Imperial College London

Investigation of Intra-Tumour Heterogeneity to Identify Texture Features to Characterise and Quantify Neoplastic Lesions on Imaging

WITH A FOCUS ON NON-SMALL LUNG CANCER AND OVARIAN MASSES

MUBARIK AHMED ARSHAD

10-20-2017

Department of Surgery and Cancer, Division of Cancer
Imperial College London, Hammersmith Campus, London, W12 0NN,
United Kingdom

A thesis submitted to Imperial College London for the degree of
Doctor of Philosophy

Declaration of Originality

This thesis is the result of my own work and investigations, except where otherwise stated. Other sources are acknowledged by explicit references.

Copyright Declaration

Acknowledgements

I would like to express my deepest gratitude to my supervisors, Professors Eric O. Aboagye and Andrea Rockall, for supporting me in the undertaking of a PhD at Imperial College London. It has indeed been an exceptional opportunity to deepen my knowledge, skills and understanding of the field, as well as a transformative journey. I take this opportunity to thank them for their support throughout, their availability, their enthusiasm and encouragement, as well as for giving me the necessary space to grow as an independent researcher.

I would also like to thank colleagues at the Comprehensive Cancer Imaging Centre (CCIC). Special thanks go to Dr Andrew Thornton for developing TexLAB 2.0 and with help with R, and Dr Ioannis Lavdas for his support on computing aspects; a massive thank you to Huang 'Jack' Liang for help with segmentation; Dr Suraiya Dubash and Kasia Kaslowski for their expertise. I would also like to acknowledge everyone else at the CCIC for interesting and stimulating discussions. It has been very pleasant to work within such a multi-disciplinary environment. From the Ovarian Cancer Action Research Centre, I would like to thank Haonan Lu for his help with the laboratory, clinical, analysis aspects of the ovarian cancer section, Dr Giacomo Avesani for his help with segmentation, and to Prof Hani Gabra, Dr Paula Cunnea and Dr Christina Fotopoulou for providing valuable insight and clinical specimens.

I would like to thank collaborators I have worked with throughout my Ph.D. Special thanks go to Dr Tara Barwick, Dr Henry Tam and Dr Kathryn Wallitt from Imperial College Healthcare, Dr Sue Chua from The Royal Marsden, Dr Gary Cook from Kings, Dr Andrew Scarsbrook and Dr Garry McDermott from Leeds University Hospital, Dr Richard O'Connor from Nottingham University Hospital and Dr Nicola Rodgers, A Masters student from Imperial for kindly providing me with PET images to develop my work.

Finally I would like to thank my family and friends. Their presence, love and encouragements have been key in helping me to negotiate the difficulties encountered on the journey, keeping balanced and not forgetting the wider perspectives on life. Very special thanks go to my parents, Muhammad and Abida for their infinite love and support throughout.

Abstract

The aim of this work was to further our knowledge of using imaging data to discover image derived biomarkers and other information about the imaged tumour. Using scans obtained from multiple centres to discover and validate the models has advanced earlier research and provided a platform for further larger centre prospective studies. This work consists of two major studies which are describe separately:

STUDY 1: NSCLC

Purpose The aim of this multi-center study was to discover and validate radiomics classifiers as image-derived biomarkers for risk stratification of non-small-cell lung cancer (NSCLC).

Patients and methods Pre-therapy PET scans from 358 Stage I–III NSCLC patients scheduled for radical radiotherapy/chemoradiotherapy acquired between October 2008 and December 2013 were included in this seven-institution study. Using a semiautomatic threshold method to segment the primary tumors, radiomics predictive classifiers were derived from a training set of 133 scans using TexLAB v2. Least absolute shrinkage and selection operator (LASSO) regression analysis allowed data dimension reduction and radiomics feature vector (FV) discovery. Multivariable analysis was performed to establish the relationship between FV, stage and overall survival (OS). Performance of the optimal FV was tested in an independent validation set of 204 patients, and a further independent set of 21 (TESTI) patients.

Results Of 358 patients, 249 died within the follow-up period [median 22 (range 0–85) months]. From each primary tumor, 665 three-dimensional radiomics features from each of seven gray levels were extracted. The most predictive feature vector discovered (FVX) was independent of known prognostic factors, such as stage and tumor volume, and of interest to multi-center studies, invariant to the type of PET/CT manufacturer. Using the median cut-off, FVX predicted a 14-month survival difference in the validation cohort (N = 204, p = 0.00465; HR = 1.61, 95% CI 1.16–2.24). In the TESTI cohort, a smaller cohort that presented with unusually poor survival of stage I cancers, FVX correctly indicated a lack of survival difference (N = 21, p = 0.501). In contrast to the radiomics classifier, clinically routine PET variables including SUV_{max}, SUV_{mean} and SUV_{peak} lacked any prognostic information.

Conclusion PET-based radiomics classifiers derived from routine pre-treatment imaging possess intrinsic prognostic information for risk stratification of NSCLC patients to radiotherapy/chemo-radiotherapy.

STUDY 2: Ovarian Cancer

Purpose The 5-year survival of epithelial ovarian cancer is approximately 35-40%, prompting the need to develop additional methods such as biomarkers for personalised treatment.

Patient and Methods 657 texture features were extracted from the CT scans of 364 untreated EOC patients. A 4-texture feature 'Radiomic Prognostic Vector (RPV)' was developed using machine learning methods on the training set.

Results The RPV was able to identify the 5% of patients with the worst prognosis, significantly improving established prognostic methods and was

further validated in two independent, multi-centre cohorts. In addition, the genetic, transcriptomic and proteomic analysis from two independent datasets demonstrated that stromal and DNA damage response pathways are activated in RPV-stratified tumours.

Conclusion RPV could be used to guide personalised therapy of EOC.

Overall, the two large datasets of different imaging modalities have increased our knowledge of texture analysis, improving the models currently available and provided us with more areas with which to implement these tools in the clinical setting.

Table of Contents

Contents

List of Figures

List of Tables

Glossary of Abbreviations

AJCC	American Joint Committee on Cancer (staging system)
AUC	Area under the curve
AUC-CSH	Area under the cumulative SUV-volume histogram curve
BSA	Body surface area
BW	Body weight
CE-CT	Contrast Enhanced Computed Tomography
CI	Confidence Interval
CR	Complete Response
CSH	Contrast SUV-volume histogram
CT	Computed Tomography
DNA	Deoxyribonucleic acid
EGFR	Epidermal growth factor receptor
EORTC	European Organisation for Research and Treatment of Cancer
ER	Oestrogen Receptor
FBP	Filtered-back projection
FIGO	Fédération Internationale de Gynécologie et d'Obstétrique
FD	Fractal Dimension
^{18F}-FDG	2-deoxy-2-(^{18F})-fluoro-D-glucose
FLAB	Fuzzy locally adaptive Bayesian (segmentation method)
FOS	First order Statistic
FWHM	Full-width half $_{max}$imum
GLCM	Grey Level Co-Occurrence Matrix
GLRLM	Grey Level Run Length Matrix
GLSZM	Grey Level Size Zone Matrix

Glut 1   Glucose transporter 1

Glut 3   Glucose transporter 3

GT   Ground Truth

ICC   Interclass correlation coefficient

LBM   Lean body mass

LOR   Line of Response

LPFS   Local progression-free Survival

MTV   Metabolically active tumour volume

MR   Magnetic Resonance

MRI   Magnetic Resonance Imaging

NGTDM   Neighbourhood Grey Tone Difference Matrix

NR   Non-responder (to therapy)

NSCLC   Non-small cell lung cancer

OS   Overall survival

OSEM   Ordered-subsets expectation-$_{max}$imum (reconstruction method)

PD   Progressive disease

PERCIST   PET response Criteria in Solid Tumours

PET   Positron Emission Tomography

PFS   Progression-free survival

PgR   Progesterone receptors

PK   Pharmacokinetics

PR   Partial response

PSF   Point spread function

PVE   Partial volume effect

R   Responder (to therapy)

RC   Recovery Coefficient

RNA   Ribonucleic acid

ROC   Receiver Operating Characteristic (Curve)

ROI   Region of Interest

RECIST   Response Evaluation Criteria in Solid Tumours

SAM   Standardised added metabolic activity

SD   Stable disease

SPECT   Single Photon Emission Computed Tomography

SUV   Standardised Uptake Value

SUV_{FD} $_{Mean}$ standardised uptake value (FD method)

Std   Standard deviation

STIR   Software for Tomographic Image Reconstruction

TAC   Time Activity Curve

TBR   Target to background ratio

TGV   Total glycolytic volume

TK1   Tyrosine kinase 1

TL   Tumour (longitudinal) length

TLA   Total lesion activity (or total lesion metabolic activity)

TLG   Total lesion glycolysis

TNM   Tumour Node Metastasis classification system

TOI   Total object intensity

TPV   Total proliferative volume

TV   Tumour volume

US   Ultrasound

Ve(s)   Volume element (s)

VOI   Volume of Interest

VOL$_{FD}$ Lesion volume (FD method)

WHO   World Health Organisation

Thesis Contribution

My contribution to the work of this thesis includes the submission of the regulatory applications including the ethics for both sections (Ethics reference numbers: 14HH1908 and 15/WM/0237). This also included further regulatory applications to two hospitals as per their local rules.

In addition to this, I was involved at all stages from the initial proposal of the project to collection, obtaining results, analysis and presentation. Personally, I collected the data from the centres (this included downloading the scans from the individual hospitals in over half the cases for the lung cancer project and curating this data to ensure it was suitable). I also collected the clinical data including patient details, histology and survival for over half these patients. I was involved in segmenting the entire collection and performed the intra- and organised the interobserver analysis. I performed analysis of the clinical data and contributed to some code-writing and development of TexLAB v.2. I did significant amounts of testing of this software to ensure that it was robust. I also wrote some code to output heatmaps, test the PET scans from various centres and performed survival analysis. I produced the analysis of intra- and intra-observer correlation, performed the texture analysis and performed some of the analysis from the collected texture analysis and made all the graphs. The method to choose the texture features formed from the training set to test on the validation set was made partly by myself. The final analysis was performed by myself. I was a major contributor to the eventual write-up from which the paper was eventually accepted.

In the ovarian cancer project, I was involved in the initial outline of the project. All of the patients in our centre who had tissue were explored to see if they had pre-treatment CT scans. This was performed by myself, and I downloaded the CT scans from the hospital and curated this data. I performed the majority of the segmentations for the primary tumours and peritoneal metastasis and nearly all the semantic data was collected by myself in combination with a consultant radiologist. I performed the texture analysis on this data set. I also performed analysis on the CT scanner data and patient data including histology type, etc. The inter-observer correlation involved myself. I helped in the analysis when the proteomic, mRNA and texture analysis was collected. I was also involved in the interpretation of the data and write-up of the paper of the eventual paper that was accepted.

Preface

Cancer is a heterogeneous entity that will affect half of the UK population born after 1960 in some form (UK, 2016, Ahmad et al., 2015, CRUK, 2017). The most common types differ between the sexes, with the 5-year survival for some cancers varying from 98% in testicular cancer to 1% for pancreatic cancer (UK, 2016, CRUK, 2017).

Currently, imaging plays a clear role in the detection, visualisation of the spread of cancer and response to treatment. The workhorses of the imaging world include CT, PET and MRI, which are used in diagnosing and measuring treatment response. There is a common misconception that as the role of imaging is well established, there is limited potential for radiology to provide any additional information. However, if these images captured relevant, reliable and objective information, it may be possible in the future to stratify patients on an individual level, with the potential of discerning the underlying cancer biology and help improve treatment effectiveness. Furthermore, imaging may aid personalised therapeutic choice, treatment monitoring and cost-effectiveness.

There is a big drive to translate personalised or precision medicine into the clinical setting. These are 'individualised' treatments tailored specifically to certain tumour characteristics. For example, Erlotinib and Gefitinib specifically target EGFR mutations in the adenocarcinoma subtype of non-small cell lung cancer (Lynch et al., 2010, Pirker et al., 2009). However, it is important to note that adenocarcinomas represent approximately half of all NSCLCs and the

frequency of this mutation varies between geographical locations, and has been estimated to be between 12-47% (Midha et al., 2015). Other potential biomarkers have failed because of the variability of the predictive indicators (Patani et al., 2013, Sölétormos et al., 2016, Waterton and Pylkkanen, 2012).

Currently, we are on the cusp of a wave, where "big data" incorporating clinical, genetic, proteomic and imaging features have the potential to discover clinically useful biomarkers for promising therapies (Noor et al., 2015). The big advantage of imaging is its non-invasive nature, which allows repeated whole tumour imaging at baseline and at regular intervals (Gillies et al., 2016). In clinical practice, invasive biopsy and molecular profiling monitor tumours. However, due to the spatial and temporal limits of biopsy (Sved et al., 2004), the complete state of the tumour fails to be captured. Disadvantages of biopsies are plenitude and include the occasional necessity of repeats, the possibility of improper sampling, expense and the potential of complications. Despite this, biopsy remains the best method of tumour sampling currently available. The imaging phenotype may contain a wealth of information, including details of the underlying genotype (Segal et al., 2007), (Yang et al., 2003, Diehn et al., 2008). The tumour environment may also provide additional information that is currently not utilised, and this may have a therapeutic potential (Gillies et al., 2016). Tumours exhibit strong phenotypic differences such as necrotic cores that are apparent on imaging. Thus, imaging with its ease of access has a large potential for precision medicine (Giardino et al., 2017).

Following the initial use of texture analysis, or radiomics, by El Naqa in 2009 (El Naqa et al., 2009), there has been increasing interest in this area and with an increasing number of publications.

The motivations for this thesis were to add, and significantly advance, our knowledge of texture analysis, so that can we could accurately and consistently describe spatial and statistical complexity. The development of these tools will measure the extent of the heterogeneity, attempt to predict survival and examine the underlying reasons for this.

Outline of the Thesis

Section 1.2 of this chapter will provide an introduction to cancer and intra-tumour heterogeneity. The role of intra-tumour heterogeneity and the need for quantitative methods for whole tumour examination will be detailed. The imaging modalities of PET and CT will be introduced. The main applications and current methods of quantification in oncology will be discussed. Current clinical practices are also briefly outlined.

Chapter 2 is a literature review of PET and CT quantification beyond its current clinical use. There is a detailed overview of texture analysis, segmentation and texture feature selection methods used in this study. Chapter 3 presents the hypothesis, aims, objectives and scope of this thesis. Chapter 4 outlines the

methods firstly for the NSCLC section and then for the Ovarian cancer study. Chapter 5 presents the results for the NSCLC study. This is sub-divided as follows: Section 5.1 will focus on the selection of the scans from different centres. Section 5.2 will showcase the clinical differences and similarities between the centres participating in the study. Section 5.3 will outline the differences in the scanners between the centres. Section 5.4 will demonstrate the results of the PET features and texture analysis linked with survival. Section 5.5 will use the accumulated database used to test other published studies on texture analysis with the accumulated dataset acting as an independent validation set. Chapter 6 will discuss the results of the NSCLC findings and also the limitations in the context of the current published data. Chapter 7 represents the second major oncology section of this thesis. An outline is presented on the use of texture analysis on pre-operative CT scans with patients with ovarian masses with links to genomic, proteomic and clinical markers. There is also a section on using the pre-operative scans to assess for peritoneal spread and compare them with surgical findings. Chapter 8 discusses the findings, with important sections on limitations and prospective work. The thesis is concluded in chapter 9.

Thesis contributions and software developed

Thesis contributions

To my knowledge, the work presented in Texture Analysis in FDG-PET in non-small cell lung cancer is one of the largest multi-centre studies of its type. It is the first to incorporate wavelet transformations and fractal dimensions in PET

images. This study has one of the largest numbers of texture features to date. It is also amongst the first to use machine learning to select the texture features and link these with survival. It has used the large dataset to interrogate previous published work that was developed with much smaller datasets.

The ovarian cancer texture project is one of the first texture analysis study on CT for this particular cancer type which links the clinical outcomes with texture, semantic features and also with underlying RNA and protein sequences.

Software developed

TexLAB 1.0 was developed in the Cancer Imaging Centre by Dr Julien Willaime as part of his PhD, utilising a MATLAB toolbox and graphical user interface for texture analysis. TexLAB 2.0 was a significant upgrade of this software by Dr Andrew Thornton. Changes included the modification of existing texture features and the addition of further texture features to include fractal dimensions and wavelet transformations, which increased the 47 features of the previous software to 665 features. The latest software update permitted batch calculation, which significantly reduced the computation time from weeks to minutes. The ability to perform texture analysis on CT images was also added. The output was designed to facilitate statistical analyses.

1 Background

In this section, an explanation of the background of the research is given. An outline of the biological basis of cancer, the possible causes of heterogeneity, and a brief overview of the roles of oncology and clinical imaging are given.

1.1 Biological basis of cancer and heterogeneity

Tumours demonstrate heterogeneity, which reflect underlying cellularity, necrosis, proliferation, neo-vascularity and hypoxia (Gerlinger and Swanton, 2010, Marusyk et al., 2012). These phenotypes are postulated manifestations of a host of underlying genetic and local environmental effects (Gerlinger and Swanton, 2010, Gerlinger et al., 2012). The presence of heterogeneity is thought to confer drug resistance, treatment failure and metastatic spread (Fisher et al., 2013). This may explain why patients with the same histological subtype of tumour and stage respond differently to the same treatment and have different survival outcomes.

Current methods to investigate heterogeneity utilise invasive methods such as resection or biopsy, which have their own limitations regarding safety (Rodriguez and Terris, 1998, Abhishek and Khunger, 2015, Tøndel et al., 2012). A potential issue that could arise is whether the biopsy accurately represents heterogeneity. Furthermore, the nature of the tumour changes during the course of the treatment, which would require multiple biopsies to determine these changes. Currently, imaging identifies the sites of disease, stage and response to treatment by assessing morphological changes.

1.1.1 Hallmarks of cancer

In 2000, Hanahan and Weinberg (Hanahan and Weinberg, 2000) proposed that six biological traits characterised cancer: sustained cell proliferation, infinite replication potential, resistance to cell death, angiogenesis, and invasion and metastases. Malignant cells evade normal cell control mechanisms and proliferate autonomously of their environment. The exact mechanisms by which cancers escape normal controls are unique to each tumour type. In 2011, Hanahan and Weinberg (Hanahan and Weinberg, 2011) added genetic instability and inflammation as two further factors to their landmark paper.

1.1.2 Multi-scale extent of biological heterogeneity in cancer

Cancer is a multifaceted, dynamic and wide-ranging process, varying across spatial and temporal scales (Gillies et al., 2010). Tumour heterogeneity is displayed at both macroscopic and microscopic environments (Basu et al., 2011, Marusyk and Polyak, 2010). At the population level, some molecular subtypes have been recognised to be important in specific cancers. For example, in breast cancer, the hormone dependency for the oestrogen receptor alpha (ERα), progesterone receptor (PgR) and human epidermal growth factor receptor 2 (HER2) expressions have been acknowledged as key markers for stratifying cancer subtype at diagnosis (Patani et al., 2013). However, to complicate matters, there are reports that a low concordance exists between the expression of molecular biomarkers measured at primary and metastatic sites for a range of

cancers (Marusyk and Polyak, 2010). Due to this heterogeneity between distant lesions, differences in response to therapy have been demonstrated (Kenny et al., 2007).

1.1.3 Intra-tumour biological heterogeneity

Individual lesions in each patient are composed of a heterogeneous pool of cells (e.g. cancer cells, immune, inflammatory cells, stroma, vascular structures, etc. (Diaz-Cano, 2012)). Virtually all tumour types display intra-tumour heterogeneity. These manifest through phenotypic expressions such as diversity of cell morphology, the unequal expression of receptors and the inhomogeneity of the tumour microenvironment (Marusyk and Polyak, 2010). Different regions of the same tumour display different degrees of vasculature, hypoxia, metabolism, and proliferation (Hanahan and Weinberg, 2011). Gerlinger (Gerlinger et al., 2012) demonstrated the spatial nature of this heterogeneity, demonstrating that different regions of the same tumour (renal cell carcinoma) contained distinct genetic properties and/or genetic expressions associated with favourable or unfavourable prognosis.

The reasons for intra-tumour heterogeneity are likely multitude and are still poorly understood. Different theories have been postulated to explain the cause of this. The "clonal evolution" (Nowell, 1976) and the "stem-cell" theory concept (Visvader and Lindeman, 2008) have proposed different evolutions in terms of inheritable traits and local changes (Marusyk and Polyak, 2010, Campbell and Polyak, 2007). The "clonal evolution" concept postulates that tumour

progression follows localised Darwinian processes where small portions of all mutations affecting the cancerous cell population in a tumour region are selected and passed onto subsequent generations. The other view, termed the "cancer stem-cell" theory postulates that a select few cells within tumours acquire the ability to proliferate infinitely and thus drive tumour progression. This theory suggests that intra-tumour heterogeneity is a non-inherited characteristic due to differentiated daughter cells. More recently, theories involving tumour cell plasticity have been suggested (Gerlinger et al., 2012). This phenotypic plasticity concept suggests that cancer cells have different "stem-cell like" potentials that are influenced by their local microenvironment (Marusyk and Polyak, 2010, Park et al., 2000, Smalley et al., 2005).

1.2 Clinical oncology: strategies, therapies and clinical endpoints

1.2.1 Cancer management strategies

1.2.1.1 Cancer staging

Cancer staging consists of assessing the extent and development of the cancer in the body (Detterbeck et al., 2013). Clinicians use cancer staging in treatment planning to aid prognosis. It is also used as a framework to interpret treatment outcome, as well as providing multi-centre comparisons. Staging is unique for each tumour type. And different assessment systems are used. Two commonly used methods are the Tumour Node Metastasis (TNM) and American Joint Committee on Cancer (AJCC). The TNM is a standard method accepted across the

world (Union for International Cancer Control, 2013). The AJCC method is based on the TNM. Briefly, these two classification methods describe:

- The primary tumour (T) e.g. T0: not visible, Tis: carcinoma in situ, T1-4: size and/or extent of the tumour;

- Lymph node (N) involvement, e.g. N0: no nodal involvement, N1-N3: nodal involvement (number and size);

- Metastasis (M) e.g. M0: no metastasis and M1: metastasis.

For each grouping there is also an X to denote unknown.

The TNM is then given an overall score (from I to IV), which indicates the severity of the cancer, also known as the overall staging grouping:

- Stage 0: isolated tumour (carcinoma in situ)

- Stage I: cancer isolated to specific area;

- Stage II and III: locally advanced cancer with lymph node involvement (specific classification depends on cancer type).

- Stage IV: cancer that has spread to another organ (s)

The staging process is performed using physical examination, imaging (X-rays, CT, MRI, PET), laboratory tests (e.g. Blood, urine, other fluids), pathology (e.g. biopsy) and surgical reports. The staging process often occurs before surgery using clinical examinations, also known as clinical staging. After surgery, a pathological staging is sometimes likely. However, these two classifications are kept separate, as the two are performed under different conditions and times (e.g. when tissue from neo-adjuvant therapy is used before resection of the tumour and pathological staging).

The guidelines for lung cancer staging for this thesis were taken from the 7th edition of TNM in Lung Cancer of the International Association for the Study of Lung Cancer (IASLC) Staging Committee in 2009. There has been a recent update of the staging in the 8th edition (Goldstraw et al., 2016), but these are very recent and were published after the collection and analysis of the data.

Gynaecological cancers are assessed by the FIGO (Fédération Internationale de Gynécologie et d'Obstétrique) system. In general, there are five stages:

Stage 0: Carcinoma in situ

Stage I: confined to the organ of origin

Stage II: Invasion of surrounding organs or tissue

Stage III: spread to distant nodes or tissue within the pelvis

Stage IV: distant metastasis (es).

The most recent staging system is from 2014 (Prat and Oncology, 2015).

1.2.1.2 Tumour grading

Tumour grading is a classification of cells based on the appearance of the tissues under a microscope (pathological examination). It indicates the degree of differentiation of different cancer cells from their original cells. It is sometimes called the pathological grading of tumours. Grading varies across cancer types.

The guidelines issued by the AJCC to grade tumours are as follows (National Cancer Institute, 2013b).

- GX: cannot be assessed (indeterminate grade);

- G1: well differentiated (low grade);

- G2: modestly differentiated (intermediate grade);

- G3: poorly differentiated (high grade);

- G4: undifferentiated (high grade).

1.2.1.3 Very brief overview of cancer therapies

There are many different types of cancer treatment. These are broadly divided into radical for curative intent and palliative for symptomatic relief. The main active treatment options include surgery, chemotherapy, radiotherapy, hormone therapy and biological therapy (Research, 2016).

Treatment options depend on a host of factors, including the availability of treatment, the spread of the lesion, the grade of the lesion(s) and the patient's overall wellbeing and co-morbidities. These are discussed in a multi-disciplinary meeting in conjunction with doctors of different specialties such as surgeons, pathologists, oncologists and radiologists.

Treatment options differ through the course of therapy, with radiotherapy and chemotherapy sometimes given before (neo-adjuvant) or after (adjuvant) surgery.

1.2.1.4 The role of Imaging

Currently imaging plays a key role in the diagnosis and staging of tumours. It is also used extensively for the follow up of cancers during and after the course of treatment (O'Connor et al., 2016). For clinical trials, RESIST is used to assess tumour size or burden on images.

PET uses semi-quantitative methods of metabolic activity, such as the maximum (SUV_{max}), mean (SUV_{mean}) standardised uptake values amongst others including the Total Lesion Glycolysis (TLG) (Wahl et al., 2009). These have shown value as biomarkers, although there are conflicting data on the outcome (Vansteenkiste et al., 1999, Okereke et al., 2009, Machtay et al., 2013).

1.2.1.5 Clinical and pathological endpoints

Predictive and prognostic markers are used for clinical management and new drug selection (Wahl et al., 2009). One of the main variables taken into account is the change in size over the course of treatment. Methods that have implemented previously in the research environment include World Health Organisation (WHO) criteria in 1981 and two versions of the Response Evaluation Criteria in Solid Tumours (RECIST) criteria (Eisenhauer et al., 2009, Therasse et al., 2000).

In February 2000, an international collaboration between the European Organisation for Research and Treatment of Cancer (EORTC), National Cancer Institute and the National Cancer Trials group jointly published guidelines. The initial guidelines, version 1.0 (James et al., 1999, Therasse et al., 2000) were updated in 2009, version 1.1, (Eisenhauer et al., 2009). In brief, the guidelines include recommendations for measuring lesions, including the minimum size to be included. Initially, in version 1.0, a recommendation of a maximum of 10 'target lesions' per patient (and 5 per organ), was made. In version 1.1, this figure was revised down to a total of 5 target lesions per patient with a maximum of 2 per organ. Other lesions, outside the criteria, so-called non-target lesions, were visually assessed to indicate the overall response.

Very briefly, the change in the target lesions determines the response. There are four possible outcomes: complete response (CR), partial response (PR), progressive disease (PD) and stable disease (SD). A complete response denotes the disappearance of all lesions and pathological lymph nodes to less than 10mm in the short axis; partial response (PR) is the reduction in the sum of the diameters of the target lesions by more than 30% (compared with the nadir); progressive disease (PD) is the increase in the sum of the diameters of the target lesions by 20% (compared with the baseline) or the appearance of new lesions. Stable disease (SD) is used for any changes between 30% decrease and 20% increase. The RECIST criteria are purely based on anatomical measurements (Hayes et al., 2016). However, a reduction in tumour volume does not always occur and is not relevant in the use of new cytostatic therapies (Michaelis and

Ratain, 2006, Nishino et al., 2013). Moreover, there is the potential of impeding chances of survival in patients who are not responding to conventional treatment by delaying a change in therapeutic strategy whilst receiving treatment with potential toxic side effects (Kenny et al., 2007).

Assessing metabolic tumour response using ^{18}F-FDG PET has been suggested as an addition to RECIST, as this method has had some success as a potential biomarker of treatment outcome (Wahl et al., 2009). Interim PET studies after two cycles of treatment have been used clinically in the assessment of all types of lymphoma as part of the Deauville criteria since 2009 (Gallamini et al., 2014, Barrington et al., 2014). There are, however, some problems in other tumours and although the reduction of tumour uptake after the first few cycles of therapy has been shown to predict tumour response (Avril et al., 2009, Wahl et al., 2009), the lack of standardisation has resulted in poor uptake in the clinical context for functional assessment for other tumour types (Eisenhauer et al., 2009). No consensus has been established with regards to timing of post treatment scans or the threshold of reduction (Avril et al., 2009).

Further work in this area has elucidated that the uptake of tracer is dependent on the therapeutic agent (Avril et al., 2009) and choice of administration protocol (Schneider-Kolsky et al., 2010). There are also criticisms of this method to assess therapy response due to its non-specific nature (Brindle, 2008). Other imaging probes have been developed that are insensitive to inflammation and

target cell proliferation, such as ^{18}F FLT (Kenny et al., 2007) or hypoxia (Weber, 2006). A comprehensive review of non ^{18}-FDG probes that have been investigated in both pre-clinical and clinical trials is given here (Kenny et al., 2007, Alam et al., 2015).

1.3 Consequences of biological heterogeneity in cancer treatment and monitoring

1.3.1 Towards more specific targeted therapies and current challenges

Personalised medicine is a term used for specific treatments tailored to individuals based on the tumour's genetic profile (Stricker et al., 2011, Maughan, 2017). Many targeted therapies have been successful as second-line in treating tumours when systemic chemotherapy has failed, including in adenocarcinoma subtypes of non-small cell lung cancer (Lim et al., 2017), renal cell (He et al., 2017), hepatocellular (Galun et al., 2017), neuroendocrine (Cidon, 2017), melanoma(de Unamuno Bustos et al., 2017) amongst others (Longo, 2012).

