, Pan D, Ma Z, Liang C, Tian J, Liang C. 2016b. Radiomics signature: a potential biomarker for the prediction of Disease-Free Survival in Early-Stage (I or II) Non-Small Cell lung Cancer. *Radiology* **281**:947–957.

Hänzelmann S, Castelo R, Guinney J. 2013. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**:7.

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**:249–264.

Joshi-Tope G, Gillespie M, Vastrik I, D'Eustachio P, Schmidt E, de Bono B, Jassal B, Gopinath GR, Wu GR, Matthews L, Lewis S, Birney E, Stein L. 2005. Reactome: a knowledgebase of biological pathways. *Nucleic Acids Research* **33**:D428–D432.

Kobayashi S, Yoneda-Kato N, Itahara N, Yoshida A, Kato JY. 2013. The COP1 E3-ligase interacts with FIP200, a key regulator of mammalian autophagy. *BMC Biochemistry* **14**:1.

Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, Forster K, Aerts HJ, Dekker A, Fenstermacher D, Goldgof DB, Hall LO, Lambin P, Balagurunathan Y, Gatenby RA, Gillies RJ. 2012. Radiomics: the process and the challenges. *Magnetic Resonance Imaging* **30**:1234–1248.

Kuo MD, Jamshidi N. 2014. Behind the numbers: decoding molecular phenotypes with radiogenomics--guiding principles and technical considerations. *Radiology* **270**:320–325.

Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, Zegers CM, Gillies R, Boellard R, Dekker A, Aerts HJ. 2012. Radiomics: extracting more information from medical images using advanced feature analysis. *European Journal of Cancer* **48**:441–446.

Leijenaar RT, Carvalho S, Velazquez ER, van Elmpt WJ, Parmar C, Hoekstra OS, Hoekstra CJ, Boellaard R, Dekker AL, Gillies RJ, Aerts HJ, Lambin P. 2013. Stability of FDG-PET radiomics features: an integrated analysis of test-retest and inter-observer variability. *Acta Oncologica* **52**:1391–1397.

Li H, Zhu Y, Burnside ES, Drukker K, Hoadley KA, Fan C, Conzen SD, Whitman GJ, Sutton EJ, Net JM, Ganott M, Huang E, Morris EA, Perou CM, Ji Y, Giger ML. 2016. MR Imaging Radiomics signatures for predicting the risk of breast Cancer recurrence as given by Research Versions of MammaPrint, Oncotype DX, and PAM50 gene assays. *Radiology* **281**:382–391.

Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. 2011. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**:1739–1740.

Liu Y, Kim J, Balagurunathan Y, Li Q, Garcia AL, Stringfield O, Ye Z, Gillies RJ. 2016. Radiomic features are associated with EGFR mutation status in lung adenocarcinomas. *Clinical Lung Cancer* **17**:441–448.

Mattonen SA, Palma DA, Johnson C, Louie AV, Landis M, Rodrigues G, Chan I, Etemad-Rezai R, Yeung TP, Senan S, Ward AD. 2016. Detection of local Cancer recurrence after stereotactic ablative Radiation therapy for lung Cancer: physician Performance Versus Radiomic Assessment. *International Journal of Radiation Oncology*Biology*Physics* **94**:1121–1128.

Menendez D, Shatz M, Resnick MA. 2013. Interactions between the tumor suppressor p53 and immune responses. *Current Opinion in Oncology* **25**:85–92.

Mirsadraee S, Oswal D, Alizadeh Y, Caulo A, van Beek E. 2012. The 7th lung Cancer TNM classification and staging system: review of the changes and implications. *World Journal of Radiology* **4**:0.

Nicolasjilwan M, Hu Y, Yan C, Meerzaman D, Holder CA, Gutman D, Jain R, Colen R, Rubin DL, Zinn PO, Hwang SN, Raghavan P, Hammoud DA, Scarpace LM, Mikkelsen T, Chen J, Gevaert O, Buetow K, Freymann J, Kirby J, Flanders AE, Wintermark M, TCGA Glioma Phenotype Research Group. 2015. Addition of MR imaging features and genetic biomarkers strengthens glioblastoma survival prediction in TCGA patients. *Journal of Neuroradiology* **42**:212–221.

O'Connor JP, Rose CJ, Waterton JC, Carano RA, Parker GJ, Jackson A, O'Connor JPB. 2015. Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome. *Clinical Cancer Research* **21**:249–257.

O'Connor JP, Aboagye EO, Adams JE, Aerts HJ, Barrington SF, Beer AJ, Boellaard R, Bohndiek SE, Brady M, Brown G, Buckley DL, Chenevert TL, Clarke LP, Collette S, Cook GJ, deSouza NM, Dickson JC, Dive C, Evelhoch JL, Faivre-Finn C, Gallagher FA, Gilbert FJ, Gillies RJ, Goh V, Griffiths JR, Groves AM, Halligan S, Harris AL, Hawkes DJ, Hoekstra OS, Huang EP, Hutton BF, Jackson EF, Jayson GC, Jones A, Koh DM, Lacombe D, Lambin P, Lassau N, Leach MO, Lee TY, Leen EL, Lewis JS, Liu Y, Lythgoe MF, Manoharan P, Maxwell RJ, Miles KA, Morgan B, Morris S, Ng T, Padhani AR, Parker GJ, Partridge M, Pathak AP, Peet AC, Punwani S, Reynolds AR, Robinson SP, Shankar LK, Sharma RA, Soloviev D, Stroobants S, Sullivan DC, Taylor SA, Tofts PS, Tozer GM, van Herk M, Walker-Samuel S, Wason J, Williams KJ, Workman P, Yankeelov TE, Brindle KM, McShane LM, Jackson A, Waterton JC, O'Connor JPB. 2017. Imaging biomarker roadmap for Cancer studies. *Nature Reviews Clinical Oncology* **14**:169–186.

Pardoll DM. 2012. The blockade of immune checkpoints in Cancer immunotherapy. *Nature Reviews Cancer* **12**:252–264.

Parmar C, Leijenaar RT, Grossmann P, Rios Velazquez E, Bussink J, Rietveld D, Rietbergen MM, Haibe-Kains B, Lambin P, Aerts HJ. 2015a. Radiomic feature clusters and prognostic signatures specific for lung and Head & Neck Cancer. *Scientific Reports* **5**:11044.

Parmar C, Grossmann P, Rietveld D, Rietbergen MM, Lambin P, Aerts HJ. 2015b. Radiomic machine-learning Classifiers for prognostic biomarkers of head and neck cancer. *Frontiers in Oncology* **5**:272.

Pencina MJ, D'Agostino RB. 2004. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Statistics in Medicine* **23**:2109–2123.

Postow MA, Callahan MK, Wolchok JD. 2015. Immune checkpoint blockade in Cancer therapy. *Journal of Clinical Oncology* **33**:1974–1982.

R Development Core Team. 2013. *R: A Language and Environment for Statistical Computing*. Austria, R Foundation for Statistical Computing. http://www.R-project.org

Rabbani MG, Hossain SA, Islam KK, Uddin SN. 2014. Constitutive photomorphogensis Protein1 (COP1) mediated p53 pathway and its oncogenic role. *Biomedical Research and Therapy* **1**:142–151.

Rizzo S, Petrella F, Buscarino V, De Maria F, Raimondi S, Barberis M, Fumagalli C, Spitaleri G, Rampinelli C, De Marinis F, Spaggiari L, Bellomi M. 2016. CT Radiogenomic characterization of EGFR, K-RAS, and ALK mutations in non-small cell Lung cancer. *European Radiology* **26**:32–42.

Rutman AM, Kuo MD. 2009. Radiogenomics: creating a link between molecular diagnostics and diagnostic imaging. *European Journal of Radiology* **70**:232–241.

Schröder MS, Culhane AC, Quackenbush J, Haibe-Kains B. 2011. Survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. *Bioinformatics* **27**:3206–3208.

Segal E, Sirlin CB, Ooi C, Adler AS, Gollub J, Chen X, Chan BK, Matcuk GR, Barry CT, Chang HY, Kuo MD. 2007. Decoding global gene expression programs in liver Cancer by noninvasive imaging. *Nature Biotechnology* **25**:675–680.

Sottoriva A, Spiteri I, Piccirillo SG, Touloumis A, Collins VP, Marioni JC, Curtis C, Watts C, Tavaré S. 2013. Intra-tumor heterogeneity in human glioblastoma reflects Cancer evolutionary dynamics. *PNAS* **110**:4009–4014.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *PNAS* **102**:15545–15550.

Theodoridis S, Koutroumbas K. 2008. *Fourth Edition*. *Pattern Recognition*. Academic Press.

Whitlock MC. 2005. Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. *Journal of Evolutionary Biology* **18**:1368–1373.

Win T, Miles KA, Janes SM, Ganeshan B, Shastry M, Endozo R, Meagher M, Shortman RI, Wan S, Kayani I, Ell PJ, Groves AM. 2013. Tumor heterogeneity and permeability as measured on the CT component of PET/CT predict survival in patients with non-small cell lung Cancer. *Clinical Cancer Research* **19**:3591–3599.

Wolchok JD, Saenger Y. 2008. The mechanism of anti-CTLA-4 activity and the negative regulation of T-cell activation. *The Oncologist* **13 Suppl 4**:2–9.

Yoon HJ, Sohn I, Cho JH, Lee HY, Kim JH, Choi YL, Kim H, Lee G, Lee KS, Kim J. 2015. Decoding tumor phenotypes for ALK, ROS1, and RET Fusions in lung adenocarcinoma using a Radiomics Approach. *Medicine* **94**:e1753.

Yuan BZ, Jefferson AM, Popescu NC, Reynolds SH. 2004. Aberrant gene expression in human non small cell lung carcinoma cells exposed to demethylating agent 5-aza-2'-deoxycytidine. *Neoplasia* **6**:412–419.

Zhao B, Tan Y, Tsai WY, Qi J, Xie C, Lu L, Schwartz LH. 2016. Reproducibility of radiomics for deciphering tumor phenotype with imaging. *Scientific Reports* **6**:23428.

CHAPTER

3

Imaging-genomics reveals driving pathways of MRI derived volumetric tumor phenotype features in Glioblastoma

Patrick Grossmann, David A. Gutman*, William D. Dunn Jr, Chad A. Holder, and Hugo J.W.L. Aerts*
*These authors contributed equally to this work

ABSTRACT

Background

Glioblastoma (GBM) tumors exhibit strong phenotypic differences that can be quantified using magnetic resonance imaging (MRI), but the underlying biological drivers of these imaging phenotypes remain largely unknown. An Imaging-Genomics analysis was performed to reveal the mechanistic associations between MRI derived quantitative volumetric tumor phenotype features and molecular pathways.

Materials and methods

One hundred fourty one patients with presurgery MRI and survival data were included in our analysis. Volumetric features were defined, including the necrotic core (NE), contrast-enhancement (CE), abnormal tumor volume assessed by post-contrast T1w (tumor bulk or TB), tumor-associated edema based on T2-FLAIR (ED), and total tumor volume (TV), as well as ratios of these tumor components. Based on gene expression where available ($n = 91$), pathway associations were assessed using a preranked gene set enrichment analysis. These results were put into context of molecular subtypes in GBM and prognostication.

Results

Volumetric features were significantly associated with diverse sets of biological processes (FDR < 0.05). While NE and TB were enriched for immune response pathways and apoptosis, CE was associated with signal transduction and protein folding processes. ED was mainly enriched for homeostasis and cell cycling pathways. ED was also the strongest predictor of molecular GBM subtypes (AUC = 0.61). CE was the strongest predictor of overall survival (C-index = 0.6; Noether test, $p = 4\text{x}10^{-4}$.

Conclusions

GBM volumetric features extracted from MRI are significantly enriched for information about the biological state of a tumor that impacts patient outcomes. Clinical decision-support systems could exploit this information to develop personalized treatment strategies on the basis of noninvasive imaging.

Keywords

Imaging-genomics – Radiomics – Glioblastoma – Volumetric – Pathways – Prediction – Noninvasive – Radiation Oncology – Neuro-imaging

BACKGROUND

Glioblastoma (GBM) is a highly invasive and diffuse WHO grade IV tumor and is the most lethal central nervous system malignancy with an annual age-adjusted incidence rate of 3.19/100,000 per population [1]. Despite aggressive surgical therapy, radiation therapy, and temozolomide administration the 2-year survival rate remains around 27 % [2]. As a result, recent investigations have focused on capitalizing on the high molecular heterogeneity of gliomas to develop personalized treatment strategies [3].

One promising avenue of these investigations involves quantitative analyses of radiographic data, where imaging modalities are used to quantify tumor phenotype non-invasively. In magnetic resonance imaging (MRI), GBM tumors exhibit strong phenotypic features such as Necrosis, Edema, Contrast Enhancement, and Tumor Bulk (Fig. 1). These properties can be captured without and with intravenous administration of gadolinium-based contrast agents including T1-weighted or FLuid-Attenuated Inversion Recovery (FLAIR) (Fig. 2). In this way, visible tumor phenotype features can be systematically quantified.
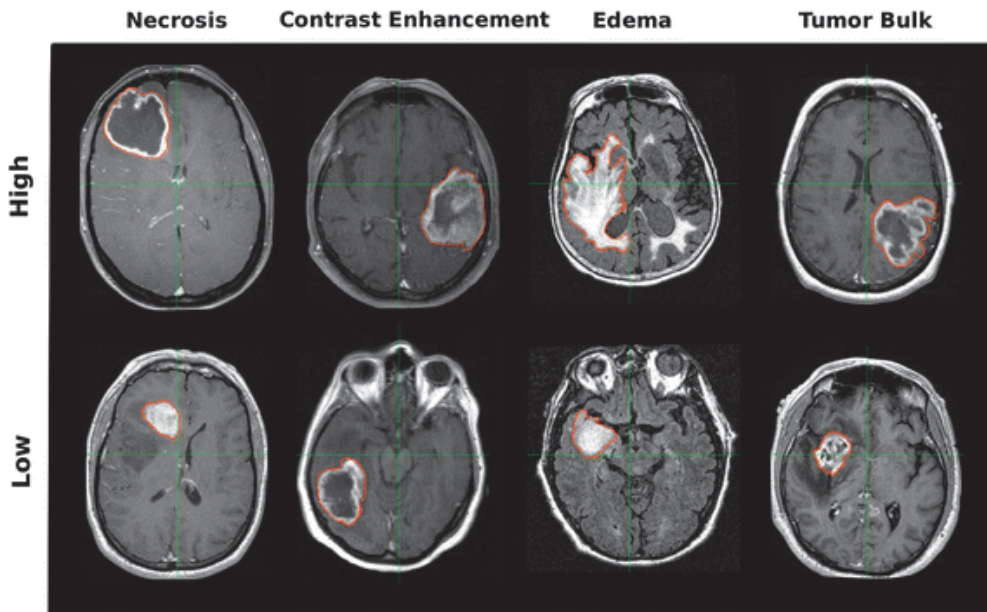
Fig. 1

Examples of volumetric tumor phenotype features. Glioblastoma (GBM) tumors show strong phenotypic differences, which can be objectively quantified with volumetrics. This figure shows examples of GBM tumors exhibiting high (top) and low (bottom) volumetric feature values for Necrosis, Contrast Enhancement, Edema, and Tumor Bulk (columns) as they appear on T1 weighted (columns 1,2, and 4) or T2-FLAIR (column 3) magnetic resonance images for different patients.
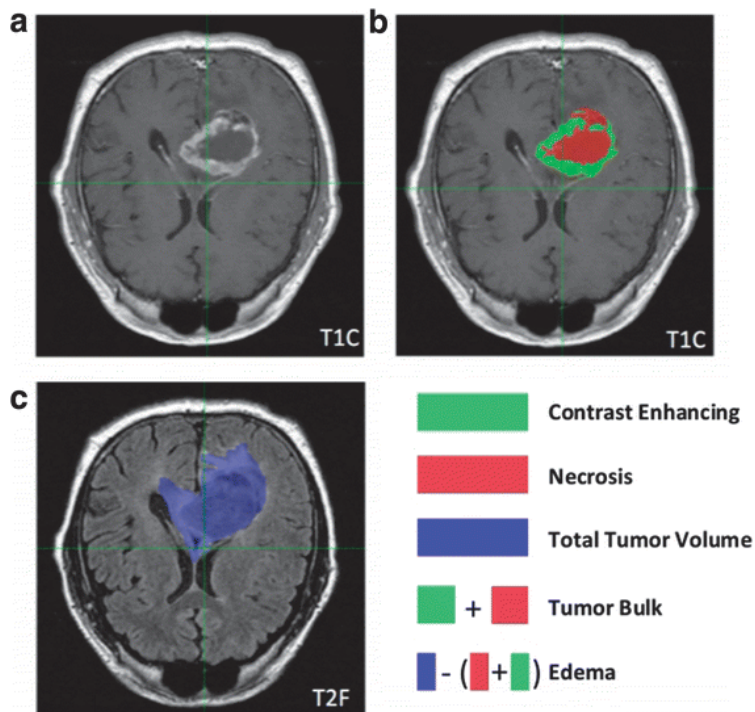
Fig. 2
Volumetric phenotype features within the same tumor. Detailed example of a glioblastoma tumor in a patient. (a,b) On T1-weighted post-Gadolinium contrast (T1C) images, a central area of Necrosis is typically surrounded by a Contrast Enhancing ring, both of which can be derived from dark and light regions, respectively. Tumor Bulk represents the addition of these tumor features. (c) The Total Tumor Volume is represented by hyperintensity extracted from T2-FLAIR images. Edema is the difference of Tumor Bulk from Total Tumor Volume.

As the underlying drivers of these phenotypes are biological in nature, recent efforts have been conducted indicating underlying genetic characteristics of imaging features. For example, tumor "Ring Enhancement" was found to be significantly associated with unmethylated MGMT promoter status [4, 5], which is known to be a biomarker for response to temozolomide and survival. Similarly, "Contrast Enhancement" and "Mass Effect" imaging features were found to be strongly correlated with expression of groups of genes involved in hypoxia and proliferation, respectively [6]. However, a systematic classification of tumor phenotype features in terms of their underlying cell biological processes on a genome-wide scale in GBM remains absent, although clinical applicability of these image features will depend on knowledge about how these features are driven by tumor biological processes that determine disease progression.

In this study, we present an Imaging-Genomics analysis to investigate the associations of a large set of biological processes and presurgical diagnostic MRI derived quantita-

tive volumetric tumor phenotype features, such as Necrosis or Edema, focusing on the publicly available GBM dataset from The Cancer Genome Atlas (TCGA). These analyses were tied to molecular subtypes in GBM and prognostics. Image based volumetric features provide noninvasive tumor phenotype information complementary to genomic technologies and clinical information, potentially allowing advances in patient stratification and clinical decision-making.

METHODS

Magnetic resonance imaging

The DICOM formatted files of presurgical T1 and T2 sequence magnetic resonance images (MRIs) were accessed and downloaded in November 2014 from TCIA (https://wiki.cancer-imagingarchive.net/ display/Public/TCGA-GBM), a large archive of medical images of cancer patients who have matched molecular data at The Cancer Genome Atlas (TCGA). Cases that had both T1 and T2-FLAIR images available, were of reasonable quality to perform tumor segmentation, and had presurgical negative status were included. As the presurgical status of an image is not explicitly included in the TCIA data, presurgical status was verified to the best of our ability by a trained neuroradiologist (CH, 17 years of experience) by examining the skull surrounding the tumor for signs of surgeries. The patients in our study were originally imaged at Thomas Jefferson University Hospital and Henry Ford Hospital.

Images of sufficient quality were next analyzed for volumetric features. Briefly, 2D masks which were annotated using FSLView, a module in the FMRIB Software Library 5.0 (FSL [7]), were applied surrounding the tumor regions on the post gadolinium (GD) contrast T1-weighted images and T2-weighted images. For the T1 images, a single contour was segmented including both the dark (Necrotic or NE) and bright (Contrast Enhancement or CE) areas, and the entire volume was referred to as Tumor Bulk (TB). The pixels contained in these masks were then clustered into dark (NE) and bright (CE) areas by K-means clustering using the FSL FEAT (fMRI Expert Analysis Tool, Version 5.0). The area volume contained within the mask of the T2 FLAIR image set encompasses the Edema (ED) envelope, including regions of hyperintense signal and inclusive of any other abnormal signal in the region previously identified on the T1 (i.e., TB), and was referred to as Total Abnormal Tumor Volume (TV). Afterwards, all masks were visually checked by a trained radiologist (CH). We did not attempt to discriminate between peritumoral edema and non-enhancing tumor, as both appear hyperintense on FLAIR. In addition to the raw volumetric features, we calculated the following feature ratios as investigated in previous studies [8, 9, 10] mainly to investigate combined T1/FLAIR signals: NE/TV, CE/TV, ED/TV, TB/TV, NE/CE, and CE/TB. A representation of the tumor volumes analyzed are displayed in Fig. 2.

Gene expression

Matching GBM gene expression (mRNA) data for the TCIA patient cohort was obtained from TCGA using the CBioPortal [11] with the 'cgdsr' R package version 1.1.33. The profile identifier 'gbm_tcga_pub_mrna_median_Zscores' was used together with the case identifier 'gbm_tcga_pub_mrna' to download the expression values of 18,055 genes given as median Z-scores across the Agilent, Affymetrix U133, and Affymetrix Exon platforms. Expression data were downloaded on April 3, 2015, for 91 patients for which also imaging data was available. Based on expression of 1740 genes, Verhaak et al. [12] classified TCGA-GBM patients into the four molecular GBM subtypes proneural, neural, classical, and mesenchymal, which were functionally annotated by presence of oncogenic events. To test predictive power for subtypes, we downloaded the classification results on TCGA patients by Verhaak et al. from https://tcga-data.nci.nih.gov/docs/publications/gbm_exp/TCGA_unified_CORE_ClaNC840.txt and calculated the multiclass area under curves (AUCs) of the receiver operator characteristic [13] of the volumetric features. Imaging and subtype data were available for 79 patients.

Pathway analysis

To quantify the association of a volumetric features with biological processes, preranked Gene Set Enrichment Analysis [14] (GSEA) version 2.2.0 was performed; gene ranks were calculated for every feature according to -$\log10(p)$ r, where r is the Spearman rank correlation coefficient, and p its p-value. GSEA was performed on the C5-BP collection version 5.0 from the Molecular Signature Database [15] (MSigDB), which contains the expert-curated Gene Ontology [16] (GO) gene sets for biological processes. Those 583 gene sets containing at least 15 and at most 500 genes were analyzed. We investigated gene sets that were significantly enriched under a false-discovery-rate (FDR) < 0.05 as specified by GSEA to account for multiple hypothesis-testing [17].

Survival analysis

Overall survival data was available for 141 patients with imaging data, and was downloaded from CBioPortal on April 3, 2015. Prognostic associations of volumetric features were assessed with the concordance index (CI) using the 'survcomp' package in Bioconductor [18]. All statistical analyses were carried out using R version 3.1.0 [19] on a Linux operating system.

RESULTS

To investigate which biological processes drive volumetric tumor phenotype features in GBM, we performed a pathway analysis based on gene expression profiles using a pre-ranked Gene Set Enrichment Analysis [14]. We compared these results to molecular sub-types in GBM and evaluated the prognostic value.

Volumetric tumor phenotype features in GBM

Based on MRI, we quantified the following volumetric features in GBM: Necrosis (NE), Contrast Enhancement (CE), Edema (ED), Tumor Bulk (TB), and Total Tumor Volume (TV). In addition, we calculated the following ratios mainly to investigate combined T1/FLAIR signals: NE/TV, CE/TV, ED/TV, TB/TV, NE/CE, and CE/TB. The areas of the tumor that these features correspond to are highlighted in Fig. 2b and Fig. 2c. In general, we found that these features were not or only moderately correlated (mean Spearman rho 0.48 and -0.41 for positive and negative correlation, respectively), however a number of features were highly positively correlated (e.g., NE and TB, rho = 0.96) and a number of features ratios were highly anti-correlated (e.g., NE/CE and CE/TB, rho = -0.98) as shown in Fig. 3.

Biological processes underlying volumetric features

In total, 64 biological processes were significantly associated in at least one of the volu-metric features or their ratios (FDR < 0.05, Fig. 4). Table 1 summarizes the biological themes associated with each volumetric feature. These features were generally negatively (anti-correlated) enriched for biological processes unless stated otherwise. NE and TB were mainly enriched for pathways involved in immune response and apoptosis, whereas CE was enriched for signal transduction and protein folding processes. ED was enriched for cell cycling, proliferation, and replication mechanisms, but also positively enriched for homeostasis. TV was associated with synaptogenesis, biogenesis, and excretion.

Volumetric feature ratios were associated with a larger number of biological processes than the original features. Signal transduction was associated with all of the ratios we computed; processes involved in immune system were found for all ratios except for CE/TV. CE/TV, TB/TV, and ED/TV were enriched for protein complex assembly. In addition, ED/TV showed positive enrichment for defense response, cytokine production, and Nf-kB. Nf-kB was also found in NE/TV and TB/TV, as well as in NE/CE and CE/TB. Notably, NE/CE and CE/TB were also inversely enriched for inflammation, immune system response path-ways, and anti-apoptosis.
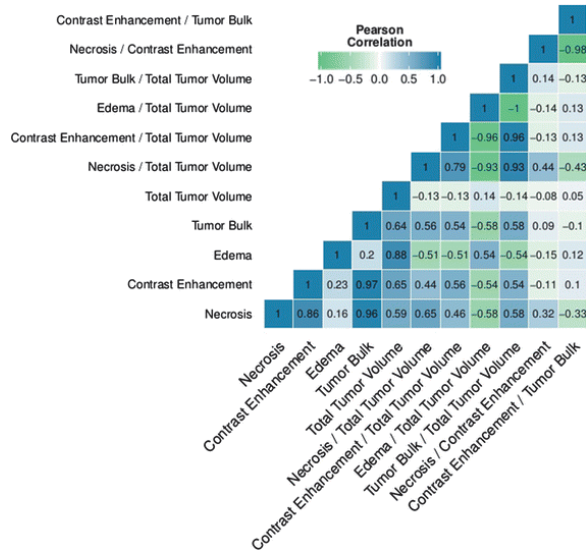
Fig. 3

Correlation map. Pairwise Pearson correlation coefficients of volumetric features. Only few volumes were highly correlated (blue) or highly anti-correlated (anti-correlated).
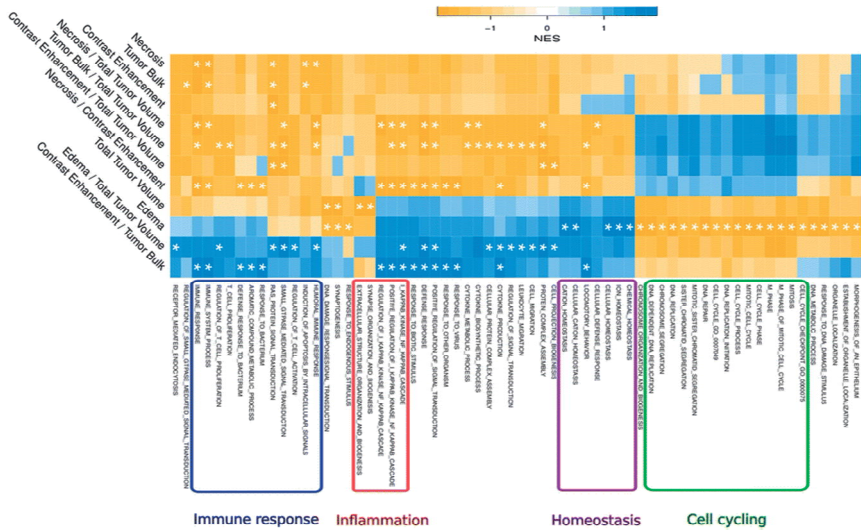
Fig. 4

Pathway enrichment analysis. In total, 64 biological processes (rows) were significantly (FDR < 0.05) enriched for at least one volumetric feature (columns) as indicated by an asterisk. Heatmap shows normalized enrichment scores (NES) calculated with Gene Set Enrichment Analysis. Positive NES (blue) correspond to correlated pathways and negative NES (yellow) correspond to anti-correlated pathways.

Table 1

Summary of pathways associated with volumetric tumor phenotype features of the original volumes (top rows) and their ratios (bottom rows).

|	Biological processes (positive correlation)	Biological processes (negative correlation)
Volume		
Necrosis		Immune response, apoptosis
Contrast Enhancement		Signal transduction
Edema	Homeostasis	Cell cycle, proliferation, replication, DNA repair, DNA metabolic process
Tumor Bulk		Apoptosis, signal transduction, immune system
Total Volume		Synaptogenesis, biogenesis, extracellular structure organization
Ratios		
Necrosis/ Total Volume		Defense response, immune response, Nf-kB, signal transduction
Contrast Enhancement/ Tumor Volume		Protein complex assembly, signal transduction, biogenesis
Edema/ Tumor Volume	Protein complex assembly, defense response, signal transduction, cytokine production, immune response, Nf-kB	
Tumor Bulk/ Tumor Volume		Signal transduction, protein complex assembly, cytokine, immune response, Nf-kB
Necrosis/ Contrast Enhancement		Response to other organism, Nf-kB, immune response, locomotory behaviour
Contrast Enhancement/ Tumor Bulk	Response to other organism, Nf-kB, immune response, locomotory behaviour	

Molecular subtypes in GBM

Based on a study by Verhaak et al. [12], patients from the TCGA-GBM cohort were classified to belong to either one of the four following molecular subtypes: proneural, neural, classical, and mesenchymal. Compared to TV, ED had the largest median size across subtypes (Fig. 5a); ED was larger in classical GBM and smaller in proneural (Fig. 5b). Other volumetric features were comparably similar in terms of median values across subtypes.
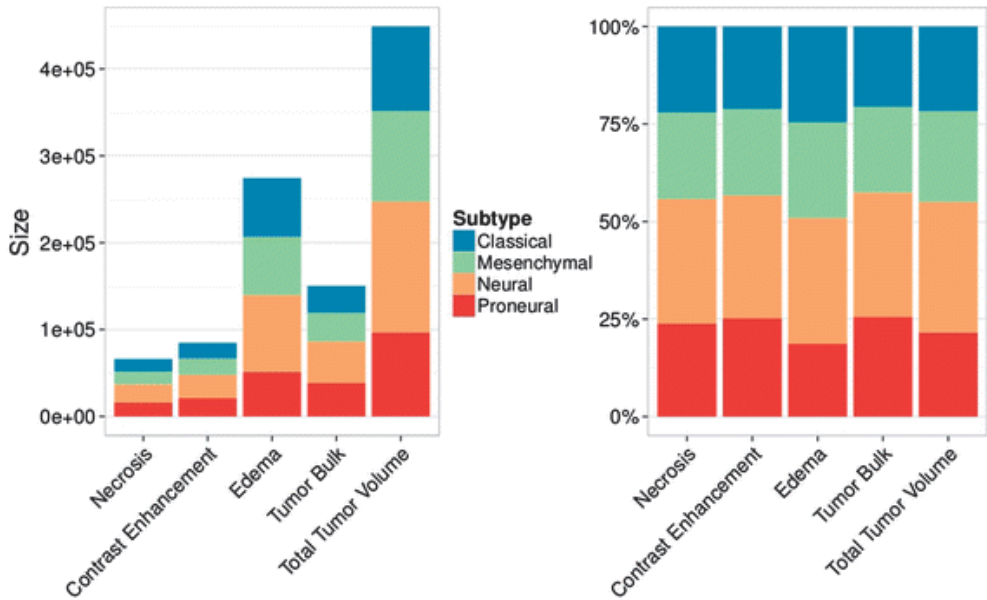
Fig. 5

*Size distribution of volumetric tumor features across molecular subtypes of GBM. (**a**) Compared to the Total Volume, Edema had the largest median size across all molecular GBM subtypes. (**b**) Classical and neural tumors showed larger Edema areas than mesenchymal and proneural tumors. Size variation of volumetric feature areas other than Edema was generally low across subtypes.*

We tested predictive value for GBM subtypes of all volumetric features by calculating the area under the curve (AUC) of the receiver operator characteristic. We found that most features performed relatively low (Table 2). ED and TV were the strongest predictors of subtypes (AUCs = 0.61). Ratios of features generally were poor predictors of subtype.

Table 2

Performances of volumetric features in predicting molecular subtypes of GBM.

Volume	Multiclass AUC
Necrosis	0.57
Contrast Enhancement	0.57
Edema	0.61
Tumor Bulk	0.57
Total Volume	0.61
Ratios	
Necrosis/Total Volume	0.56
Contrast Enhancement/Tumor Volume	0.56
Edema/Tumor Volume	0.56
Tumor Bulk/Tumor Volume	0.56
Necrosis/Contrast Enhancement	0.54
Contrast Enhancement/Tumor Bulk	0.54

Prognostic value of volumetrics

To link our pathway-imaging results to clinical patient outcome, we tested prognostic value of volumetric features for overall survival (OS). Four features (NE, CE, TB, and TV) significantly predicted OS (Noether, $p < 0.05$), but prognostic performances as measured by the concordance index [20] were only moderate (Fig. 6). Importantly, NE, CE, and TB performed significantly better than TV (one-sided t-test, $p < 0.05$). Furthermore, Kaplan-Meier and Log-Rank analyses revealed significant assessments of low and high risk survival groups by NE, CE, and TB (Additional file 1: Figure S1).

DISCUSSION

We The translation of quantitative imaging data into defined clinical settings requires knowledge of how volumetric tumor phenotype features are driven by biological processes that determine the outcome of a patient. This study presents an Imaging-Genomics analysis of presurgical diagnostic MRI derived volumetric features in GBM to evaluate if tumor phenotype features are associated with underlying tumor biology. We found different features to be enriched for different sets of biological processes. Molecular subtypes of GBM were difficult to be predicted by volumetric features. However, four out of five features showed significant prognostic value.
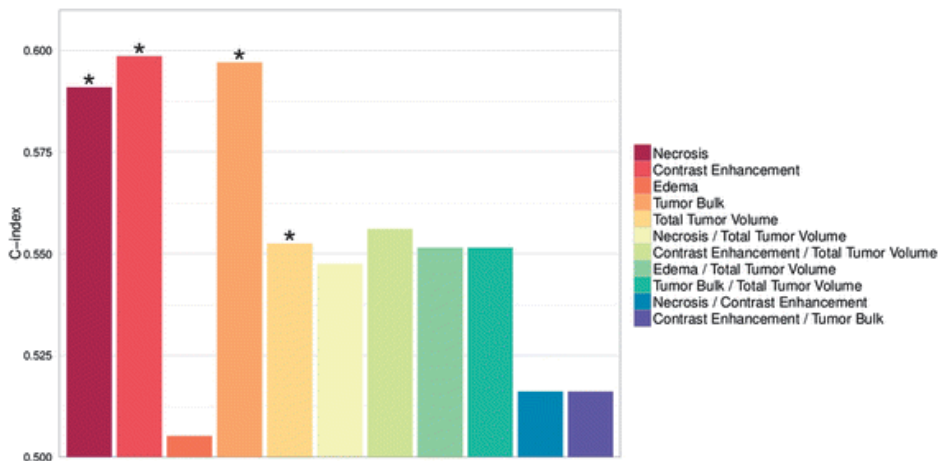
Fig. 6

Prognostic value of volumetric tumor features. Necrosis, Contrast Enhancement, Tumor Bulk, and Total Tumor Volume were significantly (asterisk) prognostic (p < 0.05). The Contrast Enhancement feature showed the highest prognostic performance as measured by the C-index.

As correlations among our volumetric features were low to moderate in general, this suggests that quantifying each of those areas individually yields complementary information about the tumor phenotype beyond the Total Tumor Volume (TV). Interestingly, we found most of the biological processes to be anti-correlated to pathway expression. The most prevalent pathways were apoptosis, immune system, and signaling pathways, which were observed mainly for Necrosis (NE), Contrast Enhancement (CE), and Tumor Bulk (TB); features that were also significantly prognostic. As those pathways are known drivers of survival outcome [21, 22, 23], this hence explains why NE, CE, and TB were found to be prognostic as well. Importantly, all of these features performed significantly better than TV, which highlights that quantification of individual imaging features should be preferred over calculating only the total tumor volume. Our finding that NE is anti-correlated with immune response and prognostic is in line with Gevaert et al. [24], who also correlated quantitative imaging features of GBM areas to molecular data and who found significant imaging associations to approximately 20 pathways. This analysis, however, differs from our analysis in that Gevaert et al. investigated a single slice of a tumor (in axial view), whereas we performed quantification using the 3D tumor volumes.

Edema (ED) was the only feature that was correlated with homeostasis, cell cycling, and proliferation pathways. Surprisingly, ED was not prognostic in our analysis, although cell cycling and proliferation are known to be involved in carcinogenesis [25]. However, using the publicly available MRI scoring scheme VASARI (https://wiki.nci.nih.gov/display/CIP/VASARI), Gutman et al. [26] found ordinal ED assessment to be not prognostic as well. Interestingly, in a related study by Diehn et al. [6], binary assessment of ED result-

ed in significant survival predictions. Similarly, a recent study indicated that an ED volume cutoff of 85'000 mm^3 is a significant prognostic factor using Kaplan-Meier analysis [8]; however, the rationale for this cutoff was not given. Prognostic performance of quantitative ED features could increase in cohorts of extended sample sizes, as ED has been reported to be a univariate predictor of survival in a large patient cohort previously [27]. Furthermore, our analysis suggests that CE and NE are prognostic. This is partially in line with Gutman et al. [26], according to whom CE is prognostic, but NE is not. The contradictions between our results and the studies by Diehn et al. and Gutman et al. could be due to the nature of the ED and NE assessments, which in our analysis were continuous, but binary and ordinal in Diehn et al. and Gutman et al., respectively. Likewise, our methodology could be compared to a study by Jamshidi et al. [28], but comparison remains challenging as their analysis focused on binary imaging traits on a relatively small cohort of patients and a subset of oncogenic pathways only.

Although ED was the only feature that was not prognostic in our analysis, it was the highest predictor of molecular subtypes in GBM instead. This is likely because ED was the only feature that expressed a different volumetric size distribution across molecular subtypes. Similar indications have been given by Gevaert et al. [24], who found three out of four features that correlated with molecular subtypes to be quantitative Edema features. In general, we found volumetric features to be only moderate predictors of subtypes suggesting that subtypes do not generally alter the size composition of tumor areas in GBM. Furthermore, we could not confirm that the proneural subtype has lower proportions of CE as suggested by Gutman et al. [26]. Poor predictability of Verhaak molecular subtypes by relative cerebral blood volume using T2-weighted MRI has been also described by Jain et al. [29].

In our analysis, ratios of volumetric features were not significantly prognostic or predictive of GBM subtypes. Generally, many more biological processes were significantly associated with the feature ratios, usually showing a trend towards a mix of pathways associated with the individual features that the ratios were composed of (e.g., NE/TV were enriched for signal transduction and biogenesis). While our study associated MRI volumetric features with biological processes, molecular subtypes, and survival outcome using genome-wide data and aimed at explaining the rationale for why MRI derived volumetric features are associated with survival on a pathway level, other studies have focused on revealing specific genetic variations between MRI features and survival [30, 31, 32, 33, 34].

Our analysis was limited to a retrospective dataset. To establish volumetric biomarkers in clinical applications, prospective evaluation of our results will be required. Biological significance could be further validated by analyses of complementary molecular data such as mutational or epigenetic data. Such analyses could provide further insight into why separate quantification of distinct volumetric tumor phenotype features yield different biological and prognostic information. We acknowledge that the prognostic and predictive performances of the volumetric features in the TCGA-GBM dataset were moderate. Heterogeneity of GBM tumors [35] could be an explanation for this, which limits the

definition of a single molecular subtype especially on the basis of single-needle biopsy [36, 37]. As imaging approaches target the entire visible tumor, we however expect in future studies that prognostic performances will drastically increase when Imaging-Genomic cohorts with even larger numbers of samples and standardized image processing become available for GBM.

While our study focused on volumetric phenotype features, alternative definitions of imaging phenotypes are available. This may, for example, include tumor location as this determines the extent of possible resection and hence is a prognostic factor in GBM [38]. In addition to such semantic phenotypes, agnostic phenotyping approaches such as radiomics could be added [39].

CONCLUSIONS

In conclusion, quantitative imaging biomarkers hold great potential, as, unlike traditional biopsies, medical imaging is noninvasive and captures the entire tumor volume. As we have shown, a relationship exists between individual volumetric phenotype features describing local, clinically relevant subareas of GBM tumors and global expression of genes. Knowledge about how these specific tumor areas are related to underlying biological cell processes may allow for advanced patient stratification and treatment decision on the basis of standard medical imaging, but efforts in optimization of prognostic and predictive performances need to continue.

ADDITIONAL FILES

Additional file 1: Figure S1. Stratification power of volumetric tumor phenotype features. Kaplan-Meier analysis of the volumetric features that showed significant prognostic value (i.e., Necrosis, Contrast Enhancement, Tumor Bulk, and Total Tumor Volume). Except for Total Tumor Volume, these features also showed significant classification in low (blue) and high (red) risk groups based on the mean feature value. Available at: https://static-content.springer.com/esm/art%3A10.1186%2Fs12885-016-2659-5/MediaObjects/12885_2016_2659_MOESM1_ESM.pdf.

REFERENCES

1.   Thakkar JP, Dolecek TA, Horbinski C, Ostrom QT, Lightner DD, Barnholtz-Sloan JS, Villano JL. Epidemiologic and Molecular Prognostic Review of Glioblastoma. Cancer Epidemiol Biomarkers Prev. 2014;23:1985–96.
2.   Omuro A, DeAngelis LM. Glioblastoma and other malignant gliomas: a clinical review. JAMA. 2013;310:1842–50.

3. Zhu J-J, Wong ET. Personalized medicine for glioblastoma: current challenges and future opportunities. Curr Mol Med. 2013;13:358–67.

4. Hegi ME, Diserens A-C, Gorlia T, Hamou M-F, de Tribolet N, Weller M, Kros JM, Hainfellner JA, Mason W, Mariani L, Bromberg JEC, Hau P, Mirimanoff RO, Cairncross JG, Janzer RC, Stupp R. MGMT gene silencing and benefit from temozolomide in glioblastoma. N Engl J Med. 2005;352:997–1003.

5. Drabycz S, Roldán G, de Robles P, Adler D, McIntyre JB, Magliocco AM, Cairncross JG, Mitchell JR. An analysis of image texture, tumor location, and MGMT promoter methylation in glioblastoma using magnetic resonance imaging. Neuroimage. 2010;49:1398–405.

6. Diehn M, Nardini C, Wang DS, McGovern S, Jayaraman M, Liang Y, Aldape K, Cha S, Kuo MD. Identification of noninvasive imaging surrogates for brain tumor gene-expression modules. Proc Natl Acad Sci U S A. 2008;105:5213–8.

7. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. FSL. Neuroimage. 2012;62:782–90.

8. Wangaryattawanich P, Hatami M, Wang J, Thomas G, Flanders A, Kirby J, Wintermark M, Huang ES, Bakhtiari AS, Luedi MM, Hashmi SS, Rubin DL, Chen JY, Hwang SN, Freymann J, Holder CA, Zinn PO, Colen RR. Multicenter imaging outcomes study of The Cancer Genome Atlas glioblastoma patient cohort: imaging predictors of overall and progression-free survival. Neuro Oncol. 2015;17(11):1525–37.

9. Wang Y, Wang K, Li H, Wang J, Wang L, Dai J, Jiang T, Ma J. Identifying the association of contrast enhancement with vascular endothelia growth factor expression in anaplastic gliomas: a volumetric magnetic resonance imaging analysis. PLoS One. 2015;10:e0121380.

10. Naeini KM, Pope WB, Cloughesy TF, Harris RJ, Lai A, Eskin A, Chowdhury R, Phillips HS, Nghiemphu PL, Behbahanian Y, Ellingson BM. Identifying the mesenchymal molecular subtype of glioblastoma using quantitative volumetric analysis of anatomic magnetic resonance images. Neuro Oncol. 2013;15:626–34.

11. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. Sci Signal. 2013;6:l1–pl1.

12. Verhaak RGW, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, Miller CR, Ding L, Golub T, Mesirov JP, Alexe G, Lawrence M, O'Kelly M, Tamayo P, Weir BA, Gabriel S, Winckler W, Gupta S, Jakkula L, Feiler HS, Hodgson JG, James CD, Sarkaria JN, Brennan C, Kahn A, Spellman PT, Wilson RK, Speed TP, Gray JW, Meyerson M, et al. Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell. 2010;17:98–110.

13. Fawcett T. An introduction to ROC analysis. Pattern Recognit Lett. 2006;27:861–74 [ROC Analysis in Pattern Recognition].

14. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102:15545–50.

15. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011;27:1739–40.

16. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene Ontology: tool for the unification of biology. Nat Genet. 2000;25:25–9.

17. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Series B Stat Methodol. 1995;57:289–300.

18. Schröder MS, Culhane AC, Quackenbush J, Haibe-Kains B. survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. Bioinformatics. 2011;27(22):3206–8.

19. R Development Core Team: R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013. URL

20. Harrell Jr FE, Califf RM, Pryor DB, Lee KL, Rosati RA. EValuating the yield of medical tests. JAMA. 1982;247:2543–6.

21. Lowe SW, Lin AW. Apoptosis in cancer. Carcinogenesis. 2000;21:485–95.

22. Gajewski TF, Schreiber H, Fu Y-X. Innate and adaptive immune cells in the tumor microenvironment. Nat Immunol. 2013;14:1014–22.

23. Levitzki A, Klein S. Signal transduction therapy of cancer. Mol Aspects Med. 2010;31:287–329 [Signal Transduction Therapy of Cancer].

24. Gevaert O, Mitchell LA, Achrol AS, Xu J, Echegaray S, Steinberg GK, Cheshier SH, Napel S, Zaharchuk G, Plevritis SK. Glioblastoma multiforme: exploratory radiogenomic analysis by using quantitative image features. Radiology. 2014;273:168–74.

25. Kaufmann WK, Kaufman DG. Cell cycle control, DNA repair and initiation of carcinogenesis. FASEB J. 1993;7:1188–91.

26. Gutman DA, Cooper LAD, Hwang SN, Holder CA, Gao J, Aurora TD, Dunn Jr WD, Scarpace L, Mikkelsen T, Jain R, Wintermark M, Jilwan M, Raghavan P, Huang E, Clifford RJ, Mongkolwat P, Kleper V, Freymann J, Kirby J, Zinn PO, Moreno CS, Jaffe C, Colen R, Rubin DL, Saltz J, Flanders A, Brat DJ. MR imaging predictors of molecular profile and survival: multi-institutional study of the TCGA glioblastoma data set. Radiology. 2013;267:560–9.

2. Lacroix M, Abi-Said D, Fourney DR, Gokaslan ZL, Shi W, DeMonte F, Lang FF, McCutcheon IE, Hassenbusch SJ, Holland E, Hess K, Michael C, Miller D, Sawaya R. A multivariate analysis of 416 patients with glioblastoma multiforme: prognosis, extent of resection, and survival. J Neurosurg. 2001;95:190–8.

28. Jamshidi N, Diehn M, Bredel M, Kuo MD. Illuminating Radiogenomic Characteristics of Glioblastoma Multiforme through Integration of MR Imaging, Messenger RNA Expression, and DNA Copy Number Variation. Radiology. 2013;270:1–2.

29. Jain R, Poisson L, Narang J, Gutman D, Scarpace L, Hwang SN, Holder C, Wintermark M, Colen RR, Kirby J, Freymann J, Brat DJ, Jaffe C, Mikkelsen T. Genomic mapping and survival prediction in glioblastoma: molecular subclassification strengthened by hemodynamic imaging biomarkers. Radiology. 2013;267:212–20.

30. Gutman DA, Dunn Jr WD, Grossmann P, Cooper LAD, Holder CA, Ligon KL, Alexander BM, Aerts HJWL. Somatic mutations associated with MRI-derived volumetric features in glioblastoma. Neuroradiology. 2015;57(12):1227–37.

31. Jain R, Poisson LM, Gutman D, Scarpace L, Hwang SN, Holder CA, Wintermark M, Rao A, Colen RR, Kirby J, Freymann J, Jaffe CC, Mikkelsen T, Flanders A. Outcome prediction in patients with glioblastoma by using imaging, clinical, and genomic biomarkers: focus on the nonenhancing component of the tumor. Radiology. 2014;272:484–93.

32. Nicolasjilwan M, Hu Y, Yan C, Meerzaman D, Holder CA, Gutman D, Jain R, Colen R, Rubin DL, Zinn PO, Hwang SN, Raghavan P, Hammoud DA, Scarpace LM, Mikkelsen T, Chen J, Gevaert O, Buetow K, Freymann J, Kirby J, Flanders AE, Wintermark M, TCGA Glioma Phenotype Research Group. Addition of MR imaging features and genetic biomarkers strengthens glioblastoma survival prediction in TCGA patients. J Neuroradiol. 2015;42:212–21.

33. Zinn PO, Majadan B, Sathyan P, Singh SK, Majumder S, Jolesz FA, Colen RR. Radiogenomic Mapping of Edema/Cellular Invasion MRI-Phenotypes in Glioblastoma Multiforme. PLoS One. 2011;6:e25451.

34. Pope WB, Chen JH, Dong J, Carlson MRJ, Perlina A, Cloughesy TF, Liau LM, Mischel PS, Nghiemphu P, Lai A, Nelson SF. Relationship between Gene Expression and Enhancement in Glioblastoma Multiforme: Exploratory DNA Microarray Analysis. Radiology. 2008;249:268–77.

35. Kim H, Zheng S, Amini SS, Virk SM, Mikkelsen T, Brat DJ, Grimsby J, Sougnez C, Muller F, Hu J, Sloan AE, Cohen ML, Van Meir EG, Scarpace L, Laird PW, Weinstein JN, Lander ES, Gabriel S, Getz G, Meyerson M, Chin L, Barnholtz-Sloan JS, Verhaak RGW. Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution. Genome Res. 2015;25:316–27.

36. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, Cahill DP, Nahed BV, Curry WT, Martuza RL, Louis DN, Rozenblatt-Rosen O, Suvà ML, Regev A, Bernstein BE. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science. 2014;344:1396–401.

37. Sottoriva A, Spiteri I, Piccirillo SGM, Touloumis A, Collins VP, Marioni JC, Curtis C, Watts C, Tavaré S. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. Proc Natl Acad Sci U S A. 2013;110:4009–14.

38. Kumar N, Kumar P, Angurana SL, Khosla D, Mukherjee KK, Aggarwal R, Kumar R, Bera A, Sharma SC. Evaluation of outcome and prognostic factors in patients of glioblastoma multiforme: A single institution experience. J Neurosci Rural Pract. 2013;4 Suppl 1:S46–55.

39.   Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiol. 2016;278:563–77.

DECLERATIONS

Acknowledgements

Funding

Availability of data and materials

MRI data is publicly available in DICOM format at TCIA (https://wiki.cancerimaging-archive.net/display/Public/TCGA-GBM). Corresponding gene expression and survival data is publically available at the same site or can be queried using the same identifiers through the cBioPortal (http://www.cbioportal.org/, see Methods).

Authors' contributions

PG and HJWLA designed the study. PG analyzed the data and drafted the manuscript. DAG performed tumor segmentation on the MRIs. DAG, WDD, and CAH quantified volumetric features. All authors wrote, reviewed, edited, and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Authors' information

No relevant additional authors' information.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Imaging, gene expression, and survival data have been de-identified by The Cancer Imaging Archive (TCIA) and The Cancer Genome Atlas (TCGA) consortia and are publicly available. Thus, no internal approval of an institutional review board was required for this study. Informed consent was collected by the TCGA Research Network.

Open Access

DECLARATIONS

CHAPTER

4

Somatic mutations associated with MRI-derived volumetric features in glioblastoma

ABSTRACT

Introduction
MR imaging can noninvasively visualize tumor phenotype characteristics at the macroscopic level. Here, we investigated whether somatic mutations are associated with and can be predicted by MRI-derived tumor imaging features of glioblastoma (GBM).

Materials and methods
Seventy-six GBM patients were identified from The Cancer Imaging Archive for whom preoperative T1-contrast (T1C) and T2-FLAIR MR images were available. For each tumor, a set of volumetric imaging features and their ratios were measured, including necrosis, contrast enhancing, and edema volumes. Imaging genomics analysis assessed the association of these features with mutation status of nine genes frequently altered in adult GBM. Finally, area under the curve (AUC) analysis was conducted to evaluate the predictive performance of imaging features for mutational status.

Results
Our results demonstrate that MR imaging features are strongly associated with mutation status. For example, TP53-mutated tumors had significantly smaller contrast enhancing and necrosis volumes ($p = 0.012$ and 0.017, respectively) and RB1-mutated tumors had significantly smaller edema volumes ($p = 0.015$) compared to wild-type tumors. MRI volumetric features were also found to significantly predict mutational status. For example, AUC analysis results indicated that TP53, RB1, NF1, EGFR, and PDGFRA mutations could each be significantly predicted by at least one imaging feature.

Conclusion
MRI-derived volumetric features are significantly associated with and predictive of several cancer-relevant, drug-targetable DNA mutations in glioblastoma. These results may shed insight into unique growth characteristics of individual tumors at the macroscopic level resulting from molecular events as well as increase the use of noninvasive imaging in personalized medicine.

Keywords
Radiogenomics – GBM – MRI – Imaging genomics – Volumetrics

INTRODUCTION

Glioblastoma (GBM) is the most common and most aggressive form of brain cancer with a median survival of less than 15 months and a 5-year survival rate of less than 10 % [1]. While factors ranging from younger age at diagnosis, cerebellar location, better cognitive performance, and more extensive tumor resection have been associated with more favorable outcome, the current standard of care treatment involving surgery, radiation, and chemotherapy ultimately fails, in part due to the proliferative and diffusely infiltrative nature of the tumor [2]. Recent molecular analyses have demonstrated significant diversity in histologically similar tumors that drive proliferation and competitive propagation [3]. In addition, integrated analyses using gene expression, copy number, methylation, and somatic mutation patterns have identified distinct GBM subtypes, some of which associated with distinct responses to treatment [4].

Methods such as magnetic resonance imaging (MRI) that can noninvasively characterize the tumor at a macroscopic scale can be of potential value, as they can provide complementary information to the tumor's molecular characterization [5]. Historically, only very basic parameters have been derived from imaging data, such as measurements of tumor size based on "representative" cross sections on a single radiology image [6, 7, 8]. While such measures are easy to perform and serve as the basis for assessing treatment response [9, 10], there is a rich set of additional visual characteristics of the tumor that can also be assessed. One effort to catalog these characteristics is the VASARI Research Project, which seeks to develop a controlled vocabulary describing the varied morphology of glioblastoma (http://cabig.cancer.gov/action/collaborations/vasari/). The VASARI feature set was developed by The Cancer Genome Atlas (TCGA) radiology working group and uses a standard lexicon with the goal of reproducibly assessing 26 imaging descriptors based on T1-weighted and T2-weighted FLAIR MRI modalities. For example, variables such as major axis length, tumor location, proportion enhancing, thickness of enhancing margin, and proportion of edema are all measured by trained radiologists in this protocol. Data obtained from this protocol has led to a number of findings demonstrating the value of adding imaging data to models predicting survival [11, 12] and molecular profile [13] in glioblastoma.

Our current work expands upon these results by using a semi-automated digital quantification technique, which recent work has shown to be more objective and lead to more robust findings than qualitative staging methods used in the past [14]. Indeed, feature measurements based on manual estimations have been shown to be subject to substantial inter- and intra-observer variability [15]. Recently, quantitative volumetric measurements of tumor subvolumes, or imaging features, such as contrast-enhancing tumor, necrosis, and tumor-associated edema, have been associated with response to treatment and overall prognosis [16, 17, 18]. Associations of tumor subvolume data and somatic mutations would be of clinical importance, as it would improve our understanding and macroscopic implications of these heavily researched molecular events [19, 20].

In this study, we investigated whether somatic mutations in genes consistently implicated in glioblastoma are associated with and can be predicted by digitally derived volumetric features from tumor MR images, including contrast enhancing, necrosis, and T2 FLAIR hyperintensity volumes, as well as combination and ratios of these features. We chose to focus on somatic mutations due to the strong literature presence and clinical relevance. We limited our analysis to an a priori selected gene set as described by Verhaak et al. [4], as well as showing mutations in at least five patients included in our cohort, resulting in the following: TP53, PTEN, NF1, EGFR, IDH1, PIK3R1, RB1, PIK3CA, and PDGFRA. Many of these genes are drug-targetable, raising the possibility of treating cancer based on noninvasively derived imaging biomarkers. In particular, several potential therapies to target mutant EGFR such as monoclonal antibodies, vaccines, or small molecule inhibitors are currently being actively investigated [21]. In addition, some PTEN nonsense mutations have shown to be targeted by drugs that inhibit PKC (byrostatin) and Raf (AZ628) [22], and IDH1 mutations have been shown to be targeted by small molecule inhibitors such as AGI-120 [23]. Noninvasive methods that can predict mutation status would thus be of great clinical importance and could guide clinical treatment decision-making, especially in situations where molecular testing or surgical biopsy is not feasible or appropriate.

MATERIAL AND METHODS

Imaging and mutation data

All Presurgical T1-weighted post-Gd contrast (T1C) and T2-weighted FLAIR sequence MR images were downloaded from The Cancer Imaging Archine (TCIA) (http://thecancerimaging-archive.net) in September of 2014. TCIA is an NCI-sponsored imaging sharing resource that houses more than three million images from 31 different institutions [24, 25]. This resource is particularly valuable because it stores a wide variety of publically available longitudinal data with associated genomic data, which is not readily obtainable at a single institution level. As the patients had been previously de-identified by TCGA and were available for public download, no Institutional Review Board approval was required. Since presurgical status of an image was not explicitly included in the TCIA data, presurgical status was verified by a trained neurologist (CH, 17 years of experience). Somatic mutation status from whole exome sequencing and clinical data were downloaded from TCGA using cBioPortal (http://www.cbioportal.org/public-portal/) queried with the "cgdsr" package version 1.1.30 in R [26]. The latest cBioPortal GBM dataset (version "provisional") was downloaded on September 25, 2014. Within TCIA, there were 185 GBM patients with both T1C and T2-weighted FLAIR images available before surgery. Of these, 76 patients had mutation data available within TCGA and were included in this analysis.

Volumetric image analysis

For T1C images, 2D masks were drawn on each MRI slice over the visible tumor using FSLView, a module in the FMRIB Software Library 5.0 (FSL [27]). For these image sets, a single contour encapsulating both the dark (necrosis) and bright (contrast enhancing) are-as was segmented, becoming the basis of what will be referred to as tumor bulk volume. Similarly, for the T2-weighted FLAIR image sets, a single contour was drawn on each slice over the visible tumor which encompassed both the region previously identified on the T1C (i.e., tumor bulk) as well as surrounding hyperintense signal including the edema envelope. To note, this markup does not attempt to differentiate between nonenhancing tumor and true edema, as they both appear hyperintense on the FLAIR image.

Following segmentation, the original mask containing the tumor region on the T1C images underwent K-means clustering using the FSL FAST tool (FMRIB's Automated Segmentation Tool) [28] to differentiate dark (necrosis) and bright (contrast enhancing) areas from one another. A subset of 5–10 machine-generated segmentations were reviewed by trained experts (CH, DG) to verify proper segmentation; we determined that the FAST algorithm module produced robust segmentations without extensive parameter optimization. This algorithm has been used routinely for more than 10 years to segment white and gray matter, and we found robust segmentation results when we applied the algorithm to segmenting "bright" versus "dark" pixels for contrast enhancement and necrosis.

Based on the annotations on the T1C images, contrast enhancing, necrosis, and tumor bulk volumes could be calculated (Fig. 1). In addition, incorporation of the T2-weighted FLAIR series allowed quantification of the total tumor volume and T2-FLAIR hyperintensity volume (total tumor volume–tumor bulk volume). For the purposes of our analyses, several derivative ratios were also calculated: necrosis/contrast enhancing, contrast enhancing/tumor bulk, contrast enhancing/total tumor, necrosis/total tumor, T2-FLAIR hyperintensity/total tumor, and tumor bulk/total tumor volumes. Individual imaging feature volumes were computed by computing the total number of voxels within the respective region and multiplying by the voxel size; these calculations were performed using FSL's fslstats module.

Imaging-genomic analysis

For each of the 11 volumes or ratios defined above, the mean values corresponding to patients with mutated genes were compared to those of the wild-type cohort. Significant differences between groups with and without mutations were tested with a two-sided Student's *t* test. Significance was defined by *p* value <0.05. Normal assumption of volumes was confirmed by *p* value <0.05 under a Shapiro-Wilk test across all patients. This analysis was limited to genes from the a priori selected gene set discussed above.
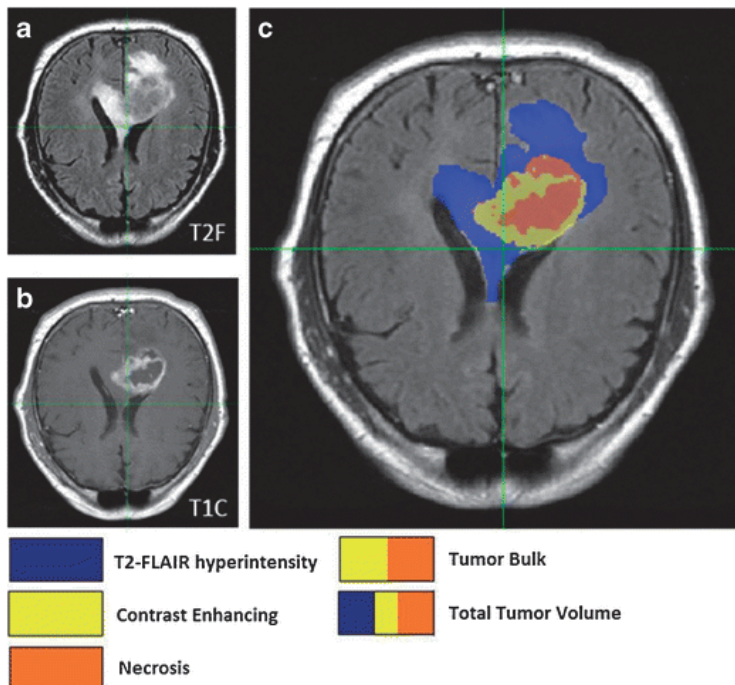
Fig. 1

*Visualization of naming conventions for the tumor volumetric features used throughout this article for TCGA-02-0033, a 54-year-old male glioblastoma patient. **a** Tumor-associated T2-FLAIR hyperintensity and total tumor volume was quantified from the T2-weighted FLAIR images. **b** Necrosis, contrast enhancing, and tumor bulk volume was quantified from T1-weighted post-Gd contrast (T1C) images. **c** This panel displays both images co-registered and overlayed on top of each other for visualization purposes. Tumor bulk is defined as the total abnormal tumor area on the T1C images: combination of contrast enhancing and necrosis volumes. Total tumor volume is defined as the combination of the tumor bulk and T2-FLAIR hyperintensity volumes.*

Predictive power of volumes to predict mutation status was assessed by the area under the curve (AUC) of the receiver operator characteristic (ROC) [29]. To make performance evaluation comparable, we calculated the absolute AUC defined as $0.5 + abs(x − 0.5)$, where x is an AUC value. All statistical analysis was carried out by the R statistical software version 3.0.2 on a Linux platform [30].

RESULTS

To investigate whether somatic mutations were associated with MRI imaging features, the GBM patients analyzed in our analysis were limited to those with mutation data available from TCGA and image data from TCIA ($N = 76$). The genes analyzed in our study were lim-

ited to an a priori selected gene set as described by Verhaak et al. [4], as well as showing mutations in at least five patients included in our cohort. This resulted in a total of nine genes included in our analysis: TP53, PTEN, NF1, EGFR, IDH1, PIK3R1, RB1, PIK3CA, and PDGFRA.

Associations between MRI-derived GBM tumor volumes

An initial visual inspection of segmented images demonstrated a wide variety of measured features across patients that could be capitalized upon in imaging genomic analyses (Fig. 2).
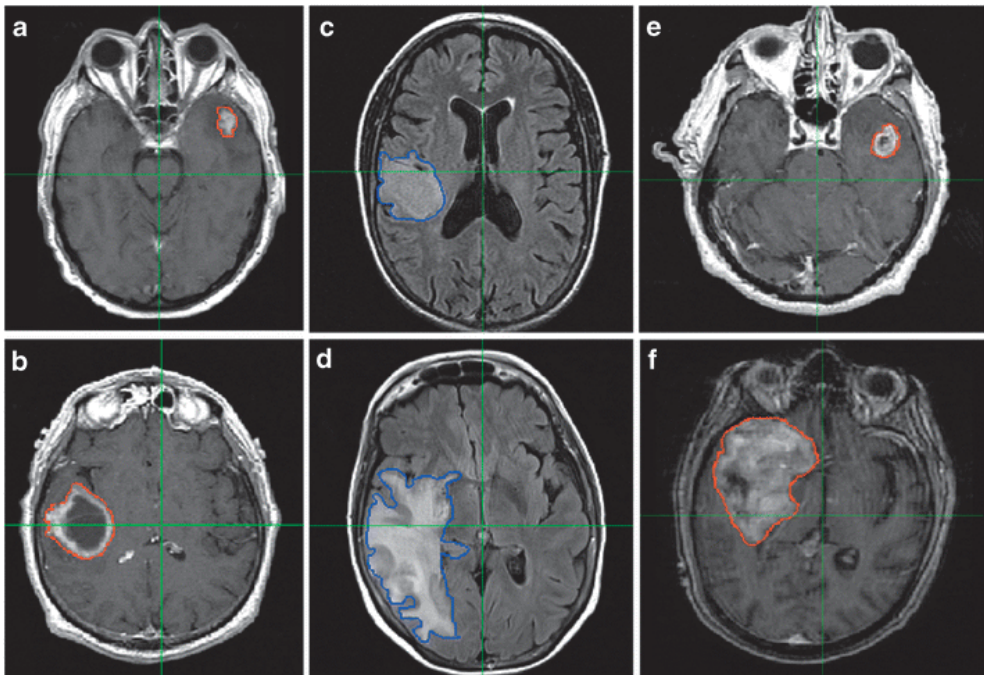
Fig. 2

*Examples of images characterized by various imaging features. Representative scans of low (**a**) and high (**b**) necrosis/total tumor volume ratios, low (**c**) and high (**d**) FLAIR/total tumor volume ratios, and low (**e**) and high (**f**) tumor bulk volumes are visualized. Masks outline areas used to determine various volumetric features used throughout the project (red for tumor bulk on T1-weighted post-Gd contrast (T1C) images, blue for T2-FLAIR hyperintensity on T2-weighted FLAIR images).*

The pairwise Pearson correlation coefficients of these 11 analyzed imaging features are shown in Fig. 3. These results demonstrated the relative independence of many of these volumetric features and ratios. While some higher correlations were noted (e.g., contrast

enhancement and necrosis, $r = 0.91$, or T2-FLAIR hyperintensity and total tumor volume, $r = 0.87$), several features also showed low correlations, indicating independence (e.g., necrosis and T2-FLAIR hyperintensity, $r = 0.07$).
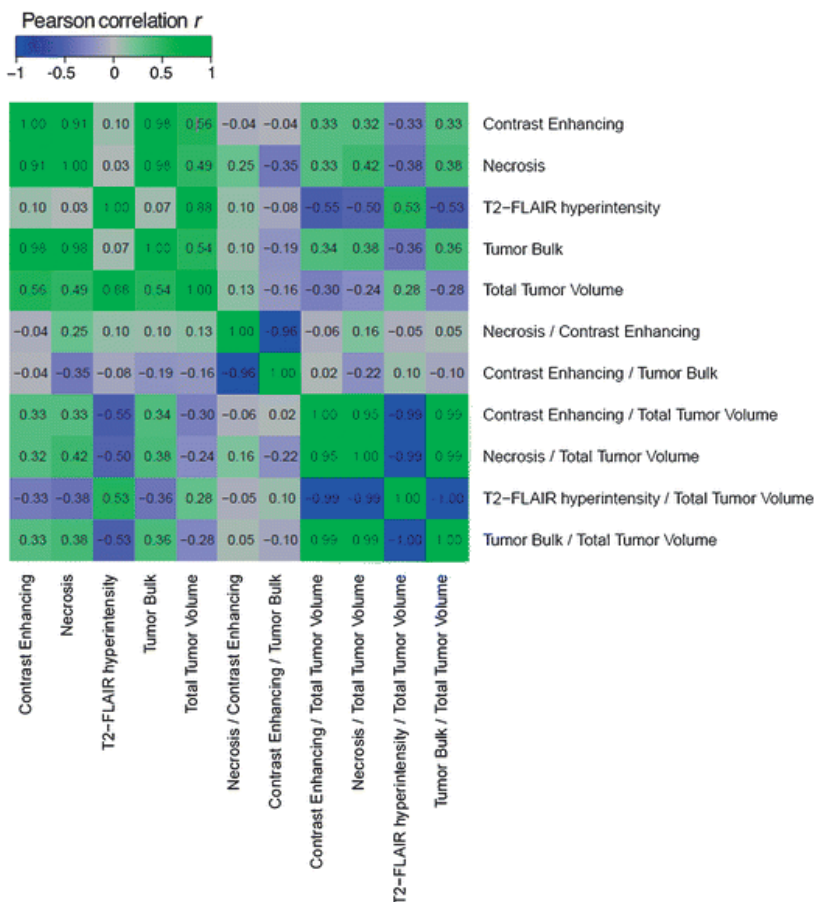
Fig. 3

Correlation coefficient matrix between the eleven imaging feature measurements used in this study. Note the high correlation between several features (e.g., T2-FLAIR hyperintensity and total tumor volume), as well as the low correlation between other features (e.g., necrosis and T2-FLAIR hyperintensity). Correlations were assessed using Pearson correlation coefficient.

In a final exploratory analysis, patients were clustered into two groups based on imaging volumes. Importantly, chi-squared tests did not indicate any significant differences of gender, disease-free status, Karnofsky performance score, and age between

these groups, suggesting that the measured imaging volumes measured are relatively independent of these clinical variables (Fig. 4).
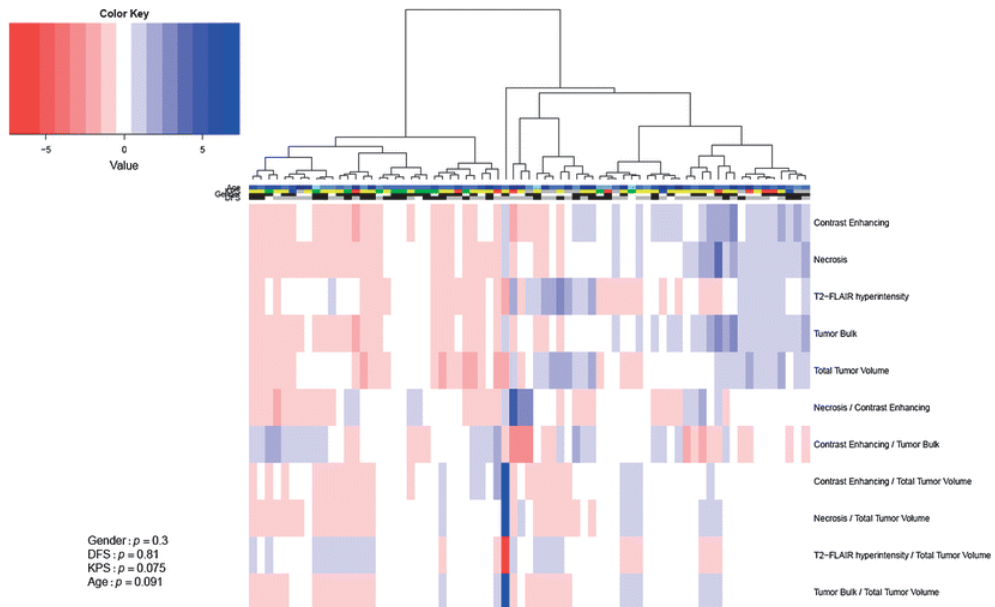
Fig. 4

Heatmap of volume values (Z-scores) and clinical parameters. Patients are clustered according to their imaging features (rows). The two main clusters show no significant association (chi-squared test) to clinical parameters gender, disease-free status (DFS), Karnofsky performance score (KPS), and age. KPS of 40, 60, 80, and 100 are indicated by blue, red, yellow, and green, respectively. Age is indicated by one darker nuance every 10 years (range 21 to 85 years). Gray bars indicate clinical parameters that were not available for a patient.

Association of MRI volumetric features with somatic mutations

In comparing the volumetric averages for the 11 measures between mutant versus wild-type tumors to investigate the associations between mutation status and volumetric features, several significant results were observed (Table 1). TP53-mutated tumors had four subvolumes that were significantly different from wild-type tumors (Fig. 5a). For example, contrast enhancing and necrosis volumes were significantly smaller for mutated tumors (8588.1 mm^3 average difference ($p = 0.012$) and 7159.2 mm^3 average difference ($p = 0.017$), respectively). EGFR-mutated tumors showed a significantly higher necrosis/contrast enhancing ratio ($p = 0.05$) and a significantly lower contrast enhancing/tumor bulk ratio ($p = 0.008$) (Fig. 5b). Furthermore, RB1-mutated tumors showed significantly smaller T2-FLAIR hyperintensity (26,354.4 mm^3 average difference, $p = 0.015$) and total tumor volume (34,467.2 mm^3 average difference, $p = 0.020$) (Fig. 5c). For the other mutations, the volumetric features did not significantly differ between mutated and wild-type tumors (Table 1, Supplemental Digital Content).

Table 1

Differences in imaging feature volumes between mutated and wild-type tumors for a subset of significant results. "Mut-WT" refers to difference in average between mutated and wild-type groups for the various volumes (in mm3) as well as differences in ratios. For each gene and imaging feature, a two-sided Student's t test was performed to measure significance of the difference and the corresponding p value is also provided. For complete set of significances in volumetric differences by mutation status, see Table 1 Supplemental Digital Content.
*Statistical significance (p < 0.05)

	Gene:	TP53	EGFR	RB1
	Number of mutations:	26	24	8
Contrast enhancement	Mut–WT difference	−8588.05	818.21	−3676.05
	t Test p value	0.012*	0.827	0.485
Necrosis	Mut–WT difference	−7159.16	3555.12	−4436.73
	t Test p value	0.017*	0.303	0.289
T2-FLAIR hyperintensity	Mut–WT difference	−11164.7	6655.63	−26354.4
	t Test p value	0.326	0.538	0.015*
Tumor bulk	Mut–WT difference	−15747.2	4373.33	−8112.78
	t Test p value	0.012*	0.533	0.387
Total tumor	Mut–WT difference	−26911.9	11028.92	−34467.2
	t Test p value	0.04*	0.402	0.02*
Necrosis/ Contrast Enhancement	Mut–WT difference	−0.69	0.14	−0.14
	t Test p value	0.492	0.05*	0.056
Contrast enhancement/ tumor bulk	Mut–WT difference	0.012	−0.046	0.038
	t Test p value	0.515	0.008*	0.145

Predicting somatic mutation based on MRI volumetric

To assess the potential of the MRI volumetric features to predict somatic mutation status noninvasively, we evaluated the predictive power using the AUC of the receiver-operating characteristic. AUC values were calculated by quantifying the performance of distinguishing between a mutated and wild-type tumor on the basis of each of the 11 imaging features (Fig. 6, Table 2, Table 2 Supplemental Digital Content).

In general, we found a tendency for volumetric features to predict mutation status of one gene specifically. For example TP53 could be significantly predicted by contrast enhancing (AUC = 0.68, p = 0.001), necrosis (AUC = 0.67, p = 0.039), as well as total tumor volumes (AUC = 0.646, p = 0.010). Of note, these three volumes were all highly correlated with each other as demonstrated in Fig. 3. Additionally, NF1 mutation status could be predicted by contrast enhancing volume (AUC = 0.68, p = 0.023) and tumor bulk volume (AUC = 0.67, p = 0.032). EGFR mutations could be predicted by the necrosis/contrast enhancing (AUC = 0.68, p = 0.001) ratio and contrast enhancing/tumor bulk (AUC = 0.68, p = 0.001) ratio. RB1 mutations could be predicted by T2-FLAIR hyperintensity (AUC = 0.66,

p = 0.022) and total tumor volume (AUC = 0.68, p = 0.011). Finally, PDGFRA could be predicted by T2-FLAIR hyperintensity/total tumor volume (AUC = 0.72, p = 0.026) and tumor bulk/total tumor volume ratios (AUC = 0.72, p = 0.026). All significant results are summarized in Table 2.
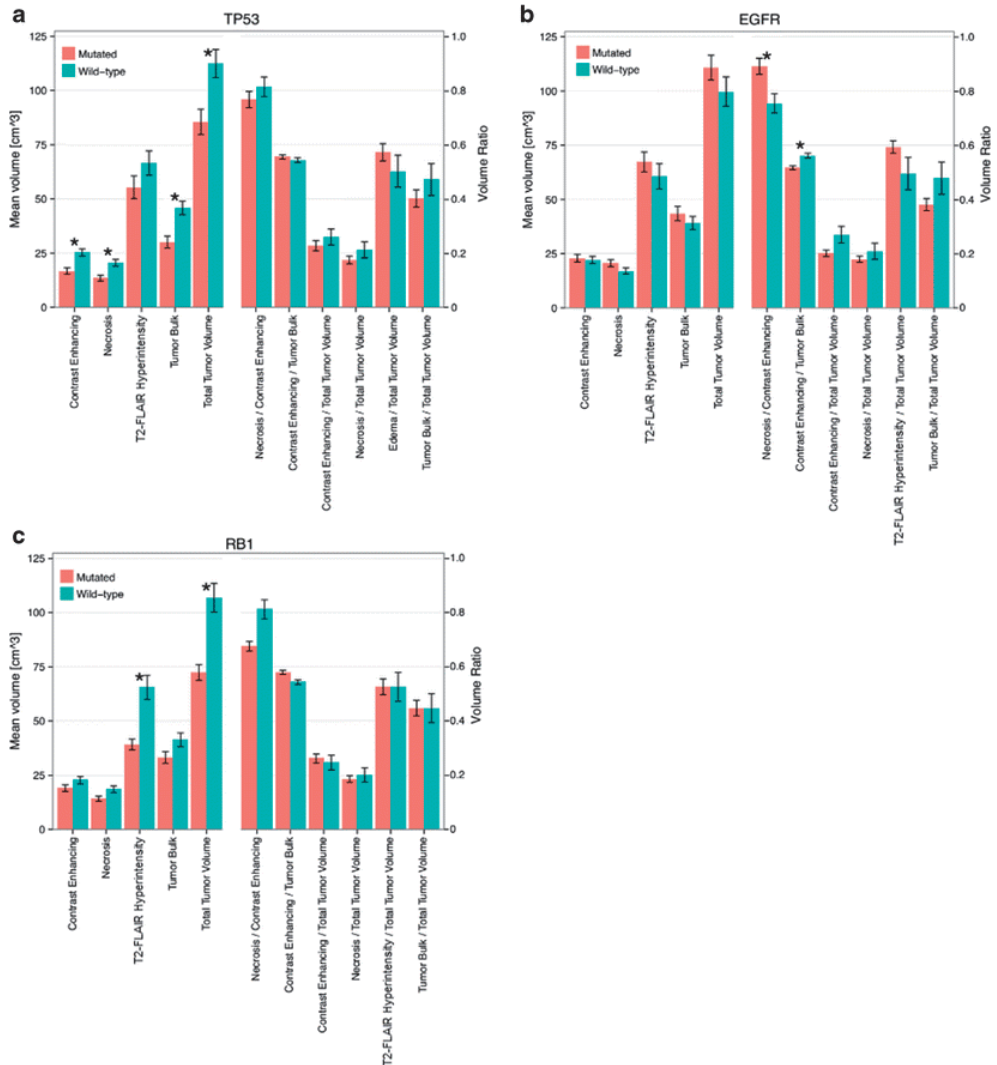
Fig. 5

Volumetric differences for mutated versus wild-type tumors for a TP53, b EGFR, and c RB1. For each plot, the left y-axis corresponds to the mean volume of the left features and the right y-axis corresponds to the volume ratio of the features on the right. The bars indicate the standard error of the mean. Note TP53-mutated tumors were found to have significantly smaller contrast enhancing, necrosis, and tumor bulk volumes compared to wild type. EGFR-mutated tumors have a significantly larger necrosis/contrast enhancing ratio, as well as a significantly smaller contrast enhancing/tumor bulk ratio. RB1-mutated tumors have significantly larger T2-FLAIR hyperintensity and total tumor volumes, compared to wild-type tumors.

DISCUSSION

Medical imaging has strong potential to stratify patients, as it is uniquely situated to non-invasively provide a macroscopic evaluation of the entire tumor volume. We quantified GBM MRI phenotypes by defining volumetric features, such as contrast enhancing, necrosis, and T2-FLAIR hyperintensity volumes, and ratios thereof. In this work, we investigated whether quantitative assessments of tumor features are significantly associated with or predictive of somatic mutation status in GBM.

We found strong associations of MRI characteristics with underlying somatic mutation patterns, such as tumor bulk and total tumor volume with TP53 and RB1 mutations. Importantly, these features were able to significantly predict mutations, such as TP53, EGFR, RB1, NF1, and PDGFRA—mutations of clinical importance in GBM [4, 31]. Although the predictive capability of the volumetric features for mutational status was not perfect (i.e., AUC = 1), performance is much higher and significantly different compared to chance (i.e., AUC = 0.50, p value ≥ 0.05), demonstrating the strong association of the imaging feature with the underlying driving biology.

Our results showed that contrast enhancing volume and necrosis volume are significantly smaller for TP53 mutants, a finding likely driven by the fact that these tumors, in previous work using the categorically defined VASARI imaging features, have been shown to be characterized by smaller volumes in general [13]. We confirmed this qualitative assessment using our digitally defined quantitative volumetric approach, and additionally showed that both the tumor bulk (on the T1C images) and total tumor volume (on the T2-weighted FLAIR images), were significantly smaller for TP53 mutants.

Our results also show that the necrosis/contrast enhancing ratio was significantly higher in EGFR mutants. These results indicate that although the total tumor volume is similar, EGFR mutants have larger necrosis volumes and smaller contrast enhancing volumes, compared to wild-type tumors. Although the tumor volume was higher for EGFR mutants in our quantitative analysis, the differences were not significant, as previous work had demonstrated [13].

Table 2

Gene mutation/volumetric imaging feature correlations for subset containing significant results. For each gene and imaging feature, area under the curve (AUC) values, corresponding pvalue, and 95 % confidence interval are provided. For complete set of correlations, see Table 2 Supplemental Digital Content.
*Statistical significance (p < 0.05)

	Gene:	TP53	NF1	EGFR	RB1	PDGFRA
	Number of mutations:	26	9	24	8	6
Contrast Enhancement	AUC (p value)	0.679 (0.001*)	0.681 (0.023*)	0.503 (0.971)	0.57 (0.489)	0.621 (0.258)
	95 % CI	0.569–0.788	0.525–0.837	0.361–0.644	0.372–0.768	0.411–0.831
Necrosis	AUC (p value)	0.666 (0.004*)	0.658 (0.063)	0.556 (0.429)	0.588 (0.365)	0.572 (0.531)
	95 % CI	0.552–0.78	0.491–0.825	0.417–0.695	0.397–0.779	0.347–0.796
T2-FLAIR hyperintesity	AUC (p value)	0.591 (0.127)	0.537 (0.479)	0.542 (0.499)	0.66 (0.022*)	0.56 (0.687)
	95 % CI	0.474–0.708	0.435–0.639	0.42–0.665	0.523–0.797	0.268–0.852
Tumor bulk	AUC (p value)	0.675 (0.002*)	0.671 (0.032*)	0.535 (0.631)	0.577 (0.441)	0.612 (0.322)
	95 % CI	0.566–0.785	0.514–0.828	0.393–0.676	0.381–0.773	0.39–0.833
Total tumor	AUC (p value)	0.646 (0.010*)	0.604 (0.103)	0.551 (0.445)	0.676 (0.011*)	0.515 (0.919)
	95 % CI	0.534–0.758	0.479–0.729	0.421–0.681	0.54–0.813	0.218–0.813
Necrosis/ Contrast ehancement	AUC (p value)	0.531 (0.655)	0.516 (0.855)	0.682 (0.001*)	0.642 (0.066)	0.612 (0.375)
	95 % CI	0.397–0.664	0.348–0.683	0.571–0.793	0.491–0.793	0.364–0.859
Contrast enhancment/ tumor bulk	AUC (p value)	0.531 (0.655)	0.516 (0.855)	0.682 (0.001*)	0.642 (0.066)	0.612 (0.375)
	95 % CI	0.397–0.664	0.348–0.683	0.571–0.793	0.491–0.793	0.364–0.859
T2-FLAIR hyperintesity/ total tumor	AUC (p value)	0.522 (0.755)	0.576 (0.256)	0.549 (0.488)	0.572 (0.526)	0.722 (0.026*)
	95 % CI	0.383–0.661	0.445–0.707	0.422–0.676	0.35–0.793	0.527–0.918
Tumor bulk/ total tumor	AUC (p value)	0.522 (0.755)	0.576 (0.256)	0.549 (0.488)	0.572 (0.526)	0.722 (0.026*)
	95 % CI	0.383–0.661	0.445–0.707	0.422–0.676	0.35–0.793	0.527–0.918

Finally, RB1 mutants were noted to have smaller T2-FLAIR hyperintensity and total tumor volumes but similar contrast enhancing and necrosis volumes compared to wild-type tumors, demonstrating the effect of RB1 on tumor-associated T2-FLAIR hyperintensity. One interpretation of these findings is that these mutations drive different growth patterns within individual tumors that are reflected in drastic differences in the imaged tumor phenotype (for example, highly necrosis/low CE vs. high CE/low necrosis). Correlations between our imaging features demonstrated that in general, the 11 features originally derived from MRI volumes are relatively independent measures of brain tumors characteristics that many have the potential to offer unique insight into tumor behavior (Fig. 3). This was also demonstrated by showing that different features predicted different mutations.

Several automatic and semi-automatic volumetric algorithms have been proposed to segment GBM tumors in relevant subvolumes [32]. For this work, we attempted a novel hybrid approach where we used a trained rater to mask the gross abnormal signal on the T1C and T2-weighted FLAIR image. The gross tumor volume on T1C is then stratified automatically into contrast enhancing and necrosis subvolumes (bright/dark pixels), using the FAST algorithm [33], improving the robustness of the segmentation process. We should note, however, that the segmentations were subsequently manually reviewed at various stages to ensure quality control.

Previous studies have investigated whether different biological subtypes confer different macroscopic properties to the images themselves and significant correlations between genetic expression and macroscopic imaging properties have been established [34, 35, 36]. Imaging genomics seeks to close the gap between genomics and neuroradiology to provide a comprehensive quantification of the tumor phenotype by applying a large number of automated image characterization algorithms [14, 20, 37]. In this paper, we applied a relatively low-dimensional feature extraction (focusing on three key volumes and derivatives for eleven features total). Future studies will expand these features and investigate the value of imaging genomics for the prediction of additional mutational patterns.

Going forward, the development of noninvasive imaging biomarkers will provide valuable insight to the clinicians to help in treatment selection and prognosis. As these biomarkers assess the entire tumor volume, they alleviate some of the concerns related to most tissue-based assessments that involve sampling only a small region of the tumor. Given the marked heterogeneity observed within tumor samples taken from even the same patient [38], a noninvasive technique that allows serial imaging (e.g., MRI) can provide valuable insight. Indeed, a limitation to our study is that in our patient set, TCGA tissue sampling was not performed under image guidance and therefore exact location of biopsy is not known. Future studies will investigate the association of intratumor mutational heterogeneity and MRI volumetric features.

In addition, it is important to note that since the TCIA imaging data used in our study was collected through a consortium of several institutions around the country, the

specific MR parameters (field strength, slice thickness, voxel size, slice gap) may not always be perfectly standardized across patients. However, our results should be relatively unaffected by issues of slice thickness, image quality, and voxel size, since we decided to analyze a set of volumetric features rather than measurements that would be more directly influenced by these subtle differences. Since the majority of the tumors were relatively large, and orders of magnitude larger than the size of an individual voxel (even accounting for voxel variability), similar to other papers published using this dataset, we therefore believe our conclusions are not largely influence by such factors. We should also note that there is no clear association between contributing site, and at least within a site the scanner(s) used were much more consistent than between contributing sites.

In conclusion, our results show that GBM mutations drive observable phenotypes that are quantifiable with MRI imaging. We demonstrate that somatic mutations are associated with macroscopic characteristics and that these clinically important mutations can be significantly predicted with high performance. These results may impact personalized medicine, as imaging is noninvasive and already applied routinely in clinical practice throughout a course of treatment. Finally, our results may shed insights into unique behavioral and macroscopically visible growth characteristics of individual tumors as a result of tumor mutational differences.