However, even the use of genetic and phenotypic profiling to develop more specific tumour classification into subtypes (e.g. in oestrogen receptor positive or negative breast cancer) has not yielded adequate targeted therapy (Basu et al., 2011, Patel et al., 2011). Examples of therapies with limited efficacy include the use of epidermal growth factor receptor (EGFR1/HER1) inhibitors in non-small

cell lung cancer (NSCLC), head and neck and colorectal cancer with low response rates between 5 and 15% (Patel et al., 2011, Fukuoka et al., 2003). Similarly, erlotinib, gefitinib and more afatinib (Keating, 2014) are tyrosine kinase (TK1) inhibitors have only shown a 10% response rate in treating non-selected NSCLC patients, where the overexpression of the EGFR was not associated with therapeutic response (Patel et al., 2011). Sub-selecting patients exhibiting EGFR mutations yielded better results (Landi and Cappuzzo, 2011, Janku et al., 2010). For example, never smokers were more likely to respond to these two aforementioned drugs than smokers (Linardou et al., 2009). A more recent study demonstrated response rates of between 50-63% in afatinib in advanced stage non-small cell adenocarcinoma patients some of whom harboured non-classical EGFR receptor mutations (Shen et al., 2017). On the other hand K-ras oncogene mutations are more frequent in the elderly and heavy smokers (Subramanian and Govindan, 2008). New breast cancer subgroups were discovered and validated by Curtis (Curtis et al., 2012) based on extensive genomic analyses between clinical outcomes.

Two major factors that have challenged biomarker development and treatment response are genetic instability and intra-tumour heterogeneity (Stricker et al., 2011). Current diagnostic techniques such as single biopsies and microarray sampling may underestimate biological heterogeneity because the localised nature may not be representative of the entire tumour (Gillies et al., 2010, Gerlinger et al., 2012, Patel et al., 2011), whereby necrotic regions differ from hypoxic and highly vascular areas. Thus, there is the possibility that these

techniques may introduce potential bias and errors during diagnosis, classification and prognosis (Patel et al., 2011, Stricker et al., 2011).

Silva (Silva and Gatenby, 2010) found in a study that even the smallest detected lesions (around 1cm^3), containing around 10^9 cells, may harbour intra-tumour heterogeneity due to genetic stability. This study was modelled on two cancer populations of the lesions, one located in the inner core of the lesion and the second located on the outside. The former was characterised by a more hypoxic, less proliferative and an acidic environment. These characteristics were thought to confer resistance to chemotherapy. The second population was located on the outer rim of the lesion and was better vascularised. This had a more chemo-sensitive profile. Two different types of therapy were used: chemotherapy to target the outer rim and a glucose competitor to deprive cells of glucose in the centre. The study found that using the glucose competitor followed by chemotherapy was the most optimal treatment method. The mutations that developed during the course of the treatment were not explored.

1.3.2 Biological heterogeneity: a dynamic and adaptive landscape

Just characterising biological intra-tumour heterogeneity is not sufficient. One of the challenges beyond this is to identify which imaging characteristics or events drive cancer growth and which ones merely demonstrate the effects of treatment or other drivers (Yap et al., 2012). Increasingly, the literature has demonstrated

that genetic instability allied with intra-tumour heterogeneity favour mechanisms of drug resistance and the natural selection of fitter resistant clones in response to targeted therapy and/or cytotoxic therapies (Gerlinger and Swanton, 2010, Gatenby and Vincent, 2003). This ability to adapt to targeted therapy has been shown as early as the first cycle of therapy (Patel et al., 2011). Additionally, the distribution of cytotoxic drugs into the targeted region is hindered by poor and disorganised vascularity, slowly proliferating cells and acidic environments (Minchinton and Tannock, 2006). Thus, studies have demonstrated that the sensitivity of cancer cells to therapy is varied, due in part to intra-tumour heterogeneity (Gatenby and Vincent, 2003). This invariably makes accurate response assessment to new therapies difficult to evaluate. In addition, the underlying mechanisms responsible for progression and treatment failure or conversely treatment success are also incompletely elucidated (Yap et al., 2012, Moiseenko et al., 2017).

1.3.3 Capturing biological heterogeneity

New methods are required to assess and quantify intra-tumour heterogeneity at the global or whole tumour level. This will allow better understanding of the complexity of cancer and, thus, the reasons for response or resistance (Alizadeh et al., 2015). As alluded to earlier, current diagnostic techniques such as biopsies and microarray sampling may not be representative of the whole tumour in the case of heterogeneous lesions. An increasing number of studies have proposed a role for non-invasive imaging methods (Gillies et al., 2010, Basu et al., 2011, Patel et al., 2011, Phelps, 2000, Heinzmann et al., 2017), through the adoption of

newly-developed quantification methods (Tomasi et al., 2012, Asselin et al., 2012, Hectors et al., 2017, Perdigones and Murtaza, 2017).

1.3.4 Imaging cancer using Positron Emission Tomography

Imaging modalities permit the non-invasive detection, staging and monitoring of cancerous lesions. Anatomical and functional methods used in both clinical and academic settings include Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Ultrasound (US), Positron Emission Tomography (PET) and Single Photon Emission Computed Tomography (SPECT). PET is a very sensitive functional imaging method that permits the study of in vivo biological processes at cellular and subcellular levels. The technique incorporates the injection of a negligible amount of a radioactive compound (pico- to femtomoles/gram) into the bloodstream to study certain biological processes without perturbing it (Phelps, 2000). PET imaging has been used in diverse areas from preclinical animal models, in the clinic and within the pharmaceutical and biological industry for drug discovery (Cherry, 2006). Clinically, [18F]-FDG PET has been used in the detection, diagnosis, staging and follow-up of patients (Heinzmann et al., 2017).

1.3.4.1 Nuclear Physics and Positron Emission Tomography

1.3.4.1.1 Positron emission and radioactive decay detection

PET detects the indirect decay of positrons that has been injected into the blood (Cherry, 2006, Phelps, 2000). A positron is a particle with the same mass as an

electron but with the opposite charge, thus it is known as an antiparticle of an electron. This particle is emitted when proton-rich radiolabelled compounds undergo radioactive decay. They travel a few millimetres in tissue and when they encounter an electron, this encounter results in annihilation giving off 511keV in opposing directions. In the PET scanner, detectors are arranged in several rings that are equipped with an electronic circuit of coincidence. The recording of a decay event occurs when two gamma photons are detected simultaneously in a line of response by scintillators. The characteristics of the detectors and electronics determine the energy resolution and maximum count rate. The recorded radioactive events can be due to true position emission, "Compton" scatter events due to the deviation of the photons' trajectory, or random coincidences. Other factors coming into play include the photoelectric energy and the non-detection of an event. Fortunately, methods exist to counter the scatter and random coincidences (Cheng et al., 2011) in PET.

1.3.4.1.2 Image reconstruction and corrections

The acquisition of PET images can be performed in static or dynamic modes. The reconstructed algorithm used is usually filtered back projection (FBP) or iterative (e.g. ordered subset expectation maximisation (OSEM)) methods (Alessio and Kidnahan, 2006). There are many factors that need to be applied prior to image analysis. These include photon attenuation corrections using a density map from a CT scan (which is the common method nowadays) or a transmission PET scan. This corrects for scattered radiation and random coincidences (Paans and van Waarde, 2002). Additional corrections for partial

volume effects (Section 1.3.4.1.3) and motion (Section 1.3.4.1.4) can also be applied. Further considerations include time of flight (Prieto et al., 2015) and the different methods for attenuation correction encountered in PET-MRI (Muzic and DiFilippo, 2014, Shah and Huang, 2015).

1.3.4.1.3 Partial volume effects and corrections

Partial volume effects (PVEs) can influence the image appearance and can result in contaminations (or 'spill-over') of tracer activity between regions and slices, such as the target region, background, and other organs (Rousset et al., 2007, Soret et al., 2007, Kjell et al., 2012). There are two concurrent effects taking place at the target level: spill out and spill in effects. The former consists of loss of apparent activity in the target to background. The latter, spill in, is the contamination of the lesions activity by background activity (Rousset et al., 2007). For example, in the case of an tracer avid lesion on a background of very low activity, the PVEs are mainly "spill-out" effects which increase in the lesion size and subsequent lower apparent radiotracer uptake (Rousset et al., 2007). Also, in lesions with necrotic centres, PVEs may indicate that there is more viable tissue than actually present. PVEs contribute greatly to the characteristic fuzziness of final reconstructed images. This is due mainly to finite spatial resolution of PET images. The final images have low spatial resolution due to the uncertainty between the exact location of radioactive decay events relative to the corresponding detected annihilation, the size of the detector, the electronics and mode of acquisition (2D versus 3D) (Rousset et al., 2007). These effects are modelled by a point-spread function (PSF) attributed to the scanner.

Another cause of PVEs, is known as "tissue-fraction effect", which is introduced to digital images, independently of the system's resolution, when quantising intensities onto the grid of the matrix (Rousset et al., 2007). This discretisation method can generate a range of intensities between adjoining structures within individual voxels. This effect is generally ignored when correcting for PVE in PET and most algorithms focus primarily on the PSF of the system (Erlandsson et al., 2012). Soret et al (2007) explored these effects and demonstrated that this 'tissue-fraction effect' could result in a decrease in the standardized uptake volume when increasing the pixel width. In his paper, a 7% reduction in SUV was seen when the pixel width was increased from 4 to 6 mm.

PVEs are complicated to correct for because they are dependent on a host of other parameters that are not and cannot be controlled for such as tumour size, shape, target to background ratios, etc. (Soret et al., 2007). Unfortunately, the spill in and spill out effects do not cancel each other out and it is difficult to predict how much the pixels are affected. In addition, PVEs are modified by reconstruction methods and parameters (e.g. number of subsets and iterations) (Rousset et al., 2007). Further PVEs can be introduced in post imaging processing when using the Gaussian kernel, which reduces the noise level.

There have been many attempts to correct for PVEs, named partial volume correction (PVC), but none of these have been standardised (Erlandsson et al., 2012). Some of the strategies suggested include pre- and post-reconstruction methods on a regional or voxel level. In pre-reconstruction, multiple indices

called recovery coefficients (RC) are calculated experimentally taking into account the lesion size, shape and location to correct for radiotracer uptake distortion (Hoffman et al., 1979). These have been subsequently developed further to take into account the background uptake of the tracer (Kessler et al., 1984). Another method known as the Geometric Transfer Matrix (GTM) has been used to account for the contribution of several tissues. In our lab, we have developed the fractal method (Willaime et al., 2014).

At the voxel level, many approaches for PVE correction have been proposed. The most widely used of these is the deconvolution method that corrects for the PSF of the system. The advantage of this over other methods is the lack of need for additional information such as anatomy, etc., however, estimating the PSF accurately becomes the critical point. In reconstruction, these methods lead to the generation of artefacts and noise amplification that need to be corrected for. Although, the PSF can vary in space, in reality this PSF can be assumed to be invariant if the preliminary data is corrected for variations in the performance of the detectors (Erlandsson et al., 2012). Other approaches included wavelet-based ones that led to high-resolution details incorporated into low-resolution images yielding interesting results (Boussion et al., 2006). Other enhanced results have been seen when PVC methods have been incorporated into the PET reconstruction e.g. non-collinearity, positron range, intercrystal scattering, and penetration effects (Rousset et al., 2007).

1.3.4.1.4 Motion Correction

Further PVEs are introduced due to motion, patient movement, cardiac and respiratory effects (Erlandsson et al., 2012). There are motion correction methods using external devices to track breathing signal (Nehmeh and Erdi, 2008). Unfortunately, the signal to noise ratio (SNR) is very low as only a few images are used to reconstruct the images. Post-reconstruction algorithms, known as reconstruction transform-average methods (RTA) (Tsoumpas et al., 2011), have been proposed to make use of all statistics ((Visvikis et al., 2006). The way this methods works is to discretise the events into discrete gated frames (within which motion is minimal) and using an optical flow technique to register and re-align the images and average the transformed gates to recover motion-corrected images. Anatomical imaging has also been suggested to incorporate the motion of the heart, lungs and other organs.

1.3.4.2 Molecular imaging probes in PET

PET radiotracers are composed of compounds that have the biological component of interest (e.g. protein, nucleic acid, antibody, or drug) radiolabelled with a radioisotope element (commonly ^{15}O, ^{18}F, ^{11}C, ^{13}N). The subsequent images display the concentration of the radiotracer in kBq/mL in tissues (Phelps, 2000). In practice, the design and production of imaging probes require intensive research and sophisticated procedures.

^{18}F-FDG (^{18}F-fluorodeoxyglucose), an analogue of glucose, is the most widely used PET tracer. Since its discovery in the 1970s (Fowler and Ido, 2002), its use

has become widespread. In the field of Oncology, it works on the Warburg effect (Gatenby and Gillies, 2004), whereby tumours have very high rates of glycolysis. FDG follows the same initial pathway as glucose but it is trapped after the initial phosphorylation step. The amount of radiotracer retained by cells is proportional to the rate of glycolysis (Phelps, 2000). Its role in cancer imaging includes detection of primary tumours, nodes, and metastases, for staging, assessing response and detecting recurrence (Papathanassiou et al., 2009). There are some drawbacks, however, including lower sensitivity and specificity in tumours with low glucose utilisation. Normal high FDG activity is seen with the brain and kidneys, and also in non-oncological process including inflammation and infection (Ertay et al., 2017). Quantitative comparisons between patients are not possible, due to the variability between the uptake of the radiotracer and measurement settings (Brindle, 2008).

More specific PET biomarkers have been developed for imaging cancer biological functions including cell proliferation, angiogenesis, etc. For example ^{11}C-carbon-choline and ^{18}F-fluoro-choline (^{18}F-FCH) have been used to investigate membrane lipid synthesis, ^{18}F-fluoroestradiol (^{18}F-FES) for imaging oestrogen receptor expression, ^{18}F-fluoro-galacto-RGD and ^{18}F-fluciclatide as markers of metastases and angiogenesis, ^{18}F-fluoromisonidazole (^{18}F-FMISO) and ^{62}Cu-diacetylbis(N4-methylthiosemicarbazone) (^{62}Cu-ATSM) for monitoring hypoxia and ^{18}F-fluorothymidine (^{18}F-FLT) as a measure of cell proliferation (Papathanassiou et al., 2009) (Kenny et al., 2007) (Weber, 2006, Contractor and Aboagye, 2009)). These tracers have as yet only had limited clinical impact (Alam et al., 2015).

1.3.4.3 Applications of PET in oncology and quantification methods

1.3.4.3.1 SUV measurements

Methods to quantify radiotracer uptake in neoplastic lesions have been proposed with the goal of differentiating between tissue types (e.g. malignant/benign), deriving prognostic indices, monitoring treatment response and measuring early response to therapy. For example, semi-quantitative analyses have enhanced visual assessment and helped distinguish between benign and malignant pathologies, and reduced inter-observer variability (Avril et al., 2009). The most commonly used index is the standardised uptake value (SUV). It is a normalised measure of the concentration of radiotracer in tissue (kBq/mL) and is corrected for the injected dose of radiotracer (MBq) and usually the body weight (BW) of the patient, with body surface area (BSA), or lean body mass (LBM) sometimes incorporated (Thie, 2004). There has been some promise in using SUV for monitoring early therapy response prior to anatomical changes (Wahl et al., 2009, Kwee et al., 2010).

1.3.4.3.2 Limitations of current quantification methods

Unfortunately, SUV parameters have not been universally adopted due to several reasons. Firstly, there is an absence of a widely accepted cut-off between malignant and benign lesions, the SUV_{BW} is very susceptible to patient body fat (Zasadny and Wahl, 1993); (Sugawara et al., 1999, Boellaard, 2009), the timing of scanning post injection, patient plasma glucose levels, and post imaging reconstructions have a strong impact on the SUV values (Keyes, 1995). There

have been efforts to standardise protocols across centres to permit multi-centre comparisons (Lasnon et al., 2017, Skougaard et al., 2013), however, the SUV indices remain strongly dependent on physical factors such as the limited spatial resolution of the scanner (PVEs) and noise (Karakatsanis et al., 2015). Due to the influence of the neighbouring structures, the mean SUV can be affected. The SUV_{max}, the value of the single most intense voxel, has been suggested as a method to overcome this limitation but this is strongly sensitive to noise. SUV_{peak}, which is the average of SUV with a small volume of interest, has been suggested as an alternative method (Wahl et al, 2009). However, the volume of interest differs between centres and different definitions associated with treatment response have hindered adoption of this parameter (Vanderhoek et al., 2012). Furthermore, the complexity of the intra-tumour heterogeneity, with regions of differential uptake, is not captured by these SUV parameters.

With this is mind, there has been a suggestion that medical images may contain more information than can be perceived visually, giving birth to the field of "radiomics" (Lambin et al., 2012, Kumar et al., 2012). Although imaging provides a wealth of anatomical, and functional data, this could be expanded to include metabolic, protein and genomic data (Lambin et al., 2012, O'Connor et al., 2016). In an era of personalised medicine, additional features extracted from medical imaging could include 1) more accurate response variables, and 2) information linked to genetic and molecular biochemical pathology (Chicklore et al., 2013). Segal (Segal et al., 2007) demonstrated that 28 imaging features from CT scans, including heterogeneity scores, predicted 78% of gene expression in primary hepatocellular carcinoma. In addition, Diehn (Diehn et al., 2008) was able to link

MRI imaging with underlying genomic data in glioblastoma multiforme. Further to this, Aerts et al. (Aerts et al., 2014b) were able to extract 440 radiomic features from 1019 CT scans of non-small cell lung and head and neck cancer to develop a prognostic model which was linked to underlying genetic patterns. The potential for this method to capture imaging descriptors with tumour variability may be critical in predicting tumour response, assisting clinical management in addition to the methods currently used. In addition, these tools could be useful for the development and assessment of new-targeted drugs ((Schneider-Kolsky et al., 2010).

1.3.5 Imaging using CT

The other imaging modality incorporated in this study is Computerised Tomography (CT). CT uses x-rays and overcomes the limitations of superposition and conspicuity, inefficient x-ray absorption, etc. found in conventional radiography (Goldman, 2007). There has been great innovation of CT technology due to technological advancement, clinical needs and faster computation times. The images are reconstructed either by filtered back projection or iterative reconstruction (Goldman, 2007). In the course of the last decade, there has been a marked reduction in slice thickness due to improved detector technology. The imaging in CT provides good spatial resolution and anatomical accuracy. It is used in combination with PET for attenuation correction and anatomical accuracy.

1.4 Background to Texture Analysis

In this section, a brief introduction to existing texture analysis methods is given. The statistical, fractal and wavelet techniques that were selected to analyse and characterise intra-tumour heterogeneity are further elaborated. A background to the previous work done in this field is also provided.

1.4.1 Texture Analysis

1.4.1.1 Introduction

Visual texture is an innate and qualitative concept that has not been uniquely defined. Different authors have adapted a definition depending on their field of application (Haralick et al., 1973, He and Li, 1991). The perceived texture of an object is similar to that experienced when one touches it. The irregularities of the surface, its prominence and magnitude give the object a characteristic identity when compared with other objects (Julesz, 1975). The visual texture is thus made up of intensity variations of the visual object, spatial organisation, and the rate at which this is distributed throughout the object.

Visual texture variations can be both a nuisance, in the process of segmenting an object in a region of interest, or a means to characterising and differentiating between natural objects. In this way, we consider, the spatial variation of intensities at scales many times smaller than the object of interest (Petrou and García Sevilla, 2006). Texture is generally scale independent.

Various texture analysis methods and descriptors have been used to analyse images by classifying them and also to segment regions of interest. Examples in

research include remote sensing, document processing, automated inspection and medical image analysis.

Different researchers and techniques have incorporated and emphasised different aspects of texture analysis. These include model-based, statistical and geometrical methods (Haralick, 1979, Castellano et al., 2004).

1.4.1.2 Model-based texture analysis approaches

Model-based methods use a model with descriptions of the intensities of the voxels. The values of the model parameters extracted are used to quantify the texture of the object of interest. Model-based texture techniques include autoregressive models, random field models (such as Markov random field) and fractals (Lopes and Betrouni, 2009). Markov random fields state that each volxel's intensity in the region/volume of interest is dependent only on the intensities of its neighbours. Thus, coefficients obtained for course texture will be similar, whereas they will differ widely for fine textures. Fractals are a class of mathematical objects that are termed 'self-similar' at all scales of measurement. This was introduced to characterise the complexity of objects so that they overcame the limitations of traditional Euclidean geometry (Mandelbrot, 1982). In practical terms, this assumption is limited by the resolution and size of the images. In order to compute the fractal dimensions of binary and gray-level images, several techniques have been developed, such as box-counting, fractional Brownian motion and area measurement methods (Lopes and Betrouni, 2009).

1.4.1.3 Geometrical and structural texture analysis methods

Another approach mentioned in the literature, termed structural methods, surmises that visual texture is made up of primitives with specific shapes and other properties and these follow certain placement rules or other criterion. The use of these technique is only suitable if the texture contains regular patterns (Petrou and García Sevilla, 2006). Geometrical approaches aim to detect textural patterns prior to texture analysis using statistical or placement rule methods.

1.4.1.4 Signal processing methods

Signal processing and transforms represent a range of techniques for performing texture analysis. The texture content can be characterised in terms of the number of edges: with fine texture associated with a high number of edges and course texture associated with fewer edges. Edge detection is achieved using spatial filtering methods (such as Robert's and Laplacian operators). Another way to describe texture is in terms of the frequency content: finer objects contain higher spatial frequency intensity variation whereas the opposite is true in courser objects. Methods include the Fourier transform and more powerful techniques that retain spatial information (such as Gabor filtering and Wavelet transforms).

1.4.1.5 Statistical texture analysis approaches

Textural images are intrinsically linked to two elements: the set of intensities ("gray-levels") and their spatial organisation. Statistical texture methods are based on the distribution of the spatial orientation of gray-levels in images using higher moments. First order statistics (FOS) consider the properties of individual

voxels whereas higher order statistics consider the relationship between pairs (second order statistics) or larger groups of voxels' intensities (higher order statistics). FOS generates gray-level distribution summary statistics (e.g. Mean, mode, median, standard deviation, skewness, etc.). FOS uses the whole volume of interest and does not provide specific inter-voxel relationships. The arrangement of the paired voxels' gray-level intensities is characterised used Gray-level Co-occurrence Matrix (GLCM) method was initially developed by Haralick in 1973 (Haralick et al., 1973) for the characterisation of 2D synthetic aperture radar images of various sandstone scenes. Higher order methods have been developed to characterise the relationship between three voxels and higher such as the Gray-level Run Length Matrix (GLRLM) (Galloway, 1975), Gray-level Size Zone Matrix (GLSZM) (Thibault et al., 2009), Neighbourhood Gray Tone Difference Matrix (NGTDM) (Amadasun and R, 1989).

In the work presented in this thesis statistical texture analysis methods and wavelets were used to characterise intra-tumour heterogeneity in PET and CT imaging.

1.4.1.5.1 Gray-level Quantisation

The initial step of statistical analysis is to normalise the imaging intensities into smaller numbers of gray-scale. The most common method used is uniform quantification, which bins or discretises the range of the original intensity into a number of discrete intervals (termed gray-levels). The formula is as follows:

$$I_{norm} = \frac{Ng - 1}{I_{max} - I_{min}} (I - I_{min}) + 0.5$$

Where *I* is the raw intensity of a voxel in the image, *Ng* is the number of quantisation levels used for normalising the intensities in the Volume of Interest (VOI), I_{max} and I_{min} are the minimum and maximising intensities in the original VOI, and [*x*] is the greatest integer no larger than *x* (floor function). Adding 0.5 ensures that the intensity is rounded to the nearest integer.

Other normalisation techniques such as the histogram equalisation, Gaussian and log transforms have also been utilised (Clausi, 2002). There is a critical trade-off between the choosing the gray-level scale that is a trade-off between retrieving relevant textural information whilst reducing the noise in the image and preserving relevant information content. In Clausi's paper (Clausi, 2002), he showed that the gray-level parameter maximising the differences between classes of images are feature-dependent although in other studies, unique parameters are used.

There has been some interest in texture analysis of PET data. Orlhac et al. (Orlhac et al., 2014) demonstrated that the methods used to calculate the texture indices substantially impacts on the eventual texture result. In addition, gray-level quantification below 32 gray-levels has a false correlation with the SUV. A further study, by Hatt (Hatt et al., 2015) demonstrated that the correlation with volume was affected by the quantisation level used. The gray-levels used in by Hatt varied from 4-256 gray-levels.

The issue with CT is different and the normalisation step is carried out over 25 Hounsfield units. This is the only normalisation carried out as per the Aerts et al. study (Aerts et al., 2014a).

1.4.1.5.2 First Order Statistics (FOS)

This set of features is derived from intensity distributions of radiotracer uptake in lesion to measure the variability of intra-tumour uptake. It is independent of gray-level quantisation.

1.4.1.5.3 Gray-Level Co-Occurrence Matrix (GLCM) method

Haralick developed the Gray-Level Co-Occurrence Matrix in his landmark paper in 1973 (Haralick et al., 1973). It was created as a tool to categorise aerial images of sandstone. The GLCM model developed in this work relies on this paper and extends it to the 3D space. Haralick's proposed methodology was to generate a matrix with dimensions equivalent to the number of occurrences between every pixel in the image. Fourteen features were then calculated from this matrix and were designed to quantify features that humans understand such as coarseness or smoothness.

Consider an image with four levels of pixel intensity. The corresponding GLCM will be 4x4 matrix wherein the number of occasions a given pixel pair occur are tabulated (Figure 1.1)
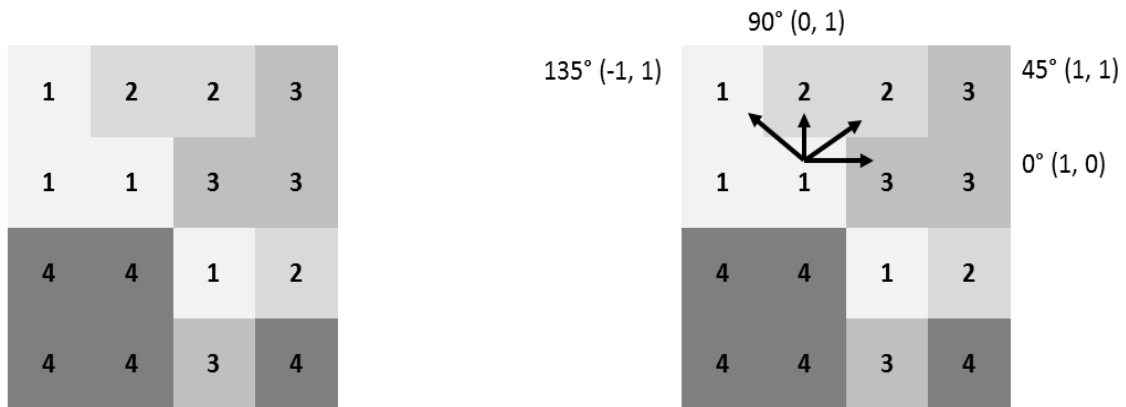
Figure 1.1 Generating the co-occurrence matrix. (left) The symmetrical gray-level co-occurrence matrix tabulating pixel pairs in the 0°. (right) The matrix is then normalised by the sum of the contents.

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

0°	1	2	3	4
1	0.083	0.083	0.042	0.042
2	0.083	0.083	0.042	0
3	0.042	0.042	0.083	0.083
4	0.042	0	0.083	0.17

Figure 1.2 The image and the GLCM. A sample image with 4 gray-levels (left) and the 4 2D directions of the pixels relationships from the measurement pixel (right).

The given examples represent the GLCM derived from the example image in the 0° direction. Both the X and Y-axis represent the gray-level of one of the pixels in the pair. For example, pixels of intensity 1 and 2 exist adjacent to each other, horizontally, on two occasions. In a 2D image, there are four directions available with the reference in the centre, covering the entire imaging space. An additional consideration is the pixel distance used in

establishing pixel pairs. In this example, as well as the GLCMs implemented in this thesis, a pixel distance of one is used. Note that the GLCM is symmetric in order to account for pixel relationships in the opposite directions. After 4 GLCMS are derived, one for each direction, the matrix is normalised by the sum of the contents. This step allows for easier interpretation of the features derived from the GLCM.

Haralick's paper detailed the GLCM methodology for 2D images, however, tomographic medical imaging including multiple slices comprise of a 3D volume. Co-occurrence matrixes were created in 3D by extending the directions from 4 to 13 (Fig 1.3)
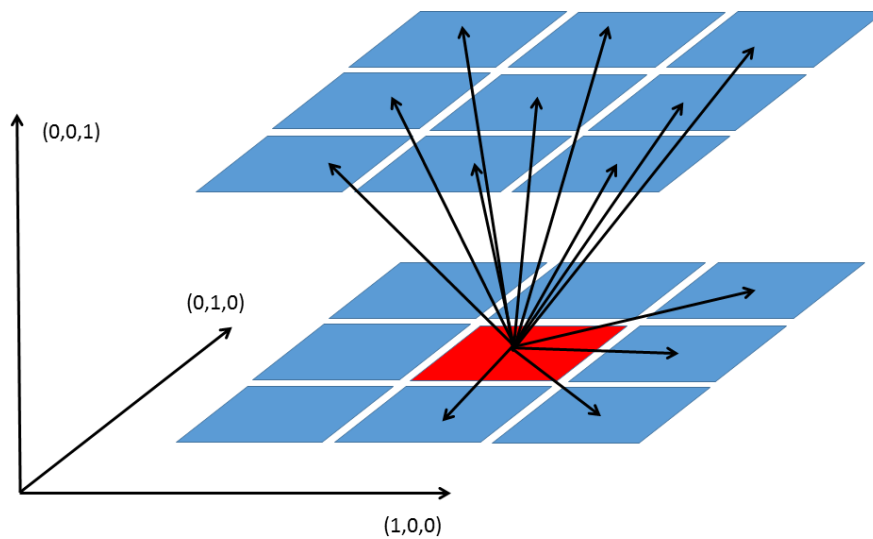
Figure 1.3 Converting the Haralick features into three-dimensions.
Thirteen directions of pixel pairs to account for a volumetric GLCM.

Symmetry ensures that the entire volume is accounted for, totalling 26 directions. Therefore, 13 normalised GLCMs are produced for a single volume. This is followed by an averaging step.

Another method used by Hatt et al. (Hatt et al., 2015) was a single matrix accounting for all 13 directions simultaneously without an averaging step. The argument given was that the first method averaged fewer matrices potentially leading to information loss and reduction of the effect of the residual noise from the quantisation process. However, there was a very high correlation between the two methods using Spearman Rank correlation (between r= 0.82 and 0.91) given in the same paper.

Haralick presented 14 local, 2nd order texture features extracted from each directional GLCM. The Aert's paper (Aerts et al., 2014a) increased this to 23 and this is the number incorporated in this thesis.