REFERENCES

1. Tran B, Rosenthal MA (2010) Survival comparison between glioblastoma multiforme and other incurable cancers. J Clin Neurosci 17(4):417–21
2. Thakkar JP, Dolecek TA, Horbinski C, Ostrom QT, Lightner DD, Barnholtz-Sloan JS et al (2014) Epidemiologic and molecular prognostic review of glioblastoma. Cancer Epidemiol Biomarkers Prev 23(10):1985–96
3. Piccirillo SGM, Colman S, Potter NE, van Delft FW, Lillis S, Carnicer M-J et al (2015) Genetic and functional diversity of propagating cells in glioblastoma. Stem Cell Rep 4(1):7–15C
4. Verhaak RGW, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD et al (2010) Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell 17(1):98–110
5. Rutman AM, Kuo MD (2009) Radiogenomics: creating a link between molecular diagnostics and diagnostic imaging. Eur J Radiol 70(2):232–41
6. Levin VA, Crafts DC, Norman DM, Hoffer PB, Spire JP, Wilson CB (1977) Criteria for evaluating patients undergoing chemotherapy for malignant brain tumors. J Neurosurg 47(3):329–35
7. Miller AB, Hoogstraten B, Staquet M, Winkler A (1981) Reporting results of cancer treatment. Cancer 47(1):207–14
8. Chinot OL, Macdonald DR, Abrey LE, Zahlmann G, Kerloëguen Y, Cloughesy TF (2013) Response assessment criteria for glioblastoma: practical adaptation and implementation in clinical trials of antiangiogenic therapy. Curr Neurol Neurosci Rep 13(5):347
9. Wen PY, Macdonald DR, Reardon DA, Cloughesy TF, Sorensen AG, Galanis E et al (2010) Updated Response Assessment Criteria for High-Grade Gliomas: Response Assessment in Neuro-Oncology Working Group. J Clin Oncol 28(11):1963–72
10. Macdonald DR, Cascino TL, Schold SC Jr, Cairncross JG (1990) Response criteria for phase II studies of supratentorial malignant glioma. J Clin Oncol 8(7):1277–80

11. Mazurowski MA, Desjardins A, Malof JM (2013) Imaging descriptors improve the predictive power of survival models for glioblastoma patients. Neuro Oncol 15(10):1389–94

12. Nicolasjilwan M, Hu Y, Yan C, Meerzaman D, Holder CA, Gutman D et al (2014) Addition of MR imaging features and genetic biomarkers strengthens glioblastoma survival prediction in TCGA patients. J Neuroradiol. doi:10.1016/j.neurad.2014.02.006

13. Gutman DA, Cooper LAD, Hwang SN, Holder CA, Gao J, Aurora TD et al (2013) MR imaging predictors of molecular profile and survival: multi-institutional study of the TCGA glioblastoma data set. Radiology 267(2):560–9

14. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Cavalho S et al (2014) Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 5:4006

15. Suzuki C, Torkzad MR, Jacobsson H, Aström G, Sundin A, Hatschek T et al (2010) Interobserver and intraobserver variability in the response evaluation of cancer therapy according to RECIST and WHO-criteria. Acta Oncol 49(4):509–14

16. Zhang Z, Jiang H, Chen X, Bai J, Cui Y, Ren X et al (2014) Identifying the survival subtypes of glioblastoma by quantitative volumetric analysis of MRI. J Neuro Oncol 119(1):207–14

17. Pope WB, Sayre J, Perlina A, Villablanca JP, Mischel PS, Cloughesy TF (2005) MR imaging correlates of survival in patients with high-grade gliomas. AJNR Am J Neuroradiol 26(10):2466–74

18. Iliadis G, Kotoula V, Chatzisotiriou A, Televantou D, Eleftheraki AG, Lambaki S et al (2012) Volumetric and MGMT parameters in glioblastoma patients: survival analysis. BMC Cancer 12:3

19. Zinn PO, Colen RR (2013) Imaging genomic mapping in glioblastoma. Neurosurgery 60(Suppl 1):126–30

20. ElBanan MG, Amer AM, Zinn PO, Colen RR (2015) Imaging genomics of glioblastoma: state of the art bridge between genomics and neuroradiology. Neuroimaging Clin N Am 25(1):141–53

21. Taylor TE, Furnari FB, Cavenee WK (2012) Targeting EGFR for treatment of glioblastoma: molecular basis to overcome resistance. Curr Cancer Drug Targets 12(3):197–209

22. Xu J, Li Z, Wang J, Chen H, Fang J-Y (2014) Combined PTEN mutation and protein expression associate with overall and disease-free survival of glioblastoma patients. Transl Oncol 7(2):196–205.e1

23. Dimitrov L, Hong CS, Yang C, Zhuang Z, Heiss JD (2015) New developments in the pathogenesis and therapeutic targeting of the IDH1 mutation in glioma. Int J Med Sci 12(3):201–13

24. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P et al (2013) The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. J Digit Imaging 26(6):1045–57

25. Prior FW, Clark K, Commean P, Freymann J, Jaffe C, Kirby J et al (2013) TCIA: An information resource to enable open science. Conf Proc IEEE Eng Med Biol Soc 2013:1282–5

26. CRAN - Package cgdsr [Internet]. [cited 2014 Oct 20]. Available from: http://CRAN.R-project.org/package=cgdsr

27. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. FSL. Neuroimage. 2012 Aug 15;62(2):782–90

28. Zhang Y, Brady M, Smith S (2001) Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans Med Imaging 20(1):45–57

29. Fawcett T (2006) An introduction to ROC analysis. Pattern Recogn Lett 27(8):861–74

30. Team RC. R: A language and environment for statistical computing. cran.case.edu; 2012; Available from: http://cran.case.edu/web/packages/dplR/vignettes/timeseries-dplR.pdf

31. Cancer Genome Atlas Research Network (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 455(7216):1061–8

32. Egger J, Kapur T, Fedorov A, Pieper S, Miller JV, Veeraraghavan H et al (2013) GBM volumetry using the 3D Slicer medical image computing platform. Sci Rep 3:1364

33. Zhang Y, Brady M, Smith S (2001) Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans Med Imaging 20(1):45–57

34. Barajas RF Jr, Hodgson JG, Chang JS, Vandenberg SR, Yeh R-F, Parsa AT et al (2010) Glioblastoma multiforme regional genetic and cellular expression patterns: influence on anatomic and physiologic MR imaging. Radiology 254(2):564–76

35. Diehn M, Nardini C, Wang DS, McGovern S, Jayaraman M, Liang Y et al (2008) Identification of noninvasive imaging surrogates for brain tumor gene-expression modules. Proc Natl Acad Sci U S A 105(13):5213–8

36. Gevaert O, Mitchell LA, Achrol AS, Xu J, Echegaray S, Steinberg GK et al (2014) Glioblastoma multiforme: exploratory radiogenomic analysis by using quantitative image features. Radiology 273(1):168–74

37. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P et al (2012) Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 48(4):441–6

38. Nobusawa S, Lachuer J, Wierinckx A, Kim YH, Huang J, Legras C et al (2010) Intratumoral patterns of genomic imbalance in glioblastomas. Brain Pathol 20(5):936–44

SUPPLEMENTARY MATERIAL

Available online at https://static-content.springer.com/esm/art%3A10.1007%2Fs00234-015-1576-7/MediaObjects/234_2015_1576_MOESM1_ESM.docx.

NOTES

Acknowledgments

Ethical standards and patient consent

We declare that this manuscript does not contain clinical studies or patient data where Institutional Review Board approval is required as patients had been previously de-identified by the TCGA and were available for public download.

Conflict of interest

We declare that we have no conflict of interest.

COPYRIGHT INFORMATION

Open Access

PART 3

Radiomics for Targeted Therapies

CHAPTER

5

Quantitative prognostic imaging biomarkers for risk stratification of patients with recurrent glioblastoma treated with bevacizumab

Quantitative prognostic imaging biomarkers for risk stratification of patients with recurrent glioblastoma treated with bevacizumab

Patrick Grossmann, Vivek Narayan, Ken Chang, Rifaquat Rahman, Lauren Abrey, David A. Reardon, Lawrence H. Schwartz, Patrick Y. Wen, Brian M. Alexander, Raymond Huang, and Hugo J.W.L. Aerts

ABSTRACT

Background
Anti-angiogenic therapy with bevacizumab is the most widely used treatment option for recurrent glioblastoma, but therapeutic response varies substantially and effective biomarkers for patient selection are not available. To this end, we determine whether novel quantitative radiomic strategies on the basis of MRI have the potential to noninvasively stratify survival and progression in this patient population.

Methods
In an initial cohort of 126 patients, we identified a distinct set of features representative of the radiographic phenotype on baseline (pretreatment) MRI. These selected features were evaluated on a second cohort of 165 patients from the multicenter BRAIN trial with prospectively acquired clinical and imaging data. Features were evaluated in terms of prognostic value for overall survival (OS), progression-free survival (PFS), and progression within 3, 6, and 9 months using baseline imaging and first follow-up imaging at 6 weeks posttreatment initiation.

Results
Multivariable analysis of features derived at baseline imaging resulted in significant stratification of OS (hazard ratio [HR] = 2.5; log-rank P = 0.001) and PFS (HR = 4.5; log-rank P = 2.1×10^{-5}) in validation data. These stratifications were stronger compared with clinical or volumetric covariates (permutation test false discovery rate [FDR] <0.05). Univariable analysis of a prognostic textural heterogeneity feature (information correlation) derived from postcontrast T1-weighted imaging revealed significantly higher scores for patients who progressed within 3 months (Wilcoxon test P = 8.8×10^{-8}). Generally, features derived from postcontrast T1-weighted imaging yielded higher prognostic power compared with precontrast enhancing T2-weighted imaging.

Conclusion
Radiomics provides prognostic value for survival and progression in patients with recurrent glioblastoma receiving bevacizumab treatment. These results could lead to the development of quantitative pretreatment biomarkers to predict benefit from bevacizumab using standard of care imaging.

Keywords
Bevacizumab – recurrent glioblastoma – radiomics – survival

IMPORTANCE OF THE STUDY

Glioblastoma is the most common and aggressive primary malignant brain tumor and recurrence is almost inevitable. At recurrence, patients are faced with poor prognosis due to limited treatment options. One widely used treatment for recurrent glioblastoma in the United States is anti-angiogenic therapy with bevacizumab. However, treatment response varies substantially and effective biomarkers for stratification are absent. To this end, we propose an emerging approach, radiomics, that translates standard radiographic images into quantitative data for in-depth analysis of radiographic tumor phenotypes. We applied radiomics to prospective clinical and imaging data from the multicenter BRAIN trial. Our analysis indicates that this approach yields significant prognostic value for OS and PFS at both baseline and follow-up imaging. These results prompt further prospective validation to develop a noninvasive pretreatment biomarker from standard of care medical images for the purpose of objectively predicting treatment response in patients with recurrent glioblastoma treated with bevacizumab.

INTRODUCTION

Glioblastoma remains the most aggressive primary malignant brain tumor, with a median survival time of 15 months and a 5-year survival of ~5% after initial diagnosis.[1–3] Despite standard of care therapy, including maximal safe surgical resection, radiotherapy, and temozolomide,[4] nearly all patients relapse.[5]

One commonly used treatment option at recurrence in the US is bevacizumab, an inhibitor of vascular endothelial growth factor developed to block angiogenesis.[6] The noncomparative randomized phase II, multicenter, open-label BRAIN trial (AVF3708g) investigating bevacizumab plus irinotecan versus bevacizumab alone[7,8] contributed to accelerated FDA approval of bevacizumab (Avastin, Genentech) for treatment of recurrent glioblastoma in 2009. Despite its current US approval for treatment of recurrent glioblastoma, phase II and III clinical trials for patients with recurrent glioblastoma,[9,10] as well as 2 recent phase III randomized clinical trials for patients with newly diagnosed glioblastoma,[11,12] found no improvement in overall survival (OS) with the addition of bevacizumab to standard therapy.

However, it is reasonable to hypothesize that there may be a patient population with a meaningful clinical benefit, given that the demonstrated activity of bevacizumab evidenced by impact on imaging-based endpoints[13–15] and the development of novel biomarkers will be critical to identifying such patients. Imaging biomarkers that could predict response to therapy, have significant prognostic value, or can identify patients unlikely to respond to any therapy would be an extremely valuable tool in this regard.

Radiomics is an emerging field that translates medical images into quantitative data by applying a large set of feature extraction algorithms to characterize tumor phenotype.[16–18] Two advantages of radiomics are its noninvasive approach and the ability to evaluate the entire visible tumor and any intratumoral heterogeneity of glioblastoma.[18–20] Previous studies have employed a radiomic approach to glioblastoma[21–24] reporting prognostic and predictive value of quantitative imaging features; moreover, associations of radiomics with underlying molecular profiles have been suggested.[25,26]

In this study, we present a radiomic analysis of patients with recurrent glioblastoma treated with bevacizumab. In an independent selection cohort, we identified a set of distinct quantitative imaging features from MRI defining radiographic tumor phenotype. We evaluated these features on prospectively acquired clinical and imaging data from the BRAIN trial, to determine whether there was an association with posttherapy survival and progression at baseline (pretreatment) and at follow-up at week 6 (posttreatment initiation). Hereby, we aimed at assessing whether such a radiomic approach can be valuable for the development of noninvasive biomarkers for risk stratification in this patient population.

MATERIALS AND METHODS

Study cohorts

The current study was covered by 13–055 Partners institutional review board (Aerts).

Selection cohort

To independently determine the set of radiomic features that define the radiographic phenotype of a tumor and to avoid false discoveries due to a high-dimensional radiomic feature space, we leveraged a retrospectively collected cohort of 126 nonconsecutive patients at our institution for unsupervised feature selection blinded from clinical data. Patients had pathologically confirmed glioblastoma of at least one recurrence and received bevacizumab either alone or with irinotecan between December 2006 and August 2014 after standard of care treatment, including surgical resection followed by radiotherapy as well as concurrent and adjuvant temozolomide. Further baseline characteristics are detailed in Table 1.

BRAIN trial

Features selected on the selection cohort were retrospectively analyzed on prospectively acquired data from the phase II, multicenter, open-label, randomized, noncomparative BRAIN trial (AVF3708g).[7] The trial randomized 167 patients to receive either bevacizumab alone ($n = 85$) or in combination with irinotecan ($n = 82$) at one of 10 treatment sites. Pa-

tients were enrolled between July 2006 and September 2007, were at least 18 years of age with histologically confirmed contrast-enhancing glioblastoma in first or second relapse, and had failed first-line standard therapy including concurrent radiotherapy and te- mozolomide. There were at least 8 weeks between completion of radiotherapy and study enrollment. Baseline Karnofsky performance status (KPS) at start of trial was ≥70. Treat- ment length of the trial was 104 weeks or until progression, death, or discontinuation. Two patients in the BRAIN cohort were treated at our institute and were therefore includ- ed in the selection cohort. Further baseline characteristics are detailed in Table 1. Data acquisition was compliant with the Health Insurance Portability and Accountability Act and had received approval of the respective local institutional review boards. Before trial par- ticipation, patients provided written informed consent.

Table 1
Patient baseline characteristics.

Characteristic	Value in Selection Cohort	Value in BRAIN Cohort
Median age, y (range)	57 (24–88)	56 (23–79)
Median KPS (range)	Not available	80 (70–100)
OS, mo	8.3	9.3
PFS, mo	3.2	4.4
Sex		
n female	48 (38%)	52 (32%)
n male	78 (62%)	113 (68%)

To reduce the effect of slice thickness variation, all images were resampled to voxels of size $3 \times 3 \times 3$ mm^3 (initial dimensions: 1–5 mm in Z direction). For further data normalization, voxel values were discretized with a binwidth of 25 and voxels included in feature extraction were constrained to an intensity value range of 3 standard deviations from the mean. From the total BRAIN cohort, MRIs were available for 165 patients. Specif- ically, postcontrast T1-weighted and FLAIR imaging was available at baseline for 160 and 152 patients, respectively, and at follow-up for 117 and 133 patients.

Radiomics

For both study cohorts, we extracted 65 quantitative imaging features each from T1 and FLAIR scans at baseline (pretreatment) and follow-up after 6 weeks (posttreatment initia- tion) using our radiomics pipeline (Fig. 1). These features can be grouped into 3 types: (i)

first-order statistics of the voxel intensity histogram, (ii) tumor shape, and (iii) tumor texture. Feature definitions are detailed in Supplementary File S1, Supplementary Table S1, and Supplementary Table S2. In addition to baseline and follow-up, we calculated the percentage difference (delta) of feature values by delta = 100 * (F–B) / B, where B and F are the values at baseline and follow-up, respectively.
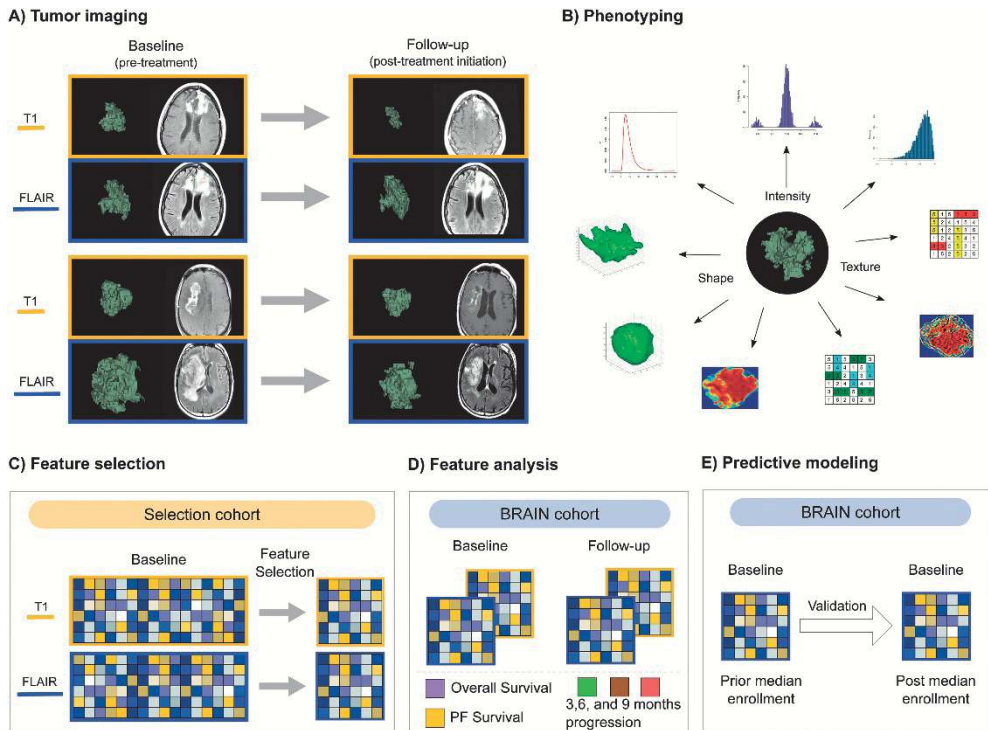
Fig. 1

Study design. (A) Patients underwent MRI at baseline (pretreatment) and follow-up at 6 weeks after treatment initiation. Tumors were segmented in 3D on the basis of postcontrast T1-weighted and FLAIR imaging. (B) From each segmentation, the tumor imaging phenotype was defined with radiomics to quantify tumor intensity, shape, and texture features. (C) To avoid multiple-hypothesis testing, we identified a canonical set of representative features with unsupervised feature selection on the selection cohort. (D) These selected features were independently evaluated on the BRAIN cohort to (E) develop prognostic biomarkers of survival and progression, trained and validated on the first and second BRAIN cohort half, respectively.

To reduce dimensionality, we leveraged the selection cohort to focus our analysis on a set of representative features. For this, we implemented an unbiased unsupervised feature selection blinded from clinical information. We performed this selection on baseline features derived from postcontrast T1-weighted and FLAIR imaging separately. First,

we applied a principal component (PC) analysis and retained those PCs that accounted for 99% of the total variance in the data. Next, we selected the unique set of features that contributed to all PCs the most (≥0.9 rank correlation) using a factor analysis.[29] Mean Spearman rank correlation of a particular feature with all other features was restricted to ≤0.9. From this remaining feature set, the 10 most variant features were selected according to the coefficient of variation. To compare radiomics to traditional radiological assessments, we calculated maximal axial diameter and total tumor volume automatically from the tumor segmentations and added these 2 volumetric size features to our feature set. In summary, we analyzed 12 features derived each from postcontrast T1-weighted and FLAIR images.

Statistical Analysis

The features selected externally from the selection cohort were independently evaluated on the BRAIN cohort. All statistical analyses were carried out with R[30] version 3.1.0 on a Linux environment. Details of library versions are given in Supplementary File S2.

Endpoints
We considered OS, progression-free survival (PFS), as well as progression at 3, 6, and 9 months as endpoints. OS was defined as time from treatment start to death.[7] PFS was defined in the BRAIN trial as the time from treatment start to progression or death, whichever occurred first.[7] Progression at 3, 6, and 9 months correspond to the 6-week scan intervals of the BRAIN trial protocol. Median OS and PFS were 9.3 and 4.4 months, respectively. To cover a comprehensive analysis, prognostic value of our feature set was assessed each for postcontrast T1-weighted and FLAIR imaging at baseline and follow-up, and delta while correcting for multiple-hypothesis testing as described below.

Univariable analysis
We assessed the prognostic value of individual features for OS and PFS with the concordance index (CCI),[31,32] which takes time censoring into account. Prognostic value for binary progression at 3, 6, and 9 months was evaluated with the area under the curve (AUC) of the receiver operator characteristic.[33]

Significances of CCIs and AUCs under the null hypothesis that predictor and outcome variables are not associated (ie, value of 0.5) were tested 2-sided using Noether's procedure.[34] *P*-values for all features derived from postcontrast T1-weighted and FLAIR images were corrected for multiple-hypothesis testing using the false discovery rate (FDR) procedure according to Benjamini and Hochberg.[35] Features with corrected significant performance were also tested for direct distributional differences between patients who progressed within 3 months with the 2-sided Wilcoxon rank-sum test.[36]

To investigate stratification power of individual features, patients were assigned to lower and higher risk groups depending on whether their feature value was higher than

the median value across all patients or not. Hazard ratios (HRs) and *P*-values were calculated using univariable Cox proportional hazards models and one-sided Wald tests of Cox variables, respectively.

Training and validation data

For an unbiased assignment of patients into a training and validation set, we used the median treatment date of the full BRAIN trial cohort. To mimic prospective validation as closely as possible, data from patients treated before the median treatment date were used for training and data from patients with later dates were used for validation. If no treatment date was available, patients were assigned to the training set (*n* = 4) for more power in fitting. This split resulted in *n* = 86 and *n* = 81 patients for training and validation, respectively. From the training (validation) data, n_t = 82 (n_v = 78), n_t = 59 (n_v = 58), and n_t = 57 (n_v = 58) samples were available for postcontrast T1-weighted imaging at baseline, follow-up, and delta, respectively; n_t = 80 (n_v = 72), n_t = 71 (n_v = 62), and n_t = 69 (n_v = 62) samples were available for FLAIR imaging at baseline, follow-up, and delta, respectively.

Multivariable analysis

The univariable analysis indicated strong prognostic value for features derived from post-contrast T1-weighted baseline imaging. To increase this prognostic performance, we aimed at combining individual features in multivariable models. To not overfit our models and to subsequently compare with volumetric and clinical variables, we selected complementary radiomic features with supervised forward feature selection. First, we ranked all 10 radiomic features with the minimal redundancy maximal relevance algorithm[37] using the training baseline data. Next, those features were sequentially added into a growing Cox proportional hazards model for PFS or OS starting with the top ranked feature. After each iteration (ie, addition of a feature) the performance was assessed with the CCI and validated with repeated random cross-validation (*n* = 1000 resamples and 80%/20% split of training data). The final model feature set was obtained by stopping addition of features when the cross-validated mean CCI did not further increase (forward selection). This procedure was performed solely on the training set and selected 2 features for PFS (Gray-level co-occurrence matrix [GLCM] information correlation and GLCM correlation) and 1 feature for OS (GLCM information correlation).

Selected features were used to fit a random-forest classifier (*n* = 500 trees) on the training set to predict one-year OS, as well as progression at 3, 6, and 9 months. Parameter optimization was done with repeated 5-fold cross-validation (*n* = 10 repeats) choosing optimal decision tree splits with the one standard error method according to Breiman et al[38] using accuracy as evaluation metric. Performance of the fitted models was evaluated with HRs and log-rank tests on the validation set.

For comparison with clinical and volumetric models, the above procedure describing random-forest model building was applied to create a model with age, sex, and KPS, and to create a model with maximal axial diameter and total tumor volume. Perfor-

mance of radiomic models were compared with these clinical and volumetric models with repeated random permutation tests ($N = 1000$ resamples) of the CCI differences; *P*-values were corrected with the FDR.

RESULTS

Identification of an Independent Radiomic Feature Set

To determine the canonical set of features able to comprehend the radiographic characteristics of recurrent glioblastoma on MRI (Fig. 1A–B), we used unsupervised feature selection on a selection cohort (Fig. 1C). This defined a subset of 10 radiomic features derived each from baseline postcontrast T1-weighted and FLAIR images. These features quantify a wider range of intensity, shape, and texture characteristics of a tumor (Supplementary Tables S1 and S2) and were compared to two standard volumetric features (ie, maximal axial diameter and total tumor volume).

This feature set was independently evaluated on the BRAIN trial cohort (Fig. 1D–E). Clustering analysis of radiomic features indicated differences between phenotypic imaging profiles of patients (Fig. 2A). Identified clusters were independent of baseline KPS; mean Spearman rank correlation to KPS was $\rho = 0.06$ and $\rho = 0.13$ for postcontrast T1-weighted and FLAIR imaging, respectively. Furthermore, features had only low pairwise correlation (mean absolute Spearman's $\rho = 0.31$), even when derived from the same imaging modality, suggesting that these features provide complementary information (Fig. 2B and Supplementary Fig. S1).

Fig. 2

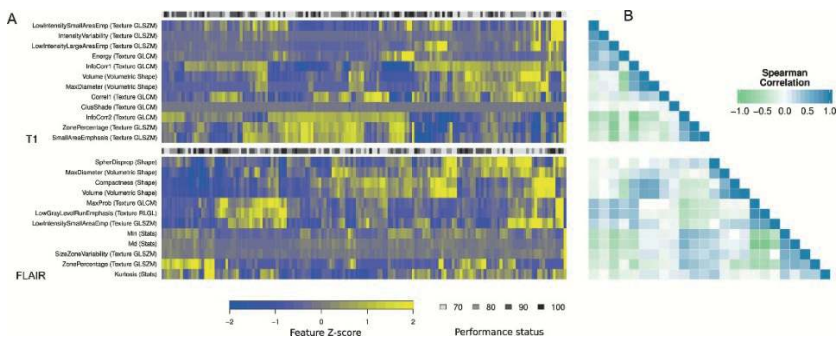*Independence of features. (A) Visualization of features derived from baseline postcontrast T1-weighted (top) and FLAIR imaging (bottom). Values are given as Z-scores and group patients based on hierarchical clustering. (B) Pairwise Spearman rank correlation between all features at baseline imaging. Correlation among features was low, indicating interfeature independence (mean absolute Spearman's $\rho = 0.31$).*

Verification of Prognostic Value of Individual Features

We evaluated the prognostic value of our feature set for OS and PFS, as well as progression within 3, 6, and 9 months for baseline and follow-up imaging and for the delta between baseline and follow-up (Fig. 3). In general, more features derived from postcontrast T1-weighted than from FLAIR imaging showed significant prognostic value (CCI, Noether FDR < 0.05). At baseline imaging, only those features derived from postcontrast T1-weighted imaging performed significantly.

　　　Moreover, distributions of prognostic features were significantly different in patients who progressed at different time points when derived from postcontrast T1-weighted imaging (Supplementary Figure S2). For example, the textural heterogeneity feature information correlation had significantly greater scores for patients who progressed within 3 months (Wilcoxon rank-sum test $P = 8.8 \times 10^{-8}$). This feature was significantly predictive of all tested endpoints under postcontrast T1-weighted baseline imaging, but performed highest when predicting patients who progressed within 6 months from follow-up imaging (AUC = 0.68, Noether FDR = 0.0019). Furthermore, at baseline imaging this feature significantly stratified patients into high- and low-risk groups of OS (HR = 1.7, Wald test $P = 0.0017$) and PFS (HR = 1.7, Wald test $P = 0.0029$). All other features showed lower discrimination at baseline and follow-up imaging.

　　　On the delta between baseline and follow-up imaging, the highest performing feature was energy of the imaging texture for progression within 6 months (AUC = 0.65, Noether FDR = 0.04) with moderate risk stratification power (HR = 0.54, 95% CI = [0.36, 0.82], Wald test $P = 0.0037$). Further stratification results are detailed in Supplementary Figure S3 and Supplementary Table S3.

Baseline Models of Survival and Progression

To develop pretreatment biomarkers for survival and progression, we assigned patients into a training and validation set based on median treatment date. First, we trained a model to predict one-year OS from baseline postcontrast T1-weighted imaging. As shown in Fig. 4, this model significantly stratified OS of patients from the validation set (HR = 2.5, 95% CI = [1.5, 4.4]; log-rank $P = 0.001$) after adjusting for age, sex, and KPS, and without adjusting (HR = 2.3, 95% CI = [1.4, 3.8]).

　　　We trained additional models to predict progression at 3, 6, and 9 months using baseline postcontrast T1-weighted imaging (Fig. 5). These models significantly stratified PFS of patients from the validation set (adjusted; 3 months progression: HR = 2.8, 95% CI =
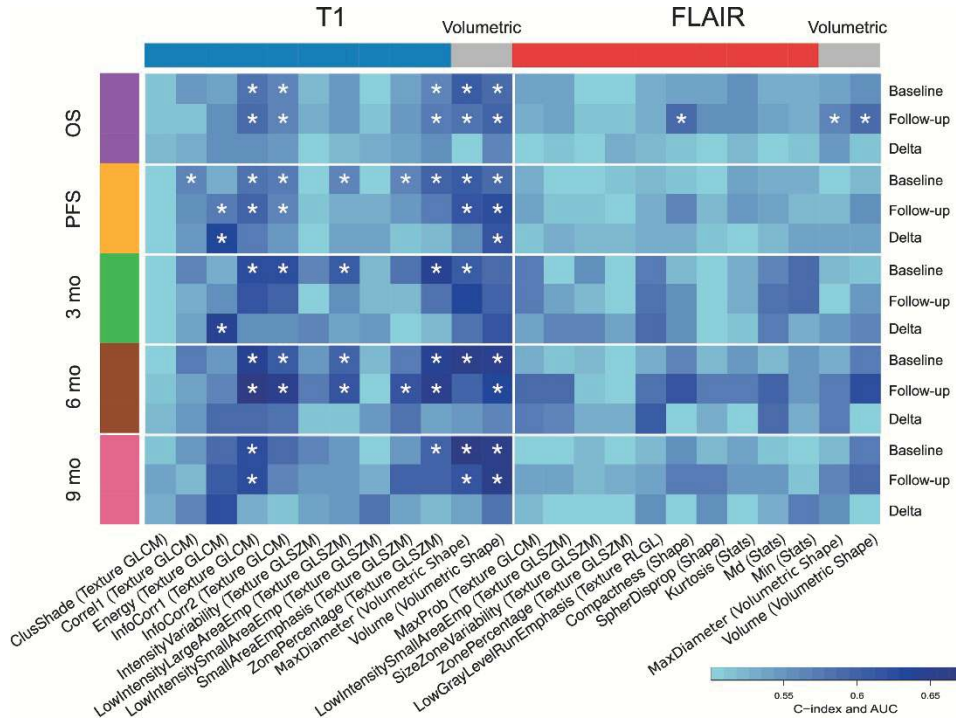
Fig. 3

Univariable performances. Prognostic value of features at treatment baseline, 6-week follow-up, and the delta difference of 10 radiomic features for T1 and FLAIR, respectively. Standard volumetric features are shown for comparison. Performances were evaluated with the CCI for OS and PFS, and with the AUC of the receiver operator characteristic for progression at 3, 6, and 9 months. Asterisks indicate significance (FDR < 0.05) after correcting for multiple-hypothesis testing according to Benjamini and Hochberg. Generally, performance of features derived from postcontrast T1-weighted imaging tended to be higher than features derived from FLAIR imaging.

[1.6, 4.8], log-rank $P = 5.9 \times 10^{-4}$; 6 months progression: HR = 3.8, 95% CI = [2.1, 6.7], log- rank $P = 2.1 \times 10^{-5}$; and 9 months progression: HR = 4.5, 95% CI = [2.2, 9.2], log-rank $P = 3.5 \times 10^{-5}$). Sensitivity and specificity for the internal classification of one-year OS, as well as for the classification of 3, 6, and 9 months progression, are reported in Supplementary Table S4. Furthermore, no impact of treatment arm was observed on the performance of these models.
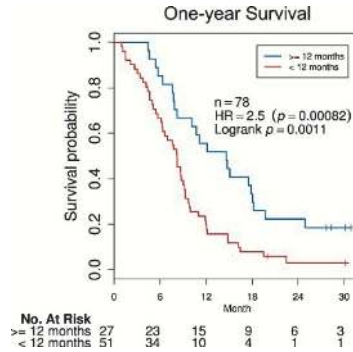
Fig. 4

Survival analysis of a multivariable radiomics model for OS. A biomarker with radiomic features was optimized to predict one-year OS. This marker showed significant pretreatment stratification power in discriminating patients at higher and lower risk in validation data held independent of training (HR = 2.5; Wald test P = 8.2 × 10^{-4}; log-rank P = 0.0011) after adjusting for age, sex, and KPS at baseline.

All 4 radiomic models performed significantly better (HR permutation test FDR < 0.05) than models of clinical covariates only (ie, age, sex, and KPS) or volumetric features only (ie, maximal diameter and tumor volume). The radiomic models also performed better than models combining clinical and volumetric covariates at slightly lower significance (FDR < 0.09). Stratification was not improved when combining volumetric and radiomic features or combining radiomic features derived from postcontrast T1-weighted and FLAIR imaging (Supplementary Figure S4A and Supplementary Figure S4B, respectively).

DISCUSSION

We present a retrospective radiomic analysis of prospectively acquired data from the BRAIN trial to develop noninvasive quantitative prognostic biomarkers for survival and progression on the basis of standard MRI for patients with recurrent glioblastoma treated with bevacizumab. Our results suggest that a radiomic approach provides prognostic insight about response of these patients, which could lead to further investigation for predictive biomarkers to meet the current need of stratifying patients, as bevacizumab has recently been suggested not to improve OS.[11,12,39]
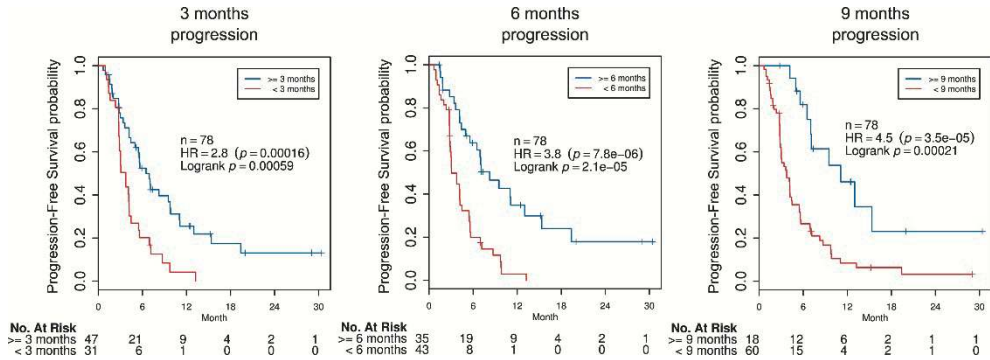
Fig. 5

Survival analysis of multivariable radiomics models for PFS. Biomarkers were optimized to predict progression at 3, 6, and 9 months. All models significantly stratified patients in validation data; hazard ratios ranged from 2.8 to 4.5 (Wald test $P \leq 1.6 \times 10^{-4}$; log-rank $P \leq 5.9 \times 10^{-4}$) after adjusting for age, sex, and KPS at baseline.

Our study provides baseline models for OS of patients, as well as baseline models for progression at 3, 6, and 9 months. On the basis of pretreatment imaging, these models were able to identify patients in the validation set who had more than 4 times the risk to progress within 6 and 9 months throughout the course of treatment. With increasing time point, these models consistently performed better, suggesting that particularly patients who experience progression late can be identified. Furthermore, these models performed substantially better than individual features alone. Importantly, radiomic features in these models were independent of volumetric features and age, sex, and KPS, which were the clinical variables available to us and which are known clinical prognostic factors for glioblastoma[2]; however, age and sex may be less prognostic for OS than KPS, as has only recently been suggested.[40]

Consequently, our results could lead to the development of a noninvasive biomarker that would identify patients who are more likely to benefit from bevacizumab even before treatment initiation. However, to determine whether any proposed radiomic biomarker also is predictive of positive treatment outcome, a prospective control arm without anti-angiogenic therapy would be necessary. Alternatively, failure of our model to predict similar endpoints in a non-bevacizumab containing cohort would offer evidence to support a predictive hypothesis. Furthermore, prospective validation of our models in follow-up studies will be required.

Our multivariable analysis included information correlation, which quantifies textural-imaging heterogeneity. Not only did this feature yield significant prognostic baseline value across all tested endpoints, but it also showed higher scores for patients who progressed early, suggesting that textural-imaging heterogeneity is a prognostic factor. Univariable analysis of our full feature set indicated further prognostic candidates; however, all features had significantly less performance compared with our fully trained models.

Both our univariable and multivariable analyses indicated that postcontrast T1-weighted imaging data provide more prognostic information than FLAIR imaging, irrespective of investigated time- and endpoints. It remains a difficulty to hypothesize why FLAIR imaging is less prognostic; however, previous studies have documented that tumor volume was not prognostic under FLAIR in recurrent glioblastoma and bevacizumab treatment.[41–43] Overall, a general tendency toward higher predictability of PFS was observed, which could be explained by a more clearly defined endpoint than OS, as patients who progressed were discontinued from study treatment and observed for OS. Differences between treatment arms in the BRAIN cohort were not expected and hence data have been pooled, as published previously.[8,44,45]

Robustness of our results is increased, given that we leveraged 2 independent cohorts to select and evaluate, respectively, our feature set; this limits false discoveries in a high-dimensional approach such as radiomics. Importantly, we blinded the feature selection from clinical data to avoid information leakage and hence increase validity even further. We chose to use our institutional cohort as the selection cohort, as the heterogeneous nature of this retrospective data implies less suitability for fitting models compared with the prospectively acquired BRAIN data. Conversely, our study benefits from evaluation of radiomic features in prospectively acquired clinical trial data of high quality, which includes predefined criteria for patient enrollment, treatment, and follow-up. As this was a multicenter trial with 10 sites, MRI protocols specifying imaging acquisition parameters and equipment have varied across institutions, further highlighting the potential generalizability of our proposed models. Especially for texture-based features we do not expect substantial variability, as those features are generally relative measurements. In these ways, our study advances the results of a recent study indicating prognostic value of radiomics in a retrospectively acquired cohort of patients with recurrent glioblastoma receiving bevacizumab,[46] which was also demonstrated by the high performance of our validated models. In addition, we provide further insights into the relationship of intratumoral heterogeneity in imaging and patient survival, as well as comparative analyses of commonly used MRI modalities and measurements of tumor burden.

Other approaches to develop imaging biomarkers were explored in comparable cohorts.[8,13,15,47–49] For example, apparent diffusion coefficient histogram analysis has been previously conducted on a subset of the BRAIN cohort totaling 97 patients.[48] As radiomics can be applied to standard imaging, complementary value with those approaches could be tested when these data become available. In addition, integration of imaging with biological assays could yield valuable insight into tumor progression[50,51] and allow early identification of resistances.[52,53] This could also contribute to unraveling the mechanistic connections between radiomics, tumor biology, and clinical outcomes.

In conclusion, we demonstrated the strong prognostic value of radiomics to predict survival and progression of patients with recurrent glioblastoma treated with bevacizumab in prospectively acquired data from the BRAIN trial. Our results could lead to further development of a noninvasive predictive biomarker based on standard of care pre-

treatment imaging to identify a subpopulation of patients who would benefit from bevaci-zumab.

SUPPLEMENTARY DATA

Supplementary material is available at *Neuro-Oncology* online.

REFERENCES

1. Ostrom QT , Gittleman H , Fulop J et al CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2008–2012. Neuro Oncol . 2015;17(Suppl 4):iv1–iv62.
2. Thakkar JP , Dolecek TA , Horbinski C et al Epidemiologic and molecular prognostic review of glioblastoma. Cancer Epidemiol Biomarkers Prev . 2014;23(10):1985–1996.
3. Ostrom QT , Bauchet L , Davis FG et al The epidemiology of glioma in adults: a "state of the science" re-view. Neuro Oncol . 2014;16(7):896–913.
4. Stupp R , Mason WP , van den Bent MJ et al ; European Organisation for Research and Treatment of Cancer Brain Tumor and Radiotherapy Groups; National Cancer Institute of Canada Clinical Trials Group. Radio-therapy plus concomitant and adjuvant temozolomide for glioblastoma. N Engl J Med . 2005;352(10):987–996.
5. Stupp R , Hegi ME , Mason WP et al ; European Organisation for Research and Treatment of Cancer Brain Tumour and Radiation Oncology Groups; National Cancer Institute of Canada Clinical Trials Group. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in gli-oblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. Lancet Oncol . 2009;10(5):459–466.
6. Ferrara N , Hillan KJ , Gerber HP , Novotny W . Discovery and development of bevacizumab, an anti-VEGF antibody for treating cancer. Nat Rev Drug Discov . 2004;3(5):391–400.
7. Friedman HS , Prados MD , Wen PY et al Bevacizumab alone and in combination with irinotecan in recur-rent glioblastoma. J Clin Oncol . 2009;27(28):4733–4740.
8. Ellingson BM , Kim HJ , Woodworth DC et al Recurrent glioblastoma treated with bevacizumab: contrast-enhanced T1-weighted subtraction maps improve tumor delineation and aid prediction of survival in a multicenter clinical trial. Radiology . 2014;271(1):200–210.
9. Wick W , Stupp R , Gorlia T et al Phase II part of EORTC study 26101: the sequence of bevacizumab and lomustine in patients with first recurrence of a glioblastoma. ASCO Meeting Abstracts . 2016;34(15_suppl):2019.
10. Wick W , Brandes AA , Gorlia T et al LB-05: phase III trial exploring the combination of bevacizumab and lomustine in patients with first recurrence of a glioblastoma: the EORTC 26101 trial. Neuro Oncol . 2015;17(suppl 5):v1.
11. Gilbert MR , Dignam JJ , Armstrong TS et al A randomized trial of bevacizumab for newly diagnosed glio-blastoma. N Engl J Med . 2014;370(8):699–708.
12. Chinot OL , Wick W , Mason W et al Bevacizumab plus radiotherapy-temozolomide for newly diagnosed glioblastoma. N Engl J Med . 2014;370(8):709–722.
13. Chang K , Zhang B , Guo X et al Multimodal imaging patterns predict survival in recurrent glioblastoma patients treated with bevacizumab. Neuro Oncol . 2016;18(12):1680–1687.

14. Ellingson BM , Sahebjam S , Kim HJ et al Pretreatment ADC histogram analysis is a predictive imaging biomarker for bevacizumab treatment but not chemotherapy in recurrent glioblastoma. AJNR Am J Neuroradiol . 2014;35(4):673–679.

15. Pope WB , Kim HJ , Huo J et al Recurrent glioblastoma multiforme: ADC histogram analysis predicts response to bevacizumab treatment. Radiology . 2009;252(1):182–189.

16. Gillies RJ , Kinahan PE , Hricak H . Radiomics: images are more than pictures, they are data. Radiology . 2016;278(2):563–577.

17. Aerts HJ , Velazquez ER , Leijenaar RT et al Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun . 2014;5:4006.

18. Lambin P , Rios-Velazquez E , Leijenaar R et al Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer . 2012;48(4):441–446.

19. Kim H , Zheng S , Amini SS et al Whole-genome and multisector exome sequencing of primary and posttreatment glioblastoma reveals patterns of tumor evolution. Genome Res . 2015;25(3):316–327.

20. Kim J , Lee IH , Cho HJ et al Spatiotemporal evolution of the primary glioblastoma genome. Cancer Cell . 2015;28(3):318–328.

21. Macyszyn L , Akbari H , Pisapia JM et al Imaging patterns predict patient survival and molecular subtype in glioblastoma via machine learning techniques. Neuro Oncol . 2016;18(3):417–425.

22. Kickingereder P , Burth S , Wick A et al Radiomic profiling of glioblastoma: identifying an imaging predictor of patient survival with improved performance over established clinical and radiologic risk models. Radiology . 2016;280(3):880–889.

23. Yang D , Rao G , Martinez J , Veeraraghavan A , Rao A . Evaluation of tumor-derived MRI-texture features for discrimination of molecular subtypes and prediction of 12-month survival status in glioblastoma. Med Phys . 2015;42(11):6725–6735.

24. Cui Y , Tha KK , Terasaka S et al Prognostic imaging biomarkers in glioblastoma: development and independent validation on the basis of multiregion and quantitative analysis of MR images. Radiology . 2015;1278(2):546–553.25.   Itakura H , Achrol AS , Mitchell LA et al Magnetic resonance image features identify glioblastoma phenotypic subtypes with distinct molecular pathway activities. Sci Transl Med . 2015;7(303):303ra138.

26. Gevaert O , Mitchell LA , Achrol AS et al Glioblastoma multiforme: exploratory radiogenomic analysis by using quantitative image features. Radiology . 2014;273(1):168–174.

27. Rios Velazquez E , Aerts HJ , Gu Y et al A semiautomatic CT-based ensemble segmentation of lung tumors: comparison with oncologists' delineations and with the surgical specimen. Radiother Oncol . 2012;105(2):167–173.

28. Huang RY , Rahman R , Hamdan A et al Recurrent glioblastoma: volumetric assessment and stratification of patient survival with early posttreatment magnetic resonance imaging in patients treated with bevacizumab. Cancer . 2013;119(19):3479–3488.

29. Lê S , Josse J , Husson F . FactoMineR: an R package for multivariate analysis. J Stat Softw . 2008;25(1):1–18.

30. R Development Core Team. R Development Core Team (2013). R: A language and environment for statistical computing . R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

31. Harrell FE Jr , Lee KL , Mark DB . Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med . 1996;15(4):361–387.

32. Schröder MS , Culhane AC , Quackenbush J , Haibe-Kains B . survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. Bioinformatics . 2011;27(22):3206–3208.

33. Fawcett T . An introduction to ROC analysis. Pattern Recognit Lett . 2006;27(8):861–874.

34. Pencina MJ , D'Agostino RB . Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. Stat Med . 2004;23(13):2109–2123.

35. Benjamini Y , Hochberg Y . Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol . 1995;57(1):289–300.

36. Bauer DF . Constructing confidence sets using rank statistics. J Am Stat Assoc . 1972;67(339):687–690.

37. Peng H , Long F , Ding C . Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell . 2005;27(8):1226–1238.

38. Breiman L , Friedman J , Stone CJ , Olshen RA. Classification and Regression Trees . Belmont, CA: Wadsworth International Group; 1984.

39. Fu P , He YS , Huang Q et al Bevacizumab treatment for newly diagnosed glioblastoma: systematic review and meta-analysis of clinical trials. Mol Clin Oncol . 2016;4(5):833–838.

40. Schaub C , Tichy J , Schäfer N et al Prognostic factors in recurrent glioblastoma patients treated with bevacizumab. J Neurooncol . 2016;129(1):93–100.

41. Ellingson BM , Cloughesy TF , Lai A , Nghiemphu PL , Mischel PS , Pope WB . Quantitative volumetric analysis of conventional MRI response in recurrent glioblastoma treated with bevacizumab. Neuro Oncol . 2011;13(4):401–409.

42. Grossman R , Shimony N , Shir D et al Dynamics of FLAIR volume changes in glioblastoma and prediction of survival. Ann Surg Oncol . 2017;24(3):794–800.

43. Schmainda KM , Prah M , Connelly J et al Dynamic-susceptibility contrast agent MRI measures of relative cerebral blood volume predict response to bevacizumab in recurrent high-grade glioma. Neuro Oncol . 2014;16(6):880–888.

44. Huang RY , Rahman R , Hamdan A et al Recurrent glioblastoma: volumetric assessment and stratification of patient survival with early posttreatment magnetic resonance imaging in patients treated with bevacizumab. Cancer . 2013;119(19):3479–3488.

45. Pope WB , Qiao XJ , Kim HJ et al Apparent diffusion coefficient histogram analysis stratifies progression-free and overall survival in patients with recurrent GBM treated with bevacizumab: a multi-center study. J Neurooncol . 2012;108(3):491–498.

46. Kickingereder P , Götz M , Muschelli J et al Large-scale radiomic profiling of recurrent glioblastoma identifies an imaging predictor for stratifying anti-angiogenic treatment response. Clin Cancer Res . 2016;22(23):5765–5771.

47. Rahman R , Hamdan A , Zweifler R et al Histogram analysis of apparent diffusion coefficient within enhancing and nonenhancing tumor volumes in recurrent glioblastoma patients treated with bevacizumab. J Neurooncol . 2014;119(1):149–158.

48. Pope WB , Qiao XJ , Kim HJ et al Apparent diffusion coefficient histogram analysis stratifies progression-free and overall survival in patients with recurrent GBM treated with bevacizumab: a multi-center study. J Neurooncol . 2012;108(3):491–498.

49. Colavolpe C , Chinot O , Metellus P et al FDG-PET predicts survival in recurrent high-grade gliomas treated with bevacizumab and irinotecan. Neuro Oncol . 2012;14(5):649–657.

50. Grossmann P , Gutman DA , Dunn WD Jr , Holder CA , Aerts HJ . Imaging-genomics reveals driving pathways of MRI derived volumetric tumor phenotype features in Glioblastoma. BMC Cancer . 2016;16:611.

51. Gutman DA , Dunn WD Jr , Grossmann P et al Somatic mutations associated with MRI-derived volumetric features in glioblastoma. Neuroradiology . 2015;57(12):1227–1237. 52. Abdulla S , Saada J , Johnson G , Jefferies S , Ajithkumar T . Tumour progression or pseudoprogression? A review of post-treatment radiological appearances of glioblastoma. Clin Radiol . 2015;70(11):1299–1312.

FUNDING

ACKNOWLEDGMENTS

CHAPTER 6

Defining a radiomic response phenotype: a pilot study using targeted therapy in NSCLC

ABSTRACT

Medical imaging plays a fundamental role in oncology and drug development, by providing a non-invasive method to visualize tumor phenotype. Radiomics can quantify this phenotype comprehensively by applying image-characterization algorithms, and may provide important information beyond tumor size or burden. In this study, we investigated if radiomics can identify a gefitinib response-phenotype, studying high-resolution computed-tomography (CT) imaging of forty-seven patients with early-stage non-small cell lung cancer before and after three weeks of therapy. On the baseline-scan, radiomic-feature Laws-Energy was significantly predictive for EGFR-mutation status (AUC = 0.67, p = 0.03), while volume (AUC = 0.59, p = 0.27) and diameter (AUC = 0.56, p = 0.46) were not. Although no features were predictive on the post-treatment scan (p > 0.08), the change in features between the two scans was strongly predictive (significant feature AUC-range = 0.74–0.91). A technical validation revealed that the associated features were also highly stable for test-retest (mean ± std: ICC = 0.96 ± 0.06). This pilot study shows that radiomic data before treatment is able to predict mutation status and associated gefitinib response non-invasively, demonstrating the potential of radiomics-based phenotyping to improve the stratification and response assessment between tyrosine kinase inhibitors (TKIs) sensitive and resistant patient populations.

INTRODUCTION

The response of tumors as measured on imaging is historically one of the oldest biomarkers used in drug discovery and clinical practice. Response has been assessed by measuring tumor burden with a number of surrogates such as unidimensional measurements according to the RECIST criteria1 bidimensional measurements as defined by the World Health Organization2. The purpose of these criteria as well as the modifications to the criteria are twofold; first, to enhance the ability of the biomarker to be prognostic or predictive and/or to improve the accuracy and reproducibility of the biomarker. Tumor burden and change in tumor burden during therapy, as measured at imaging, has been demonstrated to have value as a biomarker3.

However, medical imaging can provide more information about the tumor phenotype, beyond volumetric measurements; a process referred to as image-based phenotyping. Medical imaging is intuitively very suitable as a biomarker source to predict treatment response, as it is able to visualize and quantify time series of disease processes in a non-invasive way in individual patients. The characterization of quantitative imaging features which reflect tumor biology, physiology and tumor phenotype is increasingly being explored. Radiomics is the study of these quantitative features and their correlation with tumor phenotypes4'5'6. For example, recent studies have used CT-based radiomic signatures to successfully predict overall survival, disease free survival, and distant metastases in lung cancer4'7'8. Other examples have demonstrated that an imaging feature which could be quantified, for example, the percentage of ground-glass opacity (GGO) volume, is significantly higher in patients with the exon 21 missense mutation than in tumors with other EGFR mutation status9. This is thought be related to the fact that exon 21 missense mutation was significantly more frequent in lepidic predominant adenocarcinomas10. While these quantitative imaging features are under investigation in many cancers for their correlation with tumor phenotype and mutational status there is preliminary evidence to suggest that there may be an association between these features and both clinical outcomes and the underlying genomic signatures in lung cancer4'8'11.

In this pilot study, we look for the first time at the value of quantitative radiomic imaging features, in addition to tumor burden which was previously studied, for predicting known sensitizing EGFR mutations associated with Gefitinib response, to understand the relationship between imaging features and mutational status at baseline and especially with change in therapy in patients with and without the sensitizing mutation.

MATERIAL AND METHODS

This is a re-analysis of an existing dataset; imaging and tissue data were obtained prospectively as an exploratory analysis within a phase II trial of neoadjuvant Gefitinib in patients

with NSCLC12. Correlation results of early diameter and volume changes with EGFR mutation status was published previously3. As described previously3, all experiments were performed in accordance with relevant guidelines and regulations, and approved by the institutional review board (IRB) at Columbia University College and New York Presbyterian Hospital. Also, informed consent was obtained from all subjects included in this study.

Clinical data

At time of resection, tumor tissue was snap frozen in liquid nitrogen and stored in a −80 °C freezer. Representative areas of these specimens were pathologically reviewed to confirm the diagnosis and presence of tumor. Genomic DNA was analyzed for the most common EGFR-sensitizing mutations (exons 19 and 21) using previously described PCR-based methods13′14. EGFR wild-type (WT) tumors were also tested for KRAS mutations, which were found in a non-overlapping subset of lung adenocarcinomas that have been found to be resistant to EGFR tyrosine kinase inhibitor therapy15. If no EGFR or KRAS mutations were found, then the remaining EGFR exons were assessed by standard dideoxynucleotide sequencing. Selected specimens that were found to be EGFR/KRAS WT were submitted for more detailed mutational testing using mass spectrometry.

Tumor imaging and measurement

Baseline computed tomography of each patient was done within 2 weeks before gefitinib initiation. A follow-up computed tomography scan was done using the same imaging acquisition technique about three weeks post therapy, before surgery. Non–contrast enhanced diagnostic chest computed tomographies were done with a LightSpeed 16 scanner (GE Medical Systems) during a breath hold. High-resolution images with 1.25-mm slice thickness and lung kernel were reconstructed3. Three patients were excluded because 1.25-mm slice thickness reconstructions were not collected as required by protocol (two patients) and delineation of lesion contour did not reach radiologists' consensus during the data review of this study (one patient), leaving 47 of 50 patients remained in our analysis. Tumor contours were semi-automatically delineated using a three-dimensional segmentation algorithm3′16.

Quantitative radiomics analysis

In this work, we extracted 183 radiomic features from both baseline and follow up scan images that were resampled down to 0.25 mm resolution in all three directions. The definitions of these features are provided in Supplemental 1. The Delta dataset was defined as the pre-treatment radiomic feature values minus the post-treatment values. On the Delta dataset, the 15 most variant features were selected using the coefficient of variation. From this set, highly correlated features were removed who had a mean correlation of higher than 0.95. This procedure yielded 11 independent features; we added Volume and

maximum diameter for comparison, resulting in 13 features in total. A correlation matrix for these features was calculated using Spearman rank statistic. For every dataset, the area under the curve (AUC) of the receiver operator characteristic (ROC)[17] was calculated to assess predictive power of EGFR-sensitizing mutation. Since only 13 features were tested, correction for multiple testing was not considered. All statistical analysis was conducted using the R statistical software version 3.1.018 on a Linux operating system.

Technical validation

Technical validation of the features were conducted in a test-retest setting on the RIDER dataset, which contains of 31 patients each two lung CT scans taken approximately 15 min. apart. The RIDER lung cancer dataset[19] of the same-day repeat CT scans and the intraclass correlation coefficient (ICC) was used to assess the stability of features for test-retest. The CT imaging protocol of the RIDER data set was identical to the one used in this study[20].

RESULTS

To investigate if radiomic biomarkers are associated with mutational status and response to Gefitinib treatment, we analyzed clinical data of 47 early stage NSCLC patients whom were imaged before and after treatment. In Fig. 1 representative CT scans of an EGFR mutant and an EGFR wild-type tumor are shown before and after Gefitinib treatment, demonstrating clear phenotypic differences. To quantify these differences, we performed a radiomic analysis (see Fig. 2). The analysis was restricted to features with high and independence variance (Supplement I), resulting in eleven radiomic features and two volumetric features (volume and max diameter), that were included in our analysis (see Table 1). Using this strategy, we were able to identify a limited number of independent features, and only these features were assessed for performance to predict mutational status and associated Gefitinib response.

Baseline radiomics associations with mutational status

We investigated the correlations between the image features evaluated in our analysis. In Fig. 3 the correlations between the features extracted from the before treatment CT scan are shown. Although the GLCM features showed high positive and negative correlations, overall the correlations between those features were low (mean ± std: −0.16 ± 0.95), demonstrating independency of those features. Note the low correlation of total tumor volume with the other features (mean ± std: 0.01± 0.33).
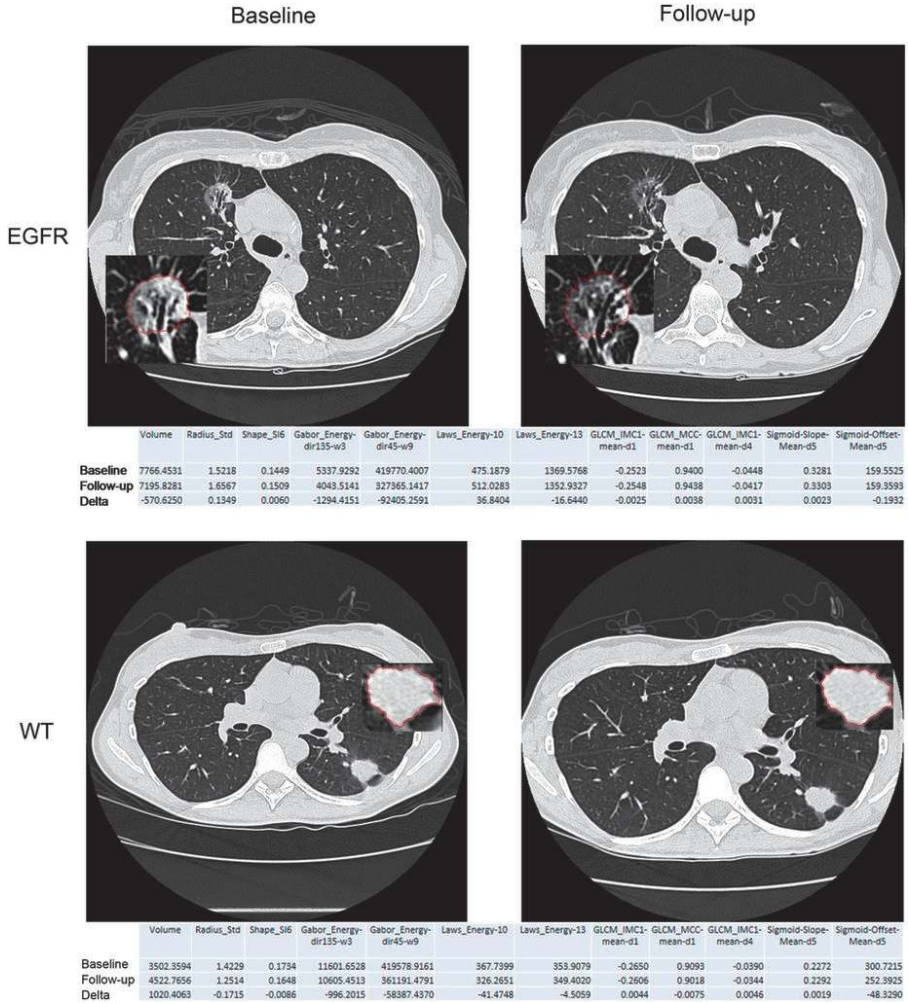
Figure 1
Example images of a patient with EGFR mutation and without (wild-type; WT) at baseline and follow-up scan. Radiomic feature values are given below each image for baseline and follow-up time points, as well as their delta differences.
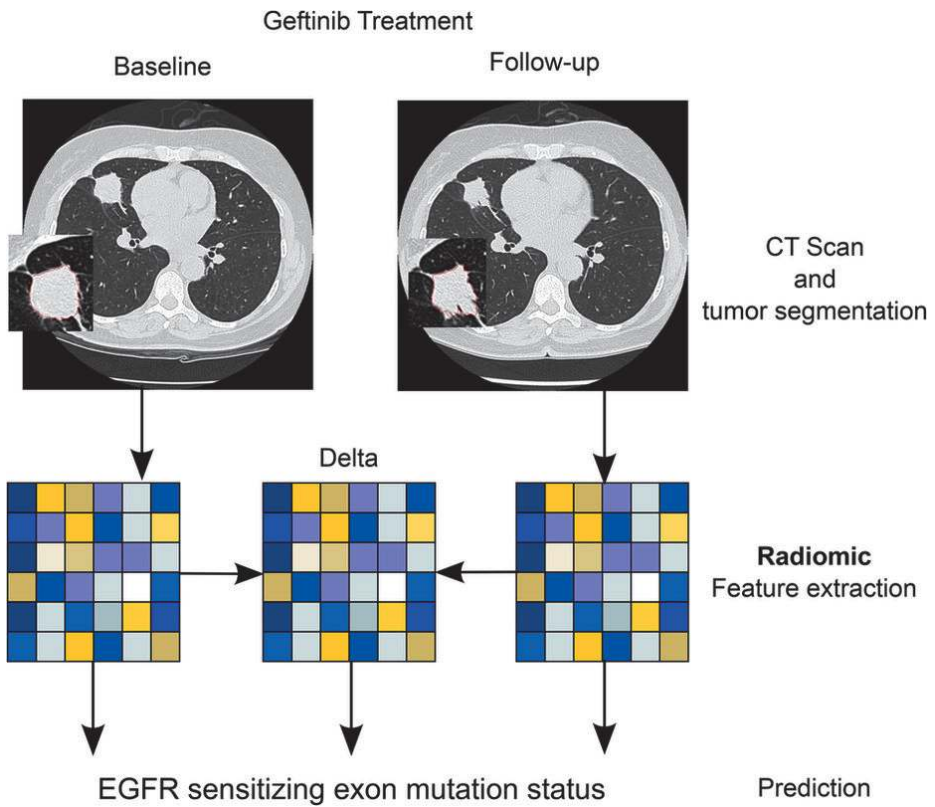
Figure 2
Patients included to the study were received Gefitinib treatment. CT scans at baseline and first follow-up were used to segment the tumor and to extract radiomic features. Baseline, follow-up, and delta radiomics (differences between baseline and follow-up) were used to assess EGFR sensitizing mutation status.

Table 1

Definitions of evaluated quantitative image features.

Feature	Description
Volume	Tumor volume is calculated by multiplying the number of tumor voxels by the image resolutions in x-, y- and z-directions.
Gabor Energy feature class	Gabor filters are linear filters designed for edge detection. This is an oriented Gaussian function modulated by a sinusoidal wave. The Gabor Energy feature is defined as the sum of the square of density over all lesion pixels on the images pre-processed by Gabor filter.
	· Gabor_Energy-dir135-w3: the Gabor Energy feature calculated on the images pre-processed using the Gabor filter built with an orientation of 135° and a wavelength of 3 pixels.
	· Gabor_Energy-dir45-w9: the Gabor Energy feature calculated on the images pre-processed using the Gabor filter built with an orientation of 45° and a wavelength of 9 pixels.
Sigmoid Function feature class	To quantify lesion margins, Sigmoid curve is used to fit density change along a sampling line drawn orthogonal to the lesion surface. Each sampling line, going through one voxel on the lesion surface, has a certain length inside and outside the lesion.
	· Sigmoid-Offet-Mean-d5: The average of the densities between a lesion and lung parenchyma on all lines. The line length is 5 mm at both sides of the lesion.
	· Sigmoid-Slope-Mean-d5: The average of the density change speed between lesion and lung parenchyma on all lines. The line length is 5 mm at both sides of the lesion.
Shape Index feature class	Local surface shape of a 3D object can be intuitively captured by 9 Shape Index features, Shape_SI1-9. The value of each Shape Index ranges from 0 to 1. The larger the value, the greater the portion of the shape on the surface.
	· Shape_SI6: Describes the saddle ridge shape.
Boundary_Radius_Std	This feature is defined as the standard deviation of the lengths of the line segments from

	the center of an object to any voxel on the surface of the object. A spherical shape has the smallest value of zero (0).
GLCM feature class	GLCM stands for grey-level co-occurrence matrix. This feature class characterizes image textures by creating a new matrix, GLCM, which is based on the frequency of image pixel pairs having particular gray-level values in a particular spatial arrangement (i.e., distance and direction). A number of statistical features can then be extracted from GLCM to characterize homogeneity, contrast, entropy and so on. In the following feature names, "mean" specifies average of feature values calculated at 13 directions. "d1" and "d4" indicate pixel pairs separated by 1-pixel distance and 4-pixel distance.
	· GLCM_IMC1-mean-d1: The average of informational measure of correlation 1 calculated at pixel pairs separated by 1-pixel distance
	· GLCM_IMC1-mean-d4: The average of informational measure of correlation 1 calculated at pixel pairs separated by 4-pixel distance
	· GLCM_MCC-mean-d1: The average of Maximum Correlation Coefficient calculated at pixel pairs separated by 1-pixel distance. calculated at pixel pairs separated by 4-pixel distance.
Laws Energy feature class	Laws' Energy emphasizes edge, spot, ripple and wave patterns through Laws filters generated by the following 5 basic raw vectors: Average $L_5 = (1, 4, 6, 4, 1)$, Edge $E_5 = (-1, -2, 0, 2, 1)$, Spot $S_5 = (-1, 0, 2, 0, -1)$, Ripple $R_5 = (1, -4, 6, -4, 1)$, and Wave. By multiplying and combining the transpose of one basic vector and/or the vector itself, 14 standard Laws filters can be built, each generating one feature. A Laws Energy feature is computed by summing the square of image pixel value over all tumor pixels on images processed by one of the 14 Laws filters.
	• Laws_Energy-10: Energy calculated on the images processed by Laws filter #10 .
	• Laws_Energy-13: Energy calculated on the images processed by Laws filter #13 $(2_* W_5 \times W_5)$
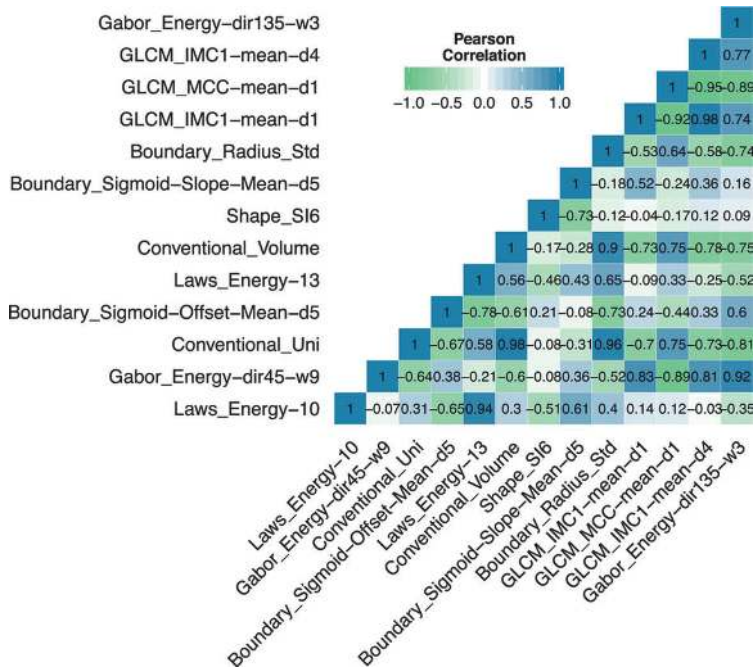
Figure 3
Correlations of Radiomics features. Correlation coefficient matrix between the 13 imaging features evaluated in the analysis. Note the overall low correlation between radiomic features. Correlations were assessed using pearson correlation coefficient.

To assess the associations of radiomic features and EGFR mutational status, we evaluated the predictive power using the AUC of the ROC. Figure 4 shows the AUC values that were measured for imaging features extracted from the scan before treatment, after treatment, and the difference between both scans (delta). Interestingly, Laws Energy 10 is the only radiomic feature that is significantly predictive for mutational status extracted from the pre-treatment scan (AUC = 0.67, p = 0.03). Note that pre-treatment tumor volume (AUC = 0.59, p = 0.27) and maximum diameter (AUC = 0.56, p = 0.46) are not predictive of mutation status. No features are significantly predictive extracted from the image scan after treatment (highest AUC = 0.64, p = 0.08, Shape SI6). Also, the remaining volume after treatment is not predictive (AUC = 0.54, p = 0.63).

Response phenotyping by radiomics feature change to predict mutational status

To assess the difference in radiomic feature values between the two scans (delta), showed strong predictability for mutation status (Fig. 4C). The strongest predictors are delta volume (AUC = 0.91, p = 10^{-25}) and delta maximum diameter (AUC = 0.78, p = 10^{-5}). However,
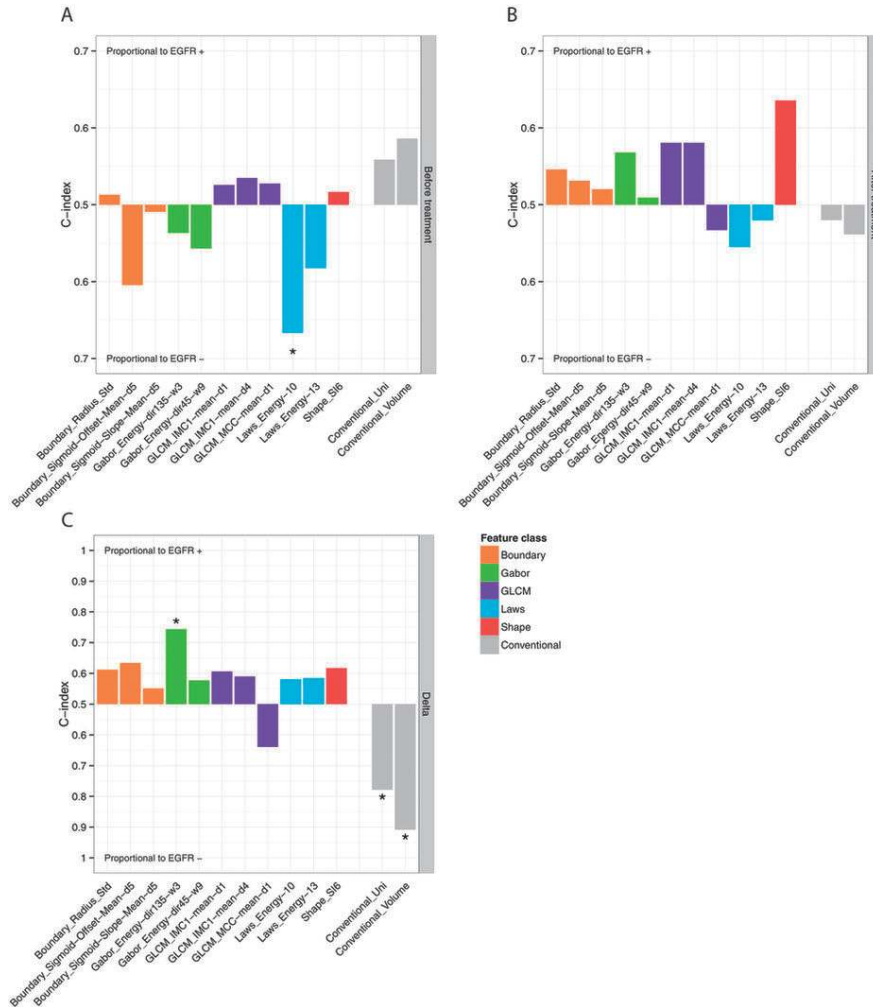
Figure 4

*Value of radiomic features to predict mutational status extracted from (**A**) before treatment scan, (**B**) post-treatment scan, and (**C**) delta (difference between the scan before and after treatment). Predictive value is assessed using the area under the curve (AUC) of the receiver operator characteristic (ROC). Asterisk denotes features that significantly predict mutation status better than random (p < 0.05). Note, before treatment only the radiomic feature Laws_Energy was significant for predicting mutation status, and conventional imaging markers showed no significant predictive power. After treatment no features are predictive, however delta feature values between the pre and post-scan show strongest predictive value.*

one radiomic feature was also significantly predictive: delta Gabor Energy (dir135-w3), which is Gabor Energy calculated at the wavelength of 3 pixels and direction of 135 grade (AUC = 0.74, p = 3 × 10^{-4}). Although, this feature is predictive, the correlations with tumor volume (r = −0.31) and maximum diameter (r = −0.43) are low (Fig. 3). These results demonstrate the predictability of radiomic features quantifying phenotypic characteristics other than volumetric features.

Technical validation of radiomic features for stability

As a technical validation, we assessed the stability of the four features that showed significantly predictive capabilities on either the pre, after, or delta scans, i.e. (I) Laws Energy, (II) Gabor Energy, (III) tumor volume, and (IV) maximum diameter (Fig. 4). For this purpose we used the independent RIDER lung cancer dataset consists of same-day repeated CT scans for 31 patients. The intraclass correlation coefficient (ICC) was used to assess the stability of each feature for test-retest. We found that all four significantly associated features with mutation status were also highly stable for test-retest (mean ± std: ICC = 0.96 ± 0.06). Feature Laws_Energy_10, the only feature significantly associated with mutational status extracted from the pretreatment scan, had a high stability of ICC = 0.87. The stability of the features significantly associated with mutational status between the two time points, was also very high: Volume (ICC = 0.99), Gabo Energy (ICC = 0.97), and maximum diameter (ICC = 0.99). These results validate that features significantly associated with mutation status are also stable for test-retest.

DISCUSSION

Biomarkers that are able to predict treatment response are crucial for clinical introduction of targeted therapies. Erlotinib and Gefitinib are examples of targeted therapies that were successfully introduced in practice, largely due to accurate predictive biomarkers (i.e. EGFR mutation status). Others have failed to become clinically approved, such as Cetuximab in NSCLC, to a great extent because an accurate response biomarker is lacking[21,22,23,24]. Medical imaging is intuitively very suitable for this purpose, as it is able to visualize and quantify time series of disease processes in a non-invasive way in individual patients. "Radiomics", the extraction and analysis of large amounts of advanced imaging features, is able to quantify tumor phenotypic properties, thereby potentially providing valuable diagnostic, prognostic or predictive information. Radiomic features have been associated with clinical outcomes[4,8,25,26,27,28,29], however the predictive capability of radiomics for response to targeted therapies remains largely unknown.

The goal of this study was to investigate if radiomic data could define a response phenotype for NSCLC patients treated with Gefitinib therapy. Our data show strong associations between radiomic features and tumors with and without sensitizing mutations. The radiomics feature Laws Energy extracted from the baseline, pretreatment CT scan showed the strongest performance to predict mutation status (AUC = 0.67, p = 0.03). In contrast, we found that tumor volume and maximum diameter were both not significantly predictive (p > 0.27). These data show that radiomics analyses can reveal more phenotypic characteristics than standard imaging features, such as size and volume, and have tremendous potential to be incorporated into precise response biomarkers. However, before clinical applicability, the predictive performance of these features have to be evaluated in large independent cohorts across institutions.

As expected, response of imaging phenotype was extremely different for patients with and without sensitizing mutations. We compared the radiomic data extracted from the change in pre- and post-treatment scans. As expected, tumor volume and maximum diameter had the highest performance; however the radiomic feature Gabor Energy was also significantly predictive. The correlation of this feature with tumor volume and maximum diameter was low. This could indicate the independent predictability for mutational status; however, future studies are needed to show if these features have complementary value in multivariate models for response prediction or may be seen earlier in treatment response. Due to the limited sample size of this study, we were not able to identify and validate multivariate biomarkers. The analysis supports CT-based response assessment as one effective tool for distinguishing sensitive and resistant tumor, but there may be other metrics that are additive and would allow a more comprehensive measure of response. The specific radiomic features may depend upon the specific mechanism of action of the drug or class of therapy.

Correlations between our selected radiomics features demonstrated that in general, the features originally derived from CT segmentation of tumors are indeed independent measures of phenotypic characteristics. Each feature has the potential to offer unique insight to the tumor behavior. This was demonstrated by the fact that different radiomic features predicted for tumors with and without EGFR mutations, and confirmed by the relative independence between most of the imaging features. However, future studies have to investigate how these patterns in the imaging data captured by these radiomic features, are associated with the underlying driving biology.

As a technical validation we assessed if the four features significantly associated with mutation status, identified by our analysis, could be stably extracted from CT scans. We evaluated this stability using another dataset with repeated test-retest CT scans for 31 patients (RIDER). All four features could be validated and had high stability for test-retest (ICC > 0.87).

Variability in CT acquisition and reconstruction parameters is inherent in clinical practice. In this study we used a prospective collected dataset with high resolution CT scans with the same imaging protocol for all included patients. However, the optimal

scanning parameters still have to be defined and more domain specific quantitative features are anticipated to be developed. The Quantitative Imaging Network (QIN) of the National Institute of Health, and others, play an important role in this process by performing phantom studies, radiomic feature definition standardizations, and segmentation challenges30. Furthermore, they establishes open and standardized protocols for image acquisition, reconstruction, and analysis30'31'32. It is expected that these efforts will improve the predictive performance of radiomic response biomarkers even further.

As this was a pilot investigation our analysis had several limitations. A main limitation was the limited sample size as we included only 47 patients in our analysis. As the data was acquired in a research study, with high quality imaging data and imaging before and after gefinitib treatment, more data was not available. Also, we limited our analysis to eleven independent radiomic features. A full radiomic analysis4, evaluating hundreds of features, could potentially achieve higher performance radiomic biomarkers for gefitinib response, however requires large independent training and validation cohorts. Furthermore, before clinical application imaging protocols have to be standardized and hence variability in CT acquisition and reconstruction parameters in clinical practice has to be reduced. The Quantitative Imaging Biomarker Alliance from the Radiological Society of North America (RSNA) and the QIN play an important role in this process by establishing standards for image acquisition and reconstruction, by conducting phantom studies, and by performing segmentation challenges. In addition, multiples studies have already documented the robustness of radiomic feature extractions in terms of reproducibility and repeatability in test/re-test settings4'19'27. However, before clinical utility, future studies have to evaluate radiomic biomarkers in independent and prospective validation cohorts with large sample sizes, and show improved performance compared to volumetric imaging features.

In conclusion, we found that radiomics features are able to define a Gefitinib response phenotype non-invasively, that is able to distinguish between tumors with and without EGFR sensitizing mutations at baseline and a quantitative change in these radiomic features at follow up. The use of radiomics based response assessment tools could improve the stratification between sensitive and resistant patient populations and the detection of response to treatment. This may provide an opportunity to improve decision-support at low aditional cost, as imaging is routinely and repeatedly used in clinical practice.

SUPPLEMENTARY INFORMATION

Available online at: https://images.nature.com/original/nature-assets/srep/2016/160920/srep33860/extref/srep33860-s1.pdf

REFERENCES

1. Therasse, P. et al. New Guidelines to Evaluate the Response to Treatment in Solid Tumors. J. Natl. Cancer Inst. 92, 205–216 (2000).
2. Organization, W. H. & Others. WHO handbook for reporting results of cancer treatment (1979).
3. Zhao, B. et al. A pilot study of volume measurement as a method of tumor response evaluation to aid biomarker development. Clin. Cancer Res. 16, 4647–4653 (2010).
4. Aerts, H. J. W. L. et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat. Commun. 5, 4006 (2014).
5. Kumar, V. et al. Radiomics: the process and the challenges. Magn. Reson. Imaging 30, 1234–1248 (2012).
6. Lambin, P. et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur. J. Cancer 48, 441–446 (2012).
7. Huang, Y. et al. Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage (I or II) Non—Small Cell Lung Cancer. Radiology 152234 (2016).
8. Coroller, T. P. et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. Radiother. Oncol. 114, 345–350 (2015).
9. Yang, Y. et al. EGFR L858R mutation is associated with lung adenocarcinoma patients with dominant ground-glass opacity. Lung Cancer 87, 272–277 (2015).
10. Lee, H.-J. et al. Epidermal growth factor receptor mutation in lung adenocarcinomas: relationship with CT characteristics and histologic subtypes. Radiology 268, 254–264 (2013).
11. Gevaert, O. et al. Non–Small Cell Lung Cancer: Identifying Prognostic Imaging Biomarkers by Leveraging Public Gene Expression Microarray Data—Methods and Preliminary Results. Radiology 264, 387–396 (2012).
12. Rizvi, N. A. et al. Molecular characteristics predict clinical outcomes: prospective trial correlating response to the EGFR tyrosine kinase inhibitor gefitinib with the presence of sensitizing mutations in the tyrosine binding domain of the EGFR gene. Clin. Cancer Res. 17, 3500–3506 (2011).
13. Pan, Q., Pao, W. & Ladanyi, M. Rapid polymerase chain reaction-based detection of epidermal growth factor receptor gene mutations in lung adenocarcinomas. J. Mol. Diagn. 7, 396–403 (2005).
14. Pao, W. et al. EGF receptor gene mutations are common in lung cancers from 'never smokers' and are associated with sensitivity of tumors to gefitinib and erlotinib. Proc. Natl. Acad. Sci. USA 101, 13306–13311 (2004).
15. Pao, W. et al. KRAS mutations and primary resistance of lung adenocarcinomas to gefitinib or erlotinib. PLoS Med. 2, e17 (2005).
16. Zhao, B., Reeves, A. P., Yankelevitz, D. F. & Henschke, C. I. Three-dimensional multicriterion automatic segmentation of pulmonary nodules of helical computed tomography images. Opt. Eng. 38, 1340–1347 (1999).
17. Fawcett, T. An introduction to ROC analysis. Pattern Recognit. Lett. 27, 861–874 (2006).
18. Team, R. C. R: A language and environment for statistical computing (2012).
19. RIDER Collections - The Cancer Imaging Archive (TCIA) Public Access - Cancer Imaging Archive Wiki. Available at: https://wiki.cancerimagingarchive.net/display/Public/RIDER+Collections. (Accessed: 10th September 2015).
20. Zhao, B. et al. Evaluating Variability in Tumor Measurements from Same-day Repeat CT Scans of Patients with Non–Small Cell Lung Cancer. Radiology 252, 263–272 (2009).
21. Pirker, R. et al. Cetuximab plus chemotherapy in patients with advanced non-small-cell lung cancer (FLEX): an open-label randomised phase III trial. Lancet 373, 1525–1531 (2009).
22. Lynch, T. J. et al. Cetuximab and first-line taxane/carboplatin chemotherapy in advanced non-small-cell lung cancer: results of the randomized multicenter phase III trial BMS099. J. Clin. Oncol. 28, 911–917 (2010).

23. Khambata-Ford, S. et al. Analysis of Potential Predictive Markers of Cetuximab Benefit in BMS099, a Phase III Study of Cetuximab and First-Line Taxane/Carboplatin in Advanced Non-Small-Cell Lung Cancer. J. Clin. Oncol. 28, 918–927 (2010).

24. Hirsch, F. R. et al. Increased EGFR Gene Copy Number Detected by Fluorescent In Situ Hybridization Predicts Outcome in Non-Small-Cell Lung Cancer Patients Treated With Cetuximab and Chemotherapy. J. Clin. Oncol. 26, 3351–3357 (2008).

25. Grove, O. et al. Quantitative Computed Tomographic Descriptors Associate Tumor Shape Complexity and Intratumor Heterogeneity with Prognosis in Lung Adenocarcinoma. PLoS One 10, e0118261 (2015).

26. Hawkins, S. H. et al. Predicting Outcomes of Nonsmall Cell Lung Cancer Using CT Image Features. IEEE Access 2, 1418–1426 (2014).

27. Balagurunathan, Y. et al. Reproducibility and Prognosis of Quantitative Features Extracted from CT Images. Transl. Oncol. 7, 72–87 (2014).

28. Parmar, C. et al. Robust Radiomics feature quantification using semiautomatic volumetric segmentation. PLoS One 9, e102107 (2014).

29. Parmar, C., Grossmann, P., Bussink, J., Lambin, P. & Aerts, H. J. W. L. Machine Learning methods for Quantitative Radiomic Biomarkers. Sci. Rep. 5, 13087 (2015).

30. Clarke, L. P. et al. The Quantitative Imaging Network: NCI's Historical Perspective and Planned Goals. Transl. Oncol. 7, 1–4 (2014).

31. Buckler, A. J. et al. Quantitative imaging test approval and biomarker qualification: interrelated but distinct activities. Radiology 259, 875–884 (2011).

32. Buckler, A. J., Bresolin, L., Dunnick, N. R. & Sullivan, D. C. & Group. A collaborative enterprise for multistakeholder participation in the advancement of quantitative imaging. Radiology 258, 906–914 (2011).

ACKNOWLEDGEMENTS

AUTHOR INFORMATION

Contributions

H.J.W.L.A., P.G., L.H.S. and B.Z. designed the analysis and prepared the main manuscript text. Y.T., G.O.F. and N.R. provided clinical and technical expertise. All authors reviewed the manuscript.

Competing interests

The authors declare no competing financial interests.

Corresponding author

Correspondence to Patrick Grossmann.

P A R T 4

Prognostic Value of Radiomics Machine Learning

CHAPTER 7

Machine learning methods
for quantitative radiomic biomarkers

Chintan Parmar*, Patrick Grossmann*, Johan Bussink, Philippe Lambin, and Hugo J.W.L. Aerts
*These authors contributed equally to this work

ABSTRACT

Radiomics extracts and mines large number of medical imaging features quantifying tumor phenotypic characteristics. Highly accurate and reliable machine-learning approaches can drive the success of radiomic applications in clinical care. In this radiomic study, fourteen feature selection methods and twelve classification methods were examined in terms of their performance and stability for predicting overall survival. A total of 440 radiomic features were extracted from pre-treatment computed tomography (CT) images of 464 lung cancer patients. To ensure the unbiased evaluation of different machine-learning methods, publicly available implementations along with reported parameter configurations were used. Furthermore, we used two independent radiomic cohorts for training (n = 310 patients) and validation (n = 154 patients). We identified that Wilcoxon test based feature selection method WLCX (stability = 0.84 ± 0.05, AUC = 0.65 ± 0.02) and a classification method random forest RF (RSD = 3.52%, AUC = 0.66 ± 0.03) had highest prognostic performance with high stability against data perturbation. Our variability analysis indicated that the choice of classification method is the most dominant source of performance variation (34.21% of total variance). Identification of optimal machine-learning methods for radiomic applications is a crucial step towards stable and clinically relevant radiomic biomarkers, providing a non-invasive way of quantifying and monitoring tumor-phenotypic characteristics in clinical practice.

INTRODUCTION

'Precision oncology' refers to the customization of cancer care, where practices and/or therapies are being tailored to individual patients. Such customization process can maximize the success of preventive and therapeutic interventions with minimum side effects. Most of the precision oncology related research has centered on the molecular characterization of tumors using genomics based approaches, which require tissue extraction by tumor biopsies. Although several genomics based approaches have successfully been applied in clinical oncology[1], there are inherent limitations to biopsy based assays. Tumors are spatially and temporally heterogeneous, and repeated tumor biopsies, which increase the risk for a patient, are often required to capture the molecular heterogeneity of tumors. These ethical and clinical challenges related to biopsy-based assays, can be addressed by medical imaging, which is a routine practice for cancer diagnosis and staging in clinical oncology. Unlike biopsies, medical imaging is non-invasive and can provide information regarding the entire tumor phenotype, including the intra-tumor heterogeneity. Furthermore, recent advances in high-resolution image acquisition machines and computational hardware allow the detailed and efficient quantification of tumor phenotypic characteristics. Therefore, medical imaging provides unprecedented opportunities for precision oncology.

"Radiomics", an emerging and promising field, hypothesizes that medical imaging provides crucial information regarding tumor physiology, which could be exploited to enhance cancer diagnostics[2]. It provides a comprehensive quantification of tumor phenotypes by extracting and mining large number of quantitative imaging features[3]. Several studies have investigated various radiomic features in terms of their prognostic or predictive abilities and reliability across different clinical settings[4,5,6,7,8,9,10]. Different studies have shown the discriminating capabilities of radiomic features for the stratification of tumor histology[6], tumor grades or stages[11], and clinical outcomes[8,12,13]. Moreover, some studies have reported the association between radiomic features and the underlying gene expression patterns[8,14,15].

"Machine-learning" can be broadly defined as computational methods/models using experience (data) to improve performance or make accurate predictions[16]. These programmable computational methods are capable of "learning" from data and hence can automate and improve the prediction process. Predictive and prognostic models with high accuracy, reliability, and efficiency are vital factors driving the success of radiomics. Therefore, it is essential to compare different machine-learning models for radiomics based clinical biomarkers. Like any high-throughput data-mining field, radiomics also underlies the curse of dimensionality[17], which should be addressed by appropriate feature selection strategies. Moreover, feature selection also helps in reducing overfitting of models (increasing the generalizability). Thus, in order to reduce the dimensionality of radiomic feature space and enhance the performance of radiomics based predictive models, different feature selection methods[18] should be thoroughly investigated. However, as radiomics is

an emerging research field, most of the published studies have only assessed the predictive capabilities of radiomic features without putting much emphasis on the comparison of different feature selection and predictive modeling methods. Only few recent studies have investigated the effect of different feature selection and machine learning classification methods on radiomics based clinical predictions[19,20], but with limited sample sizes. Furthermore, these studies lacked independent validation of the results, which may restrict the generalizability of their conclusions.

In this study, we investigated a large panel of machine-learning approaches for radiomics based survival prediction. We evaluated 14 feature selection methods and 12 classification methods in terms of their predictive performance and stability against data perturbation. These methods were chosen because of their popularity in literature. Furthermore, publicly available implementations along with reported parameter configurations were used in the analysis, which ensured an unbiased evaluation of these methods. Two independent lung cancer cohorts were used for training and validation, with in total image and clinical outcome data of 464 patients. Feature selection and predictive modeling are considered as the important building blocks for high throughput data driven radiomics. Therefore, our investigation could help in the identification of optimal machine-learning approaches for radiomics based predictive studies, which could enhance the applications of non-invasive and cost-effective radiomics in clinical oncology.

METHODS

Radiomic Features

A total of 440 radiomic features were used in the analysis. These radiomic features quantified tumor phenotypic characteristics on CT images and are divided into four feature groups: I) tumor intensity, II) shape, III) texture and IV) wavelet features. Tumor intensity based features estimated the first order statistics of the intensity histogram, whereas shape features described the 3D geometric properties of the tumor. Textural features, derived from the gray level co-occurrence (GLCM)[21] and run length matrices (GLRLM)[22], quantified the intra-tumor heterogeneity. These textural features were computed by averaging their values over all thirteen directions. Wavelet features are the transformed domain representations of the intensity and textural features. These features were computed on different wavelet decompositions of the original image using a coiflet wavelet transformation. Matlab R2012b (The Mathworks, Natick, MA) was used for the image analysis. Radiomic features were automatically extracted by our in-house developed radiomics image analysis software, which uses an adapted version of CERR (Computational Environment for Radiotherapy Research)[23] and Matlab for the preprocessing of medical

images. Mathematical definitions of all radiomic features, as well as the extraction methods, were previously described8.

Datasets

In this study, we employed two NSCLC cohorts from the two different institutes of Netherlands: (1) Lung1:422 NSCLC patients treated at MAASTRO Clinic in Maastricht. (2) Lung2:225 NSCLC patients treated at Radboud University Medical Center in Nijmegen. CT-scans, manual delineations and clinical data were available for all included patients. More details on the included datasets are described in Supplementary-A. We dichotomized the censored continuous survival data using a cutoff time of 2 years. The patients who lived beyond the cutoff time were labeled as 1, whereas the deceased ones were labeled as 0. The objective of the study was to stratify patients into these two labeled survival classes. Two-years is considered as a relevant survival time for NSCLC patients and several other studies have designed their prediction models using a survival cutoff of 2 years24'25'26. We excluded the patients, which were followed for less than 2 years. It resulted in 310 patients in training cohort (Lung1) and 154 patients in validation cohort (Lung2). All the features were normalized using Z-score normalization.

Feature Selection Methods

Fourteen feature selection methods based on filter approaches were used in the analysis (Fisher score (FSCR), Relief (RELF), T-score (TSCR), Chi-square (CHSQ), Wilcoxon (WLCX), Gini index (GINI), Mutual information maximization (MIM), Mutual information feature selection (MIFS), Minimum redundancy maximum relevance (MRMR), Conditional infomax feature extraction (CIFE), Joint mutual information (JMI), Conditional mutual information maximization (CMIM), Interaction capping (ICAP), Double input symmetric relevance (DISR)). In order to improve the readability of this manuscript, we have defined all the acronyms related to feature selection methods in Table 1. We chose these methods mainly because of their popularity in literature, simplicity and computational efficiency. Furthermore, publicly available implementations were readily available for these methods27'28, which increases their reusability. Filter methods are feature-ranking methods, which rank the features using a scoring criterion. All filter based feature selection methods can be divided into two categories: univariate methods and multivariate methods. In case of univariate methods, the scoring criterion only depends on the feature relevancy ignoring the feature redundancy, whereas multivariate methods investigate the multivariate interaction within the features and the scoring criterion is a weighted sum of feature relevancy and redundancy. Feature relevancy is a measure of feature's association with the target/outcome variable, whereas feature redundancy is the amount of redundancy present in a particular feature with respect to the set of already selected features. Further descrip-

tion regarding the theoretical formulation of feature selection problem and each of the used feature selection methods can be obtained from Supplementary-B online.

Table 1
Table defining the acronyms related to the used feature selection and classification methods.