1.4.1.5.4 GLRLM

The gray-level run length matrix (GLRLM) presented by Galloway in 1975 is an additional texture matrix similar to the GLCM, but represents lengths of pixels instead of pairs (Galloway, 1975). Each Gray-level run consists of a series of pixels of uniform intensity directed along a linear path. A length can be as short as a single pixel, or as long as the image allows. As in the GLCM, run lengths are

calculated throughout the volume according to the thirteen angles previously designated. As runs of any length may be considered, symmetry is inherent to the matrix. Generally, fine textures tend to have many small runs of similar pixel intensities while coarse textures contain longer runs of varying intensities. GLRLM features emphasise regional information instead of the highly localised information of the GLCM. From the GLRLM, it can be determined if the image contains long or short lengths of high or low intensity gray-levels.

The formation of the GLRLM is similar to the GLCM matrix in that they are both dependent on the number of gray-levels present, which is designated along the vertical axis.

The other component is the length of the gray-level intensity designated along the horizontal axis. For each intensity level along the vertical axis, the number of occurrences for a given length of runs is tabulated in the matrix (Figure 1.4).

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

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

Figure 1.4 The Run Length Matrix. (left) A sample image containing 4 gray-levels and (right) the corresponding run length matrix in the 0° direction.

Pixels of intensity 1 occur once in a length of 4 pixels, once in a length of 2 pixels, and once in which there is a single pixel present. This tabulation is performed in all directions, producing 13 matrices, as in the GLCM.

Eleven regional, high order texture features are derived from the GLRLM describing features of the image (Tang, 1998). Features describe the prevalence of shorter lengths, representing fine structure, or longer lengths, correlating with coarseness. Other features describe the distribution of pixels based on intensity, or various combinations of lengths and pixel intensities.

1.4.1.5.5 GLSZM

The GLSZM represents another texture algorithm similar to the GLRLM. Instead of calculating lengths of pixels, however, the GLSZM permits the calculation of zones of uniform pixels in an image (Thibault et al., 2009). In a 2D image, a zone consists of any adjacent pixel in the image. This model was extended to a volume by finding groups of uniform pixel in each of the 26 available directions in 3D space. The matrix tabulates the occurrence of gray-levels grouped by zones of any size. As with the GLRLM, the GLSZM characterises regional texture features. As the zones were directionally independent however, only a single matrix was produced (Figure 1.5).
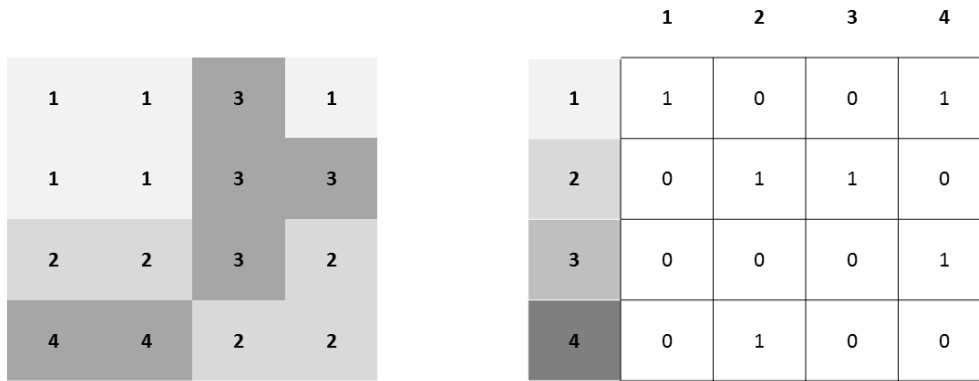
Figure 1.5 The GLSZM. (left)A sample image containing 4 gray-levels and (right) the corresponding size zone matrix.

The GLSZM is similar to the GLRLM in both design and purpose. 11 regional higher order features derived from GLRLM are applied to the GLSZM, with the zones substituted for the run lengths (see Appendix B). Analysing the distributions of uniform volumes of pixels within an image is perhaps more intuitive than extracting lengths from the volumes. A homogenous image may have a more uniform distribution of zones while a homogenous image may have a high variation in zones.

1.4.1.5.6 NGTDM

Previous algorithms described pairs or groupings of pixels and extracted features reflecting characteristics resembling what our own visual perceptions may be. The NGTDM represents a more complex set of features. Instead of grouping pixels based on intensity, the NGTDM takes an average of the surrounding pixel neighbours in order to determine the variation in texture

(Amadasun and R, 1989). In other words, this texture analysis approach examines changes in intensity between the target pixel and the surrounding neighbours. The relationship between the averaged neighbourhood values and the target values of the target pixel, can be small or large differences, and reflect the texture of an image. Originally, the NGTDM was proposed for 2D images. The functional form of the NGTDM is presented as

$$s(i) = \sum |i - \bar{A}_i|$$

Where $\bar{A}_i = \frac{1}{W-1}\left[\sum_{m=-d}^{d}\sum_{n=-d}^{d} f(k+m, l+n)\right]$

$(m,n) \neq (k,l)$

W is a normalisation factor to account for the neighbourhood size.

$$W = (2d + 1)^2$$

d represents the distance away from the centre pixel to the neighbourhood pixels. In this study, $d = 1$. For each intensity level, the surrounding neighbours with respect to the centre pixel of those intensities are added, normalised, and then compounded. This is explained in the following figure (Figure 1.6).
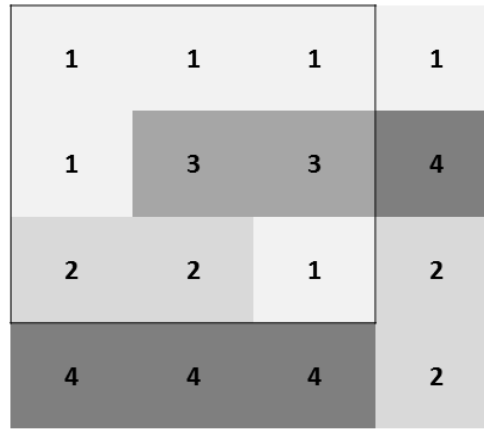
Figure 1.6 The NGTDM. A sample image with 4 gray-levels. The box indicates the neighbourhood surrounding a single pixel.

In the designated area of this image, the centre pixel has a value of 3. The total value of the neighbours is 12. In this pixel, $\bar{A}_i = \frac{12}{8}$. There are two pixels of value 3 in this image however. The NGTDM entry $s(3) = \left|3 - \frac{12}{8}\right| + \left|3 - \frac{15}{8}\right| = 2.625$. This is performed for every intensity level. The length of the matrix depends on the number of gray-levels in the image.

The NGTDM was extended to 3D in this work. Extending this matrix to 3D is not presented in the literature; however, it was a fairly straightforward process following the logic of the algorithm. In the calculation of the matrix, the number of available neighbours is 26 instead of 8, accounting for the area of a cube surrounding the centre pixel. The 3D weighting factor becomes $W = (2d + 1)^3$.

Since the calculation is directionally independent, only one matrix is produced for a 3D volume. The features presented in the original paper were created for a 2D image. Certain changes must be applied to account for 3D space, most of which substitute the 2D weighting for a 3D factor. There are 5 local, high order features derived from the NGTDM.

1.4.1.5.7 Wavelets

Wavelet transform effectively decouples textural information by decomposing the original image, in a similar manner to Fourier analysis, in low– and high frequencies. In this study a discrete, one-level and undecimated three-dimensional wavelet transform was applied to each image, decomposing the original image. Consider L and H to be a low-pass (i.e. a scaling) and, respectively, a high-pass (i.e. a wavelet) function, and the wavelet decompositions of X to be labelled as X_{LLL}, X_{LLH}, X_{LHL}, X_{HLL}, X_{HLH}, X_{HHL} and X_{HHH}. For example, X_{LLH} is then interpreted as the high-pass sub band, resulting from directional filtering of X with a low-pass filter along x-direction, a low pass filter along y-direction and a high-pass filter along the z-direction and is written as:

$$XLLH\,(i,j,h) = \sum_{p=1}^{N_L}\sum_{q=1}^{N_L}\sum_{r=1}^{N_H} L(p)L(q)H(r)X(i+p,j+q,k+r)$$

Where N_L is the length of filter L and N_H is the length of filter H. The other decompositions are constructed in a similar manner, applying their respective ordering of low or high-pass filtering in x, y, and z-direction. Since the applied wavelet decomposition is undecimated, this is shift invariant. Because of these properties, the original tumour delineation of the gross tumour volume (GTV)

can be applied directly to the decomposition after wavelet transform. In the current thesis "Coiflet 1" wavelet was applied on the original CT images.
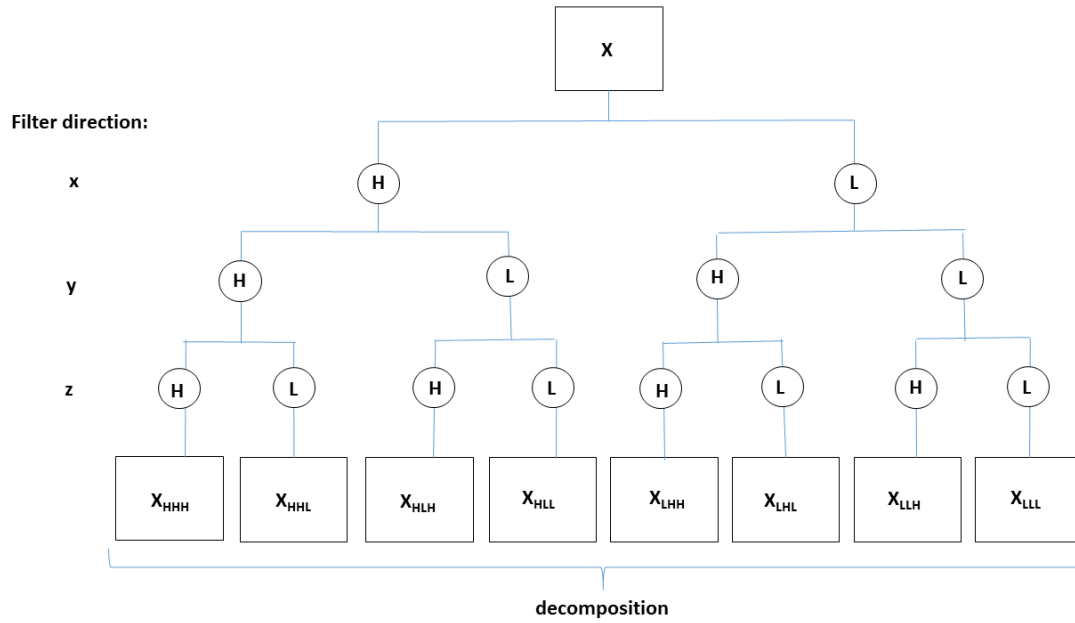
Figure 1.7 Schematic of the undecimated three-dimensional wavelet transform applied to each image. The original image X is fitered into 8 decompositions, by directional low-pass (i.e. a scaling) and high-pass (i.e. a wavelet) filtering: XLLL, XLLH, XLHL, XHLL, XHLH, XHHL and XHHH.

1.5 Textural feature analysis

So in summary, in this thesis, a total of 665 image descriptors are computed. The features were derived from: FOS, GLCM, GLRLM, GLSZM, NGTDM, Fractals and wavelets. The radiotracer concentration for each primary tumour was normalised prior to texture characterisation for a large number of quantisation levels: 4, 8, 16, 32, 64, 128, 256gl. This was a necessary step prior to applying the texture analysis.

Table 1.1 Overview of the change in texture features. A Comparison between the early generation software to the new generation software.

TexLAB 1.0	TexLAB 2.0
First order statistics (FOS) (12)	First order statistics (FOS) (15)
Grey-level co-occurrence matrix (GLCM) (17)	Grey-level co-occurrence matrix (GLCM) (23)
Grey-level size zone matrix (GLSZM) (13)	Grey-level size zone matrix (GLSZM) (13)
Neighbourhood grey- tone difference matrix (NGTDM) (5)	Neighbourhood grey- tone difference matrix (NGTDM) (5)
	Size and Shape SNS) (8)
	Gray-Level Run Length Matrix (GLRLM) (11)
	Fractal Dimension (FD) (6)

	Wavelet transformation x8
Total number of features = 47	Total number of features =665

Table 1.2 List of computed image descriptors. The texture features with full equations and explanation are detailed in Appendix B.

Technique (Total number)	Features
First order statistics (FOS) (15)	Coefficient of Variation
	Mean
	Median
	Mode
	Standard Deviation
	Minimum
	Maximum
	Range
	Skewness
	Kurtosis
	Mean Absolute Deviation
	Root mean square
	Area under the Curve
	Entropy

	Energy
Grey-level co-occurrence matrix (GLCM) (23)	Variance
	Correlation
	Information Measure of Correlation 1)
	Information Measure of Correlation 2)
	Cluster Shade
	Cluster Prominence
	Angular Second Moment
	Maximum Probability
	Entropy
	Contrast
	Dissimilarity
	Homogeneity
	Sum Average
	Sum Variance
	Sum Entropy
	Difference in Variance
	Difference entropy
	Autocorrelation
	Cluster Tendency

	Homogeneity 1
	Inverse Difference Moment Normalised
	Inverse Difference Normalised
	Inverse Variance
Grey-level size zone matrix (GLSZM) (13)	Small Zone Emphasis
	Large Zone Emphasis
	Grey-level Non Uniformity
	Size zone Non uniformity
	Zone Percentage
	Zone Low grey-level Emphasis
	Zone Low grey-level Emphasis
	Small Zone Low grey level Zone Emphasis
	Small Zone High grey level Zone Emphasis
	Large Zone Low grey level Emphasis
	Large Zone High grey level Emphasis
	Grey Level Variance
	Size-Zone Variance
Neighborhood grey- tone difference matrix (NGTDM) (5)	Coarseness
	Contrast
	Busyness
	Complex

	Strength
Size and Shape SNS) (8)	Volume
	Area
	Surface to volume ratio
	Sphericity
	Spherical disproportion
	Compactness 1
	Compactness 2
	Maximum 3d diameter
Grey-Level Run Length Matrix (GLRLM) (11)	Short Run Emphasis
	Long Run Emphasis
	Grey-Level Non-Uniformity
	Run Length Non-Uniformity
	Run Percentage
	Low Grey-Level Run Emphasis
	High Grey Level Run Emphasis
	Short Run Low Grey Level Emphasis
	Short Run High Grey Level Emphasis
	Long Run Low Grey Level Emphasis
	Long Run High Grey Level Emphasis

Fractal Dimension (FD) (6)	Mean
	Standard deviation
	Variance
	Lacunarity
	Maximum
	Minimum
Wavelet transformation x8	As above with 8 filters
Total number of features =665	

In the following chapter, there is a review of the literature and the growing evidence that demonstrate the potential for using such approaches to characterise and quantify intra-tumour heterogeneity in PET imaging. The limitations of PET quantification are also discussed.

Chapter 2 – Literature reviews, Aims and Objectives, and Methods

A systematic literature review was undertaken to find previous work on texture analysis of CT and/or PET. MEDLINE (date range: 1950-2016) and ISIS Web of Science (Date Range: 1970-2016) were interrogated using the terms listed in Appendix A1. The systematic review question was as follows:

The search was last run on 17/11/2016 (MEDLINE and Web of Science) using the following search terms and combination:

(variability OR heteroge* OR descriptor OR spatial OR feature)

AND

(characteri* OR quantificat*)

AND

(PET OR "Positron Emission Tomography" OR CT OR "Computerised

Tomography") AND (cancer OR lesion OR oncolog* OR neoplasm* OR tumour*)

Inclusion criteria were:

- The focus was on quantification methodologies of the radiotracer uptake variability in PET beyond the use of SUV metrics and variability in attenuation in CT oncology
- They addressed issues related to quantification in CT and PET imaging.

Studies were excluded if:

- PET radiotracer uptake characterisation or quantification was not examined and only briefly mentioned.

- Only detectability issues in PET were studied.

- Quantification procedures were exclusively carried out for evaluating and validating scanners' performance or correction algorithms;

- The variability in PET radiotracer uptake was not addressed;

- Texture analysis in other modalities such as MRI, radiographs, histology images or ultrasound.

- The disease studied was not cancer.

The literature review revealed that there have been different approaches in characterising tumour variability. After selecting the relevant papers (Appendix A2), studies could be further categorised as follows:

(i) Studies that used image analysis methods to characterise and quantify intra-tumour heterogeneity (N = 43);

(ii) Studies correlating inter-tumour variability across PET radiotracers or CT, imaging modalities, or techniques (e.g. after surgical resection) (N =23);

(iii) Studies investigating techniques to segment/delineate the tumour (N=6).

Within the scope of this thesis, given the large amount of work carried out in each of these categories, the focus in this chapter has been on the use of image

descriptors to characterise and quantify the spatial and intensity variability in PET radiotracer uptake and CT attenuation in oncology. The studies and their limitations will be outlined. Further brief discussions of (ii) and (iii), although essential to the field, will be briefly discussed here and in the discussion in Chapter 5.

A total of 43 papers relating to the quantification of intra-tumour variability and heterogeneity in the PET signal and CT attenuation were retrieved. Overall, within the last few years, different methods, incorporating heterogeneity of PET intensity and spatial distribution and CT attenuation differences have been proposed. These include model-based, statistical methods and/or approaches comparing the local spatial uptake of successive scans. In the subsequent paragraphs, a review of these is made.

2.1 Literature review: descriptors of intra-tumour PET radiotracer uptake

2.1.1 Model based techniques

In 2003, O'Sullivan et al. (O'Sullivan et al., 2003) developed a model-based technique, in particular a heterogeneity index, to differentiate between the deviation of local intra-tumour spatial uptake of PET radiotracer and the uptake in similarly sized homogenously filled ellipsoid phantoms. This homogenous object model was fitted to real data and gradient intensities to account for PVEs in PET imaging. This heterogeneity index was then used to predict clinical outcome in 238-treatment naïve sarcoma patients, who were scanned with [18]F-

FDG PET, prior to treatment with neo-adjuvant chemotherapy or surgical resection (Eary et al., 2008). The researchers found that their elliptical index (HE) was a strong independent predictor of patient survival.

A major limitation of the study was the use of ellipsoid models for homogenous lesions. A second limitation was this model was not spatially varied and poorly imitated the ground truth. The authors also admitted that the model poorly reflected the presence of necrosis, which is known to be associated with higher-grade sarcomas.

Beyond this work, the same group addressed the limitations of shape by employing an independent delineation step to define realistic tumour boundaries. The external boundary was accounted for by a set of contours, which were then defined for filling the lesion with a gradient of intensities. The authors developed a new descriptor, termed the surface heterogeneity (*HS*), similar to HE (except for the definition of the set of contours). In addition to this, a further index for the contours was derived, HG, which took into account the more elliptical distance of the boundary from the centre. These parameters were then compared with the results of the previous method in a group of 179 sarcoma patients. They found that *HS* and *HE* were strongly correlated with each other. In addition, HS obtained minimal further information in a multivariate model when controlling for HE. However, the authors did find that the risk of death associated with one standard deviation from these variables was significantly greater in HS than for HE (71.7% vs 36.2%). Further statistical analysis suggested that *HS* was the strongest predictor of survival for this group of sarcoma patients. O'Sullivan

et al. (O'Sullivan et al., 2011) further developed this work by incorporating global deviations from the model PET radiotracer uptake with intra-tumour heterogeneity, representing necrosis, and thought that this represented different stages of heterogeneity development.

Another model-based technique derived from fractal analysis was proposed (Dimitrakopoulou-Strauss et al., 2002) to use time activity curves (TACs) to differentiate different tissues in dynamic PET imaging. Fractal Dimensions (FD) quantify complexity across scales of measurement, and was applied in this study as an index of space-filling property of the TACs in a 2D plane.

They demonstrated promising methods by utilising this method to classify bone lesions to either benign or malignant, with a sensitivity of 71.9%, specificity of 81.6% and accuracy of 77.1%. The combination of the FD with the SUVs and kinetic parameters further improved the results. In addition, a combination of kinetic parameters were linked with short and long-term survival (set at 1 year) in metastatic colorectal cancer (Dimitrakopoulou-Strauss et al., 2004).

FD had an accuracy of 88.9% when discriminating between primary colorectal tumours compared with normal tissue (Strauss et al., 2007). However, FD did not discriminate between short- and long-term survival in advanced NSCLC (Dimitrakopoulou-Strauss et al., 2007), although there were only 42 metastatic lesions in 14 patients in this study and accurate conclusions cannot be made. In a further study with 50 patients with soft tissue sarcoma metastasis, FD failed to predict response to chemotherapy (Schmitt et al., 2011).

A major drawback to this method is that FD is derived solely from a graph of mean radiotracer uptake at successive time points, failing to account for spatial radiotracer heterogeneity. Also, the method is very sensitive to the imaging pre-processing and noise levels.

Using a different methodology, Brooks and Grigsby (Brooks and Grigsby, 2013b) proposed an index of intra-tumour heterogeneity derived from the deviation of average measure of voxel intensity from the smoothest intensity transition between any to voxels in a VOI. The research found some association with their metric to clinicians' ranking of heterogeneity and they found that some associations were significant ($p < 0.05$, Spearman Rank). The strengths of these associations were, however, not reported. There was also extensive pre-processing of the imaging prior to using their metric, which may have introduced bias. Only the largest cross-sectional image was used and the whole tumour in 3D was not assessed.

In a further development of their work, Brooks and Grigsby (Brooks and Grigsby, 2013a) modified their metric to account for lesions with complex shapes. This modification worked by ignoring the distance between two voxels in a line that were separated by some background, and instead considered that these voxels were neighbours. Although this approach may be justified for characterising tumours with necrotic cores, the example given by the authors should be interpreted cautiously. The method was illustrated on a spherical object with

decreasing intensity values from the centre to the object's border. Briefly, they found that the increasing number of decimations reflected the increasingly complex nature of the tumour and heterogeneity. However, they later found that this metric correlated strongly with volume and did not predict the lymph node involvement in increasingly complex cervical cancer primaries.

2.1.2 Texture analysis methods

Texture analysis, a non-invasive post-processing technique, permits the assessment of the variability within the whole tumour and any metastases to quantify grey-level differences between voxels (Willaime et al., 2013, Castellano et al., 2004). This technique detects differences over and above visual perceptibility (Tixier et al., 2014, Miller et al., 2003) and provides heterogeneity information beyond that available from a solitary histological sample.

A variety of methods have been used to produce textural parameters, including statistics-based, model-based and transform-based methods (Castellano et al., 2004). The most commonly used method in the literature has been the statistics-based model, which categorises different texture features into first-order (one voxel), second-order (two voxels) and higher-order (three or more voxels) statistics. First order statistical differences are calculated by comparing voxels in the whole volume of interest and result in features such as the mean, standard deviation, uniformity, entropy (the randomness of the values within the voxels), kurtosis or skewness (Willaime et al., 2013). Furthermore, complex second order statistical analyses between neighbouring voxels known as normalised grey-

level values in the co-occurrence matrix (GLCM) (Haralick et al., 1973) can also be examined. The calculations include the frequency of finding a voxel i in relation to its neighbourhood voxel j. These include measures such as local homogeneity, dissimilarity, neighbourhood entropy and correlation. Further second order statistics are derived from the grey-level size zone matrix (GLSZM) look at clusters of homogeneity (Thibault et al., 2009). Further regional features such as run-length and run-length variability are calculated from voxel alignment in a specified direction (Castellano et al., 2004, Galloway, 1975). Additionally, higher order statistics can quantify the coarseness or roughness, busyness, contrast or complexity (Amadasun and R, 1989). These utilise the neighbourhood grey-tone (intensity) differences matrices (NGTDM).

Table 2.1 PET texture analysis studies of all tumour types-Selected studies

Author	Year	Tumour type	No. patients	Prospective/ retrospective level	Quantization	Segmentation	Texture features used	Findings
Kidd et al ((Kidd and Grigsby, 2008)	2008	Cervix	72	Retrospective	Not mentioned	40-80% adaptive threshold function	Derivative of volume threshold significantly associated with tumour response to radiation and pelvic recurrence. MTV, and texture significant in multivariate analysis for pelvic recurrence	Change in volume over time tumour threshold significantly associated with tumour response to radiation and pelvic recurrence. MTV, and texture significant in multivariate analysis for pelvic recurrence
El Naqa et al. (El Naqa et al., 2009)	2009	Cervix Head & neck (H & N)	14 cervix 9 head & neck	Retrospective	Used a non-uniform quantization method (square root) although	40% threshold	Intensity volume histogram Texture based features (Energy,	Just for texture: Post radiotherapy outcome at 3 months (responder vs non-responder): AUC Energy>local homogeneity>contrast

Reference	Year	Cancer	Design	N	Method	Features	Findings
				value used not stated.		Contrast, Local homogeneity, Entropy) Shape-based features (Eccentricity, Euler number, Solidity and Extent)	Post-radiotherapy outcome for H&N (Responder vs Non Responder): AUC Local homogeneity > contrast > Entropy > Energy
Tixier et al. (Tixier et al., 2011)	2011	Oesophageal	Retrospective	64	Fuzzy local area Bayesian (FLAB)	FOS (7), GLCM (6), Run-length (11), size zone (11), NGTDM (3)	Therapy-response prediction (Responder vs. Non Responder): AUC (texture) > AUC(SUV)
Brooks, Grigsby, 2011 (Brooks	2011	Same as Kidd et al., 2008					Retracts previous findings, stating that the findings are merely a surrogate for tumour

and Grigsby, 2011)								volume.
Vaidya et al. (Vaidya et al., 2012)	2011	NSCLC	27	Retrospective	Not stated	Manual	Energy Contrast Entropy Local homogeneity	Texture features did not correlate with local or loco-regional recurrence.
Yang et al. (Yang et al., 2013)	2013	Cervix	20	Retrospective Pre-treatment, 2, 4 and 3 months post treatment	8	40% threshold	Gray-level run length (11 features) Gray-level size zone (11)	Correlation with the changes in tumour heterogeneity during therapy: Gray level run length: 1) High gray-level runs-emphasis, 2) Short runs high gray-level emphasis, 3) Long runs high gray-levels emphasis. Gray-level zone size: 4) High gray-level zones emphasis, 5)

						Short zone low gray level emphasis, 6)
						Large zone low gray-level emphasis, SUV$_{max}$, SUV$_{mean}$ in the complete metabolic response group but not with the partial response or new disease group.
Tan et al. (Tan et al., 2013)	2013	Oesophageal	20	Retrospective, pre and post treatment	Not stated	FOS (9), GLCM (9), Geometry features (15)	Prediction of pathologic response texture features better than SUV for responders.
Dong et al. (Dong et al., 2013)	2013	Oesophageal	40	Retrospective	Semi-automated using cut-off of >2.5 SUV	Semi-automated using cut-off of >2.5 SUV	Energy Entropy	Correlation with T and N stage: r (entropy) > r (energy) > r (SUV$_{max}$). Prediction of advanced stage ([IIa-IIb] vs. [IIIa-IV]): entropy only

Reference	Year	Cancer	N	Design	Grey levels	Segmentation	Texture features	Findings
Cheng et al. (Cheng et al., 2013)	2013	Oro-pharyngeal	70	Retrospective	4, 8, 16, 32, 64	Semi-automated using cut-off of >2.5 SUV	GLCM (7), NGTDM (5)	Multivariate analysis showed that age, tumour TLG, and uniformity were independently associated with progression-free survival (PFS) and disease-specific survival (DSS). TLG, uniformity significantly associated with overall survival (OS).
Bagci et al. (Bagci et al., 2013)	2013	Various	30		Not specified	Automated Random Walk Image Segmentation	Homogeneity; entropy; energy; run indices	Prediction of patient outcome or variation in uptake region characteristics: entropy, SRE (Short run emphasis)
Cook et al.	2013	NSCLC	53	Retrospective	Not stated	Manual	Coarseness	Therapy response:

Reference	Year	Tumour	Patient number	Study design		%	Texture features	Findings
(Cook et al., 2013)							Contrast Busyness Complexity	RECIST responders had lower coarseness, higher contrast and busyness vs non-responders. High coarseness associated with lower PFS, LPFS and OS. Higher contrast and busyness associated with longer LPFS and PFS.
Win et al. (Win et al., 2013)	2013	NSCLC	Feasibility: 56 feasibility 66 validation	Feasibility: Retrospective Validation: Prospective	Not stated	42% automated segmentation first order statistic)	Entropy (Not Haralick or first order not on multivariate analysis.	Entropy significant in the validation set but not on multivariate analysis.
Brooks, Grigsby, 2013 (Brooks and Grigsby, 2013	2013	Pre-treatment Cervix	85	Retrospective	Not stated	40% threshold	Sphericity Extent Shannon entropy Accrued deviation	No texture correlation between node positive and negative in stage 2B patients.