Classification method acronym	Classification method name	Feature Selection method acronym	Feature selection method name
Nnet	Neural network	RELF	Relief
DT	Decision Tree	FSCR	Fisher score
BST	Boosting	GINI	Gini index
BY	Bayesian	CHSQ	Chi-square score
BAG	Bagging	JMI	Joint mutual information
RF	Random Forset	CIFE	Conditional infomax feature extraction
MARS	Multi adaptive regression splines	DISR	Double input symmetric relevance
SVM	Support vector machines	MIM	Mutual information maximization
DA	Discriminant analysis	CMIM	Conditional mutual information maximization
NN	Neirest neighbour	ICAP	Interaction capping
GLM	Generalized linear models	TSCR	T-test score
PLSR	Partial least squares and principal componenet regression	MRMR	Minimum redundancy maximum relevance
—	—	MIFS	Mutual information feature selection
—	—	WLCX	Wilcoxon

Classifiers

In machine-learning, the classification is considered as a supervised learning task of inferring a function from labeled training data16. The training data consists of a set of examples, where each example is represented as a pair of an input vector (features) and a desired output value (target or category label). The classification algorithm (classifier) analyzes the training data and infers a hypothesis (function), which can be used for predicting the labels of unseen observations. Many classifiers belonging to different areas of computer science and statistics have been proposed in machine-learning literature29. In our study, we used 12 machine-learning classifiers arising from 12 classifier families (Bagging (BAG), Bayesian (BY), Boosting (BST), Decision trees (DT), Discriminant analysis (DA), Generalized linear models (GLM), Multiple adaptive regression splines (MARS), Nearest neighbors (NN), Neural networks (Nnet), Partial least square and principle component regression (PLSR), Random forests (RF), and Support vector machines (SVM)). The acronyms related to classifiers are defined in Table 1. All classifiers were implemented using R package caret30, which provides a nice interface to access many machine-learning algorithms in R. Furthermore, it also provides a user-friendly framework for training different machine-learning models. Classifiers were trained using the repeated (3 repeat iterations) 10 fold cross validation of training cohort (Lung1) and their predictive performance was evaluated in the validation cohort (Lung2) using area under ROC curve (AUC). We used parameter configurations that were previously defined by Fernandez-Delgado *et al.*31 in a comprehensive comparative study of 179 classifiers and 121 different datasets. We have listed the classification methods along with their parameters and corresponding R packages in Supplementary-C online.

ANALYSIS

Predictive Performance of Feature Selection and Classification Methods

In order to investigate and compare different feature selection and classification methods, we created a three-dimensional parameter grid for the analysis. For each of the 14 feature selection methods, we incrementally selected features ranging from 5 up to 50, with an increment of 5 features (n = 5, 10, 15, 20, … , 50). These subsets of selected features were then evaluated by using each of the 12 machine-learning classifiers and area under ROC curves (AUC).

Stability of Feature Selection and Classification Methods

In order to assess the stability of feature selection methods, we used a stability measure proposed by Yu *et al.*32 under the hard data perturbation settings33. We quantified the stability of a method as the similarity between the results obtained by the same feature selection method, when applied on the two non-overlapping partitions (of size N/2) of the training cohort (Lung1). To compute similarity between the two resultant feature sets, a weighted complete bipartite graph was constructed, where the two node sets corresponded to the two sets of selected features. The edge weights were assigned as the absolute Spearman correlation coefficient between the features at the nodes. We then applied the Hungarian algorithm34 to identify the maximum weighted matching between the two node sets, and then similarity (stability) was quantified as the final matching cost. For each feature selection method, we computed the stability 100 times using a bootstrap approach and reported the median ± std values in the results.

The empirical stability of a classifier was quantified using the relative standard deviation (RSD %) and a bootstrap approach. We first selected 30 representative features using the Wilcoxon based feature selection method WLCX and used them to compute the classifier stability. For each classification method, we trained the model on the subsampled training cohort (size N/2) and validated the performance on the validation cohort using AUC. Subsampling of the training cohort was done 100 times using a bootstrap approach. RSD is the absolute value of the coefficient of variation and is often expressed in percentage.

Here, it was defined as

$$RSD = \frac{\sigma_{AUC}}{\mu_{AUC}} * 100 \qquad (1)$$

where σ_{AUC} and μ_{AUC} were the standard deviation and mean of the 100 AUC values respectively. It should be noted that higher stability in the case of classifiers corresponds to lower RSD values.

Stability and Predictive Performance

In order to identify the highly reliable and accurate methods, we used the median values of AUC and stability as thresholds. We created two rank lists based on AUC & stability and cited the methods as highly accurate and reliable, which ranked in the top half of both the ranked lists. Feature selection methods having stability ≥0.735 (median stability of all feature selection methods) and AUC ≥ 0.615 (median AUC of all feature selection methods) are considered as highly reliable and accurate methods. Similarly, classification methods having RSD ≤ 5.97 (median RSD of all classifiers) and AUC ≥ 0.61 (median AUC of all classifiers) are considered as highly reliable and accurate ones.

Experimental Factors Affecting the Radiomics Based Survival Prediction

There are three main experimental factors, which can potentially affect the prediction of radiomics based survival prediction: feature selection method, classification method and the number of selected features. Multifactor ANOVA was used to quantify the variability in AUC scores contributed by these factors and their interactions. In order to compare the variability contributed by each factor, the estimated variance components were divided by the total variance.

All the analysis was done using R software (R Core Team, Vienna, Austria) version 3.1.2 and Matlab R2012b (The Mathworks, Natick, MA) with Windows 7.

RESULTS

To investigate the machine-learning approaches for prognostic radiomic biomarkers, a total of 440 radiomic features were extracted from the segmented tumor regions of the pre-treatment CT images of two independent NSCLC cohorts. Feature selection and classification training was done using the training cohort Lung1 (n = 310 patients), whereas the validation cohort Lung2 (n = 154 patients) was used to assess the predictive performance [see Fig. 1].
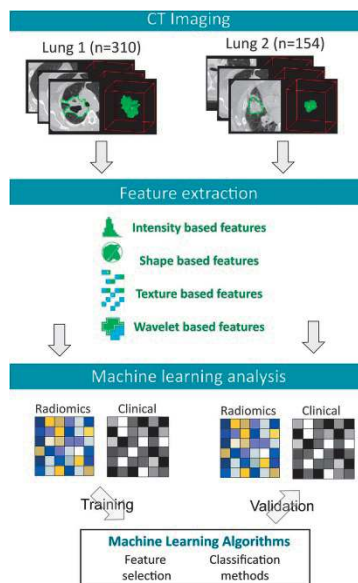

Figure 1

A total of 440 radiomic features were extracted from the segmented tumor regions of the pre-treatment CT images of 464 NSCLC patients. Feature selection and classification training was done using the training cohort Lung1 (n = 310), whereas Lung2 (n = 154) cohort was used as a validation cohort.

Predictive Performance of the Feature Selection and Classification Methods

Predictive performance of different feature selection and classification methods was assessed using the area under receiver operator characteristic curve (AUC). Figure 2 depicts the performance of feature selection (in rows) and classification methods (in columns) using 30 selected features, which are the 30 top ranked features, resulted in feature selection. For each classification method, there are 14 AUC values corresponding to the 14 different feature selection methods. We used a median of all 14 AUC values as a representative AUC of a classifier. Similarly, for each feature selection method, a median of 12 AUCs (corresponding to 12 classification methods) is used as a representative AUC. These representative AUC values for the classification and feature selection methods are given in Table 2. For classification methods, random forest (RF) displayed highest predictive performance (AUC: 0.66 ± 0.03) (median ± std), whereas decision tree (DT) (AUC: 0.54 ± 0.04) showed the lowest predictive performance. As far as feature selection methods are concerned, the Wilcoxon test based method WLCX showed highest predictive performance (AUC: 0.65 ± 0.02), whereas method CHSQ (AUC: 0.60 ± 0.03) and CIFE (AUC: 0.60 ± 0.04) had the lowest median AUCs. We repeated the above experiment by varying the number of selected features (range 5–50). Results corresponding to 10, 20, 40 and 50 representative (top ranked) features are reported in Supplementary Figures S1, S2, S3 and S4 online. Furthermore, median AUC values over each of the experimental factors (feature selection methods, classification methods and number of selected features) are depicted by the heatmaps in Supplementary Figures S5, S6 and S7 online. Here as well, random forest (RF) (classifier) and Wilcoxon test based method WLCX (feature selection) showed highest median AUCs in majority of cases.

Stability of the Feature Selection and Classification Methods

We assessed the feature selection methods in terms of their stability against data resampling using the hard data perturbation settings33. We observed that MIM was the most stable method (stability = 0.94 ± 0.02) (median ± std) followed by RELIEF (stability = 0.91 ± 0.05) and WLCX (stability = 0.84 ± 0.05), whereas GINI (stability = 0.68 ± 0.10), JMI (stability = 0.68 ± 0.05), CHSQ (stability = 0.69 ± 0.09), DISR (stability = 0.69 ± 0.05) and CIFE (stability = 0.69 ± 0.05) showed relatively low stability [Table 2].
Empirical stability of classification methods was quantified using the relative standard deviation (RSD) and a bootstrap approach. We observed that BY was the most stable classification method (RSD = 0.86%) followed by GLM (RSD = 2.19%), PLSR (RSD = 2.24%) and RF (RSD = 3.52%). BST had the highest relative standard deviation in AUC scores (RSD = 8.23%) and hence the lowest stability among the classification methods. RSD (%) values corresponding to all 12 classifiers are reported in Table 2.
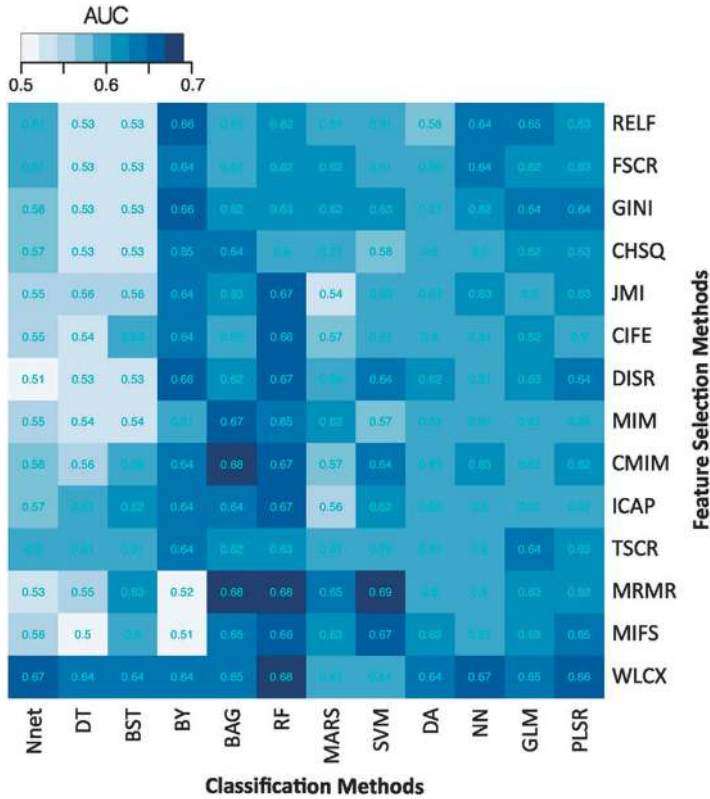
Figure 2

Heatmap depicting the predictive performance (AUC) of feature selection (in rows) and classification (in columns) methods. It can be observed that RF, BAG and BY classification methods and feature selection methods WLCX, MRMR and MIFS shows relatively high predictive performance in many cases.

Stability and Predictive Performance

Scatterplots in Fig. 3 assesses the stability and prediction performance. It can be observed that feature selection methods WLCX (stability = 0.84 ± 0.05, AUC = 0.65 ± 0.02), MIFS (stability = 0.8 ± 0.03, AUC = 0.63 ± 0.03), MRMR (stability = 0.74 ± 0.03, AUC = 0.63 ± 0.03) and FSCR (stability = 0.78 ± 0.08, AUC = 0.62 ± 0.04) should be preferred as their stability and predictive performance was higher than the corresponding median values across all feature selection methods (stability = 0.735, AUC = 0.615). Similarly for classification methods, RF (RSD = 3.52%, AUC = 0.66 ± 0.03), BY (RSD = 0.86%, AUC = 0.64 ± 0.05), BAG (RSD = 5.56%, AUC = 0.64 ± 0.03), GLM (RSD = 2.19%, AUC = 0.63 ± 0.02), and PLSR (RSD = 2.24%, AUC = 0.63 ± 0.02), the stability and predictive performance was higher than the corresponding median values (RSD = 5.93%, AUC = 0.61).

Table 2

Table describing the median values of AUC and stability for different Classification and Feature Selection methods.

Classification method	AUC	RSD %	Feature Selection method	AUC	Stability
Nnet	0.57 ± 0.04	6.41	RELF	0.61 ± 0.04	0.91 ± 0.05
DT	0.54 ± 0.04	7.89	FSCR	0.62 ± 0.04	0.78 ± 0.08
BST	0.58 ± 0.04	8.23	GINI	0.62 ± 0.04	0.68 ± 0.10
BY	0.64 ± 0.05	0.86	CHSQ	0.60 ± 0.04	0.69 ± 0.09
BAG	0.64 ± 0.03	5.56	JMI	0.61 ± 0.04	0.68 ± 0.05
RF	0.66 ± 0.03	3.52	CIFE	0.60 ± 0.03	0.69 ± 0.05
MARS	0.61 ± 0.03	6.98	DISR	0.62 ± 0.05	0.69 ± 0.05
SVM	0.61 ± 0.03	6.39	MIM	0.61 ± 0.04	0.94 ± 0.02
DA	0.61 ± 0.02	6.37	CMIM	0.62 ± 0.04	0.73 ± 0.04
NN	0.61 ± 0.02	4.08	ICAP	0.61 ± 0.03	0.72 ± 0.04
GLM	0.63 ± 0.02	2.19	TSCR	0.61 ± 0.02	0.78 ± 0.12
PLSR	0.63 ± 0.02	2.24	MRMR	0.63 ± 0.06	0.74 ± 0.03
—	—	—	MIFS	0.63 ± 0.06	0.8 ± 0.03
—	—	—	WLCX	0.65 ± 0.02	0.84 ± 0.05
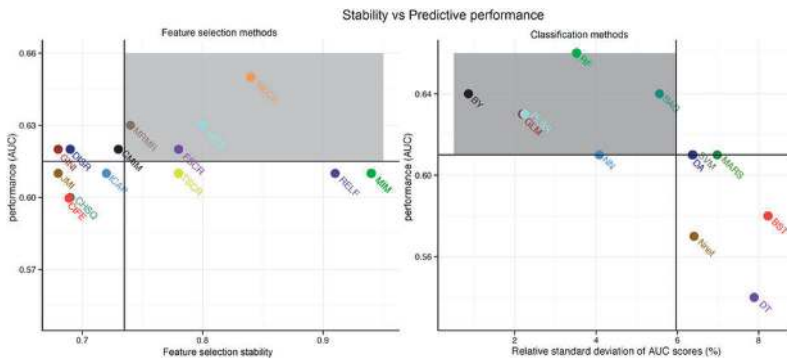
Figure 3

Scatterplots between the stability and predictive performance (AUC) of feature selection (FS) (Left) and classification methods (CF) (right). Feature selection methods having stability ≥0.735 (median stability of FS) and AUC ≥ 0.615 (median AUC of FS) are considered as highly reliable and predictive methods. Similarly, classification methods having RSD ≤ 5.97 (median RSD of CF) and AUC ≥ 0.61 (median AUC of CF) are considered as highly reliable and accurate ones. Highly reliable and predictive methods are displayed in a gray square region.

Experimental Factors Affecting the Radiomics Based Survival Prediction

To quantify the effects of the three experimental factors (feature selection methods, classification methods and the number of selected features), we performed multifactor analysis of variance (ANOVA) on AUC scores. We observed that all three experimental parameters and their interactions are the significant factors affecting the prediction performance [Fig. 4]. Classification method was the most dominant source of variability as it explained 34.21% of the total variance in AUC scores. Feature selection accounted for the 6.25%, whereas interaction of classifier & feature selection explained 23.03% of the total variation. Size of the selected (representative) feature subset only shared 1.65% of the total variance [Fig. 4].
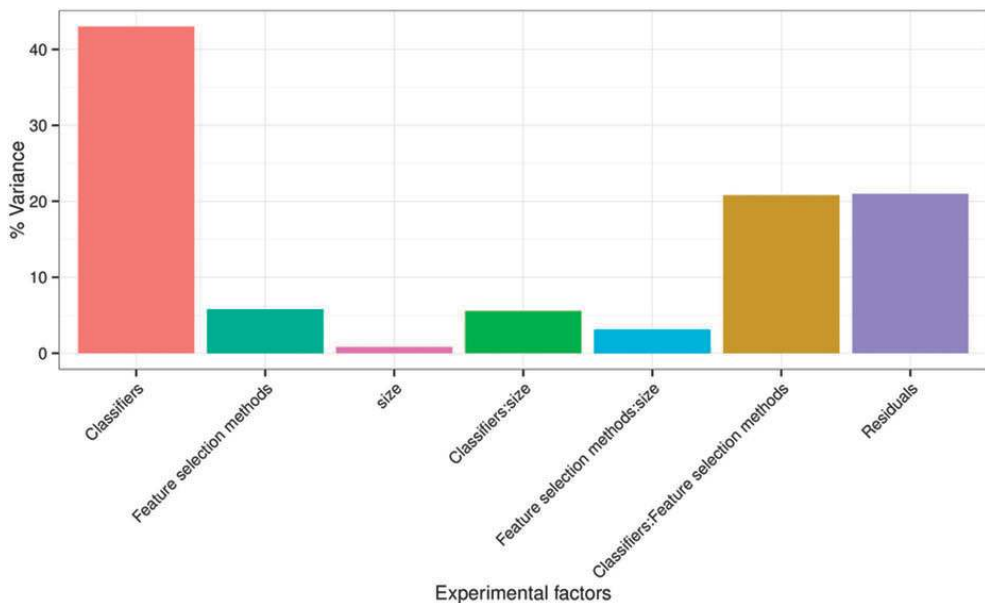
Figure 4

Variation of AUC explained by the experimental factors and their interactions. It can be observed that classification method was the most dominant source of variability. Size of the selected (representative) feature subset shared the least of the total variance.

DISCUSSION

Medical imaging is a routinely used and easily accessible source of information in clinical oncology. It serves as a non-invasive and cost-effective cancer diagnostic tool. Radiomics employs the medical imaging data for the customization of cancer care and hence adds a

new and promising dimension to precision oncology[2,3,8]. Moreover, it can also capture the intra-tumor heterogeneity, which is often considered as an important biomarker in oncology[12,35,36,37]. A number of studies have built radiomics based predictive models for various clinical factors (tumor grades, survival outcomes, treatment response, etc.)[12]. For the successful realization of radiomics based predictive analyses, it is required to evaluate and compare different feature selection and predictive modeling methods, which was the primary objective of this study.

Various feature selection methods have been employed for high-throughput data mining problems[38]. In general, feature selection methods are categorized into three main categories: (1) filter methods (2) wrapper methods and (3) embedded methods. In this study, we investigated 14 different filter based approaches for radiomics based survival prediction. We only used filter-based approaches because they are computationally more efficient and less prone to overfitting than the wrapper and embedded methods[18,27]. Furthermore, unlike wrapper and embedded methods, filter methods are classifier independent. Thus, they allow separation of the modeling and feature selection component of the predictive analysis, which increases the generalizability of each component and hence the overall analysis.

We also investigated 12 machine-learning classification methods belonging to 12 different classifier families. Many classifiers have been proposed in the machine-learning literature. Theoretically speaking, these classifiers belong to different fields (classifier families) of computer science and statistics. Therefore, it could really be difficult to understand the underlying assumptions of each and every classifier and tune the parameters in an unbiased manner. The parameter tuning could be biased by user's more (or lack of) expertise with some classifiers over the others. Usually, the studies, which propose a new classifier, only compare it to the reference classifiers of same family excluding the other classifier families. Even if classifiers belonging to different families are considered for comparison, these reference classifiers are usually implemented using simple tools and with limited parameter configurations while carefully tuning the proposed classifier. These could consequently bias the results in favor of the proposed classifiers[31]. In our study, we are not proposing any new classifier and we have used the same implementation tool (R package caret) for all the classifiers. Furthermore, to ensure unbiased usage of classifiers, we used parameter configurations that were previously defined by Fernandez-Delgado *et al.*[31], in an exhaustive study of comparing 179 classifiers over 121 different datasets. These parameter configurations were selected from the literature and have been previously validated on a large number (121) of datasets belonging to different fields. Furthermore, in our study, the parameters were tuned using the repeated cross validation of training data only. Hence, our experimental design allowed us to evaluate different classification methods in an unbiased manner.

Our results show that the Wilcoxon test based feature selection method WLCX yields the highest predictive performance with the majority of classifiers. Interestingly, WLCX is a simple univariate method based on ranks, which does not take into account the

redundancy of selected features during feature ranking. The majority of feature selection methods gave highest predictive performance when used with the random forest (RF) classifier. One could argue that with different parameter configurations, the performance of classification methods may improve further. An exhaustive parameter tuning could be investigated for evaluating the improvement of prediction performance. However, the required computational resources and high time complexity can hinder the exhaustive search. We expect that future radiomic studies focusing on different clinical outcomes and similar analysis framework could provide better understanding in this regard. A limited number of methods, which are consistently high performing across different radiomic studies, could be further assessed with an exhaustive parameter tuning. Nevertheless, It should be noted that random forests (RF) have displayed high predictive performance in several other biomedical and other domain applications as well31. These results indicate that choosing the WLCX feature selection method and/or RF classification method increases predictive performance in radiomics.

Results related to our stability analysis provide another dimension for choosing the feature selection and classification methods. Depending upon the applications, one may give importance to the predictive performance or stability and accordingly opt for the required method. Results related to multifactor ANOVA indicated that the classification method is the most dominant source of variation in the prediction performance (AUC) and hence should be chosen carefully. Size of the selected feature subset contributed the least in the total variation of AUC.

Only few studies have investigated and compared different feature selection and machine-learning modeling methods for radiomics based clinical predictions19ʹ20. Recently, Hawkins *et al.*19 have compared four different feature selection and classification methods for CT based survival prediction of NSCLC patients. This study, however, was limited by the small cohort size as the final results were obtained on only 40 patients. Furthermore, it also lacked an independent validation of the results. On the contrary, two independent radiomic cohorts of sizes 310 and 154 patients were used in our analysis and an independent validation of the results was reported.

Our radiomic analysis is focused on the prediction of two-year patient survival in NSCLC patients. It provides an unbiased evaluation of different machine-learning methods of feature selection and classification. It could be considered as a reference for the future radiomics based predictive studies. Our results indicated that choosing Wilcoxon test based feature selection method WLCX and/or random forest (RF) classification method gives highest performance for radiomics based survival prediction. Furthermore, these methods also turned out reasonably stable against data perturbation and hence they could be preferred for radiomics based predictive studies. These results should be further tested in other radiomics based predictive studies, with different imaging modalities and in different cancer types.

It has been previously shown that for NSCLC patients, statistical models based on patient's tumor and treatment characteristics provide significantly better predictions than

the human expert[24]. Moreover, several other studies have highlighted the limitation of doctors' prognostic capability for terminally ill cancer patients[39,40,41]. The predictions of human experts can suffer from inter-observer variability. On the contrary, statistical models could make the prediction system more deterministic if the parameter configurations and the training framework are fixed.

The potential clinical utility of radiomics based prognostic models has been stated in previous study[8]. With expanding radiomics cohorts and feature dimensions, we expect higher prediction performance in future radiomic studies. Furthermore, the integrative studies like radiomics-genomics in combination with standard clinical covariates could also improvise the prediction performance and further validate the utility of these methods in clinical practice. Overall, our analysis is a step forward towards the enhancements of radiomics based clinical predictions.

SUPPLEMENTARY MATERIAL

Available online at: https://images.nature.com/original/nature-assets/srep/2015/150817/srep13087/extref/srep13087-s1.pdf

REFERENCES

1. Doroshow, J. & Kummar, S. Translational research in oncology-10 years of progress and future prospects. Nat. Rev. Clin. Oncol. 11, 649 (2014).
2. Lambin, P. et al. Predicting outcomes in radiation oncology—multifactorial decision support systems. Nat. Rev. Clin. Oncol. 10, 27–40 (2013).
3. Lambin, P. et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur. J. of Cancer 48, 441–446 (2012).
4. Coroller, T. P. et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. Radiothe. Oncol. (2015), http://dx.doi.org/10.1016/j.radonc.2015.02.015 (2015).
5. Cook, G. J. et al. Are Pretreatment 18F-FDG PET Tumor Textural Features in Non–Small Cell Lung Cancer Associated with Response and Survival After Chemoradiotherapy? J. Nucl. Med. 54, 19–26 (2013).
6. Ganeshan, B. et al. Non–small cell lung cancer: histopathologic correlates for texture parameters at CT. Radiology 266, 326–336 (2013).
7. Gevaert, O. et al. Glioblastoma multiforme: exploratory radiogenomic analysis by using quantitative image features. Radiology 273, 168–174 (2014).
8. Aerts, H. J. et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat. Commun. 5 (2014).
9. Leijenaar, R. T. et al. Stability of FDG-PET Radiomics features: An integrated analysis of test-retest and inter-observer variability. Acta Oncol. 52, 1391–1397 (2013).
10. Parmar, C. et al. Robust radiomics feature quantification using semiautomatic volumetric segmentation. PLOS ONE 9, e102107 (2014).
11. Ganeshan, B., Abaleke, S., Young, R. C., Chatwin, C. R. & Miles, K. A. Texture analysis of non-small cell lung cancer on unenhanced computed tomography: initial evidence for a relationship with tumour glucose metabolism and stage. Cancer Imaging 10, 137 (2010).

12. Alic, L., Niessen, W. J. & Veenland, J. F. Quantification of heterogeneity as a biomarker in tumor imaging: a systematic review. PLOS ONE 9, e110300 (2014).

13. Jain, R. et al. Outcome prediction in patients with glioblastoma by using imaging, clinical, and genomic biomarkers: focus on the nonenhancing component of the tumor. Radiology 272, 484–493 (2014).

14. Nicolasjilwan, M. et al. Addition of MR imaging features and genetic biomarkers strengthens glioblastoma survival prediction in TCGA patients. J. Neuroradiol. (2014), 10.1016/j.neurad.2014.02.006. (2014).

15. Segal, E. et al. Decoding global gene expression programs in liver cancer by noninvasive imaging. Nat. biotechnol. 25, 675–680 (2007).

16. Mohri, M., Rostamizadeh, A. & Talwalkar, A. Foundations of machine learning. Ch. 1, 1–3, (MIT press, 2012).

17. Pękalska, E. & Duin, R. P. The dissimilarity representation for pattern recognition: foundations and applications. Vol. 64 (World Scientific, 2005).

18. Guyon, I. & Elisseeff, A. An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157–1182 (2003).

19. Hawkins, S. H. et al. Predicting Outcomes of Nonsmall Cell Lung Cancer Using CT Image Features. IEEE Access 2, 1418–1426 (2014).

20. Basu, S. et al. in Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on. 1306–1312 (IEEE).

21. Haralick, R. M., Shanmugam, K. & Dinstein, I. H. Textural features for image classification. IEEE Trans. Syst., Man Cybern. 6, 610–621 (1973).

22. Galloway, M. M. Texture analysis using gray level run lengths. Comput. Vision Graph. 4, 172–179 (1975).

23. Deasy, J. O., Blanco, A. I. & Clark, V. H. CERR: a computational environment for radiotherapy research. Med. Phys. 30, 979–985 (2003).

24. Oberije, C. et al. A prospective study comparing the predictions of doctors versus models for treatment outcome of lung cancer patients: a step toward individualized care and shared decision making. Radiothe. Oncol. 112, 37–43 (2014).

25. Hoang, T., Xu, R., Schiller, J. H., Bonomi, P. & Johnson, D. H. Clinical model to predict survival in chemonaive patients with advanced non–small-cell lung cancer treated with third-generation chemotherapy regimens based on Eastern Cooperative Oncology Group data. J. Clin. Oncol. 23, 175–183 (2005).

26. Cistaro, A. et al. Prediction of 2 years-survival in patients with stage I and II non-small cell lung cancer utilizing 18F-FDG PET/CT SUV quantifica. Radiol. oncol. 47, 219–223 (2013).

27. Brown, G., Pocock, A., Zhao, M.-J. & Luján, M. Conditional likelihood maximisation: a unifying framework for information theoretic feature selection. J. Mach. Learn. Res. 13, 27–66 (2012).

28. Zhao, Z. et al. Advancing feature selection research. ASU feature selection repository (2010).

29. Kotsiantis, Sotiris B., Ioannis, D. Zaharakis & Panayiotis, E. Pintelas. Machine learning: a review of classification and combining techniques. Artif. Intell. Rev. 26.3, 159–190 (2006).

30. Kuhn, M. Building predictive models in R using the caret package. J. Stat. Softw. 28, 1–26 (2008).

31. Fernández-Delgado, M., Cernadas, E., Barro, S. & Amorim, D. Do we need hundreds of classifiers to solve real world classification problems? J. Mach. Learn. Res. 15, 3133–3181 (2014).

32. Yu, L., Ding, C. & Loscalzo, S. in Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. 803-811 (ACM).

33. Haury, A.-C., Gestraud, P. & Vert, J.-P. The influence of feature selection methods on accuracy, stability and interpretability of molecular signatures. PLOS ONE 6, e28210 (2011).

34. Kuhn, H. W. The Hungarian method for the assignment problem. Naval Res. Logis. Q. 2, 83–97 (1955).

35. Fisher, R., Pusztai, L. & Swanton, C. Cancer heterogeneity: implications for targeted therapeutics. Br. J Cancer 108, 479–485 (2013).

36. Ng, C., Pemberton, H. & Reis-Filho, J. Breast cancer intratumor genetic heterogeneity: causes and implications. Expert Rev. Anticancer Ther. 12, 1021–1032 (2012).

37. Brown, J. R., DiGiovanna, M. P., Killelea, B., Lannin, D. R. & Rimm, D. L. Quantitative assessment Ki-67 score for prediction of response to neoadjuvant chemotherapy in breast cancer. Lab. Invest. 94, 98–106 (2014).

38. Bolón-Canedo, V., Sánchez-Maroño, N., Alonso-Betanzos, A., Benítez, J. & Herrera, F. A review of microar-ray datasets and applied feature selection methods. Inform. Sciences 282, 111–135 (2014).
39. Christakis, N. A., Smith, J. L., Parkes, C. M. & Lamont, E. B. Extent and determinants of error in doctors' prognoses in terminally ill patients: prospective cohort studyCommentary: Why do doctors overestimate? Commentary: Prognoses should be based on proved indices not intuition. Bmj 320, 469–473 (2000).
40. Glare, P. et al. A systematic review of physicians' survival predictions in terminally ill cancer patients. Bmj 327, 195 (2003).
41. Clément-Duchêne, C., Carnin, C., Guillemin, F. & Martinet, Y. How accurate are physicians in the prediction of patient survival in advanced lung cancer? Oncologist 15, 782–789 (2010).

ACKNOWLEDGEMENTS

AUTHOR INFORMATION

Contributions

H.J.W.L.A., C.P. and P.G. conceived of the project, analysed the data, and wrote the paper. J.B. and P.L. provided expert guidance, data, or analysis tools and reviewed the manuscript.

Competing interests

The authors declare no competing financial interests.

Corresponding authors
Correspondence to Chintan Parmar or Hugo J. W. L. Aerts.

CHAPTER

8

Radiomic phenotype features predict pathological response in non-small cell lung cancer

Thibaud P. Coroller, Vishesh Agrawal, Vivek Narayan, Ying Hou, Patrick Grossmann, Stephanie W. Lee,
Raymond H. Mak *, and Hugo J.W.L. Aerts *
*These authors contributed equally to this work

ABSTRACT

Background and Purpose
Radiomics can quantify tumor phenotype characteristics non-invasively by applying advanced imaging feature algorithms. In this study we assessed if pre-treatment radiomics data are able to predict pathological response after neoadjuvant chemoradiation in patients with locally advanced non-small cell lung cancer (NSCLC).

Materials and methods
127 NSCLC patients were included in this study. Fifteen radiomic features selected based on stability and variance were evaluated for its power to predict pathological response. Predictive power was evaluated using area under the curve (AUC). Conventional imaging features (tumor volume and diameter) were used for comparison.

Results
Seven features were predictive for pathologic gross residual disease (AUC > 0.6, p-value < 0.05), and one for pathologic complete response (AUC = 0.63, p-value = 0.01). No conventional imaging features were predictive (range AUC = 0.51–0.59, p-value > 0.05). Tumors that did not respond well to neoadjuvant chemoradiation were more likely to present rounder shape (spherical disproportionality, AUC = 0.63, p-value = 0.009) and heterogeneous texture (LoG 5 mm 3D - GLCM entropy, AUC = 0.61, p-value = 0.03).

Conclusions
We identified predictive radiomic features for pathological response, although no conventional features were significantly predictive. This study demonstrates that radiomics can provide valuable clinical information, and performed better than conventional imaging features.

Keywords
Radiomics – pathological response – NSCLC – biomarkers – quantitative imaging

INTRODUCTION

Radiomics is an emerging field of quantitative imaging that aims to describe tumors non-invasively using a large set of advanced imaging features [1–3]. These features can robustly create a unique phenotypic atlas for each tumor [4–6]. Associating clinical information to this atlas has enabled the identification of new, reproducible, image-based biomarkers which has been prognostic for clinical outcomes including overall survival [7–9] and distant metastasis [10]. Association was found with lung cancer patients of histology and stage [11] as well.

Lung cancer is the leading cause of cancer deaths worldwide [12]. Stage IIIA non-small cell lung cancer (NSCLC) can be treated using trimodality therapy that includes neo-adjuvant chemoradiation followed by surgery according to NCCN guidelines [13]. However, trimodality therapy is controversial, given the observed lack of survival benefit in adding surgery compared to definitive chemoradiation alone, [14,15] which underscores the importance of identifying patients who respond completely to chemoradiation and do not require additional invasive local therapy.

Pathological response is a direct measure of tumor response to neoadjuvant chemoradiation assessed at time of surgery. It has the potential to be used as a surrogate endpoint [16] for survival/local control and has been shown to be prognostic for survival in early [17] and advanced [18] stages for NSCLC patients. Predicting pathological response at an early time point would allow modification of the treatment regimen (e.g. adding surgery versus intensifying chemoradiation) based on how the tumor is likely to respond and this adaptive approach could improve patient outcomes.

Currently, tumor response is clinically assessed using RECIST [19], which classifies changes in the sum of tumor and lymph nodes diameters on CT images before and after therapy. However, the radiographic response to chemoradiation for NSCLC tumors may be slow [20], which may limit the utility of RECIST in predicting pathological response at the end of the neoadjuvant chemoradiation shortly before surgery, and hence allow very little margin for clinicians to adapt the treatment regimen consequently.

In this study we investigated the power of pre-treatment CT-based radiomic features to predict pathological response after neoadjuvant chemoradiation. We compared these results to conventional volumetric features such as tumor volume and diameter.

MATERIALS AND METHODS

Patient selection

Patients with stage II–III NSCLC treated at Dana-Farber Cancer Institute between 2001 and 2013 who were treated with neoadjuvant radiotherapy and chemotherapy (chemoradiation) prior to surgical resection were included in this study. Patients with distant metastasis at presentation or delay in surgery greater than 120 days after the completion of

chemoradiation were excluded. For all patients, CT imaging at the initiation of chemoradiation and prior to surgical resection was available. No exclusion based on histology was applied. A subset of patients received adjuvant therapy and was also included in this analysis. Finally, a total of 127 patients were included for this study.

Follow-up and endpoints

The main endpoint for this study was pathological response assessed at time of surgery. The amount of residual tumor was classified based on surgical pathology reports as pathologic complete response (pCR), microscopic residual disease (MRD) or gross residual disease (GRD). Percent residual tumor in the pathological sample was not available for this study. Three other clinical endpoints were included for this study including overall survival (OS), distant metastasis (DM) and in-field local recurrence (LR). The time associated with the endpoint was defined from treatment start date to date of first event. The last date of follow-up was used for patients with no events.

Follow-up chest CT scans with contrast (unless the patient had a contraindication to contrast, e.g. renal dysfunction or allergy) were performed every three to six months after treatment for patients at our institution based on US national guidelines [13] to assess tumor progression.

CT Acquisition and Segmentation

Planning CTs were acquired according to scanning protocol at our institution using GE "lightspeed" CT scanner (GE Medical System, Milwaukee, WI, USA). Tumor segmentation was performed on radiation therapy planning CTs using Eclipse (Varian Medical System, Palo Alto, CA, USA). The primary tumor site was retrospectively contoured guided by existing treatment planning contours. Using both soft tissues and lung windows, air, vessels, normal tissue or surrounding organs were subsequently excluded from the contours (Figure 1.A). All contours were done manually (T.P.C., V.A., Y.H.), and then all individually verified by an expert radiation oncologist by (R.H.M.).
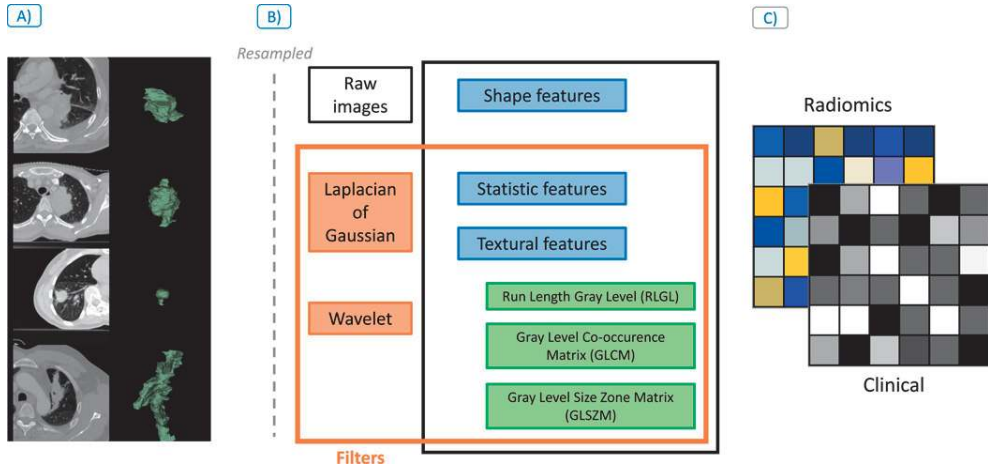
Figure 1

Radiomic analysis workflow description: A) Lung primary tumor site was manually contoured from treatment planning images (shown in green on CT image on the left and the 3D mask on the right). B) All images were subsequently resampled and from those contours the radiomic features describing tumor phenotype were extracted using three feature groups: Shape, Statistic and Textural features, with and without Wavelet and Laplacian of Gaussian filtering. C) Finally, association between radiomic features and the clinical outcomes were investigated for image-based biomarkers discovery.

Features Extraction and Selection

Radiomic features describing tumor phenotype were extracted (m=1603) from the primary tumor site with an in-house Matlab 2013 (The Mathworks Inc., Natick, Massachusetts, United States) toolbox and the software 3D Slicer 4.4.0 [21] (Figure 1.B). Average voxel spacing was (0.9mm × 0.9mm × 3mm) respectively for (x,y,z) and was resampled $3\times3\times3mm^3$ prior to feature extraction to have standardized voxel spacing across the cohort. A bin width of 25 Hounsfield units (HU) was used for textural features. All features are described in the supplement of a previous study [10].

Fifteen Radiomic features were selected based on stability and variance for this study (features selection is described in Supplement I). Additionally, we defined three conventional, pre-treatment, clinically utilized, volumetric features for comparison to advanced phenotypic features prior chemoradiation. These features consisted of tumor volume, 2D axial maximal diameter and 3D maximal diameter. 2D axial maximal diameter corresponds to the greatest diameter in the axial plane. 3D maximal diameter refers to the greatest diameter in any direction. All volume and diameter measurements were obtained from the primary tumor and did not include the sum diameters or volumes of involved lymph nodes.

Statistical Analysis

All statistical analyses were done on R software [22] version 3.1.3. Predictive performance of these remaining features were assessed using the "survcomp" package [23,24] version 1.16 from Bioconductor [25]. We computed receiver operating characteristic (ROC) area under the curve (AUC) for binary outcomes. Predictive power was reported as proportional or disproportionate to the risk of experiencing the response as the feature value is increasing.

Difference for clinical categories was assessed using chi-square or two-sided Wilcoxon-test respectively for categorical or continuous variables. Noether test was used to consider AUC significance from 0.5 (random). Survival and disease-free probability curve were computed using Kaplan-Meier analysis. A three year estimate was reported for the analysis. Log-rank test was used to assess difference in probability curves between pathological response groups. A p-value below 0.05 was considered as significant. Features with an AUC above 0.60 and a p-value below 0.05 were considered predictive.

Multivariate models were made using logistic regression for pathological response using the same subgroup for the univariate analysis to compare their performance. Three models were created with 1) Conventional (volume and axial/3D diameters), 2) Radiomics (predictive features for GRD) and 3) Combined (Conventional + radiomics) features.

We compare model performance with the validation AUC using the cross validation (CV). The cohort was split, using 80% for training and 20% for validation for each CV (for each 100 iterations). Patients were randomized using a conservative random split using the "caret" package [26]. Difference between the CV models performance was done using a permutation test. The outcome labels were randomly resampled (k=1000 times) and a new CV was computed for each random label combination. One-sided Wilcoxon test was computed for each random label models, the W_k statistic extracted and compared to W_0 (true label) to assess if a model performance was significantly greater than another.

RESULTS

127 patients with NSCLC were included in this study. The median age was 60.5 years (range 32.7 to −77.6 years), with a majority of women (53.5%) and white (92.1%). Tumor histology was predominantly adenocarcinoma (56.6%) and AJCC [27] stage IIIA (75.6%). The median follow-up was 41.8 months (range 2.7–117.2). The distribution of pathological response was 27 (21.3%), 33 (26.0%) and 67 (52.7%) respectively for complete response, microscopic and gross residual disease.

Table 1

Patient and treatment characteristics. Median (range) is reported for continuous and counts (percentage) for categorical variables. Statistical difference between complete pathological responders vs. non-complete respond-ers was computed using Chi-Square or Wilcoxon-test respectively for categorical and continuous variables.
Label: pCR = pathologic complete response, MRD=microscopic residual disease, GRD = gross residual disease.
[1]*Large Cell neuroendocrine carcinoma, Mixed NSCLC and SCLC, Adenoid cystic carcinoma, Sarcomatoid.*
[2]*RT Only, Sequential.*

Variable	Group	Median (Range) / Count (%)	pCR (n=27)	MRD & GRD (n=100)	p-value
Age [years]		60.5 (32.7 – 77.6)	61.5 (32.7 – 75.2)	60.4 (33.1– 77.6)	0.93
Performance Status	0	60 (47.2%)	10 (37.0%)	50 (50%)	0.33
	1	59 (46.5%)	14 (51.9%)	45 (45%)	
	2–3	8 (6.3%)	3 (11.1%)	5 (5%)	
Gender	Female	68 (53.5%)	11 (40.7%)	57 (57%)	0.19
	Male	59 (46.5%)	16 (59.3%)	43 (43%)	
Ethnicity	White	117 (92.1%)	25 (92.6%)	92 (92%)	0.61
	Black	5 (3.9%)	1 (3.7%)	4 (4%)	
	Hispanic	3 (2.4%)	0 (0%)	3 (3%)	
	Asian	2 (1.6%)	1 (3.7%)	1 (1%)	
Histology	Adenocarcinoma	72 (56.6%)	5 (18.5%)	67 (67%)	<0.001
	Squamous cell Carcinoma	32 (25.3%)	14 (51.9%)	18 (18%)	
	NSCLC	18 (14.2%)	6 (22.2%)	12 (12%)	
	Others1	5 (3.9%)	2 (7.4%)	3 (3%)	
Overall stages	IIA	2 (1.5%)	1 (3.7%)	1 (1%)	0.02
	IIB	8 (6.3%)	5 (18.5%)	3 (3%)	
	IIIA	96 (75.6%)	18 (66.7%)	78 (78%)	
	IIIB	21 (16.6%)	3 (11.1%)	18 (18%)	
Treatment sequence	Concurrent	111 (87.4%)	24 (88.8%)	87 (87%)	0.49
	Induction	14 (11.0%)	2 (7.4%)	12 (12%)	
	Others2	2 (1.6%)	1 (3.8%)	1 (1%)	
Radiation Dose [Gy]		54 (45–70)	54(46–70)	54(45–70)	0.32
Pathological response	pCR	27 (21.3%)	27 (100%)	0 (0%)	-
	MRD	33 (26.0%)	0 (0%)	33 (33%)	
	GRD	67 (52.7%)	0 (0%)	67 (67%)	

All treatment information can be found in the Table 1. Comparison between pathological complete response (pCR) versus microscopic (MRD) and gross (GRD) residual disease, showed no significant differences between treatment (pvalue=0.49, Chi-square test) and radiation dose (p-value=0.32). Significant differences were found between overall stages (p-value= 0.02) and histology (p-value < 0.001), likely driven by the fact that the distribution is skewed for histology and overall stage.

Relationship between clinical outcomes and pathological response subgroups was investigated (Table 2). The median (range) for overall survival, distant metastasis and local recurrence was respectively 41.8 (2.7–117.2), 10.8 (2.5–73.5) and 14 (4.7–66.5) months. No significant difference was observed for survival between pathological response (p-value =0.86, Log-rank test). However, pCR patients had significantly higher probabilities at three years for distant metastasis-free (79%, p-value = 0.036) and local recurrence-free (94%, p-value = 0.013). Kaplan-Meier curves can be found in Figure S3 and concordance index for radiomics features in Figure S4 in Supplement II.

Table 2
Three years estimate from Kaplan-Meier survival curve for each pathological response subgroup. Difference between groups was assessed with Log-Rank test.
Label: pCR = pathologic complete response, MRD=microscopic residual disease, GRD = gross residual disease.

Three years estimate probability	pCR (n=27)	MRD (n=33)	GRD (n=67)	p-value
Overall Survival	72%	53%	52%	0.86
Distant Metastasis Free	79%	59%	50%	0.036
Local Recurrence Free	94%	75%	62%	0.013

Fifteen features were selected based on stability and variance (see Supplement I) and were evaluated for performance to predict clinical outcomes. Those features included one shape, seven statistics and seven textural features. Textural features incorporated four gray-level co-occurrence matrix (GLCM) sensitive to voxel patterns and three gray-level size zone matrix (GLSZM) sensitive to flat zone (area of connecting voxel with same value). All these features are described in Table 3.

Table 3
Description of radiomic features with associated feature group and filter.

Selected Radiomic feature	Radiomic group	Filter associated	Description
Sphere Disproportionality	Shape	None	Ratio between tumor area and a sphere with the same volume as the tumor
Root Mean Square	Statistics	Wavelet HLL	Root mean square of the voxels intensity value
Range	Statistics	Wavelet LLH Wavelet LHH	The range of voxels intensity values
Energy	Statistics	Wavelet HLL	Describe the energy of the image
Mean	Statistics	Wavelet HLL	The mean voxel intensity value
Kurtosis	Statistics	LoG 3D - 5mm	Describe the shape of a probability distribution of the voxel intensity histogram
Skewness	Statistics	LoG 3D - 4mm	Describe the shape of a probability distribution of the voxel intensity histogram
Correlation	GLCM	Wavelet LHH	Correlation of the GLCM matrix
Entropy	GLCM	LoG 3D - 5mm	Complexity of the GLCM matrix (sensitive to the number of unique voxel patterns in the tumor)
Homogeneity 2	GLCM	LoG 2D - 4mm	Homogeneity of voxels patterns (similar patterns across the whole tumor)
Cluster Prominence	GLCM	LoG 3D - 3mm	Sensitive to flat zone (area of connecting voxel with same value)
Low Intensity Large Area Emphasis	GLSZM	Wavelet LHH	Sensitive to flat zone with low intensity voxel (e.g. necrotic area)
Large Area Emphasis	GLSZM	LoG 3D - 5mm	Sensitive to flat zone
High Intensity Large Area Emphasis	GLSZM	LoG 3D - 5mm	Sensitive to flat zone with high intensity voxel (e.g. calcifications, blood vessels)

Pathological response was our primary clinical endpoint. We first determined if radiomic features could identify tumors likely to respond poorly (GRD) vs. tumors likely to respond well (pCR and MRD) to the chemoradiation (Figure 2.A). The fifteen selected advanced imaging features had an AUC of 0.53 to 0.66 for GRD (described in Table S2 in Supplement II). Seven features were significantly predictive (range AUC 0.61–0.66, p-value <0.05) for GRD. From those seven predictive features, two were risk proportionate as the feature value increases (GLCM entropy and stats root mean square) and five disproportionate (mean and skewness in voxel intensity histogram, spherical disproportionality and two GLSZM large area emphasis) to experience GRD. No conventional volumetric imaging features were significant from random or above at the threshold of AUC > 0.6 (range 0.57 to 0.59, p-value >0.05) and all were disproportionate to the risk of GRD.
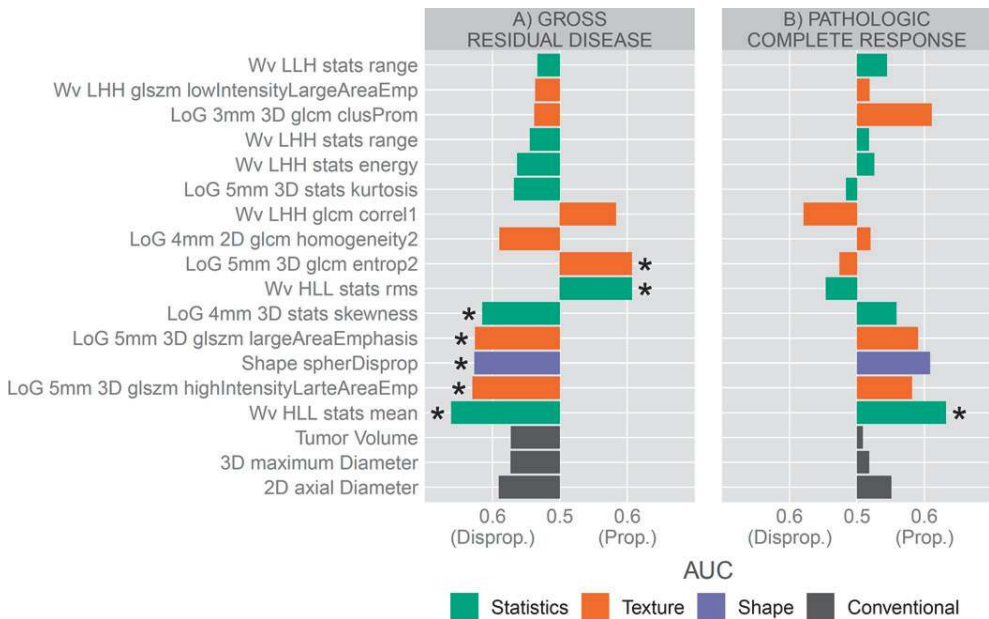
Figure 2

AUC of Radiomic features and conventional volumetric imaging features for A) poor responders (gross residual disease) vs. good responders (pathologic complete response and microscopic residual disease) and B) pathological complete responders vs. non-complete responders (microscopic and gross residual disease). Predicting power was reported as proportional or disproportionate to the risk of experiencing the response as the feature value is increasing. Features reported with a "" are significant from random (Noether test, p-value <0.05). Legend colors indicates feature group.*

We then investigated the predictive power for identifying pathologic complete response (pCR) vs. non-complete response (MRD and GRD). The AUC range of radiomic features (Figure 2.B) was 0.52–0.63 and 0.51–0.55 for conventional features (described in Table S3 in Supplement II). The best performing radiomic feature, Wavelet HLL mean, was also the only significantly predictive feature (AUC = 0.63, p-value = 0.01, Noether test) and was risk proportional. No conventional imaging features were predictive for pCR (range 0.51 to 0.55, all p-value>0.05).

Multivariate models were created for each set of features (Figure 3), including conventional (3 features), radiomics (7 predictive features for GRD) and the resulting combined set (10 features). The median AUC values of the cross validation were 0.57, 0.65 and 0.65 for GRD and 0.60, 0.61 and 0.68 for pCR respectively for conventional, radiomics and combined models. The combined and radiomics model for GRD performed significantly better on the cross validation compare to clinical features alone. For pCR, no significant difference was found between radiomics and conventional model performance. However, the combined model significantly outperformed both radiomics and conventional features for predication of pCR.

Figure 3

Comparison of multivariate models for A) Gross residual disease and B) complete pathological response. AUC from the validation is reported from the Cross-Validation (100 iterations, 80% training and 20% validation) for each model. Comparison between models were done using a permutation test. A "" is reported if the model performance is significantly greater than the other, else "ns" for non-significant.*
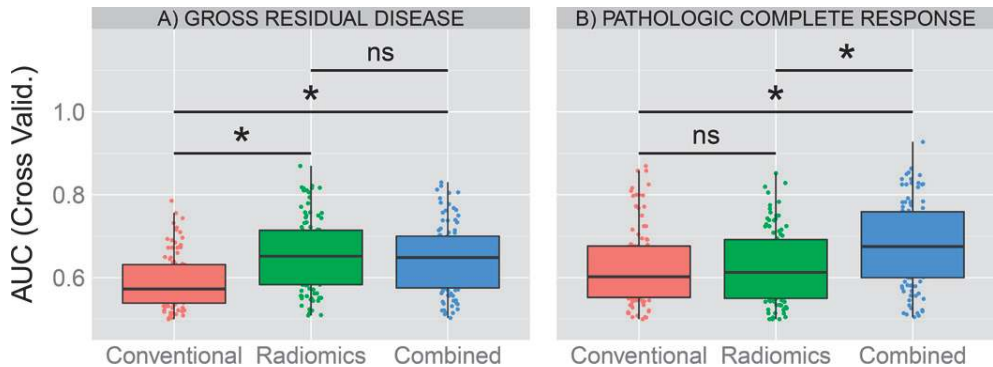
DISCUSSION

Radiomics [1] is an emerging field of quantitative imaging that aims to extract phenotypic tumor information from clinical imaging data. In this study we demonstrate the predictive power of radiomic phenotypic features for pathological response in patients with NSCLC. Pathological response is a standard endpoint, assessed at time of surgery for a direct measure of neoadjuvant chemotherapy effect. Pathologic response was significantly associated with clinical outcomes in our study (distant metastasis and local recurrence) showing the importance of predicting pathological outcome. Using data mining techniques, we selected fifteen radiomics features based on stability and variance. From those features, we identify seven features predictive of pathological gross residual disease (GRD) and one feature predictive of pathological complete responders (pCR). In contrast, no pre-treatment conventional clinically utilized features (volume or diameter) were predictive for pathological response. Lastly, the radiomics model also performed significantly better on a cross validation than the conventional model for GRD. A combined model using conventional and radiomics outperformed for pCR.

The seven significant predictive features for GRD enabled us to represent phenotypic characteristics of a lung tumor that is less likely to respond to neoadjuvant chemoradiation. Spherical disproportionality (a measure of the similarity between the tumor and a sphere with an equivalent volume) of the primary tumor site was associated with pathological response (AUC above 0.6 for both GRD and pCR). Remarkably, more complex tumors are likely to be associated with pCR while more spherical tumors are more likely to be associated with GRD. In comparison, conventional radiographic features such as volume and diameter were not predictive for pathological response (neither pCR nor GRD), and may not capture sufficient shape information as advanced shape radiomic features. It is noteworthy that larger tumor dimension (both volume and diameter) appeared to be associated with pCR. Tumor with large flat zones (area of connecting voxels of similar intensity) were associated with pCR whereas tumor with rich complex patterns (heterogeneities) were associated with GRD. Finally based on the common significant predictor for both pCR and GRD (Wavelet HLL – mean) indicated that tumor with overall lower voxel intensity (darker) were associated with GRD after Wavelet HLL filtering. Multivariate performances demonstrated that radiomics information added complementary information for pCR (significantly greater performance for the combined model) and outperformed conventional features alone for GRD.

To our knowledge this is the first study investigating pathological response to neoadjuvant chemoradiation using advanced quantitative imaging features from CT images. A previous study from Ravanelli *et al.* [28] investigated response of first line of chemotherapy (with no concurrent radiation therapy or surgery) for patients with advanced NSCLC using two textural features (uniformity and grey-level) from contrast enhanced CT images. They found an AUC of 0.74 for their multivariate model (Leave one out cross validation). Conventional features (tumor volume and diameters) have been identified as a

prognostic factor in prior studies [29–31]. However patients in those reports did not receive surgical resection as part of their therapy as did patients in our study. Conventional features may thus be more relevant as a prognostic factor in the absence of surgical intervention. Additionally, we did not find any significant association between radiomics and clinical outcomes, supporting the idea that surgical resection of the tumor undermines image-based features prediction for clinical outcomes. We believe that the underlying tumor phenotype may be more relevant for identifying pathologic response than either volume or diameter, although these may remain valuable for assessing response to therapy. Other studies [32,33] investigated the association of conventional features (tumor volume) but demonstrated lack of correlation between measured tumor volume change and pathologic response in NSCLC, supporting the fact that none of our conventional features were significant.

Limitations of this study include the cohort size (n=127). Due to a large number of available radiomic features and concern for multiple testing, dimension reduction was used with restrictive criteria to perform drastic selection, therefore excluding potential predictive features for any clinical outcomes. However despite this very restrictive data-driven approach selection (about 1% remaining after), we were able to find predictive features. Additionally, the incidence of pCR (n = 27) was much lower than GRD (n = 67), potentially explaining the differences in the predictive power for pathological complete response. Since the study period was conducted over 12 years, we recognize that there is heterogeneity in the treatments delivered to the patient cohort. However, we did not find any association between radiation dose, cycles of chemotherapy or other treatment characteristics and pathologic response. Thus, we believe the results presented remain valid despite treatment heterogeneity. A major limitation of radiomics is the lack of standardization in image acquisition that without access to raw images is difficult to control and hence standardize patient cohorts. Yet, despite the noise introduced by variation in acquisition protocols, radiomic models have shown consistent and reproducible association with outcomes in multiple independent datasets [7,8,9–11]. Furthermore, we have previously demonstrated the reproducibility [7] of the radiomics features using the RIDER [34] test / retest dataset and we resampled of imaging using a common pixel spacing (3×3×3mm3) to limit variability across patients. Finally, we acknowledge the limitations of the clinical applicability of this study. This study was initiated to find potential patterns in tumor phenotype that could predict pathological response prior to the start of therapy. Although several features were significantly predictive for pathological endpoints, we were unable to identify subgroups associated with overall survival, local recurrence or distant metastatis using radiomic features. Despite the need for further validation sets, this study provides a basis for additional research (e.g using PET-CT features) that could improve performance. We believe such radiomics based analyses can be used as a complementary method of patient stratification for NSCLC prior to the initiation of therapy as is currently being investigated in other disease sites such as breast [35,36], colorectal cancer [37], and glioblastoma [38].

In conclusion, we identified CT-based radiomic features predictive for pathological response in patient with locally advanced NSCLC who previously received trimodality therapy. This study demonstrates that radiomics can provide additional phenotypic information that may reflect underlying tumor sensitivity to chemoradiation. Predicting pathological response prior to initiation of neoadjuvant chemoradiation has significant potential clinical applications such as developing adaptive therapy based on pre-treatment tumor phenotype. These radiomic features which can be captured from clinically-available imaging modalities performed better than conventionally reported metrics.

HIGHLIGHTS

- Early prediction of response to neoadjuvant chemoradiation is crucial for improving overall treatment and patient outcomes.
- This study demonstrated an association between Radiomic features and pathological response for lung cancer patients.
- A Radiomics-Conventional combined model shown better performance for pathological response.

SUPPLEMENTARY MATERIAL

Available online at:
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4930885/bin/NIHMS778030-supplement.pdf

REFERENCES

1. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. Eur. J. Cancer. 2012;48:441–446.
2. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: the process and the challenges. Magn. Reson. Imaging. 2012;30:1234–1248.
3. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology. 2015:151169.
4. Parmar C, Rios Velazquez E, Leijenaar R, Jermoumi M, Carvalho S, Mak RH, et al. Robust Radiomics Feature Quantification Using Semiautomatic Volumetric Segmentation. In: Woloschak GE, editor. PLoS ONE. Vol. 9. 2014. p. e102107.
5. Rios Velazquez E, Aerts HJWL, Gu Y, Goldgof DB, De Ruysscher D, Dekker A, et al. A semiautomatic CT-based ensemble segmentation of lung tumors: Comparison with oncologists' delineations and with the surgical specimen. Radiother. Oncol. 2012;105:167–173.

6. Leijenaar RTH, Carvalho S, Velazquez ER, van Elmpt WJC, Parmar C, Hoekstra OS, et al. Stability of FDG-PET Radiomics features: An integrated analysis of test-retest and interobserver variability. Acta Oncol. 2013;52:1391–1397.
7. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Cavalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat. Commun. 2014:5.
8. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL. Machine Learning methods for Quantitative Radiomic Biomarkers. Sci. Rep. 2015;5:13087.
9. Leijenaar RT, Carvalho S, Hoebers FJ, Aerts HJ, van Elmpt WJ, Huang SH, et al. External validation of a prognostic CT-based radiomic signature in oropharyngeal squamous cell carcinoma. Acta Oncol. 2015:1–7.
10. Coroller TP, Grossmann P, Hou Y, Rios Velazquez E, Leijenaar RTH, Hermann G, et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. Radiother. Oncol. 2015;114:345–350.
11. Parmar C, Leijenaar RTH, Grossmann P, Rios Velazquez E, Bussink J, Rietveld D, et al. Radiomic feature clusters and Prognostic Signatures specific for Lung and Head & Neck cancer. Sci. Rep. 2015;5:11044.
12. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014: Cancer Statistics, 2014. CA. Cancer J. Clin. 2014;64:9–29.
13. National Comprehensive Cancer Network. NCCN guidelines : Non-Small Cell Lung Cancer (Version 7.2015) 2015http://www.nccn.org/professionals/physician_gls/pdf/nscl.pdf.
14. Albain KS, Swann RS, Rusch VW, Turrisi AT, Shepherd FA, Smith C, et al. Radiotherapy plus chemotherapy with or without surgical resection for stage III non-small-cell lung cancer: a phase III randomised controlled trial. The Lancet. 2009;374:379–386.
15. van Meerbeeck JP, Kramer GWPM, Van Schil PEY, Legrand C, Smit EF, Schramel F, et al. Randomized Controlled Trial of Resection Versus Radiotherapy After Induction Chemotherapy in Stage IIIA-N2 Non-Small-Cell Lung Cancer. JNCI J. Natl. Cancer Inst. 2007;99:442–450.
16. Hellmann MD, Chaft JE, William WN, Rusch V, Pisters KM, Kalhor N, et al. Pathological response after neoadjuvant chemotherapy in resectable non-small-cell lung cancers: proposal for the use of major pathological response as a surrogate endpoint. Lancet Oncol. 2014;15:e42–e50.
17. Mouillet G, Monnet E, Milleron B, Puyraveau M, Quoix E, David P, et al. Pathologic complete response to preoperative chemotherapy predicts cure in early-stage non-small-cell lung cancer: combined analysis of two IFCT randomized trials. J. Thorac. Oncol. Off. Publ. Int. Assoc. Study Lung Cancer. 2012;7:841–849.
18. Isobe K, Hata Y, Sakaguchi S, Sato F, Takahashi S, Sato K, et al. Pathological response and prognosis of stage III non-small cell lung cancer patients treated with induction chemoradiation. Asia Pac. J. Clin. Oncol. 2012;8:260–266.
19. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1) Eur. J. Cancer. 2009;45:228–247.
20. Werner-Wasik M, Xiao Y, Pequignot E, Curran WJ, Hauck W. Assessment of lung cancer response after nonoperative therapy: tumor diameter, bidimensional product, and volume. A serial CT scan-based study. Int. J. Radiat. Oncol. Biol. Phys. 2001;51:56–61.
21. Pieper S, Halle M, Kikinis R. 3D Slicer. IEEE Int. Symp. Biomed. Imaging Nano Macro 2004. 2004;1:632–635.
22. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013. ISBN 3-900051-07-0, URL http://www.R-project.org/
23. Haibe-Kains B, Desmedt C, Sotiriou C, Bontempi G. A comparative study of survival models for breast cancer prognostication based on microarray data: does a single gene beat them all? Bioinformatics. 2008;24:2200–2208.
24. Schröder MS, Culhane AC, Quackenbush J, Haibe-Kains B. survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. Bioinformatics. 2011;27:3206–3208.
25. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004;5:R80.
26. Kuhn M. Building predictive models in R using the caret package. J. Stat. Softw. 2008;28:1–26.
27. American Joint Committee on Cancer. AJCC Cancer Staging Manual. 7th. New York: Springer; 2010. Available from: American Joint Committee on Cancer. AJCC Cancer Staging Manual, 7th ed. New York: Springer 2010.

28. Ravanelli M, Farina D, Morassi M, Roca E, Cavalleri G, Tassi G, et al. Texture analysis of advanced non-small cell lung cancer (NSCLC) on contrast-enhanced computed tomography: prediction of the response to the first-line chemotherapy. Eur. Radiol. 2013;23:3450–3455.

29. Bradley JD, Ieumwananonthachai N, Purdy JA, Wasserman TH, Lockett MA, Graham MV, et al. Gross tumor volume, critical prognostic factor in patients treated with three-dimensional conformal radiation therapy for non-small-cell lung carcinoma. Int. J. Radiat. Oncol. Biol. Phys. 2002;52:49–57.

30. Alexander BM, Othus M, Caglar HB, Allen AM. Tumor Volume Is a Prognostic Factor in Non–Small-Cell Lung Cancer Treated With Chemoradiotherapy. Int. J. Radiat. Oncol. 2011;79:1381–1387.

31. Stinchcombe TE, Morris DE, Moore DT, Bechtel JH, Halle JS, Mears A, et al. Post-chemotherapy gross tumor volume is predictive of survival in patients with stage III non-small cell lung cancer treated with combined modality therapy. Lung Cancer Amst. Neth. 2006;52:67–74.

32. Cerfolio RJ, Bryant AS, Winokur TS, Ohja B, Bartolucci AA. Repeat FDG-PET After Neoadjuvant Therapy is a Predictor of Pathologic Response in Patients With Non-Small Cell Lung Cancer. Ann. Thorac. Surg. 2004;78:1903–1909.

33. Poettgen C, Theegarten D, Eberhardt W, Levegruen S, Gauler T, Krbek T, et al. Correlation of PET/CT Findings and Histopathology after Neoadjuvant Therapy in Non-Small Cell Lung Cancer. Oncology. 2007;73:316–323.

34. Zhao B, James LP, Moskowitz CS, Guo P, Ginsberg MS, Lefkowitz RA, et al. Evaluating Variability in Tumor Measurements from Same-day Repeat CT Scans of Patients with Non– Small Cell Lung Cancer 1. Radiology. 2009;252:263–272.

35. Pickles MD, Lowry M, Gibbs P. Pretreatment Prognostic Value of Dynamic Contrast- Enhanced Magnetic Resonance Imaging Vascular, Texture, Shape, and Size Parameters Compared With Traditional Survival Indicators Obtained From Locally Advanced Breast Cancer Patients: Invest. Radiol. 2016;51:177–185.[

36. Fox MJ, Gibbs P, Pickles MD. Minkowski functionals: An MRI texture analysis tool for determination of the aggressiveness of breast cancer. J. Magn. Reson. Imaging JMRI. 2015

37. Lubner MG, Stabo N, Lubner SJ, Del Rio AM, Song C, Halberg RB, et al. CT textural analysis of hepatic metastatic colorectal cancer: pre-treatment tumor heterogeneity correlates with pathology and clinical outcomes. Abdom. Imaging. 2015;40:2331–2337.

38. Lee J, Jain R, Khalil K, Griffith B, Bosca R, Rao G, et al. Texture Feature Ratios from Relative CBV Maps of Perfusion MRI Are Associated with Patient Survival in Glioblastoma. AJNR Am. J. Neuroradiol. 2016;37:37–43.

ACKNOWLEDGEMENTS

Conflict of Interest Statement

None.

CHAPTER

9

CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma

CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma

Thibaud P. Coroller*, Patrick Grossmann*, Ying Hou, Emmanuel Rios Velazquez, Ralph T.H. Leijenaar, Gretchen Hermann, Philippe Lambin, Benjamin Haibe-Kains, Raymond H.Mak, and Hugo J.W.L. Aerts
*These authors contributed equally to this work

ABSTRACT

Background and Purpose
Radiomics provides opportunities to quantify the tumor phenotype non-invasively by applying a large number of quantitative imaging features. This study evaluates computed-tomography (CT) radiomic features for their capability to predict distant metastasis (DM) for lung adenocarcinoma patients.

Materials and Methods
We included two datasets: 98 patients for discovery and 84 for validation. The phenotype of the primary tumor was quantified on pre-treatment CT-scans using 635 radiomic features. Univariate and multivariate analysis was performed to evaluate radiomics performance using the concordance index (CI).

Results
Thirty-five radiomic features were found to be prognostic (CI > 0.60, FDR < 5%) for DM and twelve for survival. It is noteworthy that tumor volume was only moderately prognostic for DM (CI=0.55, p-value=2.77×10^{-5}) in the discovery cohort. A radiomic-signature had strong power for predicting DM in the independent validation dataset (CI=0.61, p-value=1.79×10^{-17}). Adding this radiomic-signature to a clinical model resulted in a significant improvement of predicting DM in the validation dataset (p-value=1.56×10^{-11}).

Conclusions
Although only basic metrics are routinely quantified, this study shows that radiomic features capturing detailed information of the tumor phenotype can be used as a prognostic biomarker for clinically-relevant factors such as DM. Moreover, the radiomic-signature provided additional information to clinical data.

INTRODUCTION

Lung cancer is the most deadly cancer worldwide for both men and women[1]. Nonsmall cell lung cancer (NSCLC) is the most common type of lung cancer (85–90% of all lung cancers) and adenocarcinoma is the most common subtype (about 40% of all lung cancers) of NSCLC. Patients with locally advanced (stage II-III) lung adenocarcinomas are typically treated with combined modality therapy including chemotherapy with local therapy including radiation therapy and/or surgery, but overall survival remains low due to a high risk of local recurrence and distant metastasis (DM) after treatment. Despite the use of concurrent chemotherapy with local therapy, the incidence of DM after combined modality therapy is as high as 30–40% in prospective trials [2–4]. However, large randomized trials studying consolidation chemotherapy after concurrent chemotherapy and radiation therapy have not shown improvement in overall survival with additional chemotherapy[5, 6] likely because there was no selection of patients at the highest risk of DM. Therefore, developing better biomarkers to predict patients at highest risk for DM may help identify sub-groups who benefit from intensification of systemic therapy and is crucial for improving outcomes.

Due to recent technological advances in medical imaging it is possible to capture tumor phenotypic characteristics non-invasively. The most widely used imaging modality is Computed-Tomography (CT), which can quantify tissue density. In lung cancer, CT imaging is routinely used for patient management, including diagnosis, radiation treatment planning and surveillance.

Tumor phenotypic differences (e.g. shapes irregularity, infiltration, heterogeneity or necrosis) can be quantified in CT images using radiomic features. Radiomics [7–9] aims to provide a comprehensive quantification of the tumor phenotype by analyzing robustly [10–12] a large set of quantitative data characterization algorithms . Biomarkers based on quantitative features have demonstrated strong prognostic performance across a range of cancer types and investigators have reported that these features are associated with clinical outcomes and underlying genomic patterns [13–26]. Radiomics has significant clinical potential, as it can be applied to routinely acquired medical imaging data at low costs.

In this manuscript we present a radiomic analysis to identify biomarkers of DM in patients treated with chemoradiation (chemoRT) for locally advanced lung adenocarcinoma. In a discovery dataset, we extracted 635 radiomics features to identify the optimal features for predicting metastasis. Only a limited number of features with high performance for predicting DM were tested in the independent validation dataset. We evaluated the ability of radiomic features to predict DM or overall survival, and how these features compare with basic metrics (e.g. volume, diameter) as prognostic factors [27–30].

MATERIALS AND METHODS

Patient characteristics

This study is an Institutional Review Board-approved analysis of CT for treatment simulation from North-American NSCLC patients receiving chemoRT at our institution from 2001 to 2013. We limited the patient population to pathologically-confirmed lung adenocarcinoma with locally advanced disease (overall stage II-III)[30]. Patients with surgery or chemotherapy before the scheduled radiation therapy planning CT date were excluded from the study. Patients treated before July 2009 were included in the discovery Dataset1 (n=98), and after July 2009 in an independent validation Dataset2 (n=84). In total 182 patients were included in our analysis.

Clinical endpoints

Patients were followed up every three to six months after treatment, and surveillance chest CT scans with contrast (unless patient's contraindication, e.g. allergy or renal dysfunction) were performed to assess treatment response or tumor progression based on US national guidelines (NCCN). The primary endpoint of this study was distant metastasis (DM), which was defined as progression of disease to other organs as assessed in surveillance scans, and time to DM was defined as time from start of radiation to date of DM or censoring (date of last scan). Overall survival was analyzed as a secondary endpoint, and was defined as the time between the start of radiation treatment and last day of follow up or date of death.

Clinical variables

The conventional clinical prognostic factors (CPFs) used for this study included tumor grade (1-Well differentiated, 2-Moderately differentiated, 3-Poorly differentiated and 4-Not available), Eastern Cooperative Oncology Group (ECOG) performance status (PS)[31], TNM stage per the American Joint Committee on Cancer (AJCC) staging system (7th edition)[30]; CT-based measurements commonly utilized in the clinic (e.g. tumor volume and maximal tumor diameter measured on single axial slice), and treatment characteristics. Sub-group analyses of clinical variables were performed (e.g. overall stage II vs IIIA vs IIIB) and can be found in Table S1 (Supplement II.1).

CT acquisition and segmentation

Planning CT was performed according to standard clinical scanning protocols at our institution with a GE "LightSpeed" CT scanner (GE Medical System, Milwaukee, WI, USA). The most common pixel spacing was (0.93mm, 0.93mm, 2.5mm) for CT. The primary lung tu-

mor was delineated manually on Eclipse (Varian Medical System, Palo Alto, CA, USA). It was first contoured in the abdomen window to identify the boundaries with the chest wall or other soft tissues, then in the lung window to capture the maximum extent in the lung parenchyma. All contours were reviewed by an experienced radiation oncologist (R.H.M).

Radiomic features extraction

Radiomic features have the capacity to capture tumor phenotype differences by examining a large set of quantitative features (Figure 1). The feature extraction was performed in MATLAB 2013b (Mathworks, Natick, MA, USA) using an in-house developed toolbox running on the Computational Environment for Radiotherapy Research (CERR)[32]. DICOMs files (CT images + tumor contours) were imported into CERR to extract the radiomic features. The radiomic features set included is described in detail in the Supplement I.
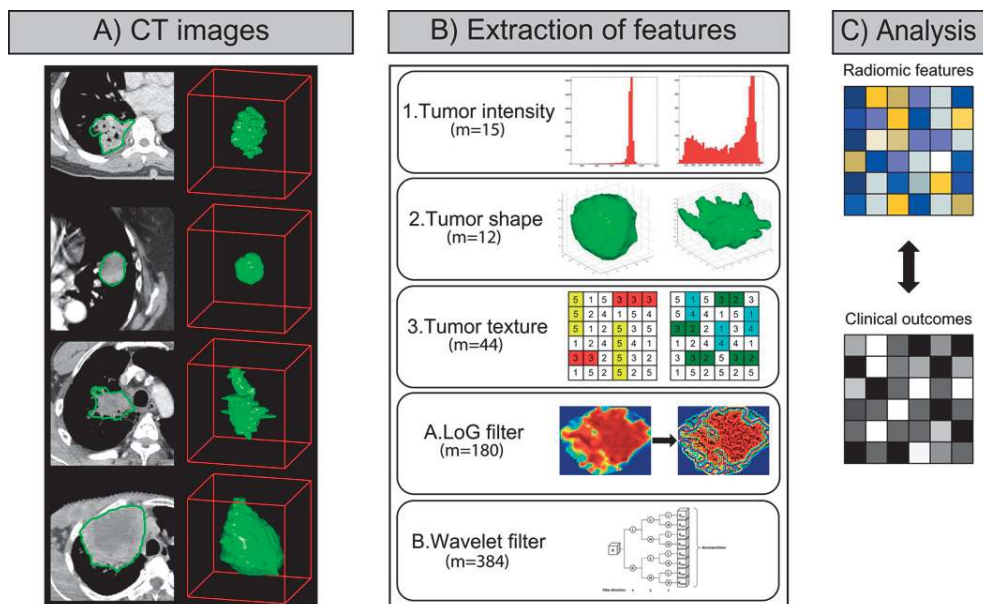
Figure 1
A) Differences between lung primary tumors with a same histology are apparent on CT images (3D model on the right and CT contours on the left). CT images of primary tumors contain critical information that can be used to predict outcomes or assess the RT treatment response. B) To quantify this information, a large set of features (m=635) is used to capture the tumor phenotype. It includes 1| intensity, 2| shape and 3| texture based features. Also, A| Laplacian of Gaussian (LoG) and B| Wavelet filtered features were investigated. C) The final step is to link radiomic information to clinical data.

Feature selection

Feature selection for the radiomic signature was performed with the minimum redundancy maximum relevance (mRMR) algorithm implemented in the mRMRe[33] package version 2.0.4 in R. The mRMR algorithm is an entropy based feature selection method, which starts by calculating the mutual information (MI) between a set of features and an outcome variable. MRMR ranks the input features by maximizing the MI with respect to outcome and minimizing the average MI of higher ranked features. Here, survival objects as implemented in R with "Survcomp" package[34] were used as outcome to select complementary features with respect to DM or survival.

Among available clinical covariates, those with $p < 0.1$ on univariate analysis of DM using a Log-Rank test were included into a multivariate clinical prognostic model.

Data analysis

Univariate and multivariate analyses were performed for this study. All analysis were performed on Dataset1, leaving Dataset2 as an independent validation cohort for evaluating the radiomic signature.

Statistical analysis was conducted using the survcomp[34] package version 1.12 and rmeta[35] package version 2.16 in Bioconductor[36]. Prognostic performances were evaluated by the concordance index[37] (CI), which is the probability that among two randomly drawn samples, the sample with the higher risk value has also the higher chance of experiencing an event (e.g. death or development of DM). CIs were either directly computed for continuous variables or on the predictions of a univariate Cox model with clinical categorical variables. Kaplan-Meier and Log-Rank statistics were used to analyze the univariate discrimination of survival and DM groups by imaging features and clinical covariates. To build the multivariate radiomic signature for DM, Cox regression models were trained on Dataset1 for selected prognostic variables and the predictions by these models were validated on Dataset2. Features were incrementally added to the model according to the relevance rank calculated by mRMR[33]. Intermediate models were tested by repeated random sub-sampling cross validation with 1,000 iterations on Dataset1. Once the mean CI of the growing model dropped, the corresponding feature set was retained selected as the final model. Only this selected model was and validated on Dataset2. Significance of CIs was assessed by bootstrapping subsamples of size 100 with 100 repetitions for A) true survival data and B) random permutations of survival data, and comparing the empirical distributions of A) and B) by an one-sided Wilcoxon signed rank test. The same procedure was used to assess if a CI was higher than another CI. To correct for multiple comparisons, we additionally adjusted P-values by the false-discovery-rate (FDR) procedure according to Benjamini and Hochberg[38]. All statistical analysis was performed using the R software[39] version 3.0.2.

RESULTS

The majority of all patients were female (62.6%) and the median age at start of treatment was 64 years (range: 35–93 years). The median follow-up time was 23.7 months (range: 1.8–119.2 months) and the median survival time was 24.7 months (range: 1.8–119.2 months). The median time to distant metastasis (DM) was 13.4 months (range: 0.3–117.5 months). Patient characteristics, clinical outcomes are shown in Table 1.

Table 1
Patient characteristics and outcomes are reported for each datasets. For categorical variables, actual numbers are reported for each category (format A/B/C). Statistical comparison between dataset 1 and 2 was computed using Chi Square (categorical variables) or Wilcoxon rank sum test (continuous variables).

Time to DM was similar between Dataset1 and Dataset2 (p-value < 0.36), as for the numbers of DM (p-value < 0.45). However, survival (p-value < 0.005) and follow-up times (p-value < 0.007) were significantly different in Dataset1.

We investigated the association of radiomics data with DM and overall survival. In Figure 2 the association of the imaging features with DM and survival in the discovery Dataset1 is shown. Of the complete radiomic features set (m=635), a total of 520 (81.88%) and 582 (91.65%) features were significant from random (FDR < 5%) for DM and survival, respectively. A total of 445 radiomic features were significant for both DM and survival. A high linear relationship was observed (R^2 =0.92, p-value < 2.7×10^{-243}), for the features significant for both DM and survival. It is noteworthy that LoG features had the highest performance compared to the other features groups.

Among all features, thirty-five radiomics features were strongly prognostic (CI > 0.60 and FDR < 5%) for DM (Table S2 in the Supplement II.3). Twelve features were found prognostic for survival. Specific details on statistic values of these features can be found in Table S3 in Supplement II.3. Between these two top performing feature sets there were four common prognostic features for both DM and survival. All of them were LoG based features (3 entropy and 1 standard deviation).

We compared the top 15 features that had the highest CIs (Top15), with tumor volume and diameter (equivalent to basic metrics). The Top15 radiomic features had notably higher CIs compared to tumor volume and diameter (Figure 3.A).
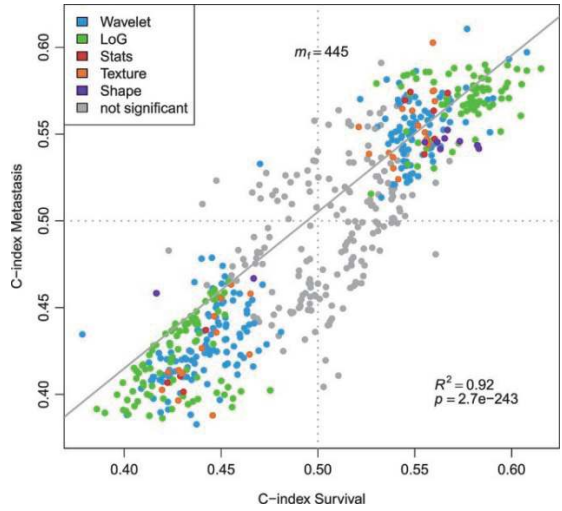
Figure 2

Univariate performances of prognostic features for Distant Metastasis (DM) and survival. Each point refers to the CI of a feature evaluating the power of feature to predict metastasis, respectively, survival. Colors refer to the type of feature. Features whose CI estimation was not significant (FDR < 5%) for both DM and survival are shown in gray. Overall, 445 of these pairs of CIs are considered to be significant estimates. Linear regression for all significant pairs of CIs yielded an R-squared value of 0.92 (F-test, p-value < 2.7e-243).
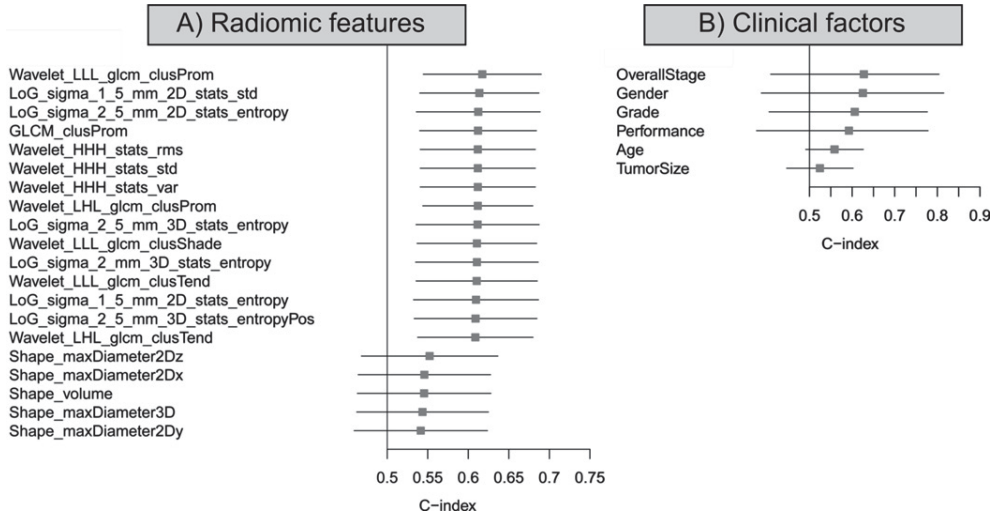
Figure 3

A) Forest plot of the 15 best performing radiomic features for Distant Metastasis on univariate analysis (Dataset1, n=98). Radiomics equivalent of basic metrics (diameter and volume) was added for comparison. B) Forest plot of the clinical factors. The absolute C-indices and their 95% confidence interval are shown.

We also investigated the association of CPFs with DM in our data set. Three clinical parameters appeared to be significant univariate prognostic factors: Overall Stage (CI=0.63, p-value < 6.78×10^{-14}), Gender (CI=0.63, p-value < 2.35×10^{-11}) and tumor grade (CI=0.61, p-value < 2.35×10^{-11}). Clinical parameters, ranked by their CI are displayed in Figure 3.B. Overall stage and gender yielded a higher CI than the radiomic features, although their 95% confidence interval is wider compared to the radiomic features.

An mRMR based feature selection on all features on Dataset1 (n=98) was performed to reduce redundancy and select a potential set of complementary and prognostic features. From this new ranking, the 15 highest mRMR-ranked features were kept after feature selection to build the radiomic signature. A multivariate Cox regression model to predict DM was developed. Features were iteratively added in order of high to low mRMR rank on Dataset1, and Dataset2 was used for independent validation. The combination that yielded the maximum CI on the discovery Dataset1 before dropping was defined as the optimal radiomic signature for predicting DM. This signature consists of three features: 1) Wavelet HHL – Skewness, 2) Gray-Level Co-occurrence Matrix – Cluster shade, and 3) LoG 5mm 2D – Skewness. Cluster shade is a textural feature sensitive to tumor heterogeneities. Skewness is a first-order feature that measures the asymmetry of the histogram from the mean, which here is associated with two different filters LoG and Wavelet.

As a final step, we compared the radiomic signature to a clinical Cox regression model containing covariates that significantly discriminated between patients with and without DM in Dataset1 in univariate analysis. The final model contained overall stage and tumor grade. This clinical model showed moderate prognostic power when applied to Dataset2 with coefficients trained on Dataset1 (CI=0.57, p-value < 1.03×10^{-7}). Combining the clinical and radiomic signature (trained on Dataset1) showed a significantly (p-value < 1.56×10^{-11}) higher association with DM when applied to Dataset2 (CI=0.60, p-value < 3.57×10^{-16}), compared to the clinical model. A median split of the patient prediction scores from applying the combined model on Dataset2 yielded a significant difference (p-value = 0.049) for metastasis-free probability estimates (Figure 4).
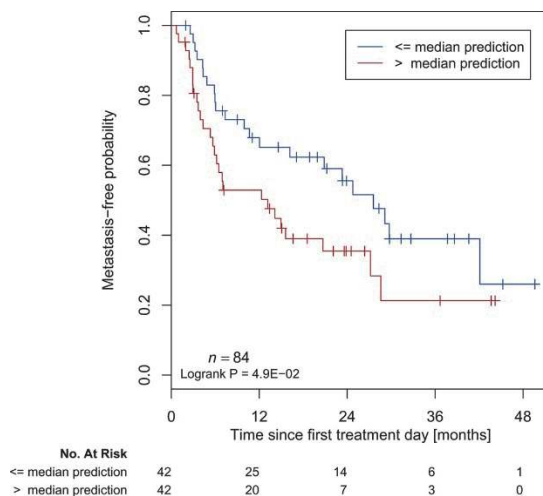
Figure 4

Kaplan Meier curves according to the combined model predicting score to predict metastasis-free probability in an independent dataset. A significant survival difference appears between patients with a high or low risk of Distant Metastasis (Dataset2, n=84, Log-Rank test, p-value < 0.049).

DISCUSSION

Medical imaging gives valuable information for diagnostic, treatment planning or surveillance of cancer patients. Routinely, basic metrics are extracted from these images to utilize as a prognostic factor [27–30], or to assess treatment response. However, there is much more tumor phenotypic information captured in these images. Radiomics are able to quantify tumor phenotypical differences from medical images by using a large set of imaging features that can be linked to clinical factors of the tumors. In this study we extracted 635 radiomic features from a total of 182 lung cancer patients treated with chemoRT to assess the ability of radiomic features as a prognostic biomarker for distant metastasis (DM), and we validated a radiomic-based signature on an independent validation dataset. Since DM remains a major cause of mortality in 30–40% of patients with locally advanced lung adenocarcinoma, early identification of patients at highest risk of developing DM would allow clinicians to adapt treatment such as incorporating consolidation chemotherapy to improve outcomes. Moreover, the theoretical benefit of consolidation chemotherapy has not been shown in large randomized studies to date. It is likely because there was no selection of patients at the highest risk of distant metastases (i.e. patients who were at low risk of distant metastases were included in these trials and would not need additional treatment). Future trial design to demonstrate benefits of con-

solidation chemotherapy will likely require stratification to identify those at the highest risk of distant metastases and may benefit most from additional treatment.

We observed strong individual correlations between clinical outcomes and quantitative imaging features. A large number of features were significant from random to predict DM (91%) and survival (82%) in univariate analysis after correction for multiple testing. Moreover, a high linear correlation was found among those 445 features that were significant factors of both DM and survival (R^2 =0.92, p-value < 2.7 × 10^{-243}). This high linear correlation is expected as there is a high correlation between DM and survival (DM greatly impact patient survival, see Table S4 in Supplement II.4). Only a small number of features, 35 for DM and for 12 survival, were prognostic, as defined by a CI > 0.6 and FDR < 5%.

Although we tested a large number of features, to minimize any risk of overfitting or bias, we performed a robust validation approach: all analysis steps, mRMR feature selection, and model fitting were performed on Dataset1 (n=98) and the results validated on an independent validation Dataset2 (n=84). With this approach we found a multivariate radiomic DM signature consisting of three features that yielded a high prognostic performance for DM in Dataset1 (CI=0.61). Combining the radiomics signature to clinical model predictors showed significant improvement (p-value < 1.56 × 10^{-11}), compared to the clinical predictors alone.

A recent study from Fried *et al*.[22] investigated DM prediction for NSCLC patients. They found a significant model DM (P-value=0.005) using both texture features and CPFs. The model used consisted of eight parameters (two CPFs and six textures). In another study, Ganeshan *et al*.[15] applied textural analysis to find univariate prognostic factors for survival. They focused on two imaging features (uniformity, associated with two LoG filter). In our analysis, these features were significant from random but lowly ranked by their CI value (184th and 146th CI-ranked features in Dataset1). However, major differences in studies design and implementation made it difficult to compare them objectively. Fields *et al*.[22] used leave-one out cross validation to validate their model instead of an independent validation dataset. Ganeshan *et al*.[15] only used one CT image slice (presenting the largest cross section) to calculate their features when we used the whole primary tumor. Finally, both these studies have a smaller patient cohort, n=54[15] and n=91[22], and had mixed histology patients. Our analysis calculated the features from the complete 3D tumor volume, contained only a single histology of NSCLC (adenocarcinoma), and is based on larger cohorts (n=182) with an independent validation dataset for the radiomic signature.

A complementary point of the study was to compare basic metrics [27–29] to radiomic features as prognostic factors for DM. The first observation made was that Shape-Maximum diameter (in every direction x/y/z) is a better univariate prognostic factors than the maximal tumor diameter on an axial slice reported by a radiologist. The advantage of the radiomic shape features is that they can be automatically acquired, reproducible[10–12], and take into account the whole tumor volume, whereas clinically assessed tumor

diameters are manually drawn on a CT slice and are therefore limited to one dimension of the tumor. Furthermore, shape or size-based features were not in the top ranked features in our study. Total tumor volume, has been associated with survival in stage I-III NSCLC patients treated with radiation therapy in a study from Etiz *et al*.[28], and a prior study from our institution by Alexander *et al*.[29] also demonstrated an association between primary tumor volume and overal survival, but not risk of distant metastasis. In our study, volume was ranked only the 405th (CI=0.55) and 224[th](CI=0.56) best univariate prognostic factor for DM and survival respectively in Dataset1. Thus, while basic metrics such as size and volume have historically been used as used in the clinical setting because such data are easily acquired, radiomic shape and size measurements can provide stronger prognostic factors.

A short-coming of our study is the variability in CT acquisition and reconstruction parameters. Our dataset includes patients from 2001 to 2013. During this time period, the standard of care for CT acquisition has evolved, differences appeared between our cohorts for some factors (Table 1). However, despite this variability in the imaging data (evolution of hardware, progress in informatics), radiomics was able to detect a strong signal to predict DM despite a temporal split. Additionally, clinical outcomes are provided by one center, which makes it hard to evaluate the generalizability of outcomes to other institutions. However, in comparison with a recent study[20] investigating clinical outcomes from another center, patient characteristics or outcomes were comparable. Future work would therefore involve studying the DM signature in other histologies and in independent validation sets from other institutions, assessing its generalizability to all NSCLC.