Reference	Year	Site	N	Study type / timing	Method / threshold	Features	Results
2013a)						from the smoothest gradients	
Kidd et al., 2013 (Kidd et al., 2013)	2013	Cervix	25	Retrospective: Pre-treatment, 2, 4 and 3 months post treatment adaptive to obtain the derivative	MTV: 40% threshold 40-80% threshold function SUV$_{max}$ MTV	Derivative of volume treatment MTV and FDG$_{hetero}$, pre-FDG$_{hetero}$ were all significantly associated with post-treatment PET response.	Week 4 SUV$_{max}$ and FDG$_{hetero}$, pre-treatment MTV and FDG$_{hetero}$ were all significantly associated with post-treatment PET response.
Yang et al., 2013 (Yang et al., 2013)	2013	Cervix	20	Retrospective: Pre-treatment, 2, 4 and 3 months post treatment	8	40% threshold	Gray-level run length (11 features) Gray-level size zone (11)	Correlation with the changes in tumour heterogeneity during therapy: HGRE, SRHGE, LRHGE, HGZE, SZHGE, LZHGE, SUV$_{max}$, SUV$_{mean}$ in the complete metabolic response group but not with the partial

Reference	Year	Cancer	N	Design	N	Method	Feature	Results
Tixier et al. (Tixier et al., 2014)	2014	NSCLC	101	Retrospective	64	FLAB (Fuzzy local area Bayesian)	Visual	response or new disease group.
								Moderate correlation between the visual & quantitative analysis.
							Entropy	
							Dissimilarity	
							Homogeneity	Higher heterogeneity: local and regional texture features (except HIE)
							Size zone homogeneity	associated with poorer prognosis (OS). In multivariate analysis
							Zone percentage	stage, MTV and texture (except HIE) were
							High Intensity emphasis (HIE)	significant. PFS was significant in MTV, TLG, texture (except HIE)
Brooks, Grigsby, 2014	2014	Cervical cancer	70	Retrospective Pre-treatment	8	40% adaptive threshold	Entropy	Entropy 5 times more sensitive to tumour volumes less than

Reference	Year	Cancer	N	Study design		Segmentation	Features	Findings
(Brooks and Grigsby, 2014)					45cm³			
Orlhac et al., 2015 (Orlhac et al., 2015)	2015	NSCLC	48	Retrospective	different	adaptive	FOS GLCM GSZLM	Absolute resampling performed better than relative resampling in differentiating malignancy from normal tissue.
Pyka et al., 2015 (Pyka et al., 2015)	2015	NSCLC	45	Retrospective, Pretreatment, T1 or T2 received primary stereotactic radiation therapy	64	2.5 isocontours of SUV	GLCM NGTDM	Entropy (HR 7.48, p=0.16) on a multivariate analysis significant for disease specific survival.
Lovinfosse et al., 2016 (Lovinfosse et al., 2016	2016	NSCLC	63	Retrospective, Treated with stereotactic	Not stated	FLAB (Fuzzy local area Bayesian)	13 global, local and regional	Univariate analysis dissimilarity (D) was associated with DSS

Study	Year	Cancer type	Study design		Segmentation	Features	Findings
et al., 2016)		body radiation therapy					(HR = 0.822, P = 0.037), multivariate analysis, dissimilarity significantly associated with DSS (HR = 0.822, P = 0.037) and with DFS (HR = 0.834, P < 0.01).
Vallieres et al., 2017 (Vallieres et al., 2017)	2017	Head and neck cancer	Retrospective	8, 16, 32 and 64	31% Drawn on CT and transferred onto PET, 69% drawn on CT using software	1615 radiomic features (quantifying tumour image intensity, shape, texture)	Predictive models developed using random forests and imbalance-adjustment strategies and tested on independent datasets. AUC on two datasets were 0.69 and 0.86
Nakajo et	2017	Thymic	Retrospective	Unknown	Unknown	4 features	High risk tumours had

Treated with stereotactic body radiation therapy

al., 2017 (Nakajo et al., 2017)	epithelial tumours		(entropy, homogeneity, intensity variability and size zone variability	higher entropy, p= 0.038, intensity variability, p = 0.041 and size zone variability, p=0.045.

AUC Area under the curve, SUV Standard uptake value, NSCLC Non small cell lung cancer, PFS progression free survival, LPFS local progression free survival, OS overall survival, MTV metabolic tumour volume, TLG total lesion glycolysis, HGRE High grey-level runs emphasis, SRHGE Short runs low grey-level emphasis, LRHGE Long runs high grey-level emphasis, HGZE high grey-level zone emphasis, SZHGE Short-zone high grey level emphasis, LZHGE Large zones high grey-level. DSS Disease specific survival.

Table 2.2 CT texture analysis studies of all tumour types-Selected studies

Author	Year	Tumour type	No. patients	Prospective/ retrospective	Quantization level	Segmentation	Texture features used	Findings
Al-Kadi, Watson, 2009 (Al-Kadi and Watson, 2008)	2009	Lung cancers	15	Retrospective		Semi-automated	Fractals	Quantitative classification of the lacunarity demonstrated 83% accuracy in distinguishing between early and late stage tumours.
Miles et al., 2009 (Miles et al., 2009)	2009	Colorectal cancer	48	Retrospective		Semi-automated	Brightness and uniformity	Phantom and clinical evaluation. Pixels between 2 and 12 pixels wide compared. CT uniformity 10 to 12 pixels had the least variability and best

Treated with stereotactic body radiation therapy

Reference	Year	Cancer	Number	Study type	Method		Texture features	Findings
Ganeshan et al., 2010 (Ganeshan et al., 2010a)	2010	NSCLC	17	Retrospective	Used fine to course detail (1.0-2.5)	Semi-automated	Mean Grey intensity (MGI), entropy (E), uniformity (U)	Course texture features correlated with tumour stage 2 (p=0.0001). SUV$_{max}$, fine tumour features correlated with tumour stage (kappa of 0.7 for tumours above stage 2 (p=0.0001). predictor of patient survival (p<0.005)
Goh et al., 2011 (Goh et al., 2011)	2011	Renal cell carcinoma	39	Retrospective	Used fine to course detail (1.0-2.5)	Semi-automated	Entropy Uniformity	Baseline and after 2 treatment cycles with tyrosine kinase inhibitor. Tumour entropy decreased by 3-45% and uniformity increased by 5-21%. Significant difference in

uniformity in KM curve (p=0.008 vs p=0.267) and Cox regression demonstrated that uniformity was independent predictor of time to progression (OR 4.02, 95% CI 1.52-10.65, p=0.005)

| Hunter et al., 2012 (Hunter et al., 2012) | 2012 | NSCLC | 25 | Retrospective | Manual and then in-house deformable image registration | GTV, image histogram, absolute gradient, co-occurrence matrix and run-length matrix | Different models tested, strong responders (normalized end of treatment GTV less than median) TPR 0.66 (0.60-0.75), FPR 0.36 (0.31-0.4), classification accuracy 64.8% |

Ganeshan et al., 2012 (Ganeshan et al., 2012)	2012	Oesophageal Cancer	21	Retrospective	Fine to course detail (1.0-2.5)	Semi-automated	Uniformity (U) Entropy (E)	U (r=0.75, p<0.0001) and E (r=0.75, p <0.0001) on unenhanced CT correlated with SUV_{mean}. Coarse uniformity independent predictor of survival OR=4.45 (95% CI: 1.08-18.4, p=0.039) (60-68%), p= 0.118 (0.055-0.215)
Zhang et al., 2013 (Zhang et al., 2013)	2013	Locally advanced squamous cell carcinoma of the head and neck who were	72	Retrospective	Fine to course detail (1.0-2.5)	Semi-automated	Entropy Skewness	Primary mass entropy and skewness associated with overall survival, in multivariate cox regression primary

previously treated with induction cisplatin, 5-fluorouracil, and docetaxel (TPF) chemotherapy							mass size, N stage and skewness were independently associated with overall survival.
Ng et al, 2013 (Ng et al., 2013b)	2013	Colorectal cancer	55	Retrospective	Fine to course detail (1.0-2.5)	Semi-automated	Entropy (E) Uniformity (U)	Entropy and Uniformity were measured at different filtration levels and largest cross sectional area was compared with whole tumour volume on contrast enhanced CT. There was a difference in U and E at all filters,

Reference	Year	Cancer type	N	Study type	Detail	Method	Features	Findings
Ng et al, 2013 (Ng et al, 2013a)	2013	Colorectal cancer	55	Retrospective	Fine to course detail (1.0-2.5)	Semi-automated	Entropy, uniformity, kurtosis, skewness, and standard deviation of the pixel distribution histogram	Fine texture features above and below their respective optimal cutoffs were associated with poorer prognosis, and were independent of each other on Cox regression and stage. and the whole tumour analysis appeared more representative of tumour heterogeneity
Aerts et al., 2014 (Aerts et al., 2014b)	2014	NSCLC and Head and neck (422	1019	Retrospective			440 features including intensity,	Developed a 4 feature texture signature which

Reference	Year	Cancer	N	Design	Method	Features	Results
					training) and (3 independent validation sets)	shape, texture and wavelet	was highly significant when tested in the independent lung and head and neck cancer datasets
Yip et al., 2014 (Yip et al., 2014)	2014	Oesophageal	36	Retrospective	Used fine to course detail (1.0-2.5) Semi-automated	Entropy, uniformity, mean gray-level intensity, kurtosis, standard deviation of medium uniformity the histogram and skewness	Post-treatment medium entropy (OS 33.2 vs 11.7 months, p=0002), coarse entropy (OS 33.2 vs 11.7, p=0.0002) and medium uniformity (OS 33.2 vs 11.7, p=0.0002) were associated with improved survival. Remained significant when adjusted for age and

Balagurunathan et al., 2014 (Balagurunathan et al., 2014)	2014	NSCLC	32	Retrospective		Semi-automatic 3D region growing algorithms	219 including size, shape, texture, wavelet	Unenhanced CT scanned 15 minutes apart demonstrated 42 features that were reproducible (>/=0.95), which were also predictive of radiological prognosis.
Knogler t al.,	2014	Hodgkin's	29	Retrospective		Run length	Texture features

stage. Pre-treatment entropy and uniformity performed better than morphological assessment for maximal wall thickness (AUC 0.77 vs 0.49, p=0.0005) and 0.80 vs 0.49, p=0.0003)

2014 (Knogler et al., 2014)		Lymphoma						were able to identify complete response with the same accuracy as CT but with higher specificity and the combination of CT and texture features had the highest accuracy, sensitivity and specificity.
Yip et al., 2015 (Yip et al., 2015)	2015	Oesophageal	31	Retrospective	Used fine to course detail (1.0-2.5)	Semi-automated	Mean grey-level intensity, standard deviation, Entropy, uniformity, kurtosis, skewness, standard deviation of histogram	Pre-treatment and post-treatment standard deviation of the histogram showed borderline association with pathological tumour response. Change in skewness was associated with

Study	Year	Cancer	N	Study type	Method	Software	Findings
Hayano et al., 2015 (Hayano et al., 2015)	2015	Soft tissue sarcoma	20	Retrospective	Used fine to course detail (1.0-2.5)	Semi-automated	Positive correlation of mean of positive pixels with microvessel density. Entropy positively correlated with VEGF receptor. improved survival.
Tian et al., 2015 (Tian et al., 2015)	2015	Soft tissue sarcoma	20	Retrospective			Arterial peak enhancement at baseline, 2 and 8 weeks post treatment with neoadjuvant bevacizumab and radiotherapy. Significant correlation of mean of positive pixels with necrosis in surgical specimen

Reference	Year	Description	N	Study type	Segmentation	Features	Results
Lubner et al., 2015 (Lubner et al., 2015)	2015	Hepatic metastases from colorectal cancer	77	Retrospective	Used fine to course detail (1.0-2.5)	entropy, kurtosis, skewness, mean, mean positive pixels (MPP), and standard deviation (SD) of pixel distribution histogram	Entropy, MPP and SD at medium filtration significantly associated with tumour grade. Skewness negative associated with KRAS mutation. Entropy at coarse filtration associated with overall survival (HR 0.65, 95% CI 0.44-0.95, p=0.03). Similar results for 2D and 3D analysis
Mattonen et al., 2015 (Mattonen et al., 2015)	2015	Post treatment lung cancer treated with	24	Retrospective	Manually delineated	Entropy	Entropy had an area under the curve of 0.7 to 0.73 to predict recurrence

		stereotactic ablative radiotherapy (SABR)					compared with radiation induced lung injury.
Lubner et al., 2016 (Lubner et al., 2016)	2016	Renal cell carcinoma	157	Used fine to course detail (1.0-2.5)	Semi-automated	Entropy, kurtosis, skewness, mean, mean of positive pixels, and SD of pixel distribution histogram	Pretreatment tumour texture analysis correlated with histology (entropy with clear cell histology and papillary cell type). Entropy, SD and MPP were associated with nuclear grade. SD, MPP and entropy were associated with time to disease recurrence and death due to disease.

Treated with stereotactic body radiation therapy

Study	Year	Cancer type	N	Detail	Segmentation	Features	Findings
Ahn et al., 2016 (Ahn et al., 2016)	2016	Liver metastases from colorectal cancer (validation)	235 (145 training and 90 validation)	Used fine to course detail (1.0-2.5)	Semi-automated	Mean grey-level intensity (Mean), standard deviation, entropy, mean of positive pixels (MPP), skewness, and kurtosis	Lower skewness, higher mean attenuation and narrower SD in the training set were independently associated with response to chemotherapy. On the validation set lower skewness (AUC=0.797) and narrower SD (AUC=0.785) showed good performance. These two were invariant to scanner.
Zhang et al., 2017 (Zhang et al., 2017)	2017	High (HG) and low-grade (LG)	105 HG and 16 LG	Used fine to course detail (1.0-2.5)	Semi-automated	Mean grey-level intensity	Low-grade urothelial cancers had lower mean,

urothelial cancer						(Mean), standard deviation, entropy, mean of positive pixels (MPP), skewness, and kurtosis	entropy and MPP on unenhanced and enhanced images enhanced images. and lower SD on enhanced images. The highest AUC was with a fine scale with AUC of 84.9% and accuracy 83.1%)
Sacconi et al., 2017 (Sacconi et al., 2017)	2017	NSCLC	68	Retrospective	Used fine to course detail (1.0-2.5)	Semi-automated	Mean grey-level intensity (Mean), standard deviation, entropy, mean of positive	0.779 (sensitivity 72.2%, specificity 0f Mean (p=0.0001), standard deviation 9 p=0.0001), and skewness (p=0. 0.0459) were found to have significant correlation with EGFR mutation positive (p=0.0001;

Treated with stereotactic body radiation therapy

						pixels (MPP), skewness, and kurtosis	p=0.0001; entropy correlated with OS (r=0.2708; p=0.0329).

TPR=true positive rate, FPR=false positive rate; OS Overall survival

2.2 PET Texture studies

Since El Naqa et al ((El Naqa et al., 2009)) first used texture analysis to quantify intra-tumour PET heterogeneity, a number of studies have been published demonstrating the predictive and prognostic power of various statistical texture analysis methods computed on baseline ^{18}F-FDG PET images (Table 2.1). The PET texture features have also been combined with CT descriptors (Yu et al., 2009a, Yu et al., 2009b, Vaidya et al., 2012). The textural descriptors (listed in the Appendix B) employed in these studies include statistics based texture.

El Naqa (El Naqa et al., 2009) implemented texture-, shape-based features and cumulative SUV-volume derived (CSH) curves to predict treatment outcomes. They tested their methods on two tumour groups imaged with ^{18}FDG PET: 14 subjects with cervix cancer and 9 subjects with head and neck cancer. In the cervix dataset, individual features were able to predict disease persistence. The texture features were correlated with overall survival in the head and neck dataset. In addition, a receiver operator curve (ROC) analysis on a subset of texture features on a multivariate analysis to predict treatment failure revealed that texture features performed the best followed by CSH features. A combination of the best individual descriptors led to an area under the ROC curve of 0.76 with the predominant contribution from the textural descriptor. Further univariate analysis, shape and CSH descriptors demonstrated the strongest correlation with clinical outcome for the head and neck cancer dataset. From this study the conclusion was drawn that class features (volume-, shape-,

or texture-based) predominate and this may depend on tumour or end-point selection. This proof of concept study was very important for introducing texture analysis to medical images, although, the small size of the datasets precludes the conclusions reached, especially on multivariate analysis. The shape descriptors may also suffer from being extra sensitive to the mode of delineation and other factors that influence shape (such as motion and PVEs).

Yu et al (Yu et al., 2009a, Yu et al., 2009b) used another method to incorporate both PET and CT data to derive parametric textural features. These features were tested against tissue grade to derive tumour targets. Manual and threshold methods for segmentation were compared. Their proposed method performed well compared with manual delineation but comparison with ground truth data was not made.

Further research (Galavis et al., 2010) was carried out to investigate the effect of PET reconstruction parameters on the texture techniques (FOS, GLCM, GLRLM, GLSZM and NGTDM) on scans of 20 subjects with various tumour types. The same scanner and protocol (time after injection) was used. Data acquisition in both 2D mode (OSEM reconstruction: 14 subsets, 2 and 4 iterations) and in 3D mode (reconstructed using iterative-Vue Point in 2 and 4 iterations) was performed. Results were computed across various matrix size (using 128 x 128 and 256 x 256 pixels) and post reconstruction filtration (3, 4, 5mm full width at half maximum [FWHM]). After computing texture features across these parameters, and found that some texture features were invariant across reconstructions, some had intermediate variability (classified here as between

10 and 25%) and a third group had a large amount of variability (more than 30%).

Tixier et al. (Tixier et al., 2011) furthered research into this field by comparing the performance of 38 statistical texture features, computed on pre-treatment ^{18}F-FDG PET images, to differentiate chemo-therapy response. Using RECIST criteria, a total of 41 patients, were classified as complete-, partial- or non-responders a month after completion of treatment. Using PET parameters, SUV_{mean} and SUV_{max} were able to discriminate between complete versus both partial and non-responders. SUV_{peak} was narrowly significant to differentiate between all three groups (p=0.045, Kruskall-Wallis test). Some textural features were good predictors of response to treatment at baseline, including $GLCM_{Entropy}$ (p=0.0006), $NGTDM_{Coarseness}$ (p= 0.0002), as well as $GLSZM_{Intensity\ variability}$ (p=0.0002). Texture features were also more sensitive (76-92%) than SUV measurements. In a follow-up study, the authors studied the repeatability of First Order Statistics and texture features derived from GLCM and GLSZM matrices in pre-treatment ^{18}F-FDG PETs of 16 oesophageal cancer patients scanned within 2 to 7 days of each other. They showed that there was much variability between the features and classified the variations into three groups: small variations of less than ±30% (e.g. $GLCM_{Entropy}$, $GLCM_{Homogeneity}$ and $GLCM_{Dissimilarity}$); medium variations of around ±50% (e.g. $GLCM_{Contrast}$, $FOS_{Skewness}$) and large variations of more than ±100% (e.g. $GLSZM_{Small\ Area\ Emphasis}$).

Further studies have investigated texture analysis in ^{18}F-FDG PET oesophageal cancers. Dong et al (Dong et al., 2013) compared two GLCM features at baseline with histological grade and stage (both TNM and AJCC) in 40 oesophageal

squamous cell carcinoma patients. Moderate associations were found between tumour stage and $GLCM_{Entropy}$ (rs = 0.69, p < 0.001, Spearman Correlation) and $GLCM_{Energy}$ (rs = – 0.47, p = 0.002). Weak associations were found between SUV_{max} and tumour stage (rs = 0.39, p = 0.013), Node stage and $GLCM_{Entropy}$ (rs = 0.50, p = 0.001), Node stage and $GLCM_{Energy}$ (– 0.41, p = 0.008), as well as with SUV_{max} (rs = 0.33, p = 0.04). The study design did not state whether their methods enhanced or complemented current clinical methods. A better study design would have compared outcomes using established predictive or prognostic factors. The stage system has been developed using thousands of patients and no advantage can be gained by using such a small cohort to develop a texture model.

Tan and co-workers (Tan et al., 2013) proposed a model based on intensity-, shape-, textural- and geometric-based features on 20 patients' pre-, post-treatment and difference (pre- minus post-treatment) ^{18}F-FDG PET images. 192 features were tested on their discrimination between response groups using the area under the ROC curve and Mann-Whitney U-Test. The descriptors with the highest area under the ROC were selected. A major limitation of this study is the small number of patients included which increases the likelihood of a type 1 error. Another limitation was that the texture features were not investigated through a repeatability study.

Vaidya et al. (Vaidya et al., 2012) combined ^{18}F-FDG PET and CT feature-based metrics for tumour assessment in NSCLC. A total of 32 texture features, selected from a mixture of first order statistics, GLCM, CSH and PET from 27 pre-therapy

scans were obtained. The statistical framework was similar to that detailed in previous work (El Naqa et al., 2009) and used univariate Spearman Rank correlation for predicting treatment outcome to radiotherapy. Clinical endpoints included the development of recurrence within the radiation field and chest. The authors found that CSH best predicted clinical outcome on the univariate analysis but the strength of the individual correlations were weak (from 0.12 to 0.37). There was slight improvement using bivariate logistic regression including two CSH volume indices: V80 on PET and V70 measured on CT images. Vx, derived from cumulative SUV-volume histogram, represented the sub-tumour with voxels having at least x% intensity of the maximum intensity in the VOI. The bivariate model led to rs = 0.4854 (p=0.0067) and rs = 0.5908 (p = 0.0013) for the two respective clinical endpoints. There was weak association between GLCM textural features and clinical endpoints (<0.3), which was not significant on univariate analysis. After correcting for motion, there was only a slight improvement. However, on careful observation of the results, the results were not borne out and only weak associations were demonstrated. For the remainder of the descriptors, the associations post motion correction weakened or were unchanged. Thus, it is still unclear whether motion correction improves the relation to clinical outcome.

George et al. (George et al., 2012) used a different method, in a study of 15 metastatic colorectal cancer patients, where they used a subspace-based framework to predict outcome and used textural analysis of primary and nodal lesions. They found that texture analysis enhanced the predictive power of the analysis compared with traditional quantification indices.

Cook and co-workers (Cook et al., 2013) applied the NGTDM only to 53 pre-treatment ^{18}F-FDG PET scan in NSCLC undergoing radical radiotherapy. Responses to treatment at 12 weeks using the RECIST criteria and survival outcome, namely local progression free (LPFS), progression free (PFS) and overall (OS), were used as outcome measures. The authors found lower NGTDM$_{Coarseness}$ (p= 0.004), higher NGTDM$_{Contrast}$ (p=0.044) and higher NGTDM$_{Busyness}$ (p=0.002) in non-responders compared with responsive patients. The other NGTDM texture feature, Complexity, and SUV measurements did not predict tumour response. In addition, NGTDM$_{Coarseness}$ was found to independently predict on multivariate analysis, while NGTDM$_{Contrast}$ and NGTDM$_{Busyness}$ were associated with PFS and LPFS. The number of gray-level discretisation was not mentioned, making it very difficult to verify or replicate these findings. The use of the same texture features for predictive and prognostic outcomes could potentially be a fallacy. In addition, the lack of further testing of these features on an independent dataset is another shortcoming.

In a break from the described texture features, Salamon et al. (Salamon et al., 2013), used the ratio of SUV$_{max}$ to SUV$_{mean}$. They compared this with the SUV$_{max}$ in a group of 50 patients with peripheral nerve sheath tumours. Using this semi-quantitative index, a significant difference was found between benign and malignant tumours (p=0.0002). However, there was a greater differentiation using the SUV$_{max}$ alone (p=0.001) and the malignant tumours were significantly bigger (p < 0.0001). Thus, the proposed differences may not be due to

heterogeneity but rather the lesion size and noise. The choice of using the SUV_{max}/SUV_{mean} was also not given.

Yang et al. (Yang et al., 2013) extended the use of texture analysis by following it up over the course of treatment in 20 cervical cancer patients treated with radio-chemotherapy. Patients were scanned pre-treatment (baseline), two and four weeks after the onset of therapy and after the completion of therapy. The authors found that regions of high uptake decreased significantly over the course of treatment in complete responders ($p < 0.001$), ie. the tumours became more homogenous. Also, the pre-treatment scan was the most accurate in predicting outcome. By contrast, the SUV indices decreased in both response and non-response groups from baseline up to four weeks post treatment commencement. The results from both GLCM and GLSZM matrices were broadly similar.

Soussan et al. (Soussan et al., 2014) investigated 3 texture features, $GLRLM_{HGLRE}$, $GLCM_{Entropy}$ and $GLCM_{Homogeneity}$, and PET features to investigate associations with prognostically poor indices in 54 ^{18}F-FDG PET pre-treatment breast cancer patients. $GLRLM_{HGLRE}$ were higher in oestrogen and progesterone negative receptors in high-grade tumours. The PET parameters were unable to identify HER-2 positive tumours, however, SUV_{max} did correlate with Ki-67 ($p<0.0004$). Multivariate analysis demonstrated that $GLRLM_{HGLRE}$ was associated with triple negative breast cancer. Combining this texture feature with SUV_{max} increased the area under the curve compared with SUV_{max} alone in identifying triple negative breast cancer. The major fallacy in this study was the small number and there were only 13 triple negative tumours in the whole sample.

In 2017, Vallieres (Vallieres et al., 2017) evaluated the usage of texture analysis to predict treatment outcomes and prognosis in head and neck cancer. 1615 texture features were extracted from pre-treatment PET and CT scans of 300 patients. Using random forests and imbalance-adjustment strategies using two of four cohorts, they developed a prediction and prognostic model that was tested in the remaining two cohorts with results AUCs of 0.69 and 0.86 for prediction and Kaplan Meier curves demonstrating survival differences (depending on the Random Forest thresholds and local and regional metastases). The strength of this study was that they had a large cohort and they tested their findings on independent datasets. They also developed their model and combined them with existing clinical factors demonstrating the additional benefit of PET and CT texture features. A major drawback to this methodology was that there were too many features and insufficient events (both recurrence and death) to include all of these in their analysis.

In summary, texture analysis and PET parameters have been used for a number of tumour types using a variety of methods and different statistical techniques. The outcomes have been different: with some linking to stage, PET parameters, treatment outcome and survival. Furthermore, the techniques have been applied to many different imaging protocols (such as scanner type and reconstruction algorithms) and cancer types in different institutions. It is difficult to apply direct comparisons due to the resolution dependent nature of texture analysis. More

studies with larger, cross-centre sites may help in answering some of these questions.

2.2.1 CT texture studies

In CT texture analysis, developments have paralleled those of the developments of PET scan. An early study by (Al-Kadi and Watson, 2008) used fractals in 15 lung cancer patients to differentiate between early and late stage disease in an early proof of concept study. There was 83.3% accuracy in distinction between the two groups.

A further texture study on 48 colorectal cancer patients with hepatic metastases and phantoms used the texture features of brightness and uniformity pixels between 2 and 12 size. They found that the largest pixel sizes correlated with patient survival. They also found that CT hepatic perfusion was not associated with survival. Although this was an early study, no subsequent studies using these findings on a validation set were performed.

The majority of texture analysis on CT has been performed using TexRAD, a propriety software (Ganeshan et al., 2010b) that looks at certain 2nd order statistics and uses filtering (Laplacian of Gaussian) of fine, medium and course detail. There have many studies (Table 2.3) using this software, for example, looking at NSCLC with clinical and PET parameters (Ganeshan et al., 2010a), RCC response prediction with tyrosine kinase inhibitors (Goh et al., 2011), oesophageal carcinoma and survival (Ganeshan et al., 2010a), previously chemotherapy treated head and neck cancer and survival (Zhang et al., 2013),

colorectal cancer and survival (Ng et al., 2013a), Hodgkins lymphoma CT texture response compared with CT (Knogler et al., 2014), soft tissue sarcoma with VEGF receptor correlation (Hayano et al., 2015) and response assessment (Tian et al., 2015). These studies have all been small studies with less than 100 patients performed retrospectively. The predictive and prognostic texture features were different even for the same tumour type. No validation studies were performed on independent datasets. Larger datasets on 157 patients with renal cell carcinoma ((Lubner et al., 2016) and 235 (145 training and 90 validation) cororectal patients with hepatic metastasis (Ahn et al., 2016) have also been performed.

Using the previously described statistical method to define texture, Aerts (Aerts et al., 2014b) have the largest study to date, with 1019 patients in total, using a 422 patient NSCLC training set and testing on 3 independent validation datasets, two of which were NSCLC and one which was head and neck cancer. A 4-feature texture analysis was found to be highly significant in predicting prognosis in all the datasets. One of the shortcomings of this study was that a test re-test dataset was used to develop the robustness of 440 texture features and this was unenhanced CT only (Balagurunathan et al., 2014).

2.3.1 Quantifying the spatial extent of PET radiotracer uptake: total lesion activity

The total lesion glycolysis (TLG) has been proposed as an additional index to quantify the distribution of radiotracer uptake in PET imaging (Larson and

Schoder, 2009, Hatt et al., 2011a, Larson et al., 1999, Willaime et al., 2014). The calculation is relatively straightforward, simply being the product of the SUV_{mean} and metabolic tumour volume. It is also known as total lesion metabolic activity (TLA) and total proliferative volume (TPV) (Hatt et al., 2010a). The lack of standardisation across centres has resulted in numerous names for the same measure, and this is also true for texture features (Orlhac et al., 2014).

The measure was first introduced by Larson et al (Larson et al., 1999) who evaluated this feature when looking at percentage change in different cancers that were treated with chemotherapy. The response assessment was evaluated by a group of experts.

Hatt et al. (Hatt et al., 2010b) studied the repeatability and reproducibility of a range of PET features, including the metabolic tumour length (MTL), mean tumour volume (MTV) and total lesion glycolysis (TLG) on an ^{18}F-FDG oesophageal dataset and an ^{18}F-FLT PET breast cancer dataset. Lesions were segmented using different methods, including fixed and adaptive thresholding, fuzzy C-means and fuzzy locally adaptive Bayesian (FLAB) (Hatt et al., 2009). The repeatability results (multiple delineation of the same lesion) yielded variability below 5% for automatic methods (adaptive thresholding, fuzzy C-means and FLAB) and from 5 to 35% for manual delineation. In this study, Hatt et al. (Hatt et al., 2011a) studied the accuracy of TLG, tumour volume and longitudinal length, to discriminate between three groups of patients with oesophageal cancer (13 non-responders, 25 partial responders and 12 complete responders). From the baseline ^{18}F-FDG PET imaging, patients were assessed for treatment response

post therapy using the RECIST criteria on CT imaging.   Two different segmentation techniques were used for comparison, the adaptive threshold method and the automatic FLB algorithm. The SUV_{mean}, SUV_{max} and SUV_{peak} did not discriminate between the three response groups. On the contrary, the other PET discriminators (MTL, MTV, TLG) predicted response on baseline ^{18}F-FDG PET scans. Sensitivity was 75% and specificity above 85%, yielding markedly better results than SUV measurements. Segmentation did not influence these results. The only other measure that differentiated between these groups was the AJCC stage. The authors combined SUV_{mean}, which by itself was not significant, with TLG and TL, which produced a stronger discrimination between response groups than using TL separately. Again, the segmentation did not change the ability of TL, TV and TLG to distinguish between the groups.

The authors further studied the prognostic value of the image descriptors with regards to survival in a pre-treatment ^{18}F-FDG PET in oesophageal cancer (Hatt et al., 2011b). They found that SUV measurements (SUV_{mean}, SUV_{max} and SUV_{peak}) did not significantly predict overall survival. TV, TL and TLG, however, were significantly prognostic but not when tested in a multivariate analysis.

The authors noted that FLAB performed better than the adaptive threshold technique with higher sensitivity, specificity, and area under the curve with smaller confidence intervals. It should be noted that the same group developed the FLAB segmentation method. The method had to be adapted when it was shown that the method could not differentiate heterogeneity well, simply two

classes (the background and foreground). The FLAB method has subsequently been improved to take into account a further class, named 3-FLAB (Hatt et al., 2010a). The 3-FLAB was compared with other automatic segmentation methods in 31 NSCLC patients in a separate institution, which found that it performed worse compared with the other methods tested (Markel et al., 2013).

Further studies by the same group (Hatt et al., 2013a), demonstrated that percentage change in TLG, imaged at baseline and after two cycles of treatment with neo-adjuvant chemotherapy, was a stronger predictor of histopathologic response compared with SUV_{mean} percentage change. Furthermore, they found that correcting for PVEs significantly impacted the feature values of baseline ^{18}F-FDG PET, however, it did not impact on the predictive values (Hatt et al., 2013b). Hatt (Hatt et al., 2013c) and Groheax (Groheux et al., 2013) both showed that TLG was a strong predictor of response in rectal and breast cancers, respectively imaged at baseline and after two cycles of neo-adjuvant chemotherapy. However, neither demonstrated that TLG was associated with treatment outcome based solely on the baseline images.

Mertens (Mertens et al., 2012) derived an index related to total lesion glycolysis, named standardised added metabolic activity (SAM). The advantage over TLG was that it was easier to derive, allowed for partial volume corrections, and did not rely on thresholding methods. The method consists of drawing two volumes of interest (VOI) around the target. The first region (VOI1) includes all the voxels including background voxels contaminated by spillover. The second region

(VOI2) includes some surrounding background activity and is drawn around VOI1. The mean background intensity is then estimated by taking into account voxels belonging to VOI2 and outside VOI1 (that is the difference between the two). Finally, the mean background activity is subtracted from the voxels' intensities in VOI1, which are then summed to obtain the SAM index. Validation was performed using phantoms of differing sizes that were filled with different target to background activity. The influence of scanning time on the SAM index was assessed in 15 subjects with primary squamous cell cancer. Further, the index change following chemotherapy was studied in 19 patients with metastatic colorectal carcinoma. The authors stated that the expected SAM index had very good accuracy in homogenous backgrounds ($\geq$ 85%), independent of lesion size. They found that scan duration (from 1 to 10 minutes) did not impact significantly on the results with squamous cell cancer. In their small cohort, SAM discriminated between 8 responders and 11 non-responders (p=0.001). Percentage change in SAM was very high (99%) but these findings would need to be confirmed in larger datasets.

Mertens (Mertens et al., 2013) then applied SAM to response assessment in metastatic colorectal liver patients. They discovered that both percentage changes in SUV_{max} and SAM were prognostic factors for PFS and OS. The authors stated that the SAM corresponds to excess metabolic uptake due to the lesion over the background rather than the TLG. The authors stated that this methodology was undertaken as the tumour tissue develops over normal tissue. However, in reality, tumours develop with a mixture of tumour cells mixed with

normal cells, stroma, etc. and this method could possibly exclude some of this. In addition, PVEs will affect lesions' voxels depending on their location, tumour size, shape and Target to Background Ratio (TBR) due to the limited spatial resolution of the scanner as well as the tissue fraction effect. The simple method used in this technique is unlikely to accurately reflect this. SAM makes the assumption that the lesion metabolism is dependent on background metabolism. However, it is possible for the same lesion embedded in a region where there are two or possibly more background tissue mean activities. If it assumed that PVEs are not affected, the TLG index, which would be identical for the lesion in both background tissues, SAM indices would be a function of the mean background tissue activity.

Therefore, SAM could be appealing due to its simplicity and ability to account for PVEs, it is a different approach from other PET quantifications tools in that the background activity impacts on the absolute values calculated.

Thus currently, there is no gold standard to measure disease extent using PET. Robust tools that are capable of segmenting heterogeneous lesions may be preferred.

2.3.2 Cumulative SUV-volume Histogram (CSH) curve

Kidd and Grigsby (Kidd and Grigsby, 2008) first proposed the use of a cumulative SUV-volume histogram (CSH) curve to quantify the intra-tumour heterogeneity of ^{18}F-FDG PET. This is also known as the intensity-volume histogram, or

volume-threshold curve. This method plots a graph to represent the lesion volume of an intensity threshold as a function of the maximum uptake in the lesion. The graph is an adaptation of the dose-volume histogram used in radiotherapy. The derivative dV/dT) of the volume-intensity threshold was used as a functional index of heterogeneity (*HET*). They tested this method on 72 patients with cervical cancer and demonstrated that HET was associated with radiotherapy response (p=0.02), as well as risk of recurrence (p=0.002). However, they also found a strong correlation between HET and lesion size in the PET image derived from thresholding segmentation approach (r^2=0.88, Spearman Rank).

In addition, lesion size was a stronger predictor of disease recurrence (p=0.0003) than the heterogeneity metric (p=0.0035). Subsequently, the same group criticised their findings (Brooks and Grigsby, 2011) that their metric was non-spatial and was directly proportional to the lesion size. Indeed, Hoang confirmed this (Hoang et al., 2013) by using the original heterogeneity metric on 40 patients with nasopharyngeal carcinoma and found a very high linear correlation between the heterogeneity index and lesion size (r = - 0.98, Spearman).

El Naqa et al (El Naqa et al., 2009) derived additional intensity and volume indices from the CSH based on volume and intensity thresholds. Their CSH features had the second highest predictive power (after texture analysis) in terms of therapy failure risk.

Van Velden et al. (van Velden et al., 2011) furthered earlier work by introducing the area under the curve (AUC) as an index to the cumulative SUV volume histogram (AUC-CSH). They then evaluated this method on PET simulations of lung lesions (accounting for attenuation and resolution effects), simulating both spatially homogenous and heterogeneous (core and rim) intra-tumour responses to treatment whilst varying SUVs and tumour size. Furthermore, they proceeded to demonstrate their method on four clinical cases, demonstrating that lower AUC-CSH was associated with greater heterogeneity using simulations and clinical data. Although, this was a good proof of concept study, further work with test-retest and larger validation sets are required.

The CSH and its derivative have been promising but have not been validated or shown to have clinical usefulness. As with all the metrics mentioned, larger independent datasets are required to show their potential utility. The need for this is to prevent published work being non-informative in future studies.

2.3.3 Assessing local tumour changes using PET imaging

Other groups have investigated the direct comparison of successive PET scans at the voxel-level on intra-tumour uptake and assessed local response to treatment. This required the accurate registration of successive PET scans using anatomical images (eg. CT). Necib et al. (Necib et al., 2008) used a method which computed the difference between two successive PET scans which were registered using the CT component. A graph of the subtracted voxels against the radiotracer

uptake of the same voxel in the original scan was obtained. The authors used a mixed Gaussian model with Expectation-Maximisation (EM) to classify the voxels into the following four groups: noise, physiological change other than in tumour, tumour voxels responding to therapy, and tumour voxels corresponding to tumour progression. The final step involved setting the groups corresponding to noise and physiological change to zero. This had the effect of displaying only the voxels representing change in SUV to be displayed.

The feasibility of this approach was tested on 8 lung tumours of two patients (Necib et al., 2008). An extra assessment of the performance of this approach was tested in 28 metabolic colorectal cancer patients (Necib et al., 2011). In total, there were 78 lesions, which included primary, liver, lung, peritoneal and other metastases. RECIST criterion was used to assess response. Global descriptors, such as percentage change in volume, SUV_{max}, SUV_{mean}, were also used and the method was compared with another response assessment method based on the guidelines of the EORTC (Young et al., 1999). They published that the technique had 100% sensitivity and 53% specificity compared with 85% and 52% for the EORTC-based classification, respectively. A major advantage of this method was the possibility of generating a map, which the heterogeneity of local intra-tumour treatment response was visualised and further, quantified. An obvious limitation is the introduction of errors due to incorrect registration due to motion and patient positioning. The parametric method did yield promising results in detecting response but a few non-responders were incorrectly

classified as responders. Clinically, it is more useful to detect non-responders earlier.

Schreibmann (Schreibmann et al., 2013) used a similar approach, employing PET/CT deformable registration, to correct for anatomical differences between scans. The researchers used a level-set framework for grouping voxels within the lesion associated with an increase or reduction in radiotracer uptake across successive scans. The method was assessed using 81 ^{18}F-FDG PET/CT of head and neck cancer patients treated with chemo-radiotherapy. The authors highlighted a level-set approach to filter out signal changes outside the tumour on the basis of spatial and intensity information compared with simple thresholding. There was however a very unsatisfactory visual assessment in the validation method, which, precludes the ability to accurately compare its results with other methods.

Further developments are needed to include tumour heterogeneity assessment.

Chapter 3 Hypothesis, Research Questions, Aims and Objectives

3.1 Hypothesis

Texture analysis on the imaging of untreated cancers can characterise and quantify lesions to provide prognostic data.

3.2 Aims and Objectives (scope of the thesis)

Different methods and approaches have investigated the complexity of PET radiotracer uptake in neoplastic lesions. However, image descriptors have only been applied to the biological function of glucose metabolism. Different tumours have been investigated including sarcoma, cervix, head and neck, colorectal, breast, oesophageal and NSCLC. Study findings have been promising in suggesting that image-based descriptors might provide additional predictive and prognostic information over and beyond that is currently available.

Questions remain regarding repeatability and the impact of other PET reconstruction parameters. PVEs have been identified as a potential factor. Questions remain regarding tumours with necrotic centres that may be falsely attributed as having more viable tissue than there is present or having an active part that may appear less aggressive than it is (Soret et al., 2007). The resultant appearance on PET would be affected by the resolution-dependent image metrics used to characterise heterogeneity (chapter 4). Metrics such as TLG are also affected by resolution through PVEs (Hatt, 2012). It is hoped that despite

this, the quantification process may capture the extent of radiotracer uptake, despite it being different between the rim and core.

The aim of this thesis is to characterise and quantify spatial properties and variability in PET radiotracer uptake in neoplastic lesions. The work will be further extended to CT and incorporate underlying genetic changes and semantic features. Robust quantification methods in oncology imaging are critical in providing clinicians with tools that will have an impact on their daily practice, so that they can detect earlier recurrence, allow for quicker identification of recurrence and resistance and also personalise treatment.

The specific objectives of this thesis were:

- To explore whether robust image descriptors gained from different scanners and from multiple centres can truly capture biological heterogeneity (at the whole tumour level) using PET and CT imaging. The previous section (chapter 1) presented the image descriptors. Within the scope of this thesis, the biological processes investigated are limited to NSCLC assessed using ^{18}F-FDG PET and CT of ovarian cancer.

- The results are presented in Chapter 5. The differences attributed to the texture features from the different scanners will be explored. The performance of these descriptors will be compared with existing clinical measures and SUV indices.

- To use the database that has been acquired to evaluate the literature to interrogate previous studies.

- To advance tumour quantification by combining attenuation and spatial descriptors in CT, and comparing with semantic features, molecular and proteomic data.

Chapter 4 – Methods for the dataset

4.1 Introduction

In this chapter, an outline of the selection for the dataset for the NSCLC and ovarian cancer texture analysis is provided. The inclusion and exclusion criteria for the scans are outlined. The methods are also detailed.

4.1.2 Materials and Methods

4.1.2.1 The Lung Cancer Dataset

The ^{18}F-FDG PET dataset is outlined below. The exclusions are also outlined. From 535 initial patients, 358 were selected (this is discussed in detail in the following chapter). The study focused on the prognosis as opposed to treatment prediction and so the individual treatment for each patient is not stressed (Law and Miles, 2013). The information for the treatment given to the patients has only partially been obtained, which is due to the multi-centred retrospective nature of this study, as information loss invariably occurs. Other pieces of information that would have been useful and have shown to provide prognostic information are performance status (Karnofsky or ECOG classification), weight loss (e.g., > 5%) and systemic inflammation (C-reactive protein or modified Glasgow Prognostic Score) (Lee et al., 2008, Detterbeck et al., 2013, Alberg et al., 2013). Again, these data were not obtained.

The inclusion criteria:

- Scanned between October 2008 to December 2013

- histological subtypes: squamous cell, adenocarcinoma or non-small cell lung cancer

- staging PET available for review

- patient had received external beam radiotherapy +/platinum based chemotherapy

- Size of tumour >5ml (Soussan et al., 2014)

- Follow-up and date of death available

Exclusion criteria:

- Follow up scans not available

- Metastatic disease (to fulfil the M criteria on the TNM) present

- Not non small cell lung cancer

- Too small (<5ml tumour volume)

- No scan available

- Primary not visualised

- No clinical details

4.1.2.2 Outcomes

As the patient data were collected from various centres, there was a need to limit follow-up imaging bias, and thus only the overall survival data was obtained. Follow-up imaging bias occurs when treated patients have a reoccurrence detected when they are reimaged or are only reimaged if symptomatic, and thus a more frequent follow-up imaging rate (Wang, 2015) will detect more recurrences. For lung cancer, overall survival (i.e. date of death) is relatively

straightforward to collect because the five-year survival is poor. In addition, the focus of this study is on prognosis rather than prediction and is thus treatment independent. The exact cause of death was not determined as this was difficult to obtain from the various participating organisations. Thus, for the purpose of this study, the overall survival or date of death was from any cause, even if it was unrelated to lung cancer.

Overall survival was defined as number of months from commencement of treatment to date of death. Patients who were alive were censored at last follow-up.

4.1.2.3 Ethics and participating centres

This retrospective observational study evaluated the relationship between texture features of pre-treatment PET-CT scans and survival outcomes in patients diagnosed with non-small cell lung cancer (NSCLC). Institutional ethics agreement (14HH1908) was obtained.

Scans from 7 separate centres were obtained:

- Imperial College Healthcare NHS Trust, London
- Kings College (St Thomas' Hospital), London
- The Royal Marsden Hospital, London.
- Leeds (St. James' University Hospital).
- Mount Vernon Hospital, Northwood, London.

- Cancer Imaging Archive (https://public.cancerimagingarchive.net/ncia/login.jsf)

  Databases LUAD and LUSC. Last accessed 30/6/2015.

- University Hospital, Nottingham

The division of the scans from the different centres to make the training, validation and test sets are outlined in Figure 4.2.

4.1.2.4 Image acquisition and reconstruction

Image acquisition was performed on different scanners with different reconstruction protocols.

The following describes the imaging protocol at Imperial: ^{18}F-FDG PET/CT imaging was performed on a Siemens Biograph 64 PET/CT scanner (Siemens AG, Erlangen, Germany) following the standard protocols at Imperial. After 6 hours of fasting, patients first underwent an unenhanced CT from the base of skull to upper thighs for the purpose of attenuation correction (50mAs/slice). 400MBq of ^{18}F-FDG (mean 393.1 MBq, range 306-454) were then injected intravenously (plasma glucose level was < 11 mmol/L at the time of FDG injection). Emission data were acquired for 3 minutes per bed position after a 60-minute uptake period (mean 60.9, range 58-68 minutes). The CT parameters were 120 KVp, 50 mAs, pitch 0.8, 5 mm slices with 3 mm separation. Images were reconstructed using ordered subset expectation maximization with 4 iterations, 8 subsets and a

Gaussian filter of 5mm FWHM. The images were attenuation-corrected using the

CT data.

The imaging acquisitions in the other institutions were obtained from the DICOM

headers from the scans. For more details of the Kings dataset please see Cook

(Cook et al., 2013).

4.1.2.5 Comparing the datasets

As the datasets were obtained from different centres, it was important to

investigate the differences between them and to account for them to ensure the

differences in texture analysis are due to the underlying histology and not due to

scanner or clinical details. The differences included an investigation into the

following: 1. Differences in clinical information between the centres including

age, sex, tumour, histology and stage 2. Differences between the machine

manufacturer, model, matrix and slice thickness as it has been demonstrated that

these can cause a difference in the texture readout (Kumar et al., 2012, Leijenaar

et al., 2015).

4.1.2.6 PET analysis

Central analyses of all PET/CT data were conducted at Imperial College London

by a semi-automated adaptive threshold method. The primary tumour was

delineated using an initial threshold of 40% of the SUV_{max} on semi-automated

software (Hermes Gold3; Hermes Medical Solutions Ltd, London, UK), as previously described (Uto et al., 2010, Cook et al., 2013). In cases where the 40% threshold did not clearly encompass the entire tumour, the threshold was lowered (Biehl et al., 2006). All segmentation was made by the same operator who was a radiologist with 4 years' of experience with tumour delineation. The operator made adjustments if the incorrect areas were included or excluded by the automated software. The SUV_{mean}, SUV_{max}, SUV_{peak}, metabolic tumour volume (MTV) and total lesion glycolysis (SUV_{mean} x MTV)(TLG) were recorded. The minimum size of the volume of interest (VOI) was selected as 5mls, according to Soussan (Soussan et al., 2014). The same volumes of interest (VOIs) were extracted and used on the radiomics software.

To assess the intra- and inter-observer variability of the segmentation method, 18 random patients were selected and segmentation of the tumour was performed (at 128 gray-levels) while blinded to the original results and clinical data. The two different observers had 6 and 10 years' experience of tumour delineation. Lymph nodes or metastases were excluded from the statistical analyses because they are biologically different from primary tumours.

4.1.2.7 MatLAB Transfer

The VOIs, defined by the aforementioned semi-automatic segmentation were extracted as Cartesian co-ordinates. In-house software was developed and used to convert the VOIs so that they could be implemented on MatLAB. The VOIs and

the PET scans were used as input for the in-house textural analysis software package (TexLAB 2.0), developed on MATLAB 2015b (Mathworks Inc., Natick, Massachusetts, USA). SUVs were calculated for each voxel of the VOIs.

4.1.2.8 Radiomics analysis

Radiomics analysis was performed at seven different quantisation/gray levels – 4, 8, 16, 32, 64 128 and 256 - on TexLAB 2.0. From each primary tumour, 660 radiomic features (Section 1.6 and Appendix B) were extracted from segmented VOIs using local, regional, global, fractal and wavelet techniques. These included intensity features, shape features and texture features (gray level co-occurrence matrix and gray level run length matrix, Neighbourhood Gray Difference Matrix) with or without wavelet transformation as previously reported (Aerts et al., 2014a, Willaime et al., 2013). Radiomics features were determined from 133 PET scans (Training set) using TexLAB 2.0.

4.1.2.9 Feature selection and radiomics signature discovery

It was important to reduce the total number of features for prediction purposes in order to avoid over-fitting and instead learn the true basis of a decision. We initially identified highly correlated features for elimination using heatmaps, as highly correlated features suggested that some feature reduction could be undertaken without information loss. Heatmaps were created in the R software (http://www.r-project.org/; version 3.03 Vienna, Austria).

Initially individual texture features from the FOS and GLCM matrix were used to test the training, validation and test set. The optimal extracted features were used to create a composite feature and this was tested on the different sets and on multivariate analysis using stage.

From the 660 sets of features at each gray-level, we used Least-Absolute-Shrinkage-and-Selection-Operator (LASSO) regression analysis for data dimension reduction and radiomics feature vector (composite feature) discovery. LASSO is a form of penalised regression used to reduce the problem of multi-collinearity. Briefly, the non-contributory variables were assigned zero-weighting and numerous iterations were performed to link the non-zero contributory variables to the chosen clinical outcome (in this example, overall survival) (Tibshirani, 1997). Analyses were conducted with the R software and the packages in R used for our analysis are indicated in Supplemental Material. Two-sided statistical significance levels were used and $P \leq 0.05$ was considered statistically significant. SPSS version 22 (SPPs v22, IBM, New York, US) was used for interclass correlation and 2-way ANOVA.

The most predictive feature vectors (FVX) were computed by linear combination of selected statistical features weighted by their respective coefficients and by comparison with overall survival (OS). Survival curves were plotted using Kaplan Meier (KM) methods, stage-specific or median values in the case of feature vecture, termed FVX. The Kaplan Meier curves were created suing median and the best cut-off on the receiver operator curve from the median

survival (Youden's J). The survival curves were evaluated using a log-rank test (Cox Regression). Multivariate analysis of the FVX, stage and volume were compared with each other using a stepwise backward procedure to determine significantly independent survival indicators. P values of less than 0.05 were considered statistically significant, and 95% confidence intervals were calculated.

4.1.3.10 Independent validation and Testing

Performance of the optimal FV (FVX) and stage were tested by comparison to OS in an independent validation set of 204 patients, and a further independent set of 21 (TESTI) patients. Similar survival comparisons were made with routine PET variables including SUV_{mean}, SUV_{max}, SUV_{peak} and TLG. An overview of the methods are given in Figures 4.1 and 4.2.

4.1.3.11 Statistical analyses

Statistical analyses have been detailed above. The analysis was performed on SPSS for Windows version 22 (IBM, Chicago, IL, USA); interclass correlation and 2-way ANOVA were used.

The Principal Component Analysis (PCA) analysis was performed using 'devtools' and 'ggbiplot' and 'vqv' packages. 'The 'gplots' package was used to draw heatmaps with a Spearman Correlation. LASSO binary logistic regression

and Kaplan Meier curves were performed using the 'doParallel', 'Matrix' 'glmnet

2.0-5', packages using statistical language R (version 3.03, Vienna, Austria).

a) Images segmented from PET   b) Segmentations converted to matrices   c) Texture features extracted   d) Analysis with link to survival
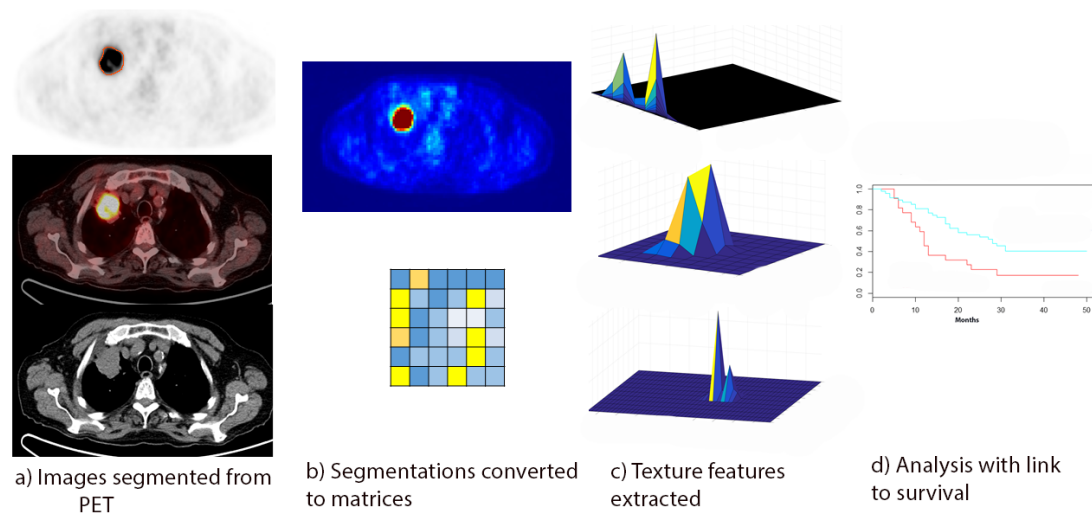
Figure 4.1 Overview of the texture analysis process:

A) The images are obtained from PET and are segmented, b) the imaging analysis

software utilises the segmentation to develop different matrices, c) the output

for various texture features at different grey levels is obtained, d) these are

linked with clinical features such as survival.
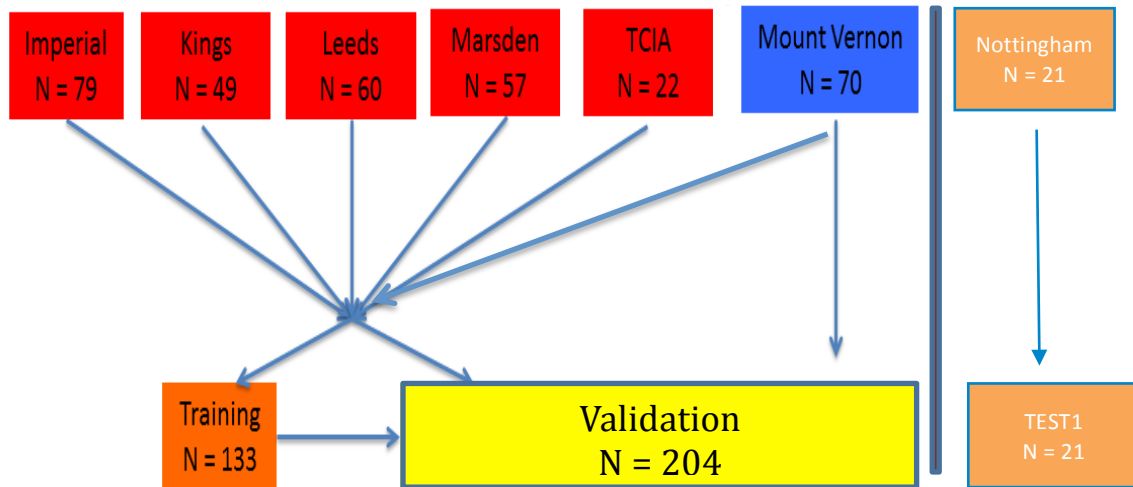
Figure 4.2 Overview of the number of different number of scans obtained from the various centres.

Overview of the texture analysis process: a) The images obtained from PET and are segmented, b) the imaging analysis software utilises the segmentation to develop different matrices, c) the output for various texture features at different gray levels is obtained, d) these are linked with clinical features such as survival, e) Overview centres contributing to the study and how the data were randomly divided into training, validation or independent test set. The Nottingham data were assigned as the independent test set, TEST1. TCIA The Cancer Imaging Archive.

4.1.3.12 Comparison with previous literature

There literature on use of texture analysis for imaging is substantial. Unfortunately, the small datasets utilised and lack of validation has meant that there is a high likelihood of type 1 errors. Using the published literature on

texture analysis and prognosis, the current dataset was used to verify if the results, when FVX's from other studies were used, are valid. Papers on different tumour types and modalities were also included.

4.2 Statistical analyses over the course of the PhD

4.2.1 Change of course as the research progressed

The development of the final methods presented in this thesis has been built on earlier work that was undertaken at the beginning of the research. An overview of this is presented in Figure 4.3. The initial results (not presented) did give very promising insights but the use of small numbers and the statistical use of (PCA) were felt to be non-reproducible in larger studies. Thus, a change in direction occurred where larger datasets, including multiple centres and the use of emerging statistical techniques of machine learning including penalised regression was pursued. In addition, there was a parallel development of improving the texture analysis software with increasing numbers of features as more publications were released. In addition, further information became available with regards to the limits of earlier published studies in terms of the minimal size of the primary and which gray-level quantisation were most useful.
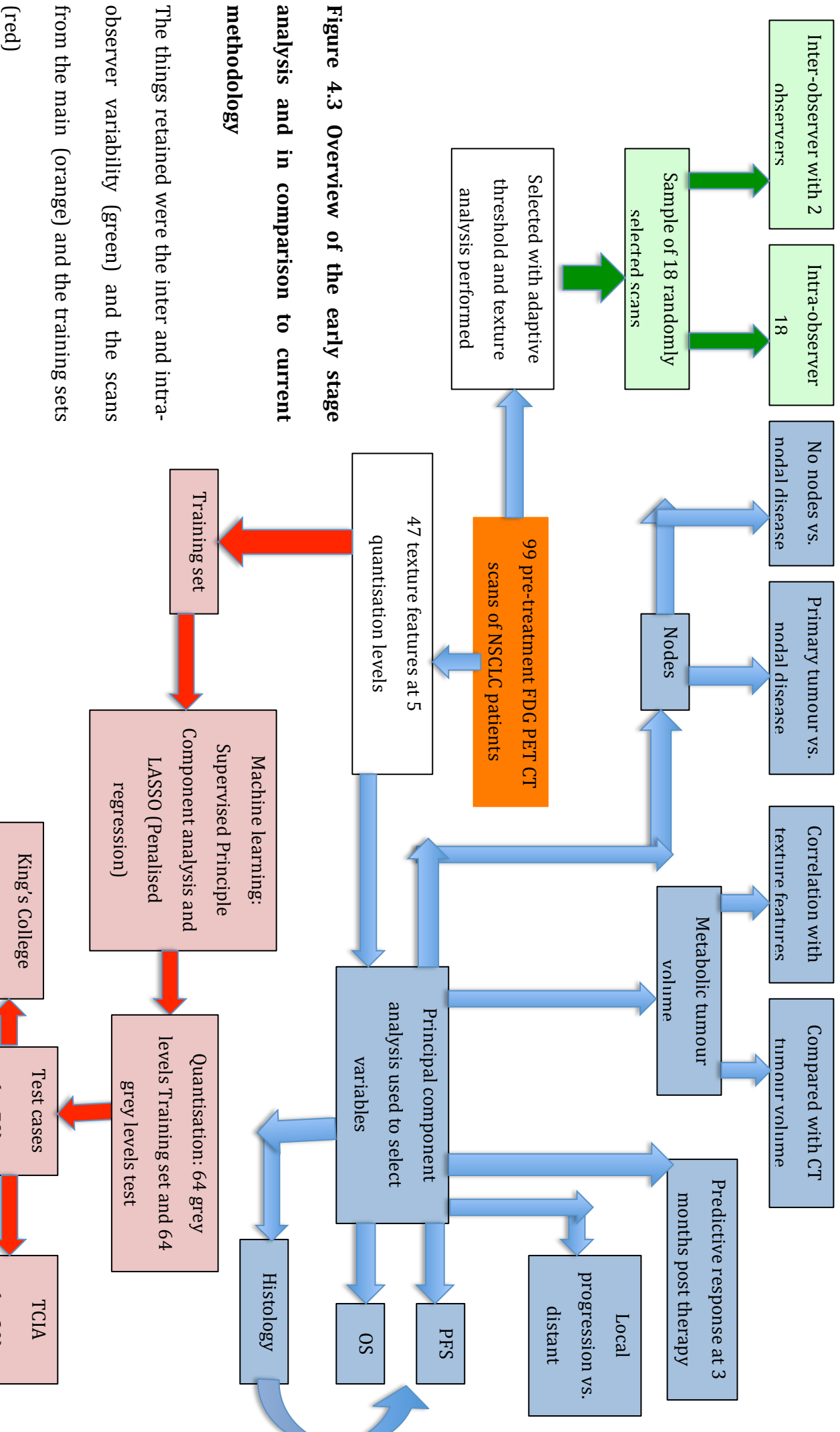
Figure 4.3 Overview of the early stage analysis and in comparison to current methodology

The things retained were the inter and intra-observer variability (green) and the scans from the main (orange) and the training sets (red)

4.3 Ovarian Study Aim of Study

The aim of this study is to investigate if radiomics and CT semantic data correlate with the underlying biology using proteomic and molecular information.