In conclusion, this study demonstrated strong association between radiomic features and DM for patients with locally advanced adenocarcinoma; and presented an independently validated radiomics signature for DM. This signature would allow early identification of patients with locally advanced lung adenocarcinoma at risk of developing DM, allowing clinicians to individualize treatment (such as intensification of chemotherapy) to reduce the risk of DM and improve survival.

HIGHLIGHTS

- Early prediction of patients that will develop distant metastasis (DM) is crucial for improving overall treatment and patient outcomes.
- This study demonstrated an association between radiomic features and DM for lung cancer patients.
- A combined signature with clinical and radiomic features was able to predict DM in an independent validation dataset.

SUPPLEMENTARY MATERIAL

Available online at:
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4400248/bin/NIHMS667506-supplement.docx

ACKNOWLEDGEMENTS

Disclaimer
None

Conflict of interest
None

REFERENCES

1. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014: Cancer Statistics: 2014. CA Cancer J Clin. 2014;64:9–29.
2. Albain KS, Swann RS, Rusch VW, Turrisi AT, Shepherd FA, Smith C, Chen Y, Livingston RB, Feins RH, Gandara DR, Fry WA, Darling G, Johnson DH, Green MR, Miller RC, Ley J, Sause WT, Cox JD. Radiotherapy plus chemotherapy with or without surgical resection for stage III non-small-cell lung cancer: a phase III randomised controlled trial. Lancet. 2009;374:379–386.
3. Curran WJ, Paulus R, Langer CJ, Komaki R, Lee JS, Hauser S, Movsas B, Wasserman T, Rosenthal SA, Gore E, Machtay M, Sause W, Cox JD. Sequential vs. concurrent chemoradiation for stage III non-small cell lung cancer: randomized phase III trial RTOG 9410. J Natl Cancer Inst. 2011;103:1452–1460.

4. Bradley JD, Paulus R, Komaki R, Masters GA, Forster K, Schild SE, Bogart J, Garces YI, Narayan S, Kavadi V, Nedzi LA, Michalski JM, Johnson D, MacRae RM, Curran WJ, Choy H Radiation Therapy Oncology Group. A randomized phase III comparison of standard-dose (60 Gy) versus high-dose (74 Gy) conformal chemoradiotherapy with or without cetuximab for stage III non-small cell lung cancer: Results on radiation dose in RTOG 0617. ASCO Meet Abstr. 2013;31:7501.

5. Hanna N, Neubauer M, Yiannoutsos C, McGarry R, Arseneau J, Ansari R, Reynolds C, Govindan R, Melnyk A, Fisher W, Richards D, Bruetman D, Anderson T, Chowhan N, Nattam S, Mantravadi P, Johnson C, Breen T, White A, Einhorn L. Phase III Study of Cisplatin, Etoposide, and Concurrent Chest Radiation With or Without Consolidation Docetaxel in Patients With Inoperable Stage III Non–Small-Cell Lung Cancer: The Hoosier Oncology Group and U.S. Oncology. J Clin Oncol. 2008;26:5755–5760.

6. Kelly K, Chansky K, Gaspar LE, Albain KS, Jett J, Ung YC, Lau DHM, Crowley JJ, Gandara DR. Phase III trial of maintenance gefitinib or placebo after concurrent chemoradiotherapy and docetaxel consolidation in inoperable stage III non-small-cell lung cancer: SWOG S0023. J Clin Oncol Off J Am Soc Clin Oncol. 2008;26:2450–2456.

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

8. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, Forster K, Aerts HJWL, Dekker A, Fenstermacher D, Goldgof DB, Hall LO, Lambin P, Balagurunathan Y, Gatenby RA, Gillies RJ. Radiomics: the process and the challenges. Magn Reson Imaging. 2012;30:1234–1248.

9. Lambin P, van Stiphout RGPM, Starmans MHW, Rios-Velazquez E, Nalbantov G, Aerts HJWL, Roelofs E, van Elmpt W, Boutros PC, Granone P, Valentini V, Begg AC, De Ruysscher D, Dekker A. Predicting outcomes in radiation oncology--multifactorial decision support systems. Nat Rev Clin Oncol. 2013;10:27–40.

10. Rios Velazquez E, Aerts HJWL, Gu Y, Goldgof DB, De Ruysscher D, Dekker A, Korn R, Gillies RJ, Lambin P. A semiautomatic CT-based ensemble segmentation of lung tumors: Comparison with oncologists' delineations and with the surgical specimen. Radiother Oncol. 2012;105:167–173.

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

12. Leijenaar RTH, Carvalho S, Velazquez ER, van Elmpt WJC, Parmar C, Hoekstra OS, Hoekstra CJ, Boellaard R, Dekker ALAJ, Gillies RJ, Aerts HJWL, Lambin P. Stability of FDGPET Radiomics features: An integrated analysis of test-retest and inter-observer variability. Acta Oncol. 2013;52:1391–1397.

13. Ganeshan B. Non_Small Cell Lung Cancer: Histopathologic Correlates for Texture Parameters at CT. Radiology

14. Davnall F, Yip CSP, Ljungqvist G, Selmi M, Ng F, Sanghera B, Ganeshan B, Miles KA, Cook GJ, Goh V. Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice? Insights Imaging. 2012;3:573–589.

15. Ganeshan B, Panayiotou E, Burnand K, Dizdarevic S, Miles K. Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: a potential marker of survival. Eur Radiol. 2012;22:796–802.

16. Ganeshan B, Abaleke S, Young RCD, Chatwin CR, Miles KA. Texture analysis of nonsmall cell lung cancer on unenhanced computed tomography: initial evidence for a relationship with tumour glucose metabolism and stage. Cancer Imaging. 2010;10:137–143.

17. He X, Sahiner B, Gallas BD, Chen W, Petrick N. Computerized characterization of lung nodule subtlety using thoracic CT images. Phys Med Biol. 2014;59:897–910.

18. Skogen K, Ganeshan B, Good C, Critchley G, Miles K. Measurements of heterogeneity in gliomas on computed tomography relationship to tumour grade. J Neurooncol. 2013;111:213–219.

19. Ravanelli M, Farina D, Morassi M, Roca E, Cavalleri G, Tassi G, Maroldi R. Texture analysis of advanced non-small cell lung cancer (NSCLC) on contrast-enhanced computed tomography: prediction of the response to the first-line chemotherapy. Eur Radiol. 2013;23:3450–3455.

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

21. Chae H-D, Park CM, Park SJ, Lee SM, Kim KG, Goo JM. Computerized Texture Analysis of Persistent Part-Solid Ground-Glass Nodules: Differentiation of Preinvasive Lesions from Invasive Pulmonary Adenocarcinomas. Radiology. 2014:132187.

22. Fried DV, Tucker SL, Zhou S, Liao Z, Mawlawi O, Ibbott G, Court LE. Prognostic Value and Reproducibility of Pretreatment CT Texture Features in Stage III Non-Small Cell Lung Cancer. Int J Radiat Oncol. 2014

23. Vaidya M, Creach KM, Frye J, Dehdashti F, Bradley JD, El Naqa I. Combined PET/CT image characteristics for radiotherapy tumor response in lung cancer. Radiother Oncol. 2012;102:239–245.

24. Aerts HJWL, Bussink J, Oyen WJG, van Elmpt W, Folgering AM, Emans D, Velders M, Lambin P, De Ruysscher D. Identification of residual metabolic-active areas within NSCLC tumours using a pre-radiotherapy FDG-PET-CT scan: A prospective validation. Lung Cancer. 2012;75:73–76.

25. Van Elmpt W, Das M, Hüllner M, Sharifi H, Zegers CML, Reymen B, Lambin P, Wildberger JE, Troost EGC, Veit-Haibach P, De Ruysscher D. Characterization of tumor heterogeneity using dynamic contrast enhanced CT and FDG-PET in non-small cell lung cancer. Radiother Oncol. 2013;109:65–70.

26. Balagurunathan Y, Gu Y, Wang H, Kumar V, Grove O, Hawkins S, Kim J, Goldgof DB, Hall LO, Gatenby RA, Gillies RJ. Reproducibility and Prognosis of Quantitative Features Extracted from CT Images. Transl Oncol. 2014;7:72–87.

27. Ball DL, Fisher RJ, Burmeister BH, Poulsen MG, Graham PH, Penniment MG, Vinod SK, Krawitz HE, Joseph DJ, Wheeler GC, McClure BE. The complex relationship between lung tumor volume and survival in patients with non-small cell lung cancer treated by definitive radiotherapy: A prospective, observational prognostic factor study of the Trans-Tasman Radiation Oncology Group (TROG 99.05) Radiother Oncol. 2013;106:305–311.

28. Etiz D, Marks LB, Zhou S-M, Bentel GC, Clough R, Hernando ML, Lind PA. Influence of tumor volume on survival in patients irradiated for non—small-cell lung cancer. Int J Radiat Oncol Biol Phys. 2002;53:835–846.

29. Alexander BM, Othus M, Caglar HB, Allen AM. Tumor Volume Is a Prognostic Factor in Non–Small-Cell Lung Cancer Treated With Chemoradiotherapy. Int J Radiat Oncol • Biol • Phys. 2011;79:1381–1387.

30. Mirsadraee S. The 7th lung cancer TNM classification and staging system: Review of the changes and implications. World J Radiol. 2012;4:0.

31. Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, Carbone PP. Toxicity and response criteria of the Eastern Cooperative Oncology Group. Am J Clin Oncol. 1982;5:649–655.

32. Deasy JO, Blanco AI, Clark VH. CERR: a computational environment for radiotherapy research. Med Phys. 2003;30:979–985.

33. De Jay N, Papillon-Cavanagh S, Olsen C, El-Hachem N, Bontempi G, Haibe-Kains B. mRMRe: an R package for parallelized mRMR ensemble feature selection. Bioinformatics. 2013;29:2365–2368.

34. Schröder MS, Culhane AC, Quackenbush J, Haibe-Kains B. survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. Bioinformatics. 2011;27:3206–3208.

35. Lumley T. Rmeta. 2012

36. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004;5:R80.

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

38. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B Methodol. 1995;57:289–300.

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

PART 5

Discussion and Future Perspectives

CHAPTER

10

General discussion and future perspectives

GENERAL DISCUSSION

Precision medicine is becoming the new paradigm in medical treatment, particularly in oncology. This paradigm aims at tailoring therapies on the basis of the individual markup of a patient to increase the probability of positive treatment outcome and reduce toxicity [1,2]. The fundamental backbone of precision medicine is the clinical availabilities of accurate biomarkers to map a patient to an optimized treatment plan [3]. While molecular biomarkers stratify patients based on their genomic profiles or expression signatures of a set of genes [4,5], to date these markers can usually only be measured when tumor tissue is obtained with invasive procedures, such as biopsies. In contrast, standard medical imaging can provide the foundation for non-invasive biomarkers [6].

Radiomics is a non-invasive imaging biomarker approach to automatically characterize the radiographic phenotype of a tumor on the basis of quantitative imaging features [7–9]. To achieve radiomic based tumor diagnostic and treatment response assessment, medical images (e.g., CT, MRT, or PET) are converted into quantitative data. Subsequently, these data are objectively mined for signaling patterns that indicate a tumor subtype. As this process is usually conducted on the entire volumetric segmentation of a tumor, radiomics allows a more comprehensive view as compared to tissue sampling. Although multiple studies have documented predictive and prognostic value of radiomics [10–13], clinical translation will require understanding of the underlying biological foundation of radiomics.

To this end, this thesis unraveled the mechanistic connections of radiomics and molecular biology of tumors and related those results to clinical disease progression of patients. In particular, we focused on phenotype-genotype interactions and on how these determine clinical factors of a tumor. Thereby, we achieved to develop a holistic view of cancerogenesis, which can be leveraged to inform clinical decisions. Furthermore, this thesis embedded radiomics in a wider perspective as to how radiomics can effectively be integrated into a clinical workflow to augment cancer diagnostics and treatment monitoring. Finally, data and code were made publicly available where possible to ensure reproducibility of the presented results and to support the growing radiomics community.

The results of this thesis were structured in three ways. First, the most extensive part of this thesis was uncovering the underlying molecular biology that drives quantitative radiomic phenotypes. Second, we show how radiomics can be used to develop non-invasive imaging biomarkers for targeted drugs. Finally, we presented novel results on the impact of key machine learning methods on prognostic value of radiomics

Understanding the Biological Principles of Radiomics

Imaging-genomics combines the fields of medical imaging and genomics to gain better understanding of the relationships of imaging and underlying genome driven events. A key question in imaging-genomics is whether imaging phenotypes can be explained in terms of

these underlying molecular processes. Understanding the imaging-genomic relationship translates into understanding of phenotype-genotype interactions in tumors and would allow using imaging as a surrogate for genomic interrogation in situations where biospecimen derived diagnostics are not feasible, such as in monitoring drug-targetable DNA mutations over time [6].

Different types of imaging-genomic studies can be conducted, depending on which imaging and genomic modalities are included. For example, radiomic or volumetric imaging approaches can be utilized, and different types of biological material can be investigated such as the expression or mutation of genes by RNA or DNA sequencing, respectively. The core of this dissertation is the integrative analysis of the biological underpinnings of advanced radiomic features and is presented in Chapter 2 [14]. In addition, we investigated how the underlying segmented tumor volumes behave as a function of molecular events [15,16]. Therefore, we analyzed how biological processes and somatic DNA mutations drive volumetric features and presented these results in Chapter 3 and 4, respectively.

The biological basis of radiomics

One common criticism of radiomics is that the underlying tumor biology is poorly understood, although biological reasoning of radiomic predictions is crucial for successful translation and approval into real world clinical applications. Therefore, in Chapter 2 we designed and carried out an integrated analysis published in the journal *eLife* [14]. With this analysis, we aimed at uncovering the mechanistic connections between radiomics, molecular biology, and clinical factors. We achieved this goal by including 1) a large number of advanced radiomic features that cover a wide range of metrics, 2) validation in novel and independent cohorts with large sample sizes to increase statistical power, 3) robust CT imaging of lung tumors, and 4) multivariable analysis to understand complementary effects of radiomic, genomic, and clinical data. To substantially increase the value of this study, we are also the first to validate key results of our analysis by immunohistochemical staining and to share data underlying the study, as well as the analysis source code, to the general scientific community.

In our analysis, we implemented a novel bi-clustering strategy to associate radiomic features with molecular pathways on the basis of global gene expression. Using this approach, we identified 13 modules each of which contained features and pathways that yielded coherent expression patterns. We extended these radiomic-pathway results to clinical factors by associating the radiomic features of all modules to tumor stage, histology, and overall survival. Strikingly, these radiomic-clinical associations robustly matched the radiomic-pathway results, closing the gaps between the combined connections of radiomics, biology, and clinical factors. For example, in one module textural entropy and dispersion features were prognostic and also associated with immune response pathways

and cell proliferation; this confirms the radiomic-pathway-clinical links as the prognostic implications of these pathways have been widely described [17–21].

Some of our findings have been previously indicated by preliminary studies of cohorts with limited scope. For example, a variety of our textural features were associated with stage and histology. Indeed, Ganeshan et al. [22] previously suggested textural CT features could predict whether a tumor was stage II or above; however, the features used in this study were derived from a single 2D image slice whereas our features were derived from the entire 3D tumor volume. Similarly, our results indicated that intensity features were associated with mitochondrial processing of oxygen, a key process in tumor hypoxia [23]. Indeed, Ganeshan et al. [24] also previously suggested that CT pixels correlate with hypoxia markers, such as Glut-1 or pimonidazole.

In addition to unraveling the relationship between radiomics, tumor biology, and clinical factors, we also developed radiomic activity predictors for specific pathways. For example, a tumor intensity based feature significantly predicted the autodegration of E3 Ubiquitin ligase COP1 (a direct regulator of p53) was also a prognostic factor. Similarly, tumor sphericity significantly predicted NFkB activation and was also a significant predictor of tumor stage and histology. These results further confirm the combined associations of radiomics, biology, and clinical factors. The clinical utility of such imaging based pathway predictors is still visionary, but could have its most effectiveness in areas where local response to targeted drugs administered over time has to be monitored longitudinally.

Finally, we addressed another key question of radiomics, which is whether radiomics provides an added prognostic benefit to genomic and clinical information. For this, we retrained and validated a previously published prognostic radiomic signature [10]. By integrating this radiomic with a previously published prognostic gene signature [25], we could demonstrate that the prognostic performance of the combined model exceeds the performance of the individual data types alone. Moreover, we integrated the radiomic and gene signature with clinical prognostic factors and convincingly observed the highest prognostic performance was achieved only by combining all three types of data. To further increase confidence in the validity of these results, we swapped the radiomic and gene signatures with other radiomic and gene signatures and consistently observed the same trend that the highest prognostic performance is yielded only by models that combine all three types of data.

It is important to note that our study benefits from multiple facets of validation. First, we included a thorough statistical design in which our putative results were generated on a discovery cohort and subsequently validated on an entirely independent validation cohort. Both cohorts were of large sample size and, importantly, these cohorts contained patients from two different continents (i.e., North America and Europe), which additionally suggests general applicability of our results. Second, we validated radiomic predictors of key cancer pathways involved in immune response and inflammation with immunohistochemical experiments. While reasonable statistical validation is becoming more

widely adapted, validating computational results biologically is still rather unique in radiomics and should be considered more frequently [6].

Although our study provides a considerable high standard of validation and reproducibility following several international guidelines [6,26], a few limitations can be noted as well. Primarily, the study cohorts were retrospectively collected. This implies that standardized protocols for imaging acquisition and registration were not available, although prognostic and predictive performances could benefit from standardized images [27]. Hence, further prospective studies will be required to assess clinical utility of our results and illustrate feasibility in clinical workflows. Furthermore, our cohorts focused on lower stage tumors and care has to be taken when generalizing to later stage tumors or even metastasized areas. Also, biological material analyzed in our study was obtained via single-needle biopsies; as lung cancer is a heterogenous disease [28,29] interpretation of all derived associations is challenging. To mitigate this issue, conducting multiple steps of statistical and biological validation were of particular importance for this study.

In conclusion, the radiogenomic study presented in Chapter 2 significantly advances our preliminary understanding of how molecular biology drives radiomic phenotypes and how these associations relate to clinical outcomes of patients; in particular, we show that radiomics contains complementary prognostic value compared to genomic and clinical data. Our study provides a rigorous classification of a broad set of radiomic features in terms of biological pathways and is the first study to validate these results by immunohistochemical staining and to release underlying data and analysis code to foster further translational radiomic research. Understanding the biological rationale of radiomics will be critical to justifying the development of prognostic and predictive imaging biomarkers, especially for targeted therapies.

Imaging-genomics for MRI derived volumetric features in glioblastoma

In the previous Chapter 2, we studied the combined relationships between advanced radiomic CT features in NSCLC, molecular pathways, and clinical parameters [14]. We aimed at extending this work to understand the biological rationale of volumetric image analysis in MRI. Therefore, in Chapter 3 and 4 we uncovered the connections between MRI derived volumetric GBM features, biological processes, molecular subtypes, and clinical characteristics of patients [15,16]. The volumetric tumor phenotype features for these studies quantified necrosis, contrast enhancing, and edema volumes in tumors, and were segmented from MR images with presurgical T1-weighted post-gadolinium contrast (T1w) and T2-weighted FLuid-Attenuated Inversion Recovery (FLAIR) sequences.

In Chapter 3 we investigated how these volumetric features are driven by biological processes based on global expression of genes, and how these associations can be interpreted in terms of molecular subtypes of GBM and overall survival of patients [15]. Generally, we wound that different volumetric features were related to distinct sets of biological processes. Interestingly, these were mostly inversely correlated. In particular,

necrosis, contrast-enhancement, and tumor bulk were highly enriched for well established cancer processes, including apoptosis, immunological responses, and signaling pathways. These processes are also known to impact patient survival [30–33], which may explain why volumetric features associated with these processes were also found to be prognostic in our analysis. Notably, these features were also significantly stronger prognostic factors than just the entire tumor volume. Considering that all volumetric features generally had only low to moderate correlation with each other, this highly suggests that there is indeed an added value to quantifying distinct volumetric features compared to quantifying just the tumor volume, as frequently proposed for prognostication [7].

In terms of volumetric associations with molecular subtypes of GBM, only edema was observed to be a volumetric predictor. Likely, this is because edema was the only feature that had a different size distribution across molecular subtypes. Recently, Gevaert et al. [34] presented similar indications by demonstrating that three edema features correlated with molecular subtypes. Interestingly, edema was the only volumetric feature that was not a significant prognostic factor. This finding is in line with previously published results that did not find significant prognostic value in edema as well, using an alternative, manual MRI scoring scheme referred to as VASARI feature set ((https://wiki.nci.nih.gov/display/CIP/VASARI). Our study advances those preliminary studies by analyzing tumors in 3D rather than in 2D, and by implementing automated quantification of those tumors.

We extended the volumetric-pathway results presented in Chapter 3 to volumetric-mutation associations in Chapter 4. In particular, in Chapter 4 [16] we Investigated how these volumetric GBM features are connected to somatic DNA mutations of genes that have consistently been implicated for GBM [35], including TP53, PTEN, NF1, EGFR, IDH1, RB1, and PIK3CA. Astonishingly, we found that volumetric features had very selective predictive power for specific mutations. For example, the contrast enhancing feature was predictive for TP53 and NF1 (AUC > 0.67, p < 0.04), but not for any other of the tested genes. Hereby, the mutated contrast enhancing volume was significantly smaller than the wild-type volume (p = 0.012). Similarly, the edema region was predictive only for RB1 (AUC = 0.66, p = 0.022) but no other tested gene, and was also significantly smaller in the mutated tumors (p = 0.015). This selectivity points to different growth patterns induced by different mutations that lead to different imaging phenotypes. An analysis of clinical covariates also revealed that the volumetric features were independent of gender, disease-free status, Karnofsky performance status, and age.

The smaller volumes for tumors carrying a TP53 mutation was previously suggested by a manual, labour intense imaging-genomic approach using manually derived MRI features referred to as VASARI features [36]. Furthermore, our study suggested that EGFR mutants have larger necrosis volumes, but this difference was statistically not significant, which is in line with a previously published study. Compared to the study in Chapter 2, the scope of the current analyses in Chapter 3 and 4 were more limited. First, only vol-

umetric measurements were considered rather than advanced first-order statistics, tumor shape, and textural properties of the image; furthermore, the number of of only eleven volumetric measurements was relatively low. Second, validation in independent datasets was not feasible and the analyzed cohorts were of comparably limited sample size, to some extent restricting generalizability. Third, it is reasonable to assume that features based on MRI are less robust to perturbation than imaging features based on CT, as Hounsfield units used in CT imaging are an absolute and well-defined measurement [37] . Fourth, due to cohort limitations, multivariable analysis could not be conducted.

In conclusion, the studies presented in Chapter 3 and 4 revealed a relationship between MRI derived volumetric features of GBM phenotypes and biological processes induced by global gene expression, molecular GBM subtypes, overall survival of patients, and somatic DNA mutations of well established oncodriver genes. We performed our analysis on publicly acquired imaging and genomic data from two well established consortia, The Cancer Imaging Archive (TCIA, https://www.tcia.org/) and The Cancer Genome Atlas (TCGA, https://cancergenome.nih.gov/). The main advantage of these multi-center datasets is its public availability and hence that full reproducibility can be ensured by independent groups. A limitation of this public dataset is the relatively small sample size of less than 100 patients with imaging for whom data on DNA mutations and gene expression were available, as well as the relatively low number of eleven investigated features. Furthermore, these features only quantified the volume of the segmented tumor areas, hence no properties such as tumor shape, texture, or intensity have been investigated. In addition, validation in external datasets could not be conducted due to unavailability of comparable datasets; this also made multivariate analyses unfeasible. To overcome these limitations, we designed and carried out an integrated analysis of advanced radiomic features in large, novel, and independent datasets containing several layers of data of the same tumors.

Radiomics for Targeted Therapies

Targeted therapies are developed for very specific patient populations that share particular properties of tumors that can be exploited [38]. Often, these properties are defined as biological events, such as the existence of a genetic mutation that renders the tumor relatively more sensitive to a chemical, biological, or ionizing compound compared with tumors that do not contain this mutation. As a result, clinical translation of targeted therapies inherently rely on the availability of accurate and robust biomarkers that indicate the existence of the biological property that should be targeted [39].

As the previous chapters demonstrated, radiomics has the potential to uncover underlying tumor biology [14–16]. Therefore, it is reasonable to believe that radiomics can be a viable tool in developing predictive and prognostic biomarkers to effectively identify patients who would benefit from a specific targeted treatment. As such imaging biomarkers do not require an invasive procedure as traditional biological biomarkers, these

biomarkers can be utilized over time to consistently monitor treatment response. To this end, in Chapters 5 and 6 [40,41]we aimed at defining radiomic response phenotypes in both brain and lung cancer treated with bevacizumab (Avastin™) and gefitinib (Iressa™), respectively.

Imaging biomarkers to stratify recurrent glioblastoma treated with bevacizumab

One of the few remaining treatment options for patients with glioblastoma who experience recurrence is administration of bevacizumab, an anti-VEGF inhibitor designed to antagonize angiogenesis [42,43]. Bevacizumab has been granted accelerated approval by the FDA for treatment of recurrent glioblastoma, following a seminal phase II clinical trial, the BRAIN trial (AVF3708g) [44]. However, recent studies, including large clinical phase III trials, suggested that bevacizumab does not improve OS of this patient population [45–47], but it is reasonable to hypothesize that this is due to absence of effective biomarkers that are capable of identifying those patients that do experience improved OS.

To this end, we developed prognostic radiomic imaging biomarkers to stratify survival of patients with recurrent glioblastoma at different timepoints [40], particularly before and after treatment initiation. We achieved this mainly by retrospectively analyzing prospectively collected data from the BRAIN trial. In particular, we investigated utility of a radiomic approach for this patient population, focusing on multiple clinical endpoints including overall survival (OS) and progression-free survival (PFS). Hereby, to limit false-discoveries we decoupled the selection of radiomic features and their analysis using two independent selection and evaluation cohorts, respectively. The results of this study have been published in the journal Neuro-Oncology [40] and are presented in Chapter 5.

Our study resulted in baseline (i.e., pretreatment) models that predicted disease progression within three, six, and nine months posttreatment initiation. Interestingly, the performances of these predictions increased with time, suggesting that particularly patients who experience progression later than others can be identified effectively. Hereby, multivariable models consistently outperformed individual radiomic features and were able to stratify patients in validation data that had 4 times higher progression risk (i.e., hazard-ratio HR > 4). A model optimized to predict OS as endpoint, resulted in significant stratification in validation data with a hazard-ratio of 2.5 and Log-Rank *p*-value of 1.1×10^{-3}. Furthermore, our analyses revealed that those models were independent of volumetric measurements, including total tumor volume and maximal tumor diameter, and were also independent of clinical prognostic covariates, including Karnofsky performance status, age, and gender.

Both our multivariable and univariable analyses suggested that radiomic features generally yield more prognostic value when derived from postcontrast enhanced T1-weighted images compared to features derived from FLAIR images. While it is difficult to interpret why this is the case, previous studies have documented similar conclusions [48–50]. These analyses also suggest that PFS can be predicted more accurately than OS, which

could be explained by the fact that PFS was a more clearly defined endpoint in the BRAIN trial as the trial protocol specified follow-up MR imaging every six weeks to identify progression of the disease and discontinue those patients from the trial. The univariable analysis of our models allowed us to gain understanding of which features had the highest prognostic value. Accordingly, those are features that quantify textural heterogeneity; for example, the textural-heterogeneity features information-correlation was significantly prognostic across all tested endpoints using baseline data and also had higher scores for patients with early progression.

These results presented in Chapter 5 could lead to the development of predictive imaging-based biomarkers for bevacizumab in patients with recurrent glioblastoma. To verify that the prognostic predictions are also predictive of treatment outcome, a treatment-negative arm would be necessary. Alternatively, failure of our models in treatment-negative cohorts could indicate predictive character of our proposed radiomic imaging biomarkers [51]. As patients were treated with either bevacizumab only or bevacizumab and irinotecan (i.e., chemotherapy), differences in those treatment arms could be suspected but have not been suggested by previous studies [52–54]. Hence, to increase statistical power we followed the approach of those studies and pooled all samples.

Robustness of our radiomic approach to imaging biomarkers in this study is ensured by the facts that we reduced the high-dimensional radiomic space by unsupervised feature selection blinded from clinical data, we utilized two independent cohorts to select and evaluate radiomic features, respectively, we developed our models from prospectively collected, multicenter trial data of high quality with standardized enrollment protocol, and we validated our proposed multivariable models in training-independent data. In this way, we aim at ensuring that the number of potential false-discoveries is drastically limited. Validation in multiple external cohorts following the same treatment combination would add statistical validity, but availability of these datasets is very limited as bevacizumab only has regulatory approval for recurrent glioblastoma in a few countries including the United States.

While our study focused on developing imaging biomarkers by utilizing radiomic approaches for standard MRI, combining MRI radiomics with other imaging modalities could be feasible. For example, recent studies have suggested that apparent diffusion coefficient histogram analysis has prognostic value in a subset of the BRAIN cohort [54,55]. Similarly, FDG-PET imaging could predict survival in high-grade gliomas treated with the same combination therapy as in the BRAIN trial [56]. Future studies need to determine whether these alternative imaging modalities, which are widely used in cancer care, have additive value and can complement MRI radiomic predictions to increase prognostic performances. If those studies cannot verify complementarity, applying to radiomics directly to T1-weighted MRI may ultimately benefit current clinical workflows as MRI is standard to care of patients with glioblastoma in modern cancer centers. Furthermore, integration with biological biomarkers could be considered; in particular, highly prognostic molecular

markers in glioblastoma such as methylation status of MGMT promoter [57] or other markers described in Chapter 4.

In conclusion, our study presented in Chapter 5 presented independently validated prognostic pretreatment radiomic imaging biomarkers to stratify patients with recurrent glioblastoma treated with bevacizumab. These results could translate into the development of predictive radiomic biomarkers to specifically predict response benefits of bevacizumab for this patient population.

A radiomic response phenotype for gefitinib treatment in lung cancer

To investigate whether radiomic approaches could identify patients who are sensitive to gefinitib using CT imaging pre (baseline) or post (follow-up) administration, we analyzed data from a clinical phase II trial in Chapter 6 [41]. Gefitinib is a specific inhibitor of EGFR signaling [58], implying that gefitinib will have high effectiveness for patients whose tumors contain a EGFR sensitizing mutation and no effectiveness otherwise. Therefore, it is imperative to correctly detect EGFR mutations for this patient population. Genetic tests are available to assess EGFR mutation status, but the vast majority of these tests rely on biopsy-based tissue and thus cannot be used to monitor mutation status over time [59]. To this end, we retrospectively applied radiomics to a prospectively collected phase II clinical trial cohort [60] of patients with and without EGFR mutation who all received treatment with gefitinib.

Our results demonstrate that radiomics can predict EGFR mutation status from pre- and post-treatment images in patients with NSCLC treated with gefitinib. Predictability of EGFR via radiomics has also been suggested by related research [61,62]. As EGFR is the explicit target of gefitinib, this leads to the conclusion that radiomics also predicts response to treatment with gefitinib. In particular, "laws texture energy", a feature that quantifies gray-scale heterogeneity, yielded the strongest performance (AUC = 0.67, $p = 0.03$) at pre-treatment imaging. It is important to note that none of the traditional volumetric features (i.e., tumor volume and maximal diameter) were predictive at baseline imaging before treatment has been initiated ($p > 0.27$). This illustrates that radiomics provides predictive capabilities in cases where standard clinical measurements fail to indicate the presence of a drug target. Moreover, our analysis revealed that advanced radiomic features and standard volumetric features correlate only low, reinforcing our earlier finding from Chapter 5 that radiomic features and standard volumetric features may provide complementary, clinically relevant information. It is reasonable to assume that individual radiomic features offer unique insight into tumor phenotypic characteristic and hence are complementary to each other as well, as their pairwise correlation was low too. To estimate the degree of these complementarities, however, large prospective, randomized, multi-center trials have to be conducted.

Although the availability of biomarkers before treatment has been initiated is required to augment treatment decisions, incorporating post-treatment imaging into predic-

tive models allows monitoring. Thus, we compared our baseline results to follow-up imaging, in particular by assessing the delta of radiomic features between pre- and post-scans. As expected, we found that volumetric measurements had strongest performance in predicting EGFR mutations as geftinib should have the highest effect on tumor size (i.e., a surrogate of response) in patients that harbor the sensitizing EGFR mutation. However, we also found that the radiomic features "gabor" energy had significant performance, as well. Here as well, these features had only low correlation with tumor volume and diameter, indicating independent information.

To rule out batch effects due to variability of imaging acquisition and reconstruction parameters inherent in clinical protocols, we conducted a technical validation. We assessed robustness of those radiomic features that significantly predicted EGFR mutations, by evaluating feature stability in another, independent test/retest dataset. All radiomic predictors could be validated with high stability (ICC > 0.87). The optimal parameters, however, have to defined by independent consortia, such as the QIN (Quantitative Imaging Network) and the QIBA (Quantitative Imaging Biomarker Alliance), that include multiple stakeholders. As the sample size of the cohort used to defined radiomic predictors was relatively low, further studies with more samples need to be conducted to confirm predictive capabilities for EGFR. A negative control arm of patients could additionally quantify how specific the predictive performance is to gefitinib, but ethical concerns may arise as conducting such as a study would require not treating patients that have a EGFR sensitizing mutation with gefitinib although its effectiveness is known.

In conclusion, the study presented in Chapter 6 shows that radiomics can define a gefitinib response phenotype that indicates the presence of an EGFR sensitizing mutation. Similar to the results presented in Chapter 5, these results of Chapter 6 prompt for further translational research towards the development of predictive CT imaging biomarkers to assess response to gefitinib in patients with lung cancer.

Machine Learning to Improve Prognostic Value of Radiomic Approaches

The previous chapters revealed the biological foundation of radiomics (Chapters 2-4) and how these phenotype-genotype interactions could translate into imaging biomarkers for targeted therapies (Chapters 5-6). However, to enable accurate and robust imaging biomarkers, advanced machine learning approaches are required. Therefore, we presented additional research in Chapters 7-9 in which we for the first time described and applied an extensive range of machine learning algorithms that are applicable to predicting overall survival of patients with non-small cell lung cancer (NSCLC) [63,64]. Furthermore, we used these algorithms to train both univariate and multivariate models to predict the pathological response to neoadjuvant chemotherapy [65] and to predict the future development of distant metastases in patients with NSCLC [12]. In particular, we discovered novel radiomic signature that improved clinical prognostic models, results some of which were also validated in independent data. We thereby demonstrated that radiomics has to be regarded

as a data science, following stringent statistical study designs, in order to fulfill its role as a clinical decision support system [66,67].

Machine learning methods for radiomics

Machine learning comprises a large panel of statistical algorithms that aim at fitting coefficients of classification models using training data in order to provide accurate predictions on novel data [68]. Machine learning is widely employed in biomedical sciences to develop predictive and prognostic biomarkers, especially in the field of gene expression signatures [69] as a majority of machine learning methods require quantitative data. This makes machine learning particularly applicable to radiomics, where all imaging features are quantitative. In radiomics, machine learning is mostly used to fit models to predict survival of patients, as survival is considered the gold standard among clinical endpoints for biomarkers [70]. However, most radiomic studies do not consider the choice of a particular machine learning method as part of their statistical design, although the final results are highly subjective to the choice of the training algorithm. Furthermore, radiomics operates in high-dimensional space and as such underlies the 'course of dimensionality' [71]. According to the course of dimensionality, the required amount of data to obtain statistically sound results grows exponentially with the number of features; however, data volume and quality are usually the limiting factor in clinical cohorts [72,73]. As a result of considering too many features without sufficient data, 'overfitting' can occur in which classification models are fit to spurious signals in the training data and thus do not generalize to novel data (i.e., the actual goal of predictive models) [74]. To mitigate this problem, feature selection methods can be used to identify an irredundant subset of features that best represents the information encoded in the entire data [75]. This also allows quicker training of models and easier interpretation of model outputs.

To support future radiomic studies employing machine learning methods in their analysis, we conducted a comprehensive and comparative study of a large panel of 12 and 14 of the most popular and broadly used classification and feature selection methods, respectively, and present the results in Chapter 7 [63]. These classification methods [76] included generalized linear model (GLM), support vector machine (SVM), and random-forest (RF). The feature selection [77] methods included the minimum redundancy maximum relevance algorithm (MRMR), and Chi-Square. mutual information maximization (MIM). We investigated the prognostic capabilities of the combinations of these classification and feature selection methods in terms of predicting two-year overall survival, also against data perturbation. Importantly, parameters of all classifiers were set to unbiased values proposed by a large and comprehensive study by Fernández-Delgado et al., who compared classification algorithms across 121 datasets [78]. The study was conducted on two large and independent cohorts totalling 464 patients with NSCLC. The goal of our comparative study was to propose more suitable methods for radiomic prognostication to justify a particular choice of algorithms.

A Wilcoxon feature selection generally yielded the highest power when combined most classifiers. However, one influential proposal from our study is the suitability of mRMR for effective radiomic feature selection. MRMR selects the highest performing features that are also complementary to each other at the same time [79]. Various radiomic studies have been published that employ mRMR [12,14,40,80,81]. Furthermore, with this study we propose random-forest classifiers as applicable choice for predictions of overall survival at two years. This is a reasonable finding, considering that random-forests have been widely used to achieve state-of-the art performances.

One apparent limitation is that we only investigated survival at two years. Although this binary endpoint is considered a gold standard in lung cancer research as in research of other cancer types [70], further investigation has to reveal whether the proposed methods are also feasible with ordinal or continuous endpoints. Furthermore, this study focused on overall survival of patients with NSCLC; hence, conclusions about utility of these methods for other cancer types is difficult to draw. However, in a follow-up study we achieved verification of these methods for patients with head & neck cancer [64].

Radiomics for pathological response

To directly measure response to neoadjuvant chemotherapy in NSCLC treated with trimodality (i.e., surgery, chemotherapy, and radiation), pathological response is often assessed after surgical resection [82]. Previous research has suggested that pathological response is a prognostic factor [83,84] and could be used as surrogate endpoint for survival [85]. Predicting pathological response prior to surgery would allow omitting surgery and potentially intensify chemoradiation for patients who would otherwise not benefit from added surgery.

In Chapter 8, we investigated whether radiomics can be used to predict pathological response for neoadjuvant chemotherapy in NSCLC by developing an unsupervised feature selection algorithm [65]. We applied this algorithm to a comprehensive radiomic feature set of 1,603 features to identify a canonical subset of features. In this way, the radiographic phenotype of tumors was still represented but on a significantly smaller feature dimension (n=15 in this case). Reduction of dimensionality is crucial to data mining of radiomic data, as spurious results could otherwise be derived [86]. This especially holds for cases in which independent validation is not part of the analysis. Unsupervised feature selection techniques particularly avoid data leakage, as they do not take the classification target (i.e., clinical endpoints) into account.

Among the 15 unsupervised selected features in the study presented in Chapter 8, seven were significantly prognostic and predicted gross residual disease (i.e., poor response); however, only one of those features also significantly predicted pathologic complete response (i.e., complete response). Interestingly, spherical disproportionality was one of the predictors and indicated that more spherical tumors are more likely to be associated with gross residual disease. It is important to note that conventional volumetric

features, in particular tumor volume and diameter, did not significantly predict pathological response in this study; these results confirm prior results [87,88] and once again indicate that advanced radiomic features can provide disease information beyond traditional imaging metrics. Multivariate analysis indicated that combining radiomic and volumetric features may outperform volumetric features alone.

Interestingly, however, radiomic features in this study did not significantly predict overall survival, local recurrence, or distant metastasis; an explanation for this could be the relatively data collection period of 12 years which could influence these treatment outcomes. Further limitations of this study are the lack of independent validation of the results and the relatively low number of patients who experienced complete pathological response (27 out of 127). In this study presented in Chapter 8, however, we for the first time applied unsupervised feature selection techniques on a vast number of features and thereby demonstrated the utmost importance of dimensionality reduction in radiomic studies utilizing machine learning methods.

Radiomics machine learning for distant metastases

In current radiomic literature, machine learning methods are widely employed to predict overall survival of patients. As in many cancers, however, a major reason for cancer-related death in NSCLC is the development of distant metastasis (DM). The incidence of DM is estimated to be as high as 40% [12]. despite combined treatment modalities, including surgical resection, chemotherapy, and radiation therapy. Identifying patients at highest risk of developing DM could help intensifying systemic therapy and therefore potentially improve outcome. To this end, we applied the machine learning methodologies from Chapter 7 to develop a generalizable radiomic-based signature, which is presented in Chapter 9.

In the current Chapter 9, we conducted univariate and multivariate analyses of 635 radiomic features in two cohorts of a total of 182 patients with NSCLC, more specifically locally advanced adenocarcinoma treated with chemo-radiation therapy [12]. As expected, we found strong correlations between DM and overall survival; the linear Pearson correlation coefficient R between the C-indices for DM and overall survival of all features was $R^2 = 0.92$ ($p = 2.7 \times 10^{-243}$). Univariately, we found that 15 features had strong prognostic power for DM (C-index > 0.60 and *FDR* < 0.05); in comparison, only 12 features met this criteria for overall survival.

In a multivariate analysis, we leveraged the proposed supervised mRMR algorithm from Chapter 7 to select a non-redundant yet complementary set of features. We used this canonical set of three features to fit a multivariate Cox proportional-hazards regression model to calculate a risk score of developing DM. We added significantly performing clinical covariates (i.e., overall stage and tumor grade) to this model, hereby demonstrating the added value of a radiomics model to a clinical model alone, validated on independent data (C-index 0.60 vs C-index 0.57; difference by bootstrapping $p < 1.56$ x

0^{-11}). Furthermore, this model significantly stratified patients into a high and low risk group (log-rank $p < 0.049$).

In this study, our proposed model for DM substantially benefitted from the sophisticated feature selection algorithm, mRMR, we employed. This step contributed in avoiding over-fitting when validating the predictions on the independent cohort. Our careful design allowed to investigate both univariate and multivariate performances of features and clinical covariates. An important finding in our study is also that the volumetric maximal three-dimensional slice was a stronger prognostic factor than the conventionally used two-dimensional maximal axial slice. At the time of publication, two other studies have been published that suggested prognostic value in CT derived textural features for DM. On the one hand, Fried et al. [89] suggested prognostic value for DM of a model with CT texture and clinical covariates; however, the relatively small sample size of this study (n = 54) makes it difficult to conclude general validity. Similarly, Ganeshan et al. [90] investigated radiomic features extracted from two-dimensional imaging slides (i.e., the largest cross sections); here, comparison with our study cannot directly be drawn as we considered the entire tumor volume. Future studies have to conclude whether three-dimensional analyses are superior to single-slice analyses.

As we also discussed for other chapters, the current study was limited by its retrospective nature. Patients included in the analysis were treated between 2001 and 2013, during which CT imaging protocols and other acquisition parameters may have changed. However, training and validation data were temporally separated; as our models validated against this temporal split, our results indicate generalizability beyond temporal variability. Furthermore, generalizability across different cancer types and institutes has to be investigated.

In conclusion, with this Chapter 9 we for the first time showed that machine-learning based radiomic approaches are viable tools to develop prognostic imaging biomarkers for the prediction of DM, which is a highly complex molecular process. Hence, with this work we laid the groundwork to investigating how prognostication via radiomics is justified biologically.

FUTURE PERSPECTIVES

This thesis aimed at closing translational gaps in understanding how radiomics, tumor biology, and clinical outcome of patients are linked to each other. We have shown that associations between radiomic features and molecular pathways can guide towards prognostic biomarkers for targeted therapies using advanced machine learning algorithms. Ultimately, this body of doctoral research can be used to develop predictive imaging biomarkers that would allow identification of patients who would benefit most (or worst) from a specific treatment. To this end, several prerequisites must be accomplished.

Obviously, regulatory authorities require any proposed biomarker to be validated through prospectively executed clinical trials. Within those trials, however, standardized protocols for image acquisition and processing can also be deduced. These protocols should be adhered by future radiomic studies to ensure reproducible interpretation of the results across different research laboratories. Standardization will also be crucial in the definition of radiomic features. Currently, it is unclear which parameter settings, including image transformation and interpolation, yield optimal performances and thus variance in those parameters is high among current radiomic literature [91]. Standardization of feature definitions will be of particular importance as automated feature definitions using deep learning techniques become more popular in medical image analysis [92]. Several organizations now aim at producing consensus guidelines for image standardization, including the Quantitative Imaging Network (QIN) [93] from the U.S.National Cancer Institute (NCI), the Quantitative Imaging Biomarker Alliance (QIBA) [94], the European Organisation for Research and Treatment of Cancer (EORTC) [95], the European Society of Radiology (ESR) [96], and the Cancer Research UK (CRUK) [97]. These guidelines should also include recommendations about tumor delineation and segmentation software; to date, it is an open question how accurate tumor delineations have to be drawn.

In addition to image standardization, other technical hurdles need to be overcome when planning to integrate a novel technology into existing clinical workflows. To integrate radiomic applications in future clinical cancer management, robust and secure infrastructure has to be developed, allowing storage of massive amounts of sensitive patient data, model development and validation, as well as transparent data and method sharing to foster collaborative research [98]. To facilitate collaborative research, novel concepts such as distributed and continuous learning could be valuable tools [66,99,100]. In particular, innovative and open source software, such as PyRadiomics for radiomic feature extraction (https://pyradiomics.readthedocs.io), the 3D Slicer suite for managing image data [101], and BraTumIA for automatic segmentation [102], can contribute in reproducing independent results in the growing radiomic research community. Besides technical advancements, it could be advisable to offer additional training for clinicians to become more accustomed to data-driven diagnostics [103].