4.3.1 Ovarian masses dataset

4.3.1.1 Ovarian masses Patients

Ovarian cancer affects around 7000 women in the UK alone (2017) and has a poor prognosis as it presents late and is often disseminated at presentation. There is a high relapse rate and a ten-year survival of only 35%. In addition to the late stage at presentation, suboptimal de-bulking and chemotherapy resistance are cited as possible causes for this poor survival (Noer et al., 2017). Although several clinical and biochemical biomarkers (such as CA125, kallikreins, etc) (Sölétormos et al., 2016) (Oikonomopoulou et al., 2008) have been suggested as possible predictive/prognostic markers, none have crossed into routine clinical use in the clinical environment (O'Connor et al., 2016).

4.3.1.2 Outcomes

All the patients included used in this study underwent their surgery at Imperial College Healthcare NHS Trust. The patients had their CT imaging either at their referring centres or at Imperial College Healthcare NHS Trust. Because the scanners used in the study were from different manufacturers and often had different matrices and slice thickness, we investigated the effect of this on the texture parameters.

4.3.1.2.1 Clinical follow-up

Patients were followed up regularly at the end of their treatment with CA-125 and clinical assessment every three months for the first two years and then six monthly thereafter. Imaging in the form of CT/MRI was ordered if there was any abnormality in the above examinations. An isolated increase in the CA-125 was not regarded as a recurrence.

4.3.1.3 Computerised tomography Imaging

4.3.1.3.1 Computerised tomography schedules

The patients had their preoperative scans either at their host institutions or at Imperial College NHS Trust. The patients were excluded if a pre-operative CT scan was not performed prior to the operation. Non-contrast scans were also excluded.

The different centres that were part of the gynaecological network as part of this study were as follows (known as the Hammersmith cohort, HH):

Internal:

Imperial

1. Charing Cross Hospital, Fulham Palace Road, London W6 8RF

2. Hammersmith Hospital, Du Cane Road, Shepherd's Bush, London W12 0HS

3. St Mary's Hospital, Praed Street, London W2 1NY.

External

1. Alliance Medical, 136 Harley St, Marylebone, London W1G 7JZ

2. BMI Chiltern, London Road, Great Missenden HP16 0EN

3. Clementine Churchill, Sudbury Hill, Harrow HA1 3RX

4. Derriford Hospital, Derriford Road, Crownhill, Plymouth PL6 8DH

5. Ealing Hospital, Uxbridge Road, Southall, UB1 3HW

6. Heatherwood Hospital, High Street, Ascot SL5 8AA

7. Hillingdon Hospital, Pield Heath Road, Uxbridge UB8 3NN

8. Lancashire Teaching Hospital, Sharoe Green Lane North, Fulwood, Preston PR2 9HT

9. Lister Hospital, Coreys Mill Lane, Stevenage SG1 4AB

10. Mount Vernon, Paul Strickland Scanner Centre, Rickmansworth Road, Northwood HA6 2RN

11. Northwest London Hospitals: Central Middlesex Hospital, Acton Lane, London NW10 7NS and Northwick Park Hospital, Watford Road, Harrow HA1 3UJ.

12. Royal Shrewsbury, Mytton Oak Road, Shrewsbury, SY3 8XQ.

13. St George's Hospital, Blackshaw Road, London, SW17 0QT

14. University Hospital Wales, Eastern Ave, Cardiff CF14 4XW

15. West Middlesex, Twickenham Road, Isleworth, TW7 6AF

16. Wellington Hospital, 8A Wellington Place, London NW8 9LE

17. Wexham Park Hospital, Wexham, Slough SL2 4HL

18. Velindre Hospital, Velindre Road, Cardiff CF14 2TL.

The inclusion criteria and exclusion criteria were as follows:

- Age range 18-99

- All ovarian histological subtypes including benign, borderline and invasive malignant;

- Staging CT available for review;

- Staging CT with contrast in the portal venous phase

- Primary visualised with no obscuration

- All patient had received surgery

Exclusion:

- Age outside 18-99

- CT not performed.

- Ct available is for recurrent disease

- Follow up data not available

- Not high grade serous pathology

- Received neo-adjuvant therapy prior to surgery

External Set

The same criteria were also applied to the validation set which was completely independent and was obtained via a public database TCGA ovarian (http://www.cancerimagingarchive.net/; last accessed June 2016).

4.3.1.3.2 Final numbers

4.3.1.3.2.1 Training and validation sets

There were three datasets in the final used in the final analysis. The Hammersmith cohort was split into training and validation sets and the TCGA was used as a further validation set.

4.3.1.4 Image acquisition and reconstruction

Different scanners with different reconstruction protocols were used to perform image acquisition of the archived data.

An example of the CT acquisition protocol at St Marys Hospital was as follows:

The CT was acquired using a 256-slice MDCT system (Phillips ICT Brilliance; Phillips Healthcare, Best, Netherlands) with a collimation of 64 x 0.625 mm and a reconstruction thickness of 1mm. A voltage of 120 kV and 180 mAs was used for patients who were normal sized. 90 ml of intravenous contrast containing 400mg l^{-1} Iodine (Iomeron-400; Bracco, Milan, Italy) was administered via a peripheral cannula and triggered to scan with a region of interest over the portal vein (at 70 seconds).

Within Imperial College NHS Healthcare Trust there were at least three different manufacturers of CT scans and various different models. All scans were only

selected if they were acquired with contrast enhancement in the portal venous phase.

The imaging acquisitions in the other institutions were obtained from the DICOM headers from the scans (see the results chapter 5).

4.3.2 Segmentation

The DICOM images were imported into ITK snap Version 3.2, 2015 retrieved from http://www.itksnap.org/pmwiki/pmwiki.php?n=Downloads.SNAP3. (Yushkevich et al., 2006). Two radiologists who had at least four years' experience of tumour delineation manually outlined the primary tumour(s), metastases, and mesenteric fat. Then, an experienced radiologist, with more than fifteen years' worth of gynaecological imaging experience, checked all the segmentations (Figure 4.4).

The entire primary mass was included in the segmentation. If there were bilateral adnexal masses, both were included separately in the final analysis. The entire primary mass including cystic, solid and calcified components were included. Ascites was excluded. The segmentations only included tissue that was considered highly likely to be cancer by the expert reader and doubtful areas were not included.
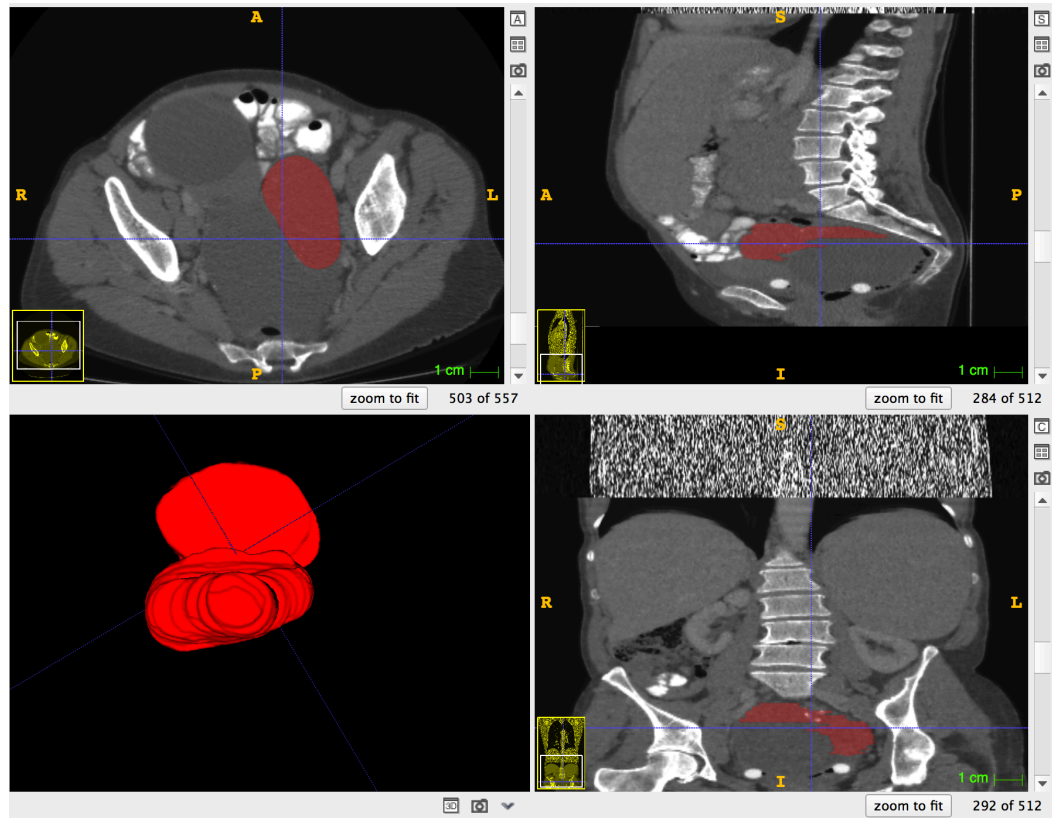
Figure 4.4 An example of segmentation performed on the segmentation software.

The tumour is visualised in three planes and a 3D model is also created.

4.3.3 Methods detailed: inter-observer variability

Two radiologists carried out the initial segmentation of the primary tumours and metastatic lesions. To assess the stability of the segmentation, 30 scans from the TCGA dataset were randomly selected and the primary tumours were re-segmented by both observers, both of whom were blinded to the clinical details and segmentations of the other.

4.3.4 CT Proforma and Peritoneal Cancer index (PCI)

Furthermore, all the CT scans were re-read and a proforma was used with a standardised report to describe the primary tumour and extent of the metastases. (Appendix E1 and E2).

Another section of this included the peritoneal cancer index to delineate the extent of the disease on CT and compare this with surgical findings. This was originally developed in colorectal cancer (Gilly et al., 2006). This compared the spread of peritoneal disease on CT and at surgery and linked it to prognosis. This index was extended to ovarian cancer. Please see Appendix E1 and E2.

Patient demographics, tumour related characteristics and follow up data were obtained from the hospital records. Three radiologists (with 5, 5 and 16 years' of experience in gynaecological imaging tumour delineation) independently reviewed and anonymised preoperative CT scans and reached a consensus on the PCI score being blinded to any clinical or outcome information.

The CT-PCI score was calculated using only the peritoneal disease that was visualised (the primary tumour(s) and lymph node disease were not included). The abdomen was into 13 regions, initially proposed by Sugarbaker (Jacquet and Sugarbaker, 1996) (Appendix E2). Anatomical regions were adapted for CT, using the ribs and iliac crests as anatomical landmarks: above the ribs were sections 1, 2 and 3; between ribs and iliac crests were sections 0, 4 and 8; and

below iliac crests were sections 5, 6 and 7 (see Appendix E2). The small bowel, an area known to be difficult to be certain of involvement, was only described as involved if there was reasonable confidence of involvement. Each section was considered independently of any adjacent section; thus, if the omental involvement straddled between different sections, each section was scored separately in order to reflect the burden of disease. The score for each region was assessed as follows: 0 = no disease; 1 = largest peritoneal nodule <0.5 cm; 2 = largest nodule between 0.5 and 5 cm; 3 = biggest nodule was >5 cm or if confluent disease was present. The total CT-PCI was obtained by totalling up the score of each region.

The correlation between the CT-PCI and surgical outcome was performed using logistic regression and Receiving Operator Characteristic (ROC) curve analysis.

The CT-PCI was correlated with OS and PFS in two different independent analyses: the first considered the absolute value of CT-PCI and the second one considered the regions of involvement, independently by the value of CT-PCI. For the first analysis patients were divided into 5 groups based on the value of the CT-PCI score: 1) CT-PCI=0; 2) CT-PCI between 1 and 5; 3) CT-PCI between 6 and 10; 4) CT-PCI between 11 and 20; 5) CT-PCI more than 20. OS and PFS of these groups were analysed using Kaplan-Meier estimator for univariate analysis and Cox-regression model for multivariable analysis (also considering age, stage, histology and residual disease after surgery). For the second part of the analysis, patients were divided into 4 groups on the basis of the abdominal region involved by cancer: group 1) no detectable disease; group 2) involvement only of

the lower region (region 5, 6 and 7); group 3) involvement of any upper region but excluding the bowel (region 0, 1, 2, 3, 4 and 8); group 4) involvement of the bowel (region 9, 10, 11 and 12).

If the necessary information for statistical analysis were not available, the patient was excluded from that specific analysis.

4.3.5 Genomic and Proteomic Studies

The proteomic (Hennessy et al., 2010) and targeted RNA (Leong et al., 2015) sequencing were performed by another laboratory and are beyond the scope of this thesis to describe the methods implemented.

4.3.6 Implementation

The VOIs defined by the aforementioned manual segmentation were used as an input for the in-house textural analysis software package (TexLAB 2.0), developed on MATLAB 2015b (Mathworks Inc., Natick, Massachusetts, USA). Using the methodology developed by Aerts (Aerts et al., 2014a), the gray-level normalisation led to a single value, which made further statistics easier relative to those of PET.

4.3.7 Unsupervised clustering and texture signature discovery

An analysis was carried out at first using principal component analysis to see if there was any batch effect, that is, variability between the manufacturer, model and slice thickness from the different centres. Then a simple spectral cluster analysis using the radiomic data from those patients with both radiomic and genomic copy-number data available was performed to see if there was any diagnostic or prognostic potential. A hierarchical clustering dendrogram was used to group patients into three clusters; patients in a given cluster tended to share radiomic profiles.

The number of genes affected by the copy-number alteration (CNA) was calculated for each tumour and the distribution of the logarithm of these numbers could be compared from patients from each of these clusters.

A Kaplan Meier curve was drawn to determine progression free and overall survival using the 'survival' package in R. The statistical significance of the survival difference between the three clusters was calculated using the log-rank test in the 'survdiff' function.

Survival difference was also tested between the different slice thicknesses. When the slice thickness demonstrated differences, further comparison was made with known prognostic indicators, such as the presence of ascites (as determined on CT) and the stage as determined at surgery. Two subsequent groups were

determined and these were compared using known clinical markers, namely CA-125 blood tumour marker, and also stage.

Unsupervised hierarchical clustering of the radiomic profiles was performed using 'hclut' and 'cutree' in R.   The raw radiomic data was scaled by mean and centred. Pearson correlation-based distance and complete linkage was used to obtain the final clusters. A repeat of the clustering analysis using Euclidean's diatnace was performed to conform the resulting clusters.

Least absolute shrinkage and selection operator (LASSO) analysis was performed to build a prognostic model for overall survival using the radiomic data. The training set was first selected and a prognostic model of overall survival was generated adjusting for stage, slice thickness and residual disease. It was not possible to decide which side was more related to prognosis in bilateral tumours, and the differences between the two were found to be minimal, both tumours were included in the model-building stage. The radiomic features with a false discovery rate of less than 5% were used as an input for LASSO regression using the glmnet package in R. The 'Cox' model was selected with ten-fold cross validation performed to select the minimum lamba for the minimum cross-validation error using cv.glmnet.

The result was a four-texture feature with coefficient weightings that were used to calculate a predictive Radiomic Prognostic Vector (RPV). Subsequesnt continuous Cox regression and Kaplan-Meyer analysis with PFS and OS were

then calculated. With patients with bilateral tumours, the higher RPV was selected as it performed better than the lower RPV. K_{means} clustering to applied to split the resulting RPVs into three subgroups (Low risk: min-0.0950, Medium risk: 0.0950-0.658, High risk: 0.658-max). The same criteria were applied to the two validation sets. Only cases with complete clinical information (stage, age and post operative residual disease) and slice thickness were included in the multivariable Cox regression analysis.

For generating the gene expression prognostic value, a similar method was used with modifications. From the TCGA study [https://genome-cancer.ucsc.edu/) and TCGA data portal [https://cancergenome.nih.gov/), the gene expression profile from the Affymatrix HT human Genome U133a (Level 2) and Agilent 244K custom gene expression G4502A_07_3 (Level 3) were downloaded. The list of genes that correlated with the RPV (FDR <0.25 for Affymatrix and FDR < 0.1 for Agilent) was generated using Spearman correlation. The subsequent list was used to perform feature selection and linear regression with the RPV.

4.3.8 Statistical analyses

Statistical analysis was performed on SPSS for Windows version 22 (IBM, Chicago, IL, USA) and LASSO penalised regression (using glmnet) statistical language R (version 3.03, Vienna, Austria). Multiple testing was corrected with False Discovery Rate (FDR) method.

Chapter 5-Results NSCLC Study 1: selection of the scans

5.1 Chapter overview

In this chapter, the following are described: first of all, the number of scans selected for each institution and the reasons for the exclusions are given. Secondly, the different thresholds for the segmentations are given. Thirdly, the segmentation method is scrutinised with intra- and inter-observer variability examined.

5.2 Inclusion and exclusion of scans from the different centres

For the separate institutes, exclusions and final numbers (Table 5.1):

1. Imperial

112 patients were initially identified who underwent radiotherapy and had pre-treatment PET scans. 12 were excluded due to the incorrect pathology (small cell cancer). 1 was excluded due to the presence of M (metastatic) disease. 20 patients were excluded because the size of the primary tumour was less than 5mls. 79 were selected.

2. Kings

Of the 53 from this centre, 1 was excluded because no scan was available and 4 had primary tumours that were less than 5mls volume. 48 were elected for the further analysis.

3. Leeds

Of the 105 patients identified from this centre, 13 were excluded because they had metastatic disease and 27 were too small. 2 were excluded because the primary was not visualised. The final number was 63.

4. Marsden

Of the 91 patients, 10 had tumours that were too small, 22 had scans that were not available and 4 where the primary was not visualised. 55 were selected for further analysis.

5. TCGA

From the two subdirectories of LUAD and LUSC, 29 patients were selected in which 1 was excluded due to metastasis and 6 because the primary was less than 5mls. The final number was 22.

6. Mount Vernon

Of 102 patients, 4 were excluded due as not NSCLC, 9 were too small, 5 no tumour was visualised, 11 with no clinical details or histology and 3 had metastases. The final number was 70.

7. Nottingham

Of 43 patients initially selected 15 were excluded, as no scans were available. One was too small and 6 were excluded on the basis of their treatment, which was either palliative or surgical from the outset.

Final numbers

535 patients were originally identified and of these, (with percentages in brackets), 16 were excluded due to the incorrect pathology (9.0%), metastases 18 (10.2%), 77 that were too small (43.5%), scan not available 38 (21.5 %), tumour not visualised 11 (6.2%), no clinical details 11 (6.2 %). 177 were excluded (33.0%) to give the final number of 358 (Figure 5.1). As no clinical details were known at the time of selecting the suitable scans, deliberate bias was not introduced. An overview of this is provided in Figure 5.1. Table 5.2 provides information of the treatment regimes for those that were available and provided by the different centres.

Table 5.1 The reasons for exclusions for scans from different centres

	Imperial	Kings	Leeds	Marsden	TCGA	Nottingham	Mount Vernon	Total
Original number	112	53	105	91	29	43	102	535
Wrong histology	12						4	16
Metastatic	1		13		1		3	18
Too small	20	4	27	10	6	1	9	77
No scan available		1		22		15		38
Primary not visualised			2	4			5	11
No clinical details							11	11
Different treatment						6		6
Total excluded	33	5	42	36	7	22	32	177
Total included	79	48	63	55	22	21	70	358

```
                    ┌─────────────────────────┐
                    │ (1) Proven Lung cancer by│
                    │         biopsy           │
                    └─────────────────────────┘

┌──────────────────┐                          ┌──────────────────┐
│ Incorrect Pathology│                         │   Metastases     │
│                   │                          │                  │
│  N= 16 (9.0%)     │                          │ N= 18 (10.2 %)   │
└──────────────────┘                          └──────────────────┘

┌──────────────────┐                          ┌──────────────────┐
│ PET Scans unavailable│                       │   No clinical    │
│                   │                          │ details/follow-up│
│  N= 23 (21.5%)    │                          │  N=11 (6.2%)     │
└──────────────────┘                          └──────────────────┘

                    ┌─────────────────────────┐
                    │ (2) Selected for segmentation│
                    │    N= 424 (86.2%)        │
                    └─────────────────────────┘

┌──────────────────┐                          ┌──────────────────┐
│    Too small      │                         │   Tumour not     │
│                   │                          │   visualised     │
│  N=77 (43.5%)     │                          │  N=11 (6.2%)     │
└──────────────────┘                          └──────────────────┘

                    ┌─────────────────────────┐
                    │ Selected for texture analysis│
                    │    N= 337 (68.5%)        │
                    └─────────────────────────┘
```
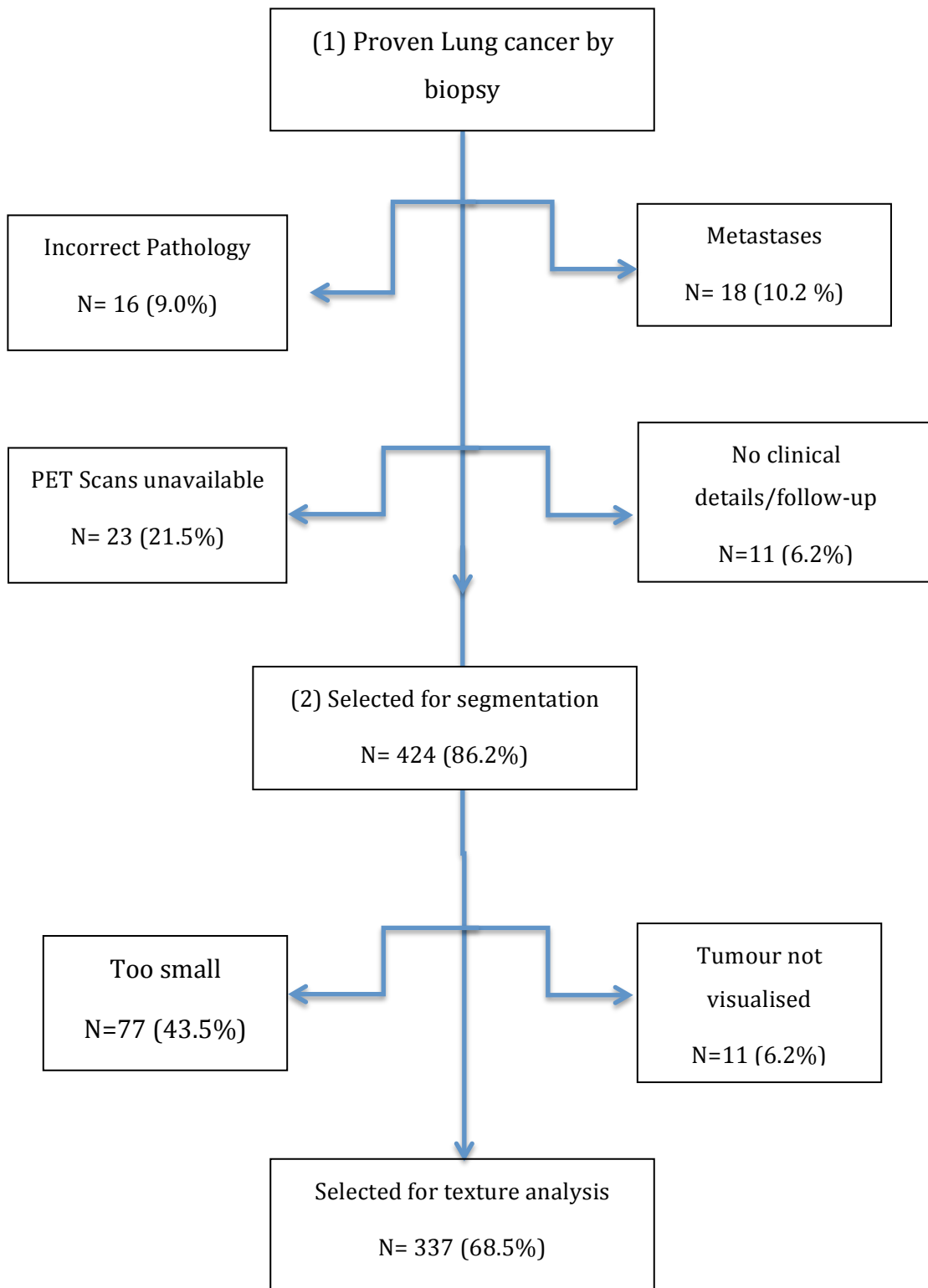
Figure 5.1 Overview demonstrating the selection of the final number of scans

Table 5.2 The treatment for each of the known centres. All patients included in the study received radiotherapy. The information for concurrent chemotherapy if it was given, is only partially available and is described above.

	Imperial	Kings	Marsden	Nottingham
Total number	79	48	55	21
Radiotherapy (number, %)	79 (100)	48 (100)	55 (100)	21 (100)
Median dose (range)	55 (55-66)	64 (55-64)	64 (55-64)	55 (50-66)
Chemotherapy (number, %)	43 (54)	48 (100)	34 (61.8)	Unknown but at least 4
Number of cycles completed	4 (range 1-6)	4 (range, 1-6)	4 (range, 1-6)	Unclear
Unknown had chemo (number, %)	10 (12.7)	0	3 (5.5)	17 (81)

5.3 Segmentation

The segmentation for all the scans were performed by a single observer with 4 years' experience of tumour delineation. The method used was an adaptive threshold method and the different segmentation thresholds are given below (Table 5.3).

Table 5.3 The different thresholds for all the scans used in this study.

The highest threshold was around 40% and the next highest threshold was 30%. Together these accounted for 75.4% of all the thresholds. 9.5% required manual adjustment.

Threshold	All (%)
10%	3 (0.8)
12%	1 (0.3)
15%	18 (5)
19%	1 (0.3)
20%	47 (13.1)
25%	14 (3.9)
30%	99 (27.7)
35%	3 (0.8)
40%	171 (47.8)
50%	1 (0.3)
Manual adjustment made	34 (9.5)
Total	358

5.3.1 Intra and inter-observer variability

There was very high intra- and inter-observer variability in the texture features.

The results are demonstrated in the Table 5.4 and 5.5. The individual breakdown of the texture features is given separately.

Table 5.4. Overview of intra-observer variability.

0.2-0.4 fair agreement, 0.4-0.6 moderate, 0.6-0.8 substantial and >0.8 almost perfect agreement (Matthews, 2002).

Texture features	ICC	95% CI
FOS	0.97	0.94-0.99
GLCM	0.82	0.69-0.90
GLSZM	0.93	0.85-0.97
NGTDM	0.91	0.80-0.96
PET	1	0.99-1
Overall Texture	0.90	0.62-0.93
Texture + PET	0.92	0.85-0.96

Table 5.5. Overview of inter-observer variability.

Texture feature class	Average ICC	95% CI	Obs 1 vs obs 2 P value	Obs 1 vs obs 3 P value	Obs 2 vs obs 3 P value
FOS	0.93	0.84-0.97	0.81	0.80	0.95
GLCM	0.90	0.79-0.96	0.85	0.94	0.79
GLSZM	0.81	0.60-0.93	0.96	0.94	0.99
NGTDM	0.80	0.23-0.87	0.80	0.76	0.76
PET	0.99	0.98-1	0.92	0.41	1
Total Texture features	0.86	0.62-0.93	0.85	0.86	0.87
Total texture features + PET	0.88	0.69-0.94	0.87	0.77	0.90

Using ICC with breakdown between observers with repeated measures ANOVA. Obs = observer. A higher score means a higher correlation. 0.2-0.4 fair agreement, 0.4-0.6 moderate, 0.6-0.8 substantial and >0.8 almost perfect agreement.(Matthews, 2002)

There was an overall intra-observer variability of 0.9 when using texture features and 0.92 when adding PET features. The overall inter-observer variability was 0.86 for the texture features and 0.88 when PET parameters were added. There were some features that were the same for all observers. The FOS

and PET parameters in both intra- and inter-observer groups demonstrated very high correlation.

Thus, in conclusion of this chapter, an outline of the reasons for the final numbers has been given. An overview of the segmentations performed used. The intra- and inter-observer results provide justification for further analyses, as there is high agreement.

5.4 Results II: Comparison between the different centres

5.4.1 Overview

The clinical details of the seven centres are provided (Table 5.1). It can be seen that the age ranges and other clinical variables are broadly similar. After this, some of the clinical data were used in various figures to show how the stage (Section 5.4.2.1), histology (Section 5.4.3) and volume (Section 5.4.4) compare between different centres and with survival.

5.4.2 Comparing the datasets: individual centres

The clinical information from the seven centres is detailed in Table 5.6. It can be seen that the majority of the patients were male. The median age was around the late 60s. The majority of the cases (62.0%) were of Stage 3. This is understandable in terms of how late the cancer usually presents and the fact that all the patients in this study had radiotherapy, which is given to locally advanced tumours. The median overall survival was 22 months and almost 70% of the patients in this study died (Table 5.6).

Table 5.6 The individual clinical details from the different centres.