This thesis proposed a number of radiomic and radiogenomic biomarkers. A novel concept for clinical trial validation could include blinded validation. Therefore, future studies would fix the parameters of the underlying models of such a biomarker and transfer this fixed signature to multiple independent groups for validation. This multi-center effort could also include additional biological validation or association studies to emerging diagnostic approaches such as liquid biopsies [104]. A meta-analysis of the individual datasets could provide further confidence in the results [105]. Furthermore, while this thesis mainly focused on radiomics derived from CT and MR imaging, further imaging modalities and analyses could be explored for applicability in biomarker development, including PET and proton imaging [106–109], as well as semantic and other qualitative approaches to image

analysis [110–112]. Similarly, radiomics can be applied to disease sites other than lung and brain [64,80,113–117] .

In conclusion, this dissertation advances our understanding of the biological basis of radiomics and the implications for novel concepts to cancer diagnostics. On the basis of the non-invasive imaging and machine learning, future studies could build on these results to develop and validate radiomic biomarkers for targeted therapies. Ultimately, this could routinely augment decision-making in daily clinical practice and thus lead to improved patient outcome.

REFERENCES

1. Ashley EA. Towards precision medicine. Nat Rev Genet. 2016;17: 507–522.
2. Shin SH, Bode AM, Dong Z. Precision medicine: the foundation of future cancer therapeutics. npj Precision Oncology. Nature Publishing Group; 2017;1: 12.
3. Sakka AP, Whiteside JR. Biomarker Discovery and Medical Diagnostic Imaging. Biomarker Validation. Wiley-VCH Verlag GmbH & Co. KGaA; 2015. pp. 59–73.
4. Hanash SM, Baik CS, Kallioniemi O. Emerging molecular biomarkers--blood-based strategies to detect and monitor cancer. Nat Rev Clin Oncol. 2011;8: 142–150.
5. Hulka BS. Biological markers in epidemiology [Internet]. 1990. Available: http://www.sidalc.net/cgi-bin/wxis.exe/?IsisScript=COLPOS.xis&method=post&formato=2&cantidad=1&expresion=mfn=001721
6. O'Connor JPB, Aboagye EO, Adams JE, Aerts HJWL, Barrington SF, Beer AJ, et al. Imaging biomarker roadmap for cancer studies. Nat Rev Clin Oncol. 2016; doi:10.1038/nrclinonc.2016.162
7. Aerts HJWL. The Potential of Radiomic-Based Phenotyping in Precision Medicine: A Review. JAMA Oncol. 2016; doi:10.1001/jamaoncol.2016.2631
8. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012;48: 441–446.
9. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology. 2016;278: 563–577.
10. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;5: 4006.
11. Grove O, Berglund AE, Schabath MB, Aerts HJWL, Dekker A, Wang H, et al. Quantitative Computed Tomographic Descriptors Associate Tumor Shape Complexity and Intratumor Heterogeneity with Prognosis in Lung Adenocarcinoma. PLoS One. 2015;10: e0118261.
12. Coroller TP, Grossmann P, Hou Y, Rios Velazquez E, Leijenaar RTH, Hermann G, et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. Radiother Oncol. 2015;114: 345–350.
13. Gevaert O, Xu J, Hoang CD, Leung AN, Xu Y, Quon A, et al. Non–Small Cell Lung Cancer: Identifying Prognostic Imaging Biomarkers by Leveraging Public Gene Expression Microarray Data—Methods and Preliminary Results. Radiology. 2012;264: 387–396.
14. Grossmann P, Stringfield O, El-Hachem N, Bui MM, Rios Velazquez E, Parmar C, et al. Defining the biological basis of radiomic phenotypes in lung cancer. Elife. 2017;6. doi:10.7554/eLife.23421
15. Grossmann P, Gutman DA, Dunn WD Jr, Holder CA, Aerts HJWL. Imaging-genomics reveals driving pathways of MRI derived volumetric tumor phenotype features in Glioblastoma. BMC Cancer. 2016;16: 611.

16. Gutman DA, Dunn WD Jr, Grossmann P, Cooper LAD, Holder CA, Ligon KL, et al. Somatic mutations associated with MRI-derived volumetric features in glioblastoma. Neuroradiology. 2015; doi:10.1007/s00234-015-1576-7

17. Feitelson MA, Arzumanyan A, Kulathinal RJ, Blain SW, Holcombe RF, Mahajna J, et al. Sustained proliferation in cancer: Mechanisms and novel therapeutic targets. Semin Cancer Biol. 2015;35 Suppl: S25–54.

18. Evan GI, Vousden KH. Proliferation, cell cycle and apoptosis in cancer. Nature. 2001;411: 342–348.

19. López-Sáez JF, de la Torre C, Pincheira J, Giménez-Martín G. Cell proliferation and cancer. Histol Histopathol. 1998;13: 1197–1214.

20. Strausberg RL. Tumor microenvironments, the immune system and cancer survival. Genome Biol. 2005;6: 211.

21. Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat Med. 2015;21: 938–945.

22. Ganeshan B, Abaleke S, Young RCD, Chatwin CR, Miles KA. Texture analysis of non-small cell lung cancer on unenhanced computed tomography: initial evidence for a relationship with tumour glucose metabolism and stage. Cancer Imaging. 2010;10: 137–143.

23. Eales KL, Hollinshead KER, Tennant DA. Hypoxia and metabolic adaptation of cancer cells. Oncogenesis. 2016;5: e190.

24. Ganeshan B, Goh V, Mandeville HC, Ng QS, Hoskin PJ, Miles KA. Non–Small Cell Lung Cancer: Histopathologic Correlates for Texture Parameters at CT. Radiology. 2013;266: 326–336.

25. Hou J, Aerts J, den Hamer B, van IJcken W, den Bakker M, Riegman P, et al. Gene Expression-Based Classification of Non-Small Cell Lung Carcinomas and Survival Prediction. PLoS One. 2010;5: e10312.

26. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM, et al. REporting recommendations for tumor MARKer prognostic studies (REMARK). Breast Cancer Res Treat. 2006;100: 229–235.

27. Ellingson BM, Bendszus M, Boxerman J, Barboriak D, Erickson BJ, Smits M, et al. Consensus recommendations for a standardized Brain Tumor Imaging Protocol in clinical trials. Neuro Oncol. 2015;17: 1188–1198.

28. Chen Z, Fillmore CM, Hammerman PS, Kim CF, Wong K-K. Non-small-cell lung cancers: a heterogeneous set of diseases. Nat Rev Cancer. 2014;14: 535–546.

29. Errico A. Lung cancer: Heterogeneity in space and time. Nat Rev Clin Oncol. 2014;11: 684.

30. Mak TW, Yeh W-C. Signaling for survival and apoptosis in the immune system. Arthritis Res. 2002;4 Suppl 3: S243–52.

31. Kalimuthu S, Se-Kwon K. Cell survival and apoptosis signaling as therapeutic target for cancer: marine bioactive compounds. Int J Mol Sci. 2013;14: 2334–2354.

32. Ichim G, Tait SWG. A fate worse than death: apoptosis as an oncogenic process. Nat Rev Cancer. 2016;16: 539–548.

33. Hsu DS, Kim MK, Balakumaran BS, Acharya CR, Anders CK, Clay T, et al. Immune signatures predict prognosis in localized cancer. Cancer Invest. 2010;28: 765–773.

34. Gevaert O, Mitchell LA, Achrol AS, Xu J, Echegaray S, Steinberg GK, et al. Glioblastoma multiforme: exploratory radiogenomic analysis by using quantitative image features. Radiology. 2014;273: 168–174.

35. Verhaak RGW, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell. 2010;17: 98–110.

36. Gutman DA, Cooper LAD, Hwang SN, Holder CA, Gao J, Aurora TD, et al. MR Imaging Predictors of Molecular Profile and Survival: Multi-institutional Study of the TCGA Glioblastoma Data Set. Radiology. 2013;267: 560–569.

37. Razi T, Niknami M, Alavi Ghazani F. Relationship between Hounsfield Unit in CT Scan and Gray Scale in CBCT. J Dent Res Dent Clin Dent Prospects. 2014;8: 107–110.

38. Baudino TA. Targeted Cancer Therapy: The Next Generation of Cancer Treatment. Curr Drug Discov Technol. 2015;12: 3–20.

39. Patel JN, Ersek JL, Kim ES. Lung cancer biomarkers, targeted therapies and clinical assays. Transl Lung Cancer Res. 2015;4: 503–514.

40. Grossmann P, Narayan V, Chang K, Rahman R, Abrey L, Reardon DA, et al. Quantitative Imaging Biomarkers for Risk Stratification of Patients with Recurrent Glioblastoma Treated with Bevacizumab. Neuro Oncol. 2017; doi:10.1093/neuonc/nox092

41. Aerts HJWL, Grossmann P, Tan Y, Oxnard GG, Rizvi N, Schwartz LH, et al. Defining a Radiomic Response Phenotype: A Pilot Study using targeted therapy in NSCLC. Sci Rep. Nature Publishing Group; 2016;6: 33860.

42. Narita Y. Bevacizumab for glioblastoma. Ther Clin Risk Manag. 2015;11: 1759–1765.

43. Chamberlain MC. Bevacizumab for the treatment of recurrent glioblastoma. Clin Med Insights Oncol. SAGE Publications; 2011;5: 117.

44. Friedman HS, Prados MD, Wen PY, Mikkelsen T, Schiff D, Abrey LE, et al. Bevacizumab alone and in combination with irinotecan in recurrent glioblastoma. J Clin Oncol. 2009;27: 4733–4740.

45. Gilbert MR, Dignam JJ, Armstrong TS, Wefel JS, Blumenthal DT, Vogelbaum MA, et al. A randomized trial of bevacizumab for newly diagnosed glioblastoma. N Engl J Med. 2014;370: 699–708.

46. Chinot OL, Wick W, Mason W, Henriksson R, Saran F, Nishikawa R, et al. Bevacizumab plus Radiotherapy–Temozolomide for Newly Diagnosed Glioblastoma. N Engl J Med. 2014;370: 709–722.

47. Fu P, He Y-S, Huang Q, Ding T, Cen Y-C, Zhao H-Y, et al. Bevacizumab treatment for newly diagnosed glioblastoma: Systematic review and meta-analysis of clinical trials. Mol Clin Oncol. 2016;4: 833–838.

48. Ellingson BM, Cloughesy TF, Lai A, Nghiemphu PL, Mischel PS, Pope WB. Quantitative volumetric analysis of conventional MRI response in recurrent glioblastoma treated with bevacizumab. Neuro Oncol. 2011;13: 401–409.

49. Grossman R, Shimony N, Shir D, Gonen T, Sitt R, Kimchi TJ, et al. Dynamics of FLAIR Volume Changes in Glioblastoma and Prediction of Survival. Ann Surg Oncol. 2017;24: 794–800.

50. Schmainda KM, Prah M, Connelly J, Rand SD, Hoffman RG, Mueller W, et al. Dynamic-susceptibility contrast agent MRI measures of relative cerebral blood volume predict response to bevacizumab in recurrent high-grade glioma. Neuro Oncol. 2014;16: 880–888.

51. Ballman KV. Biomarker: Predictive or Prognostic? J Clin Oncol. 2015;33: 3968–3971.

52. Ellingson BM, Kim HJ, Woodworth DC, Pope WB, Cloughesy JN, Harris RJ, et al. Recurrent glioblastoma treated with bevacizumab: contrast-enhanced T1-weighted subtraction maps improve tumor delineation and aid prediction of survival in a multicenter clinical trial. Radiology. 2014;271: 200–210.

53. Huang RY, Rahman R, Hamdan A, Kane C, Chen C, Norden AD, et al. Recurrent glioblastoma: volumetric assessment and stratification of patient survival with early posttreatment magnetic resonance imaging in patients treated with bevacizumab. Cancer. 2013;119: 3479–3488.

54. Pope WB, Qiao XJ, Kim HJ, Lai A, Nghiemphu P, Xue X, et al. Apparent diffusion coefficient histogram analysis stratifies progression-free and overall survival in patients with recurrent GBM treated with bevacizumab: a multi-center study. J Neurooncol. 2012;108: 491–498.

55. Rahman R, Hamdan A, Zweifler R, Jiang H, Norden AD, Reardon DA, et al. Histogram analysis of apparent diffusion coefficient within enhancing and nonenhancing tumor volumes in recurrent glioblastoma patients treated with bevacizumab. J Neurooncol. 2014;119: 149–158.

56. Colavolpe C, Chinot O, Metellus P, Mancini J, Barrie M, Bequet-Boucard C, et al. FDG-PET predicts survival in recurrent high-grade gliomas treated with bevacizumab and irinotecan. Neuro Oncol. 2012;14: 649–657.

57. Zhang K, Wang X-Q, Zhou B, Zhang L. The prognostic value of MGMT promoter methylation in Glioblastoma multiforme: a meta-analysis. Fam Cancer. 2013;12: 449–458.

58. Rosell R. What new therapeutic targets exist for EGFR-mutant NSCLC? Lancet Oncol. 2014;15: 1184–1185.

59. Ellison G, Zhu G, Moulis A, Dearden S, Speake G, McCormack R. EGFR mutation testing in lung cancer: a review of available methods and their use for analysis of tumour tissue and cytology samples. J Clin Pathol. 2013;66: 79–89.

60. Rizvi NA, Rusch V, Pao W, Chaft JE, Ladanyi M, Miller VA, et al. Molecular characteristics predict clinical outcomes: prospective trial correlating response to the EGFR tyrosine kinase inhibitor gefitinib with the presence of sensitizing mutations in the tyrosine binding domain of the EGFR gene. Clin Cancer Res. AACR; 2011;17: 3500–3506.

61. Rios Velazquez E, Parmar C, Liu Y, Coroller TP, Cruz G, Stringfield O, et al. Somatic Mutations Drive Distinct Imaging Phenotypes in Lung Cancer. Cancer Res. 2017;77: 3922–3930.

62. Liu Y, Kim J, Balagurunathan Y, Li Q, Garcia AL, Stringfield O, et al. Radiomic Features Are Associated With EGFR Mutation Status in Lung Adenocarcinomas. Clin Lung Cancer. 2016; doi:10.1016/j.cllc.2016.02.001

63. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL. Machine Learning methods for Quantitative Radiomic Biomarkers. Sci Rep. 2015;5: 13087.

64. Parmar C, Grossmann P, Rietveld D, Rietbergen MM, Lambin P, Aerts HJWL. Radiomic Machine Learning Classifiers for Prognostic Biomarkers of Head & Neck Cancer. Front Oncol. Frontiers; 2015;5. doi:10.3389/fonc.2015.00272

65. Coroller TP, Agrawal V, Narayan V, Hou Y, Grossmann P, Lee SW, et al. Radiomic phenotype features predict pathological response in non-small cell lung cancer. Radiother Oncol. 2016; doi:10.1016/j.radonc.2016.04.004

66. Lambin P, Zindler J, Vanneste BGL, De Voorde LV, Eekers D, Compter I, et al. Decision support systems for personalized and participative radiation oncology. Adv Drug Deliv Rev. 2016; doi:10.1016/j.addr.2016.01.006

67. Lambin P, van Stiphout RGPM, Starmans MHW, Rios-Velazquez E, Nalbantov G, Aerts HJWL, et al. Predicting outcomes in radiation oncology—multifactorial decision support systems. Nat Rev Clin Oncol. 2013;10: 27–40.

68. Van Dijck G, Van Hulle MM. Review of machine learning and signal processing techniques for automated electrode selection in high-density microelectrode arrays. Biomed Tech . 2014;59: 323–333.

69. Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. Nat Rev Genet. 2015;16: 321–332.

70. Driscoll JJ, Rixe O. Overall survival: still the gold standard: why overall survival remains the definitive end point in cancer clinical trials. Cancer J. 2009;15: 401–405.

71. P?kalska E, Duin RPW. The Dissimilarity Representation for Pattern Recognition: Foundations and Applications. World Scientific; 2005.

72. McFadden E. Management of Data in Clinical Trials. John Wiley & Sons; 2007.

73. Cai L, Zhu Y. The challenges of data quality and data quality assessment in the big data era. Data Science Journal. Ubiquity Press; 2015;14. Available: http://datascience.codata.org/articles/10.5334/dsj-2015-002/

74. Hawkins DM. The problem of overfitting. J Chem Inf Comput Sci. 2004;44: 1–12.

75. Hira ZM, Gillies DF. A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data. Adv Bioinformatics. 2015;2015: 198363.

76. Kotsiantis SB, Zaharakis ID, Pintelas PE. Machine Learning: A Review of Classification and Combining Techniques. Artif Intell Rev. Norwell, MA, USA: Kluwer Academic Publishers; 2006;26: 159–190.

77. Guyon I, Elisseeff A. An Introduction to Variable and Feature Selection. J Mach Learn Res. 2003;3: 1157–1182.

78. Fernández-Delgado M, Cernadas E, Barro S, Amorim D. Do we need hundreds of classifiers to solve real world classification problems. J Mach Learn Res. 2014;15: 3133–3181.

79. Radovic M, Ghalwash M, Filipovic N, Obradovic Z. Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. BMC Bioinformatics. 2017;18: 9.

80. Wei J, Tian J. Radiomics method to predict loss of heterozygosity on chromosome 1p/19q of oligodendroglial tumor. J Clin Orthod. American Society of Clinical Oncology; 2017;35: 2043–2043.

81. Tiwari P, Prasanna P, Wolansky L, Pinho M, Cohen M, Nayate AP, et al. Computer-Extracted Texture Features to Distinguish Cerebral Radionecrosis from Recurrent Brain Tumors on Multiparametric MRI: A Feasibility Study. AJNR Am J Neuroradiol. 2016;37: 2231–2236.

82. Pisters KM, Kris MG, Gralla RJ, Zaman MB, Heelan RT, Martini N. Pathologic complete response in advanced non-small-cell lung cancer following preoperative chemotherapy: implications for the design of future non-small-cell lung cancer combined modality trials. J Clin Oncol. 1993;11: 1757–1762.

83. Mouillet G, Monnet E, Milleron B, Puyraveau M, Quoix E, David P, et al. Pathologic complete response to preoperative chemotherapy predicts cure in early-stage non-small-cell lung cancer: combined analysis of two IFCT randomized trials. J Thorac Oncol. 2012;7: 841–849.

84. Isobe K, Hata Y, Sakaguchi S, Sato F, Takahashi S, Sato K, et al. Pathological response and prognosis of stage III non-small cell lung cancer patients treated with induction chemoradiation. Asia Pac J Clin Oncol. Wiley Online Library; 2012;8: 260–266.

85. Hellmann MD, Chaft JE, William WN Jr, Rusch V, Pisters KMW, Kalhor N, et al. Pathological response after neoadjuvant chemotherapy in resectable non-small-cell lung cancers: proposal for the use of major pathological response as a surrogate endpoint. Lancet Oncol. 2014;15: e42–50.

86. Han J, Pei J, Kamber M. Data Mining: Concepts and Techniques. Elsevier; 2011.

87. Cerfolio RJ, Bryant AS, Winokur TS, Ohja B, Bartolucci AA. Repeat FDG-PET after neoadjuvant therapy is a predictor of pathologic response in patients with non-small cell lung cancer. Ann Thorac Surg. 2004;78: 1903–9; discussion 1909.

88. Poettgen C, Theegarten D, Eberhardt W, Levegruen S, Gauler T, Krbek T, et al. Correlation of PET/CT findings and histopathology after neoadjuvant therapy in non-small cell lung cancer. Oncology. 2007;73: 316–323.

89. Fried DV, Tucker SL, Zhou S, Liao Z, Mawlawi O, Ibbott G, et al. Prognostic Value and Reproducibility of Pretreatment CT Texture Features in Stage III Non-Small Cell Lung Cancer. International Journal of Radiation Oncology*Biology*Physics. 2014;90: 834–842.

90. Ganeshan B, Panayiotou E, Burnand K, Dizdarevic S, Miles K. Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: a potential marker of survival. Eur Radiol. 2012;22: 796–802.

91. Lu L, Ehmke RC, Schwartz LH, Zhao B. Assessing Agreement between Radiomic Features Computed for Multiple CT Imaging Settings. PLoS One. 2016;11: e0166550.

92. Shen D, Wu G, Suk H-I. Deep Learning in Medical Image Analysis. Annu Rev Biomed Eng. Annual Reviews; 2017;19: 221–248.

93. Clarke LP, Nordstrom RJ, Zhang H, Tandon P, Zhang Y, Redmond G, et al. The Quantitative Imaging Network: NCI's Historical Perspective and Planned Goals. Transl Oncol. 2014;7: 1–4.

94. Sullivan DC, Obuchowski NA, Kessler LG, Raunig DL, Gatsonis C, Huang EP, et al. Metrology Standards for Quantitative Imaging Biomarkers. Radiology. 2015;277: 813–825.

95. Hunter AJ. The Innovative Medicines Initiative: a pre-competitive initiative to enhance the biomedical science base of Europe to expedite the development of new medicines for patients. Drug Discov Today. 2008;13: 371–373.

96. European Society of Radiology (ESR). ESR statement on the stepwise development of imaging biomarkers. Insights Imaging. 2013;4: 147–152.

97. Leach MO, Brindle KM, Evelhoch JL, Griffiths JR, Horsman MR, Jackson A, et al. The assessment of antiangiogenic and antivascular therapies in early-stage clinical trials using magnetic resonance imaging: issues and recommendations. Br J Cancer. 2005;92: 1599–1610.

98. Deist TM, Jochems A, van Soest J, Nalbantov G, Oberije C, Walsh S, et al. Infrastructure and distributed learning methodology for privacy-preserving multi-centric rapid learning health care: euroCAT. Clinical and Translational Radiation Oncology. 2017;4: 24–31.

99. Jochems A, Deist TM, van Soest J, Eble M, Bulens P, Coucke P, et al. Distributed learning: Developing a predictive model based on data from multiple hospitals without data leaving the hospital - A real life proof of concept. Radiother Oncol. 2016;121: 459–467.

100. Eastman P. IOM Report Urges Transformation of US Health Care into Continuous Learning System. Oncology Times. 2012;34: 16–18.

101. Yip SSF, Parmar C, Blezek D, Estepar RSJ, Pieper S, Kim J, et al. Application of the 3D slicer chest imaging platform segmentation algorithm for large lung nodule delineation. PLoS One. 2017;12: e0178944.

102. Meier R, Knecht U, Loosli T, Bauer S, Slotboom J, Wiest R, et al. Clinical Evaluation of a Fully-automatic Segmentation Method for Longitudinal Brain Tumor Volumetry. Sci Rep. 2016;6: 23376.

103. Brazas MD, Lewitter F, Schneider MV, van Gelder CWG, Palagi PM. A quick guide to genomics and bioinformatics training for clinical and public audiences. PLoS Comput Biol. 2014;10: e1003510.

104. Wan JCM, Massie C, Garcia-Corbacho J, Mouliere F, Brenton JD, Caldas C, et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. Nat Rev Cancer. 2017;17: 223–238.

105. Kleijnen J, Antes G. Systematic Reviews in Health Care. Meta-analysis in Context.M Egger, G Davey Smith, Doug Altman (eds). London: BMJ Books, 2001, pp. 487, £50.00. ISBN: 072791488X. Int J Epidemiol. Oxford University Press; 2002;31: 697–697.

106. Yip SS, Kim J, Coroller T, Parmar C, Rios Velazquez E, Huynh E, et al. Associations between somatic mutations and metabolic imaging phenotypes in non-small cell lung cancer. J Nucl Med. 2016; doi:10.2967/jnumed.116.181826

107. Leijenaar RTH, Carvalho S, Velazquez ER, van Elmpt WJC, Parmar C, Hoekstra OS, et al. Stability of FDG-PET Radiomics features: An integrated analysis of test-retest and inter-observer variability. Acta Oncol. 2013;52: 1391–1397.

108. van Timmeren JE, Leijenaar RTH, van Elmpt W, Reymen B, Oberije C, Monshouwer R, et al. Survival prediction of non-small cell lung cancer patients using radiomics analyses of cone-beam CT images. Radiother Oncol. 2017;123: 363–369.

109. Parekh V, Jacobs MA. Radiomics: a new application from established techniques. Expert Review of Precision Medicine and Drug Development. 2016;1: 207–226.

110. Yip SSF, Liu Y, Parmar C, Li Q, Liu S, Qu F, et al. Associations between radiologist-defined semantic and automatically computed radiomic features in non-small cell lung cancer. Sci Rep. 2017;7: 3519.

111. Velazquez ER, Meier R, Dunn WD Jr, Alexander B, Wiest R, Bauer S, et al. Fully automatic GBM segmentation in the TCGA-GBM dataset: Prognosis and correlation with VASARI features. Sci Rep. Nature Publishing Group; 2015;5: 16822.

112. Flechsig P, Zechmann CM, Schreiweis J, Kratochwil C, Rath D, Schwartz LH, et al. Qualitative and quantitative image analysis of CT and MR imaging in patients with neuroendocrine liver metastases in comparison to 68Ga-DOTATOC PET. Eur J Radiol. 2015;84: 1593–1600.

113. Braman NM, Etesami M, Prasanna P, Dubchuk C, Gilmore H, Tiwari P, et al. Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. Breast Cancer Res. 2017;19: 57.

114. Nie K, Shi L, Chen Q, Hu X, Jabbour SK, Yue N, et al. Rectal Cancer: Assessment of Neoadjuvant Chemoradiation Outcome based on Radiomics of Multiparametric MRI. Clin Cancer Res. 2016;22: 5256–5264.

115. Stoyanova R, Takhar M, Tschudi Y, Ford JC, Solórzano G, Erho N, et al. Prostate cancer radiomics and the promise of radiogenomics. Transl Cancer Res. 2016;5: 432–447.

116. Permuth JB, Choi J, Balarunathan Y, Kim J, Chen D-T, Chen L, et al. Combining radiomic features with a miRNA classifier may improve prediction of malignant pathology for pancreatic intraductal papillary mucinous neoplasms. Oncotarget. 2016; doi:10.18632/oncotarget.11768

117. Leijenaar RTH, Carvalho S, Hoebers FJP, Aerts HJWL, van Elmpt WJC, Huang SH, et al. External validation of a prognostic CT-based radiomic signature in oropharyngeal squamous cell carcinoma. Acta Oncol. 2015;54: 1423–1429.

Summary

SUMMARY

Radiomics is a promising field that aims at leveraging existing medical imaging data with advanced computational methods to provide novel quantitative and actionable data for clinical decision support. Notably, radiomic approaches enable integration of these previously untapped sources of information in a noninvasive, objective, reproducible, and cost-effective way. Numerous studies have suggested prognostic and predictive associations of radiomics and clinical endpoints, including overall survival. However, the molecular biology underlying these associations were largely unknown. To this end, this dissertation contributed in unraveling the connections between radiomic, molecular, and clinical data. Furthermore, these results were put in context to radiomic predictions for targeted therapies, as well as in context of optimizing these predictions with advanced machine learning methods.

This thesis was organized in five parts, where Part 1 introduced and outlined the objectives, Parts 2 to 4 contained the main research results, and Part 5 discussed these results. We briefly summarize the content of the thesis below.

Part 1: Introduction

In **Chapter 1** we introduced the concept of radiomics and explained the implications of radiomics for precision medicine in cancer. In this chapter, we outlined the objectives of this thesis to uncover the underlying biology of radiomics and therefore reason how radiomics can be used as imaging biomarkers for targeted therapies. We motivated these objectives by the fact that radiomics has been suggested to contain prognostic and predictive value, but that biological reasoning of this has not been achieved yet.

Part 2: Radiomics and its Underlying Biology

Chapter 2 presented a broad radiogenomic study that revealed mechanistic connections of a large set of radiomic CT features, molecular pathways, and clinical factors in lung cancer. Importantly, all associations were validated in independent data. Additionally, key associations were also validated biologically. Furthermore, we for the first time demonstrated that radiomics provides complementary prognostic value to traditional genetic and clinical predictors.

In **Chapter 3** and **Chapter 4** we extended the results from **Chapter 2** to associations between volumetric phenotype features, molecular pathways, and somatic gene mutations in brain cancer. In this way, we complement **Chapter 2** by sheding light on how imaging-genomic connections behave in another aggressive cancer type assessed by an alternative imaging modality, namely glioblastoma and MRI, respectively.

Part 3: Radiomics for Targeted Therapies

Understanding genotype-phenotype interactions in tumors as aimed for in Chapters 2-4 allows reasoning about radiomic phenotype predictors for targeted therapies that exploit genetic properties of tumors. Therefore, in **Chapters 5 and 6** we aimed at developing novel imaging biomarkers to identify patients who would respond best to bevacizumab and gefitinib in brain and lung cancer, respectively. In both studies, we leveraged prospectively acquired data of previously published clinical phase II trials.

Particularly, in **Chapter 5** we achieved to develop (and validate) radiomic biomarkers for patients with recurrent glioblastoma that stratified overall survival, as well as progression-free survival, at multiple timepoints including pre-treatment and follow-up at six and twelve weeks post-treatment initiation of bevacizumab. This study strongly suggests that radiomics can be used to develop predictive biomarkers for stratifying this patient population prior to treatment decision, which could be crucial as recent phase III clinical studies failed to suggest improvement of overall survival due to absence of effective biomarkers.

In **Chapter 6** we explored radiomic predictors for gefitinib, an EGFR inhibitor, by assessing predictive value of radiomics for EGFR mutations. Similar to Chapter 5, we investigated imaging data pre- and post-treatment, as well as the delta change of features values between those timepoints. While we observed predictive value of radiomic features at baseline (i.e., before treatment), traditional volumetric features such as tumor volume and diameter showed no predictive value at baseline.

Part 4: Prognostic Value of Radiomic Machine Learning

As we have observed from the previous chapters, radiomics is moving towards a data science, critically dependent on efficient machine learning methods. Therefore, in our final research **Chapters 7-9** we investigated applicability of a wide range of machine learning algorithms to optimize radiomic predictions.

In **Chapter 7** we present a study that evaluated a host of potential machine learning algorithms for radiomic prognostication. Specifically, we investigated 14 feature selection and 12 classification methods, and compared their performance in predicting overall survival at two years. With this study, we aimed at providing recommendations as to which popular machine learning methods are the most promising ones for the development of radiomic prognostic biomarkers.

Finally, wih **Chapter 8-9** we used unsupervised feature selection to identify a canonical set of radiomic features that could potentially predict pathological response in lung cancer following neoadjuvant chemoradiation. In **Chapter 8** we used a similar methodology to

derive features that predicted the development of distant metastasis, while optimizing the predictive model with supervised feature selection.

Part 5: Discussion and Future Perspectives

This dissertation concluded with a discussion and outlook in **Chapter 10**. Our research uncovered a large proportion of the molecular biology underlying radiomics, results that are necessary to justify translation of radiomics into clinical applications. We now gained better understanding about the high prognostic potential of combining radiomics with genomics and clinical records, as well as the implications for targeted therapies. Furthermore, we propose advance machine learning algorithms to increase prognostic and predictive performance of radiomic models. In conclusion, our work demonstrates the significance of systematically analyzing routine medical images with radiomics, as well as its integration with molecular and clinical data, to develop predictive tools for cancer diagnostics.

Valorization addendum

VALORIZATION ADDENDUM

The central objective of this dissertation was to enable biological reasoning of radiomic prognostic models to facilitate the development of imaging-based biomarkers for clinical decision support. Such biomarkers will be crucial in the success of targeted therapies. As this thesis has shown, radiomics has unprecedented potential to provide actionable data for noninvasive and cost-effective biomarker for specific cancer treatments. Furthermore, this thesis has also demonstrated that advanced machine learning algorithms are necessary to provide the level accurateness and robustness required for clinical applications.

To translate research of this thesis into clinical tools, further validation of the proposed models must be performed. To this end, transparacy in the processed data and methods used is critical. Therefore, we made data and code publicly available to the wider scientific community wherever possible. In particular, we released radiomic, genomic, and clinical data, as well as open sourced our analysis code from Chapter 2 online (https://elifesciences.org/articles/23421), which is the core of our results that associate radiomic with biological data. In addition, results from Chapter 2 were extended in Chapters 3 and 4, and data underlying Chapter 3 and 4 were collected through The Cancer Genome Atlas (TCGA, https://cancergenome.nih.gov/) and The Cancer Imaging Archive (TCIA, http://www.cancerimagingarchive.net/), which are public repositories. Hence, reproducibility of those results can be assured, as well. Obviously, releasing data is not possible for all studies due to intellectual property; for instance, this was the case for data underlying Chapters 5 and 6, which analyzed data from clinical trials. Similar holds for the remaining Chapters 7-9. Hence, more efforts should be performed to achieve global data sharing.

One solution to facilitate data sharing could be distributed learning, an emerging infrastructural concept in which machine learning models are distributed to different sites (i.e., clients such as hospitals) to learn from data of each site. In this way, data never leaves a site and hence only learned model coefficients are shared across sites. Obiously, however, full transparency in underlying data can only be achieved with access to raw data. Regarding transparency in methods, three parts have to be considered: 1) Segmentation, 2) Feature extraction, and 3) Analysis. In this thesis, methods for segmenting tumors were already publicly available either commercially (e.g, Eclipse^{TM}) or open source (e.g., 3D Slicer). Methods for feature extraction were recently shared through an open-sourced Python package at pyradiomics.readthedocs.io. All algorithms used for the analyses in this thesis were implemented in 'R', an open source package for statistical analysis.

General employment of open source methods can significantly support reproducibility analysis even in clinical studies, because reasons for deviating results can more easily be deduced. This can also have an impact on standardization of radiomic methods, which also must be achieved to gain regulatory approval for clinical use. Currently, major international consortia, such as QIN, QIBA, EORTC, and ESR, are developing guidelines for standardization in image acquisition, registration, and analysis.

Results of this dissertation can also be leveraged to improve existing biomarker approaches, as our thesis also suggests complementary prognostic value of radiomics to volumetric, genomic, and clinical assessments. To achieve effective combinations of radiomics and traditional methods, one radiomic signature should be fixed first. This radiomic signature can then be combined with genomic or clinical signatures that have been extensively described in recent literature to understand how this combination can be reproducibly and robustly optimized for a predefined clinical endpoint. Ideally, training and validation of such novel combinations should be performed by independent groups, ensuring absence of data leakage and thus generating overoptimistic results. Should radiomic approaches continue to show complementary value, combined biomarkers will lead to more accurate and cost-effective diagnostic strategies, and hence safer treatments. Inherently, more accurate biomarkers can also save considerable amount of funds required to conduct a clinical trial, as the target population can be identified easier and thus less patients need to be recruited to gain statistical significance.

Major parts of this dissertation have gained broad attention for its innovative approaches. For example, parts of the study prestend in Chapter 2 were honored with the *Best in Physics* award at the Annual Meeting of the American Association of Physicists in Medicine (http://www.aapm.org/meetings/2015AM/PRAbs.asp?mid=99&aid=29354), and was also announced with an online press release about Hot Topics (http://www.aapm.org/meetings/2015AM/VirtualPressroom/NewsReleases.asp#HotTopics). In addition, the significance of this work has also been suggested by a recent press release: https://www.sciencedaily.com/releases/2017/08/170803135210.htm. Altogether this emphasizes the anticipated impact that radiomics will make in clinical decision support. Finally, it should be noted that radiomic approaches have translational potential beyond oncology; as radiomic processes standard of care imaging techniques, any disease that can routinely be assessed by medical imaging can also be a potential application of radiomics, including cardiovascular, neurodegenerative, and bone diseases.

Acknowledgments

Expression of Thanks

I would like to express my sincere gratitude to everyone who supported me during my dissertation, professionally and personally. My first thank you goes to my supervisor and co-promoter, Dr. Hugo Aerts; Hugo when I came to Boston for the first time in 2013 you inspired me with both your enthusiasm and eagerness for scientific discovery. Without you, I would have never learned about the "business" of academic research and about the importance of connecting to people through networking, which now is one of my favorite activities. My gratitude also goes to my promoter, Prof. Dr. Philippe Lambin; Philippe, even though I conducted my research far from the actual university you always kept track of the progress of my research and continuously encouraged me to go for my promotion. It was an honor to have one of the distinguished personalities who coined the term "Radiomics" supervise my dissertation. I wish both of you, Hugo and Philippe, many successful academic years and accomplishements.

Clearly, however, there is one single person who deserves by far the biggest share of my most genuine gratefulness, appreciation, and acknowledgments. Kathy, my Love, you are the most wonderful woman to me and I could not be any more thankful for your sincere patience and endurance in bearing with all the time we had to live a transatlantic relationship. We knew upfront it would not always be easy – but you made it the easiest possible, the liveliest possible, and the happiest possible. This PhD would mean nothing without you and you deserve half of the title; also mehr als Danke für alles! Du bist mein Traum-Traum und Ich freue mich sehr auf die kommenden Zeiten mit dir ♡.

A huge fraction of my thanks goes to my all the friends and colleagues who I became friendly with during my time in the US. I am particularly thankful to the Boston-German community; I cannot count how many of you I have become close friends with over the entire period of four and a half years, but be certain that without you my stay in the US would have not been half as fun, inspiring, and motivating. Especially, I would like to (in alphabetical order) say *thank you* to Andi, the countless Christians and Christina, Elena, Frederik, Friederike, Hendrik, Jens, Johanna, Kristin, Luzia, Martin, Michael, Nico, Robert, Steffi, Thomas, Timo, Uli. Furthermore, thanks to Celine, Fred, Lulu, Manka, Orley, Rose and Kate, and Tarique. And very special thanks to you, Matty!! Having you as a roomie was like living with a brother. Special thank you to my GAC fellows Greg, Fabian, Fabien, Tim, and Sonja; I am so proud having built this conference with you and am even more delighted to still being connected this well with you beyond our time abroad. You all were wonderful people to me and I would not give away any single moment we spent with each other.

Frankly, the latter paragraph was not quite a ready exercise for me to write as it is impossible to enumerate all the amazing people who accompanied me throughout my time in the US. So please do not feel offended if I did not manage to include you in printed name – you very well know that I never forget to include you in my cherished memories.

I could not write these Acknowledgments without mentioning the people at Dana-Farber's PGA and PGSAO. It was such an awesome experience being part of fostering

the Longwood and Boston graduate student and postdoc community! Particularly you, Jen Molina, I cannot imagine how my personal DFCI experience would have looked like without you! Thank you also to Sonal Jhaveri and Lorraine Barnes for your open guidance to an expat. And thanks to Tim and Nadia who I had so much fun with!

I would also like to take the chance and thank Dr. Paul Yu and his group for being such an inspiring research lab. Paul, your sincerity about science, patients, industry and politics has been an unparalled role model; one can easily identify your mentor qualities. I will never forget our discussions and your helpful guidance with my PhD track.

I do not want to forget to mention the endless fellow researchers and scientists from the radiomic community who I enjoyed so much collaborating with, discussing scientific discoveries, or just talking about God and the world and complaining about how tough a PhD is 😉. Specifically, I would like to say thank you to Chintan, Vivek, Emmanuel, and Thibaud – the first batch of awesome radiopoetric scientists!! Moreover, thank you to Ralph and Wouter for helping me finalize my thesis!

Any Acknowledgment of mine would not be complete without distinctly thanking those who have always been the closest to me – my family. Papa, Mommsy, and Dini, your continuous support has not gone unnoticed by me and be sure that I value every welcoming thought of yours. This also holds for my "extended" family – Beate, Siggi, and Julian; ever since Kathy and I found each other I consider having found you, as well. Thank you equally for all your support throughout the course of my time in the US and beyond.

Very special thanks to Yali who designed this outstanding thesis cover!!! Keep up with your talent and leverage it as much as you can.

Thank you to everyone who was part of my Boston-Harvard PhD adventure! Cheers~

Curriculum Vitae

CURRICULM VITAE

Patrick Großmann was born on January 6th, 1989, in Bad Säckingen, Baden-Württemberg, Germany. He grew up in proximity to Switzerland and France and attended school in Grenzach-Wyhlen, Germany. After he obtained his Abitur with a natural science consolidation at the Liste-Meitner-Gymnasium in 2008, Patrick moved to the city of Bielefeld, Nordrhein-Westfalen, Germany, to study Bioinformatics & Genome Research at the University of Bielefeld. During his Bachelor of Science program, he accomplished a research stay at ETH-Zurich in the group of Prof. Dr. Joerg Stelling funded by a PROMOS stipend and completed his stay in 2011 with a thesis entitled "Sequence analysis of yeast integral membrane protein properties with respect to a potential species-specific correlation of signale peptide and first transmembrane domain".

He continued with Bioinformatics to obtain a Master of Science degree from the University of Bielefeld and wrote his Master thesis entitled "Stochastic modeling of the impact of (super)drivers and the mutator phenotype on tumor progression" in 2013, as an official Visiting Student at ETH-Zurich in the group of Prof. Dr. Niko Beerenwinkel funded by ERASMUS. During his time at the University of Bielefeld, he worked as a graduate research assistant at the Computational Metagenomics group, where he developed methods to analyze novel metagenomic and single cell datasets. He complemented his Master training with a research stay at Dana-Farber Cancer Institute, Boston, MA, USA, to conduct a meta-analysis of all >700 published human breast cancer gene signatures with respect to prognostic value distributed among molecular subtypes.

Motivated by this third external research stay, Patrick decided to return to Boston in January of 2014 to conduct doctoral research in the Computational Imaging and Bioinformatics group headed by Associate Prof. Dr. Hugo J.W.L. Aerts at the Dana-Farber Cancer Institute, Brigham and Women's Hospital, and *Harvard Medical School*. His dissertation was supervised by Prof. Dr. Phlippe Lambin at Maastricht University, The Netherlands, in co-supervision with Dr. Aerts and focused on integrated statistical methods applied to medical imaging, genomics, and clinical records with the goal of developing predictive biomarkers that may allow selection of optimal strategies in treatment of cancer. Patrick is Otto-Bayer alumni and was featured as "pioneer of tomorrow" on the official fellowship site. His doctoral research about integrating radiomics and genomics was awarded "Best of Physics" at the AAPM Annual Meeting 2015 and received a press-release about current "Hot Topics". Patrick's volunteering activities in Boston included Head of Innovation at the largest student-led, transatlantic conference, the *German-American Conference at Harvard*. He is now a fellow at the Collège des Ingenieurs headquarted in Paris and consults The Linde Group in digital strategies using Predictive Maintenance in Munich.

List of publications

SCIENTIFIC PUBLICATIONS

1. **Grossmann P**, Stringfield O, El-Hachem N, Bui MM, Rios Velazquez E, Parmar C, et al. Defining the biological basis of radiomic phenotypes in lung cancer. Elife; 2017;6. Available from: http://dx.doi.org/10.7554/eLife.23421.

2. **Grossmann P**, Narayan V, Chang K, Rahman R, Abrey L, Reardon DA, et al. Quantitative Imaging Biomarkers for Risk Stratification of Patients with Recurrent Glioblastoma Treated with Bevacizumab. Neuro. Oncol.; 2017; Available from: http://dx.doi.org/10.1093/neuonc/nox092

3. Aerts HJWL *, **Grossmann P ***, Tan Y, Oxnard GG, Rizvi N, Schwartz LH, et al. Defining a Radiomic Response Phenotype: A Pilot Study using targeted therapy in NSCLC. Sci. Rep. Nature Publishing Group; 2016;6:33860.

4. **Grossmann P** *, Gutman DA *, Dunn WD Jr, Holder CA, Aerts HJWL. Imaging-genomics reveals driving pathways of MRI derived volumetric tumor phenotype features in Glioblastoma. BMC Cancer; 2016;16:611.

5. Fischer M, **Grossmann P**, Padi M, DeCaprio JA. Integration of TP53, DREAM, MMB-FOXM1 and RB-E2F target gene analyses identifies cell cycle gene regulatory networks. Nucleic Acids Res.; 2016;44:6070–86.

6. Holst F, Hoivik EA, Gibson WJ, Taylor-Weiner A, Schumacher SE, Asmann YW, et al. Recurrent hormone-binding domain truncated ESR1 amplifications in primary endometrial cancers suggest their implication in hormone independent growth. Sci. Rep.; 2016;6. Available from: http://dx.doi.org/10.1038/srep25521.

7. Coroller TP, Agrawal V, Narayan V, Hou Y, **Grossmann P**, Lee SW, et al. Radiomic phenotype features predict pathological response in non-small cell lung cancer. Radiother. Oncol.; 2016; Available from: http://dx.doi.org/10.1016/j.radonc.2016.04.004

8. Wu W, Parmar C, **Grossmann P**, Quackenbush J, Lambin P, Bussink J, et al. Exploratory Study to Identify Radiomics Classifiers for Lung Cancer Histology. Front. Oncol.; 2016;6. Available from: http://dx.doi.org/10.3389/fonc.2016.00071.

9. Smirnov P, Safikhani Z, El-Hachem N, Wang D, She A, Olsen C, et al. PharmacoGx: an R package for analysis of large pharmacogenomic datasets. Bioinformatics.; 2016;32:1244–6.

10. Parmar C *, **Grossmann P ***, Rietveld D, Rietbergen MM, Lambin P, Aerts HJWL. Radiomic Machine Learning Classifiers for Prognostic Biomarkers of Head & Neck Cancer. Front. Oncol. Frontiers; 2015;5. Available from: http://journal.frontiersin.org/article/10.3389/fonc.2015.00272/abstract.

11. Gutman DA, Dunn WD Jr, **Grossmann P**, Cooper LAD, Holder CA, Ligon KL, et al. Somatic mutations associated with MRI-derived volumetric features in glioblastoma. Neuroradiology; 2015; Available from: http://dx.doi.org/10.1007/s00234-015-1576-7

12. Parmar C *, **Grossmann P ***, Bussink J, Lambin P, Aerts HJWL. Machine Learning methods for Quantitative Radiomic Biomarkers. Sci. Rep.; 2015;5:13087.

13. El-Hachem N *, **Grossmann P ***, Blanchet-Cohen A, Bateman AR, Bouchard N, Archambault J, et al. Characterization of Conserved Toxicogenomic Responses in Chemically Exposed Hepatocytes across Species and Platforms. Environ. Health Perspect.; 2016;124:313–20.

14. Parmar C, Leijenaar RTH, **Grossmann P**, Velazquez ER, Bussink J, Rietveld D, et al. Radiomic feature clusters and Prognostic Signatures specific for Lung and Head & Neck cancer. Sci. Rep.; 2015;5. Available from: http://dx.doi.org/10.1038/srep11044.

15. Coroller TP *, **Grossmann P ***, Hou Y, Rios Velazquez E, Leijenaar RTH, Hermann G, et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. Radiother. Oncol.; 2015;114:345–50.

16. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, **Grossmann P**, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat. Commun.; 2014;5:4006.