	Imperial	Kings	Leeds	Marsden	NIH	Mount Vernon	Nottingham	Total
N	79	48	63	55	22	70	21	358
Male (%)	53 (67.1)	29 (60.4)	34 (54)	41 (74.5)	12 (54.5)	39 (55.7)	10 (47.6)	206 (57.5)
Median Age (range)	68 (35-86)	66 (46-88)	70 (48-89)	74 (42-88)	72 (47-83)	73 (48-91)	71 (53-85)	70 (35-91)
Volume in mls, MTV (Range)	27 (5.13-255)	54.6 (10.3-525)	44.7 (5.0-500)	44 (6.02-330)	18.8 (5-83.6)	28.6 (5.3-268)	30.8 (7.03-230)	217.7 (5.0-525)
Stage 1 (%)	11 (13.9)	3 (6.3)	18 (28.6)	5 (9.1)	6 (27.3)	14 (20)	4 (19.0)	61 (17.0)
Stage 2 (%)	17 (21.5)	4 (8.3)	12 (19.0)	15 (27.3)	8 (36.4)	15 (21.4)	4 (19.0)	75 (20.9)
Stage 3 (%)	51 (64.6)	41 (85.4)	31 (49.2)	37 (67.3)	8 (36.4)	41 (58.6)	13 (61.9)	222 (62.0)
Histology: SCC (%)	40 (50.6)	23 (47.9)	24 (38.1)	24 (43.6)	16 (72.7)	37 (52.9)	14 (66.7)	178 (49.7)

Histology: Adeno (%)	27 (34.2)	19 (39.6)	18 (28.6)	25 (45.5)	6 (27.3)	23 (32.9)	5 (23.8)	123 (34.4)
Histology: NSCLC NOS (%)	11 (13.9)	6 (12.5)	12 (19.0)	6 (10.9)	0 (0)	8 (11.4)	2 (9.5)	45 (12.6)
Histology: Other (%)	1 (1.3)	0 (0)	7 (11.1)	2 (3.6)	0 (0)	2 (2.9)	0 (0)	12 (3.4)
Median survival in months (range)	26 (3-85)	22 (2-50)	20 (0-70)	27 (4-65)	9.5 (0-71)	20 (1-49)	20 (1-48)	22 (0-85)
Death (%)	62 (78.5)	33 (68.8)	42 (66.7)	40 (72.7)	10 (54.5)	46 (65.7)	17 (81.)	250 (69.8)
SUV$_{mean}$ (range)	8.11 (2.11-16.3)	9.5 (2.51-17.9)	7.7 (2.82-18.9)	7.13 (1.78-18.0)	6.91 (2.52-13.6)	7.2 (3.11-16.8)	9.55 (2.73-23.7)	8.38 (1.78-23.7)
SUV$_{max}$ (range)	14.6 (3.26-38.2)	19.7 (5.75-46.5)	17.5 (5.99-41.7)	13.7 (4.9-47.2)	12.1 (4.44-22.3)	13.5 (7.43-39.2)	18.7 (5.11-49.5)	16.5 (3.26-49.5)
SUV$_{peak}$ (range)	13.4 (2.9-29.9)	16.8 (4.81-42.3)	14.3 (5.37-39.3)	10.7 (3.8-42.8)	9.98 (3.61-20.6)	11.8 (6.66-34)	15.5 (3.93-43.1)	14.2 (2.9-43.1)
TLG (range)	274 (16.2-	554 (46.2-	367	326	127	275	260	298

	3.91×10^3	5.7×10^3	$(30.9\text{-}4.46 \times 10^3)$	$(29.6\text{-}5.45 \times 10^3)$	$(15\text{-}853)$	$(21.9\text{-}2.59 \times 10^3)$	$(32.3\text{-}4.73 \times 10^3)$	$(15\text{-}5.7 \times 10^3)$

5.4.2.1 Comparing the datasets: Overall Survival

The survival of the different centres has been plotted in a Kaplan-Meier curve (Fig 5.2) to show that generally the survival differences between the different centres were similar. This will be expanded in later in this chapter.
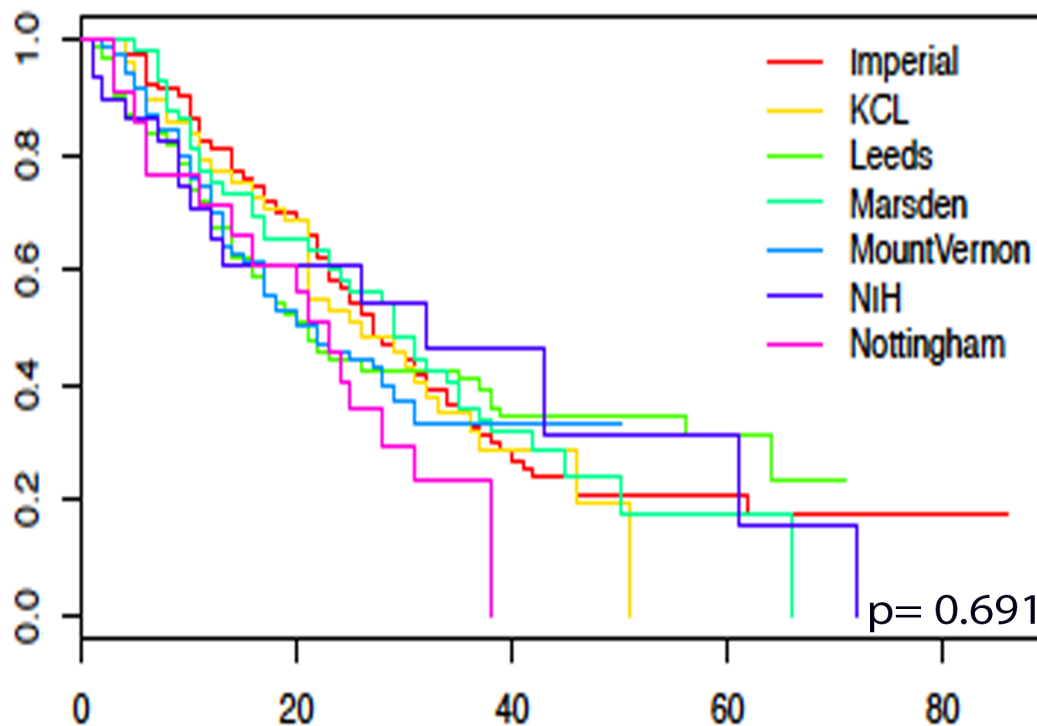
Figure 5.2 KM curve for the survival of all the different centres. There was no significant difference with a p=0.691. KCL=Kings

To further expound on the above figure, the median survival between the different centres is provided below (Table 5.7). Median survival does differ between the centres and so the use of randomly mixing the groups to create a training set would be justified.

Table 5.7 Table to demonstrate the different median survival and the range

Centre	Median Survival	Range (95% CI)
Imperial	27	(23-34)
Kings	26	(21-37)
Leeds	21	(16-39)
Marsden	29	(23-37)
NIH	32	(12-NA)
Mount Vernon	21	(17-31)
Nottingham	23	(14-NA)

NA = Not reached.

5.4.3 Comparing the datasets: Histology

The majority of the subtype of NSCLC in the study was squamous cell carcinoma (Figure 5.3). Some tumours could not be classified further into adenocarcinoma and squamous cell carcinoma and were labeled as *nos.* - not otherwise specified.
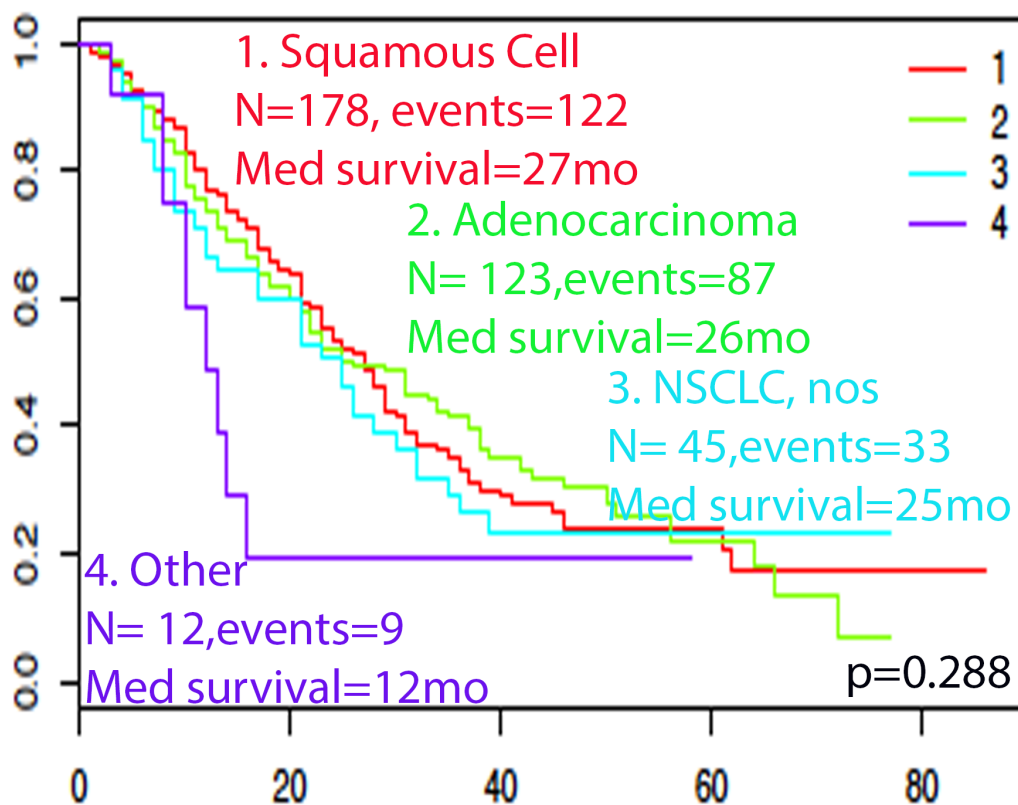
Figure 5.3 Histological subtype and survival.

Histological Subtype: 1= squamous cell, 2 = adenocarcinoma, 3= NSCLC, nos (not otherwise specified), 4 = other or mixed. For the record: group 4. Other comprised of 10 large cell tumours and 2 mixed squamous-adeno carcinomas.

The survival curves from the different subtypes of tumours were broadly similar except for group 4 (Figure 5.3). Overall, there was no significance. However, the

small group 4 was vastly different from the others (although much smaller in number) and another KM curve was done with this group excluded.
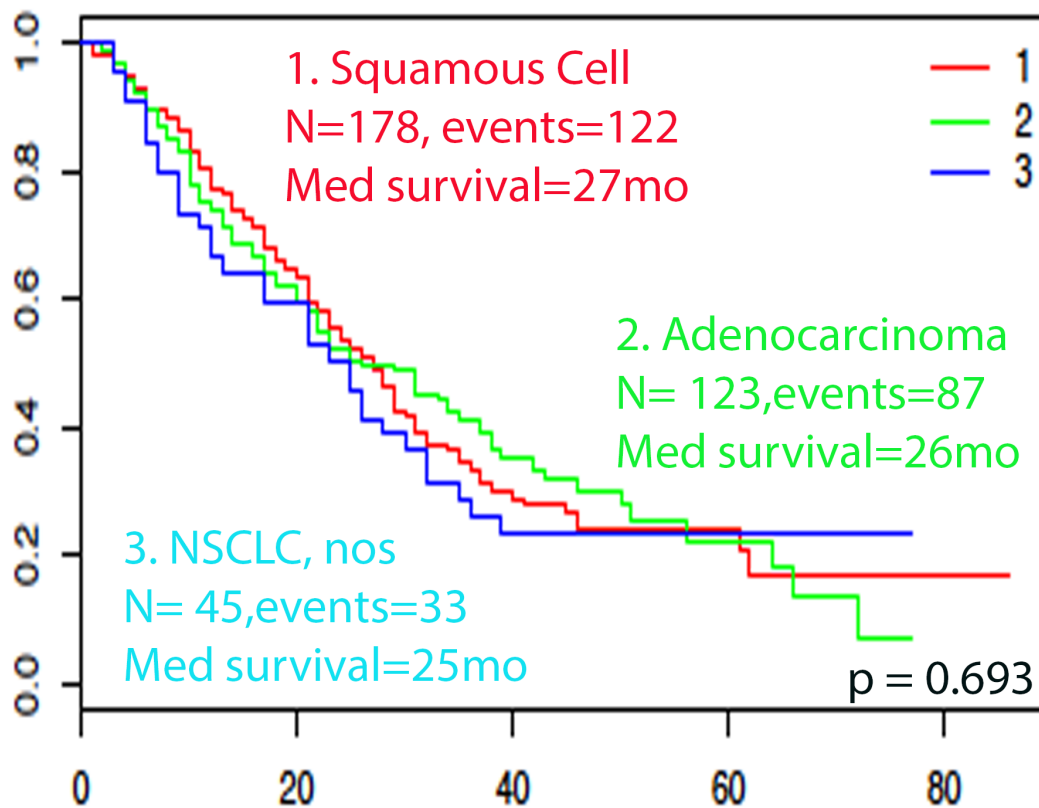
Figure 5.4 Histological subtype and survive with group 4 excluded. Histological Subtype: 1= squamous cell, 2 = adenocarcinoma, 3= NSCLC

In the above-corrected KM curve (Figure 5.4), the removal of group 4, demonstrates that there is no significant difference between histological subtype and survival. This is important in terms of texture, as it will be shown in due course that the texture features and traditional histology examine different variables.

5.4.4 Volume and survival

In the literature, as detailed earlier, there exists a strong correlation between texture features and volume. Volume has been shown to correlate with survival. In this thesis, this phenomena of the correlation between volume and survival has also been demonstrated. Two different KM curves and multivariate analysis have been performed to demonstrate the relationship. The volume in the calculation was taken from the SNS_vol from the segmentations.
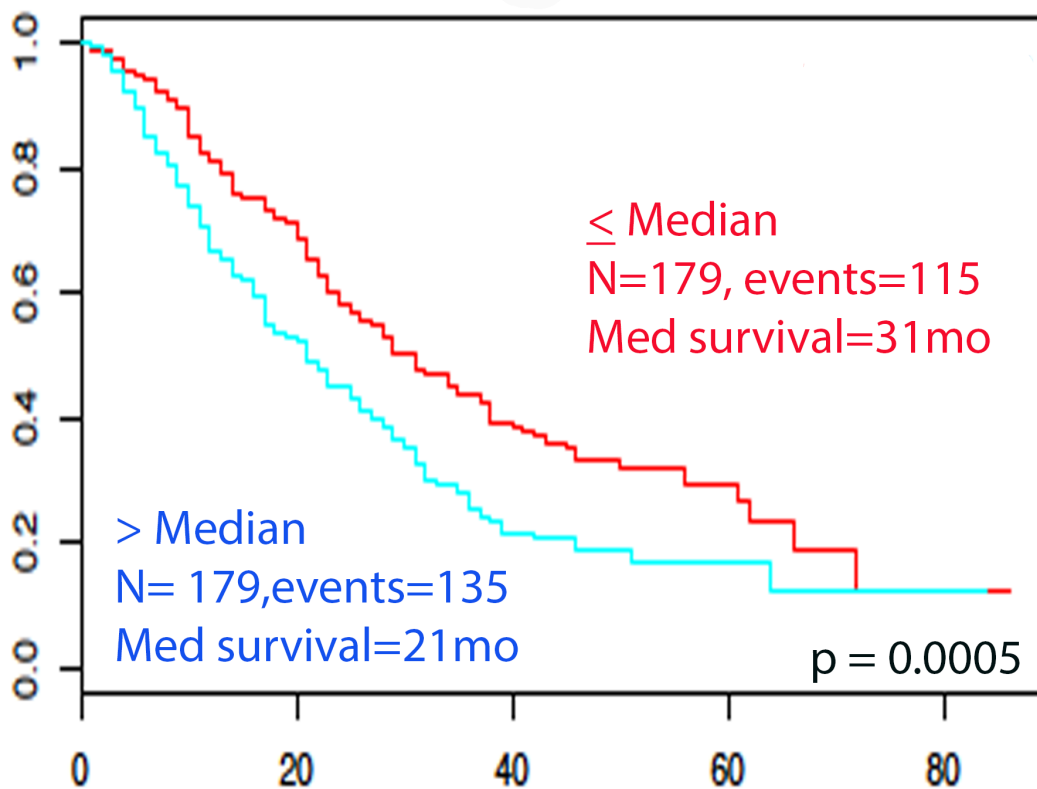
Figure 5.5. KM curve of median volume and survival. This demonstrates significance.
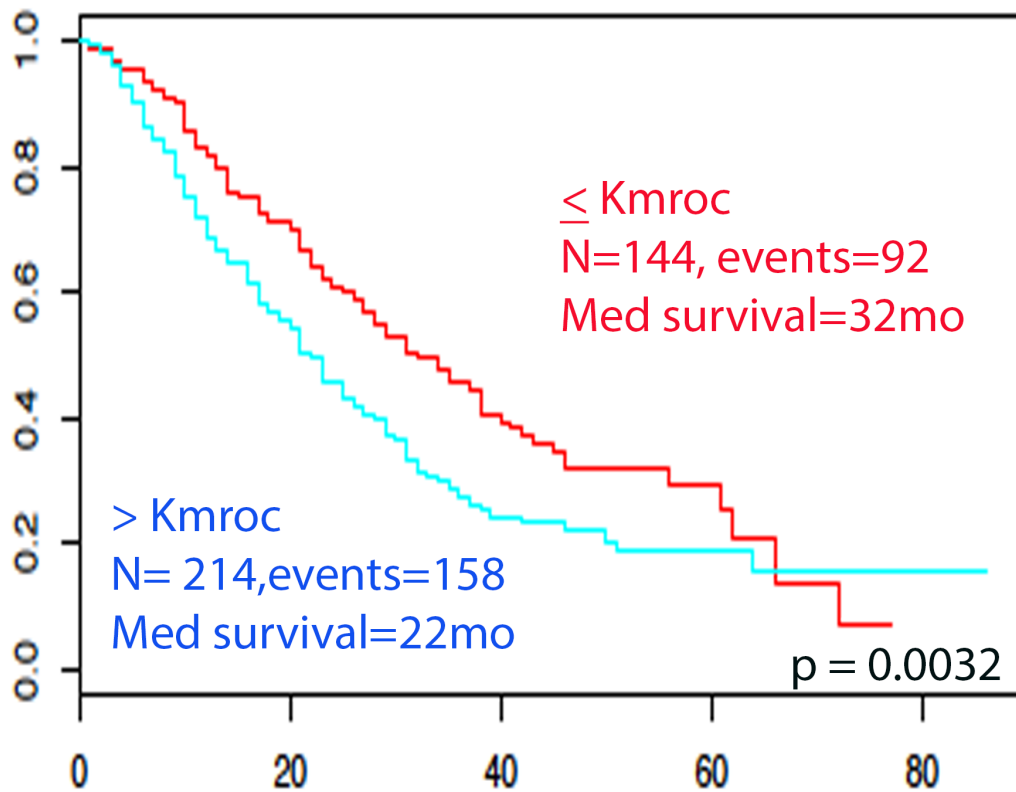
Figure 5.6 KM curve of survival and volume divided by the receiver operator curve (ROC) at 25 months. This again demonstrates significance.

Previous studies have suggested that volume is a major predictor for survival, and this was also borne out in our study. To examine this further, Cox regression analysis was performed with the inclusion of volume and stage (Table 5.8).

Table 5.8 Cox Regression analysis with the inclusion of Pathological Stage and volume.

In this two variable multi-variable analysis, both the stage and volume are significantly correlated with survival. The hazard ratios are given (Exp(coef)). Overall p= 4.16 x 10^{-5}. N = 358, events=250.

	Coef	Exp (coef)	Se (coef)	x	p
> median SNS-vol	0.3407770	1.40604	0.1369289	2.488715	0.0128206
Pathological Stage	0.2456637	1.27847	0.0908019	2.705490	0.0068204

When we explored how many variables highly correlated with volume (taken here as the SNS_vol) we found that 458 of 660 (69%) were strongly correlated with volume (spearman rank >0.7). As we could only inspect linear relations with survival it was thought it would be beneficial to divide though, i.e. normalise, by SNS_vol during feature analysis.

5.5 Summary

This section demonstrates that there are some survival differences between the centres but this is not significant. The histological subtype was not significantly different in terms of survival. The volume - using two different methods to split the data -in a multivariate analysis did appear to contribute to survival. An important step as a result of this is the normalisation of all texture features that correlated with volume.

Section 5.6: Results III Comparing the datasets: Scanners

5.6 Overview

The differences in the clinical factors have been detailed in the previous subsection. In this section, the differences in the manufacturers, models, slice thickness and matrix are demonstrated. The evaluation was performed by using principal component analysis on all the texture features to calculate how much variation the first and second principal components account for.

5.6.2 The scanners from the different centres.

The different centres used different scanners and different settings. Therefore, a degree of heterogeneity exists in the data, however, it was important to see if these differences in scanner type and acquisition contributed significantly to the variation in output. The rationale for this was that these differences could be controlled for in the analysis (Table 5.9). The Principal component analysis of the Manufacturer (Figure 5.7), Model (Figure 5.8), slice thickness (Figure 5.9) and rows (Figure 5.10) for 64 gray levels are presented. Further PCI analyses for all the other gray levels are provided in Appendix C.

.

Table 5.9 of the different manufacturers, models, slice thickness and matrix size from the different centres.

	Imperial (%)	KCL (%)	Leeds (%)	TCGA (%)	Marsden (%)	Nottingham (%)	Mount Vernon (%)	All (%)
Number	79	48	63	22	55	21	70	358
Manufacturer								
CPS				8 (36.4)				8 (2.2)
CTI				1 (4.5)				1 (0.3)
GE medical systems		48 (100)	63 (100)	10 (45.5)			67 (95.7)	188 (52.5)
Phillips Medical Systems					50 (90.1)		1 (1.4)	51 (14.2)
Siemens	79 (100)			3 (13.6)	5 (9.09)	21 (100)	2 (2.9)	110 (30.7)
Model								
CPS 1023				7 (31.8)				7 (2.0)

								Total
CPS 1024				1 (4.5)				1 (0.3)
Siemens 1080				3 (13.6)		21 (100)		24(6.7)
Phillips Allegro Body (C)					44 (80)			44 (12.3)
Siemens Biograph128 mCT					4 (7.3)			4 (1.1)
Siemens Biograph64	79 (100)						1 (1.4)	1 (0.3)
Siemens Biograph64 mCT					3 (5.5)		1 (1.4)	4 (1.1)
GE Discovery ST		24 (50%)	58 (92.1)	5 (22.7)			30 (42.9)	117 (32.7)
GE Discovery STE		24 (50)	5 (7.9)	5 (22.7)			37 (52.9)	116 (32.4)
CTI ESCAT HR+				1 (4.5)				1 (0.3)
Phillips GEMINI TF TOF 16					4 (7.3)			4 (1.1)

Phillips GEMINI TF TOF 64								
							1 (1.4)	1 (0.3)
Matrix								
128 x 128	79 (100)	48 (100)	63 (63)	19 (86.4)			67 (95.7)	276 (77.1)
144 x 144					49 (89)			50 (14.0)
168 x 168			3 (13.6)			21 (100)	25 (7.0)	
169 x 169					1 (1.8)		1 (1.4)	1 (0.3)
200 x 200					5 (9.1)		1 (1.4)	6 (1.7)
Slice thickness								
2				2 (9.1)				2 (0.6)
3					7 (12.7)		1 (1.4)	8 (2.2)
3.27		48 (100)	63 (100)	16 (72.3)			67 (95.7)	194 (54.2)

3.38		2 (9.1)			2 (0.6)
4			48 (87.3)	1 (1.4)	49 (13.7)
5	79 (100)	1 (4.5)		21 (100)	102 (28.5)
5.15		1 (4.5)			1 (0.3)
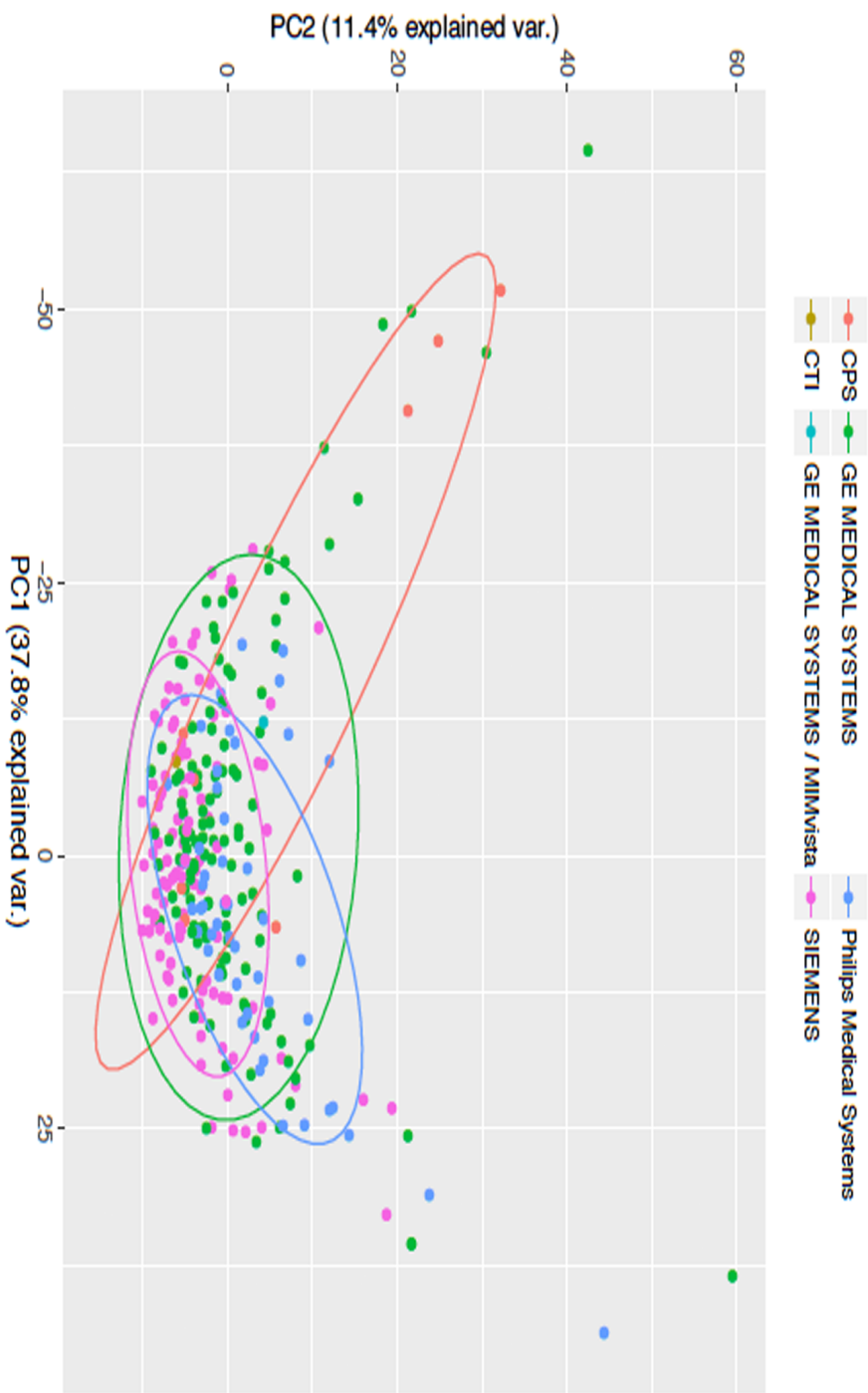
Figure 5.7 Principal Component Analysis demonstrating the variability of texture features across manufacturer type. The majority of the variation is very tightly clustered despite the different manufacturers used.
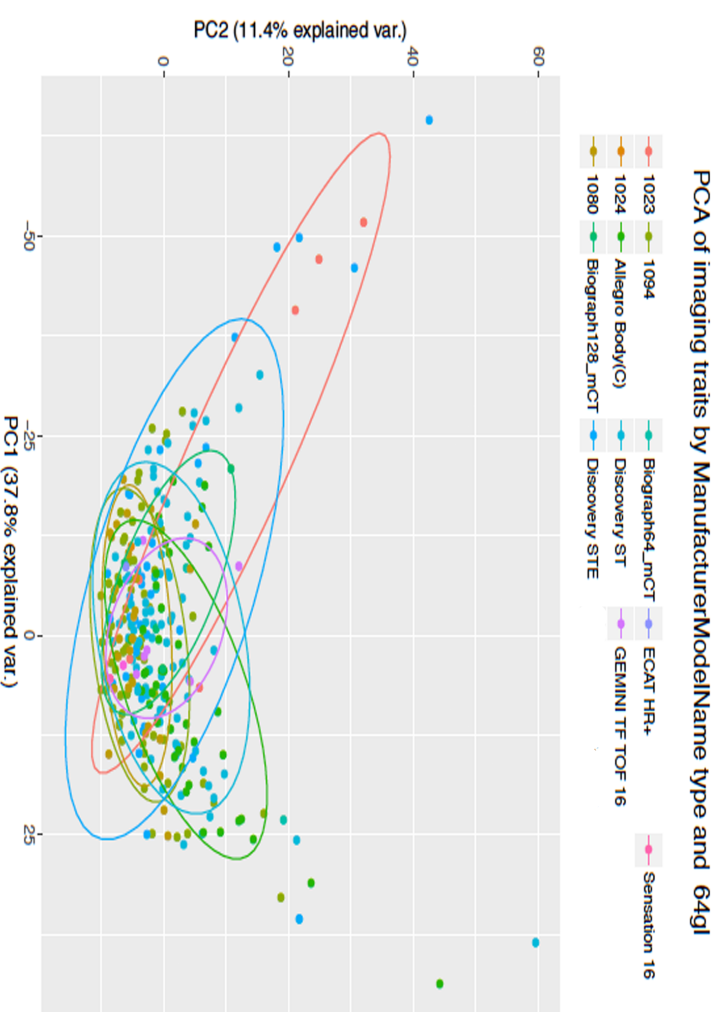
Figure 5.8 Principal Component Analysis demonstrating the variability of texture features across model type. Notice the tight clustering of the groups. CPS 1023, CPS 1024, Siemens 1080, Phillips Allegro Body (C), Siemens Biograph128 mCT, Siemens Biograph64, Siemens Biograph64 mCT, GE Discovery ST, GE Discovery STE, CTI ESCAT HR+, Phillips GEMINI TF TOF 16, Phillips GEMINI TF TOF 64
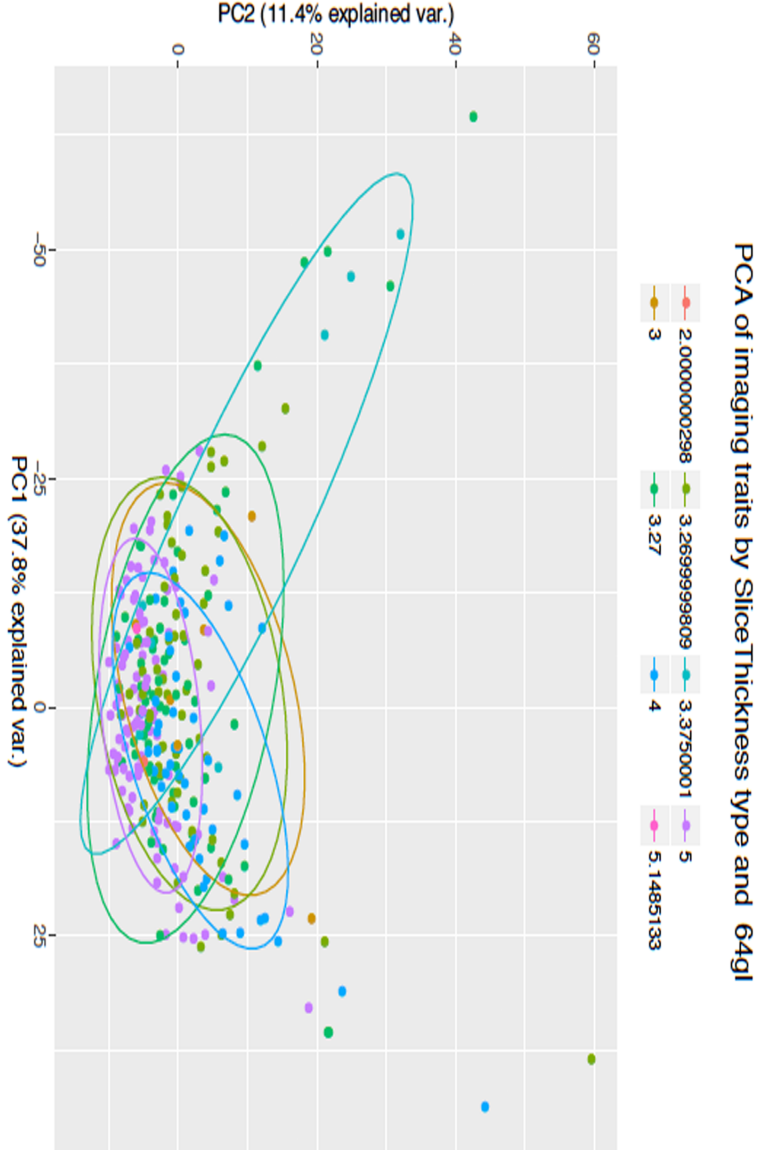
Figure 5.9 Principal Component Analysis demonstrating the variability of texture features across slice thickness. The majority of the points are clustered together. Numbers denote slice thickness in mm.
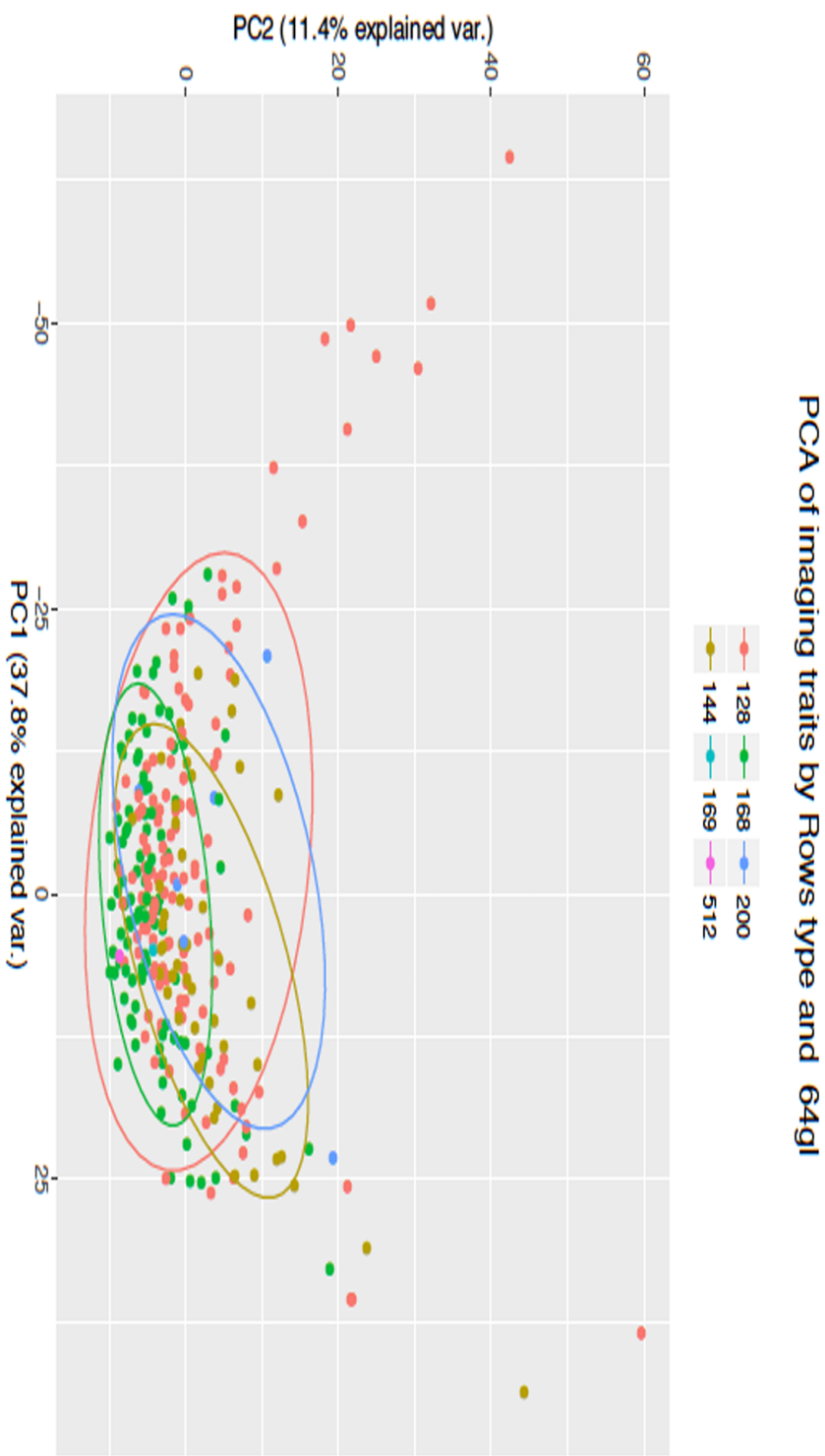
Figure 5.10 Principal Component Analysis demonstrating the variability of texture features across rows type.

The matrix is a completely symmetrical, thus the column PCA appears exactly the same. Numbers denote number of rows: 128, 144, 168, 169, 200 and 512.

The figures demonstrate that the points of variation for manufacturer, model type, slice thickness and matrix size (taken here as the row type), are minimal. In other words, the above factors contribute minimally to the variation demonstrated and thus the variation that we observe is likely attributed to intrinsic properties of the tumour.

It is important to note that the figures given above have been detailed for 64 gray-levels. The other PCI charts relating to other gray-levels are demonstrated in the supplementary material (Appendix C).

Thus in conclusion, sections 5.5 and 5.6 have shown that there are minimal differences between the different centres in terms of different scanners. The volume factor has been accounted for in further calculations.

Chapter 5.7: Results IV-PET and texture analysis for Lung cancer.

5.7.1 Overview

This chapter is broadly divided into three sections. First, the training, validation and TEST1 clinical differences were demonstrated (Table 5.10). Secondly, using the three datasets, stage and PET features were tested to see if there was a possibility to predict prognosis. Thirdly, two separate tests were used to test the texture features.

5.7.2 Comparing the datasets: Training, testing and validation

As has been described in chapter 2, the data were divided into three sets (Table 5.10). The training set was made of 133 patients and the validation set was completely independent of the training set in terms of not being used to develop the model. A separate small set was also used to act as a test set.

Table 5.10 The differences between the training, validation and test sets.

	Training set	Validation Set	Test I
Number	133	204	21
Mean age (Range) years	69 (35-89)	71 (42-91)	71 (53-101)
Male (%)	82 (61.7)	126 (61.7)	10 (47.6)
Stage I (%)	24 (18)	33 (16.2)	4 (19)
Stage II (%)	34 (25.6)	37 (18.1)	4 (19)

Stage III (%)	75 (56.4)	134 (65.7)	13 (61.9)
Histology: SCC (%)	69 (51.9)	95 (46.7)	14 (66.7)
Histology: Adeno(%)	41 (30.8)	77 (37.7)	5 (23.8)
Histology: NSCLC *NOS* (%)	18 (13.5)	25 (12.3)	2 (9.5)
Histology: Other (%)	5 (3.8)	7 (3.4)	0
SUV_{mean} (range)	8.25 (1.78-17.4)	8.44 (2.11-23.7)	7.75 (4.44-16.8)
SUV_{max} (range)	16.5 (4.9-42.8)	15.9 (3.26-49.5)	13.6 (6.66-39.2)
SUV_{peak} (range)	14.2 (3.8-35.4)	14.2 (2.9-43.1)	12.5 (6.26-34)
MTV (range) mls	40.4 (5.13-467)	33.7 (5.27-525)	30.8 (7.03-230)
TLG (range)	344 (16.2–5.45x10^3)	315.2 (19.4–5.7x10^3)	266 (40.5-2.59 x10^3)
Median overall survival (Months)	25 (0-83)	21.0 (0-85)	20 (2-37)
Number of deaths (%)	88 (66.2)	145 (71.1)	17 (81%)
Length of Follow up (median + IQR in months)	26 (12-39)	22.0 (11-36)	21 (8-31)

SCC Squamous Cell Carcinoma, *Adeno* Adenocarcinoma, *NSCLC* Non-small cell lung cancer, *MTV* Mean tumour volume, *TLG* Total lesion glycolysis, *IQR* Interquartile Range

5.7.3 Comparing the datasets: Training, testing and validation: Stage

Stage as mentioned in the earlier chapters is a known prognostic marker. The KM curves below were divided by stage (Figure 5.11).

c)
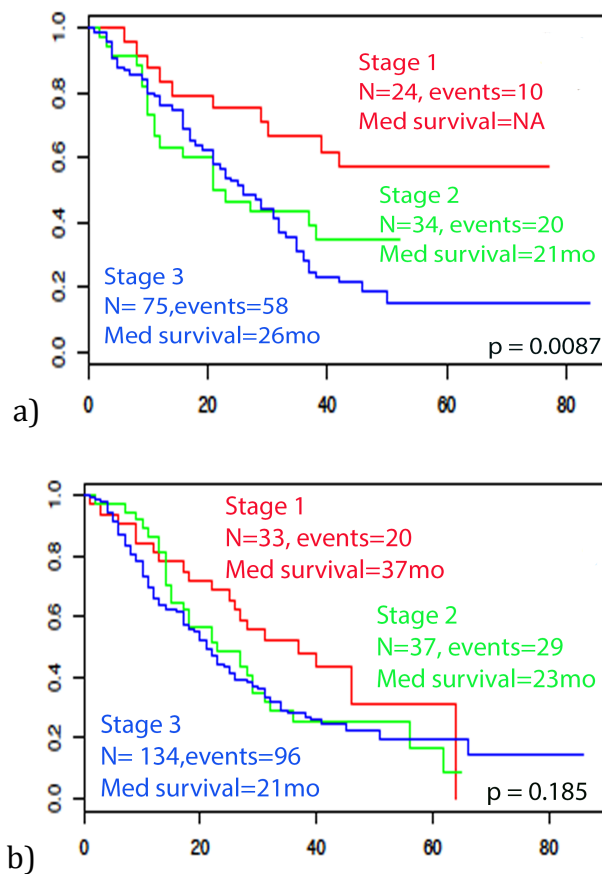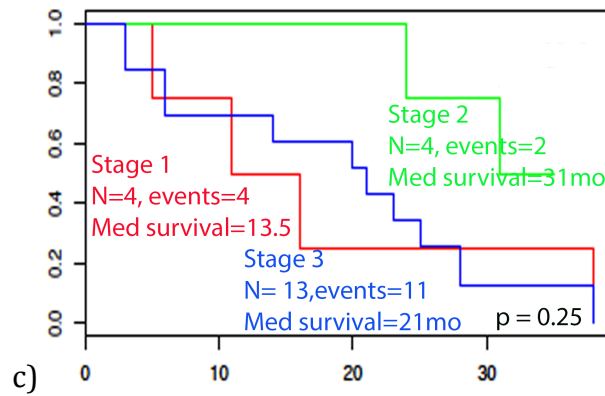
Figure 5.11 KM curves for stage for a) training, b) validation and c) TEST1.

Although the training set is significant, the validation set and TEST1 sets are not.

5.7.4.1 Comparing the datasets: Training, validation and test SUV$_{max}$

The SUV$_{max}$ is not used clinically for prognosis. This marker (Figure 5.12), as well as other PET markers, SUV$_{mean}$ (Figure 5.13), SUV$_{peak}$ (Figure 5.14) and TLG (Figure 5.15) were used to demonstrate current broad clinically available endpoints, as this variable is readily calculated in the clinical setting. In order to generate Kaplan Meier curves a dichotomous end-point is required, even in the case of continuous data. The cut-off for all the PET parameters is the maximum cut-off from the ROC curve (Youden's J) at 25 months survival (Babyak, 2004).

a)

b)

c)
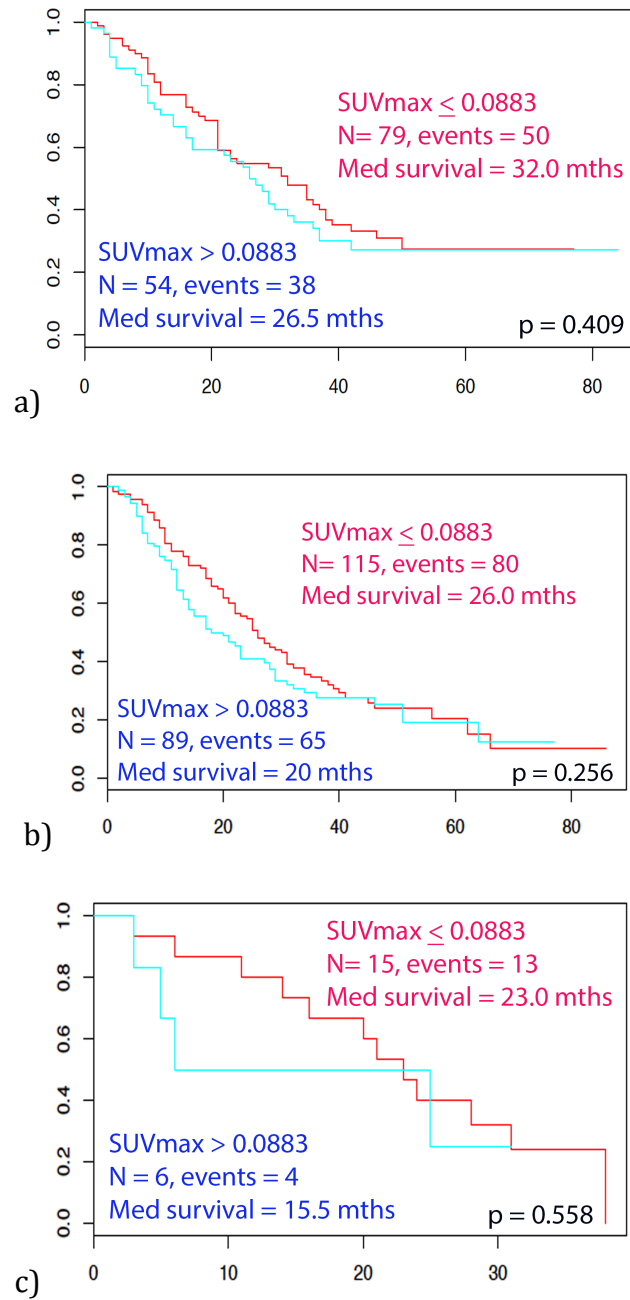
Figure 5.12 KM curves for SUV$_{max}$ for a) training, b) validation and c) TEST1. No significance demonstrated.

5.7.4.2 Comparing the datasets: Training, validation and test SUVmean

Another PET marker the SUV_{mean} has been utilized.

SUVmean > -0.337
N = 60, events = 35
Med survival = 36 mths

SUVmean ≤ -0.0337
N= 73, events = 53
Med survival = 23 mths

p = 0.033

a)

SUVmean ≤ -0.0337
N= 100, events = 70
Med survival = 25.0 mths

SUVmean > -0.337
N = 104, events = 75
Med survival = 22 mths

p = 0.648

b)

SUVmean ≤ -0.0337
N= 14, events = 12
Med survival = 22.5 mths

SUVmean > -0.337
N = 7, events = 5
Med survival = 23 mths

p = 0.688

c)
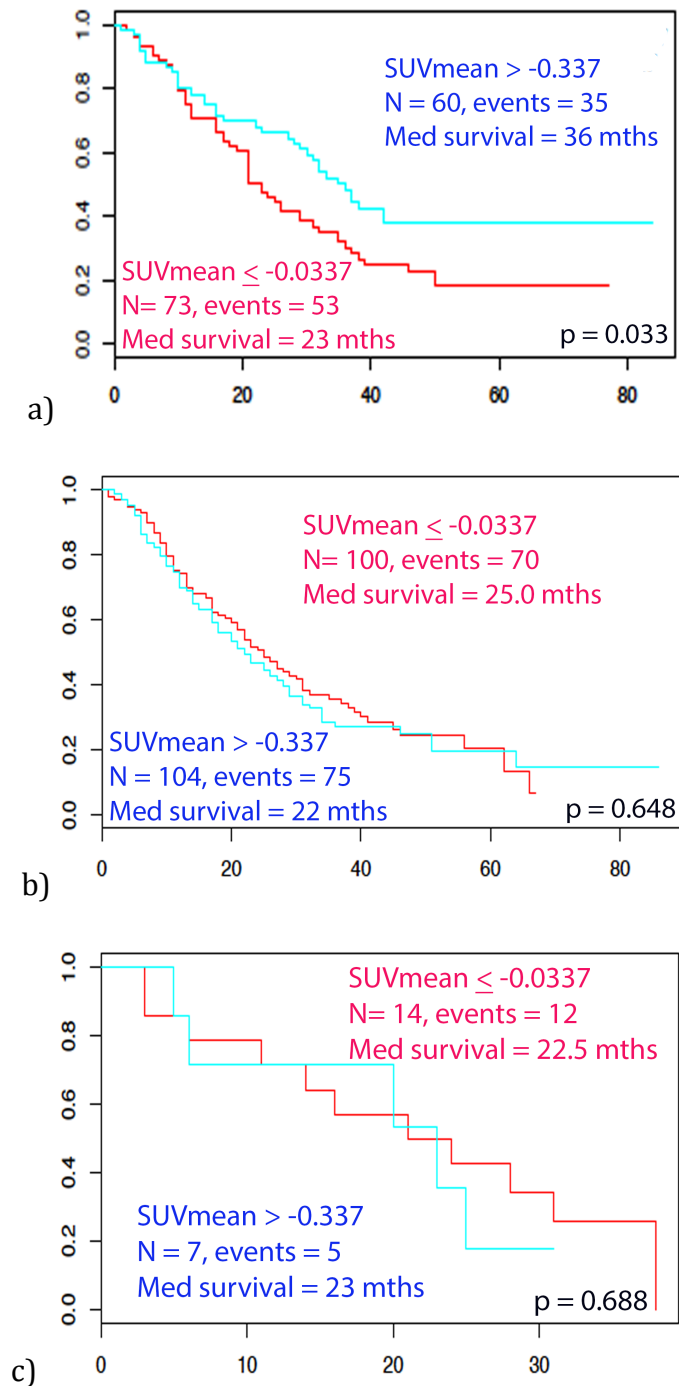
Figure 5.13 KM curves for SUV$_{mean}$ for a) training, b) validation and c) TEST1. A significant outcome in the training set was observed but in the same was not true for the other sets.

5.7.4.3 Comparing the datasets: Training, validation and test SUV$_{peak}$

The SUV$_{peak}$ was also tested.

a)

b)

c)
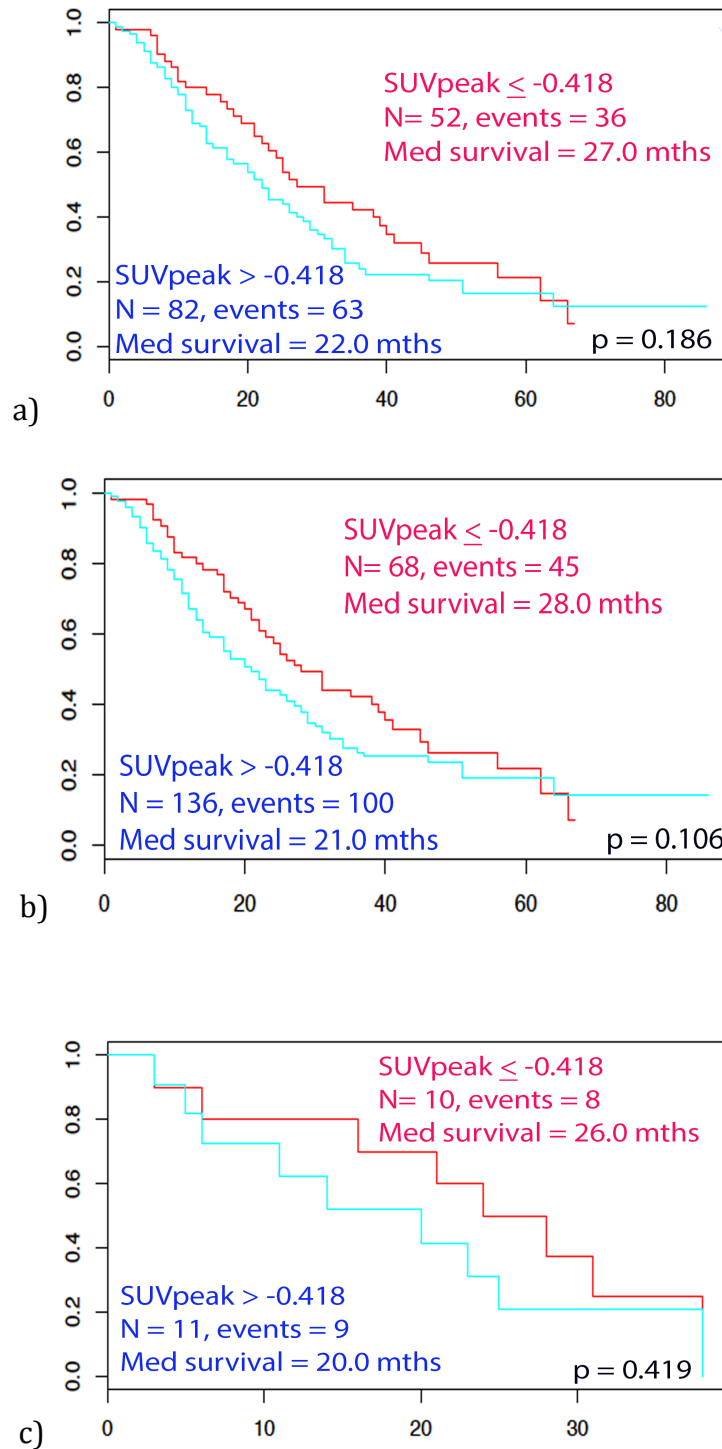
Figure 5.14 KM curves for SUV$_{peak}$ for a) training, b) validation and c)

TEST1. There were no significant results.

5.7.4.4 Comparing the datasets: Training, validation and test TLG

a)

TLG ≤ -0.344
N = 62, events = 34
Median survival = 38 months

TLG > -0.344
N = 71, events = 54
Median survival = 25 months   p = 0.00762

b)

TLG ≤ -0.344
N = 97, events = 61
Median survival = 31 months

TLG > -0.344
N = 107, events = 84
Median survival = 17 months   p = 0.00176

c)

TLG > -0.344
N = 9, events = 6
Median survival = 31 months

TLG ≤ -0.344
N = 12, events = 11
Median survival = 18.5 months   p = 0.217
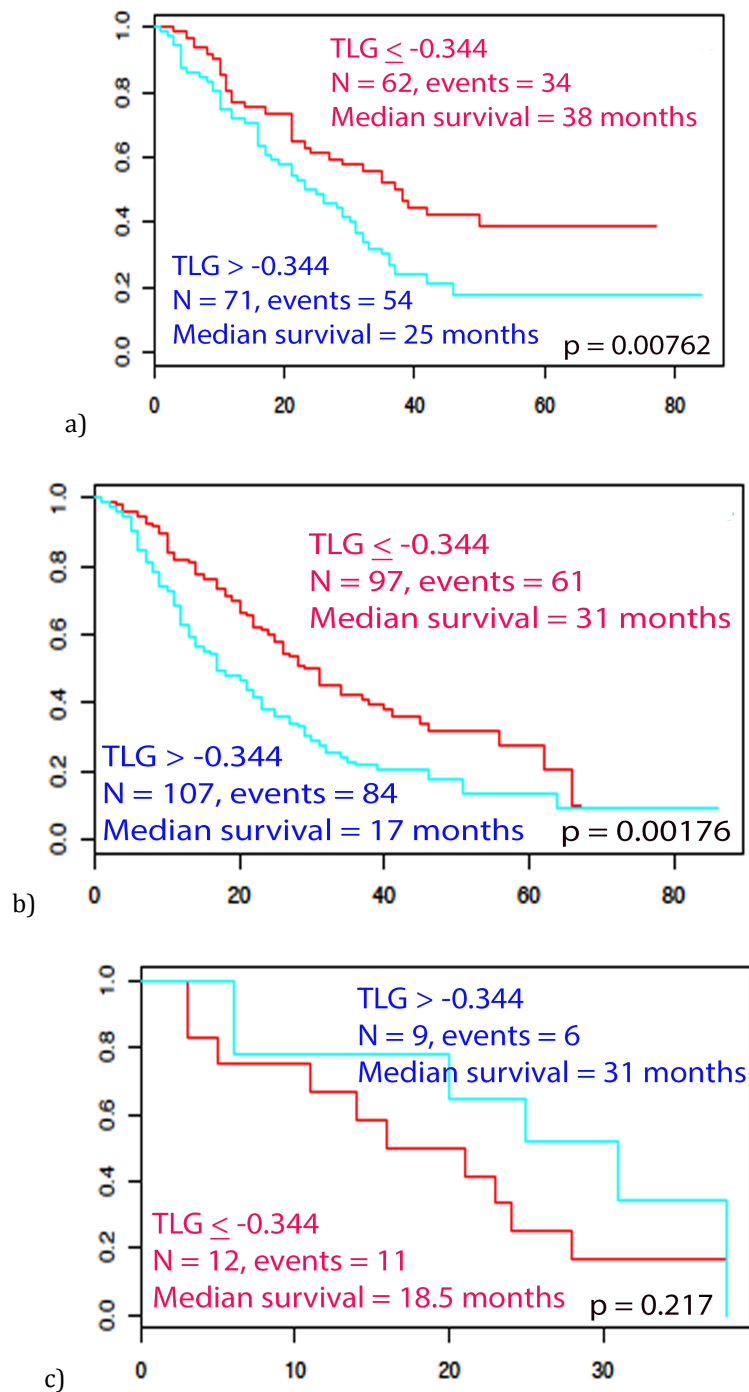
Figure 5.15 KM curves for TLG for a) training, b) validation and c) TEST1.

There were significant results in the training and validation sets.

5.8 Texture analysis: Initial analysis with individual FOS and GLCM features

A substantial amount of early work was presented in the literature (see Chapter 2: literature review) on individual texture features, some of which were found to be significant. We tried to replicate some of this work by testing two sets of texture features – FOS and GLCM-which are the two widely tested texture features in the literature. The results are detailed below (Table 5.11).

Table 5.11 Using FOS and GLCM texture features in the training, validation and test sets.

Texture feature	cut	Training	Validation	Test1
FOS_CV	0.0647	0.374	0.97	0.348
FOS_lmean	0.915	0.224	0.91	0.744
FOS_lmedian	0.854	0.0338	0.792	0.744
FOS_lmode	-.0801	0.122	0.102	0.611
FOS_lstd	-0.0237	0.0393	0.429	0.35
FOS_lmin	0.17	0.0111	0.575	0.834
FOS_lmax	0.0601	0.0213	0.132	0.284
FOS_Range	0.83	0.016	0.143	0.577
FOS_Skew	-0.244	0.269	0.84	0.204
FOS_Kurt	0.11	0.101	0.931	0.178
FOS_lmeanAbsDev	-0.0667	0.0262	0.429	0.35
FOS_RMS	0.865	0.0406	0.998	0.744
FOS_Entr_64gl	0.202	0.876	0.0929	0.312

FOS_Ener_64gl	-0.0789	0.136	0.437	0.436
GLCM_Varian_64gl	-0.0378	0.006	0.777	0.181
GLCM_Correl-64gl	-0.278	0.0216	0.328	0.22
GLCM_InfCo1_64gl	0.052	0.00835	0.0189	0.16
GLCM_InfCo2_64gl	0.299	0.00915	0.0144	0.0584
GLCM_ClShad_64gl	-0.273	0.0357	0.827	0.0789
GLCM_ClProm_64gl	0.324	0.00363	0.88	0.255
GLCM_Angsmo_64gl	-0.509	0.000258	0.0111	0.0287
GLCM_MxProb_64gl	-0.548	0.00241	0.0025	0.0287
GLCM_Entrop_64gl	0.392	0.0141	0.000296	0.0519
GLCM_Contra_64gl	-0.434	0.0197	0.807	0.402
GLCM_Dissim_64gl	-0.289	0.0226	0.743	0.402
GLCM_Homoge_64gl	-0.555	0.00689	0.0775	0.604
GLCM_sumAvg_64gl	0.00589	0.0712	0.805	0.998
GLCM_sumVar_64gl	0.288	0.00157	0.62	0.0547
GLCM_sumEnt_64gl	0.656	0.0198	2.13×10^{-5}	0.0752
GLCM_difVar_64gl	-0.28	0.0209	0.885	0.611
GLCM_difEnt_64gl	0.00977	0.00579	0.608	0.308
GLCM_AutoCorrel_64gl	-0.0426	0.0973	0.686	0.889
GLCM_ClTend_64gl	0.288	0.00157	0.62	0.0547
GLCM_Homoge1_64gl	-0.524	0.00681	0.11	0.709
GLCM_IDMN_64gl	0.4	0.00335	0.757	0.402
GLCM_IDN_64gl	0.141	0.00859	0.969	0.403
GLCM_invVar_64gl	0.0129	0.00496	0.718	0.217

FOS First Order Statistics, GLCM Gray Level Co-occurrence Matrix. For a full breakdown of the abbreviations please see Appendix B and chapter 2.

As there was multiple testing the adjusted p value (Bonferroni correction) was 0.0013514, of which none of the factors in the testing was significant. A composite feature of the 4 best features were compared:

GLCM_Angsmo_64gl

GLCM_sumVar_64gl

GLCM_invVar_64gl

GLCM_difEnt_64gl

Using this composite feature for analyzing the training set, the subsequent KM curve demonstrated p=0.000449 (not shown). Using the Kmroc cutoff, this was significant on the validation set (0.0394, again not shown), but when placed in a multivariate analysis (Cox Regression) including Stage, the composite feature was not significant.

5.8.1 Texture analysis: Using LASSO

As has been alluded previously, there is a strong correlation of the texture features with each other. The heatmap below (Figure 5.16) demonstrates this. All the areas that are coloured green or red have strong correlations with each other (Figure 5.16).
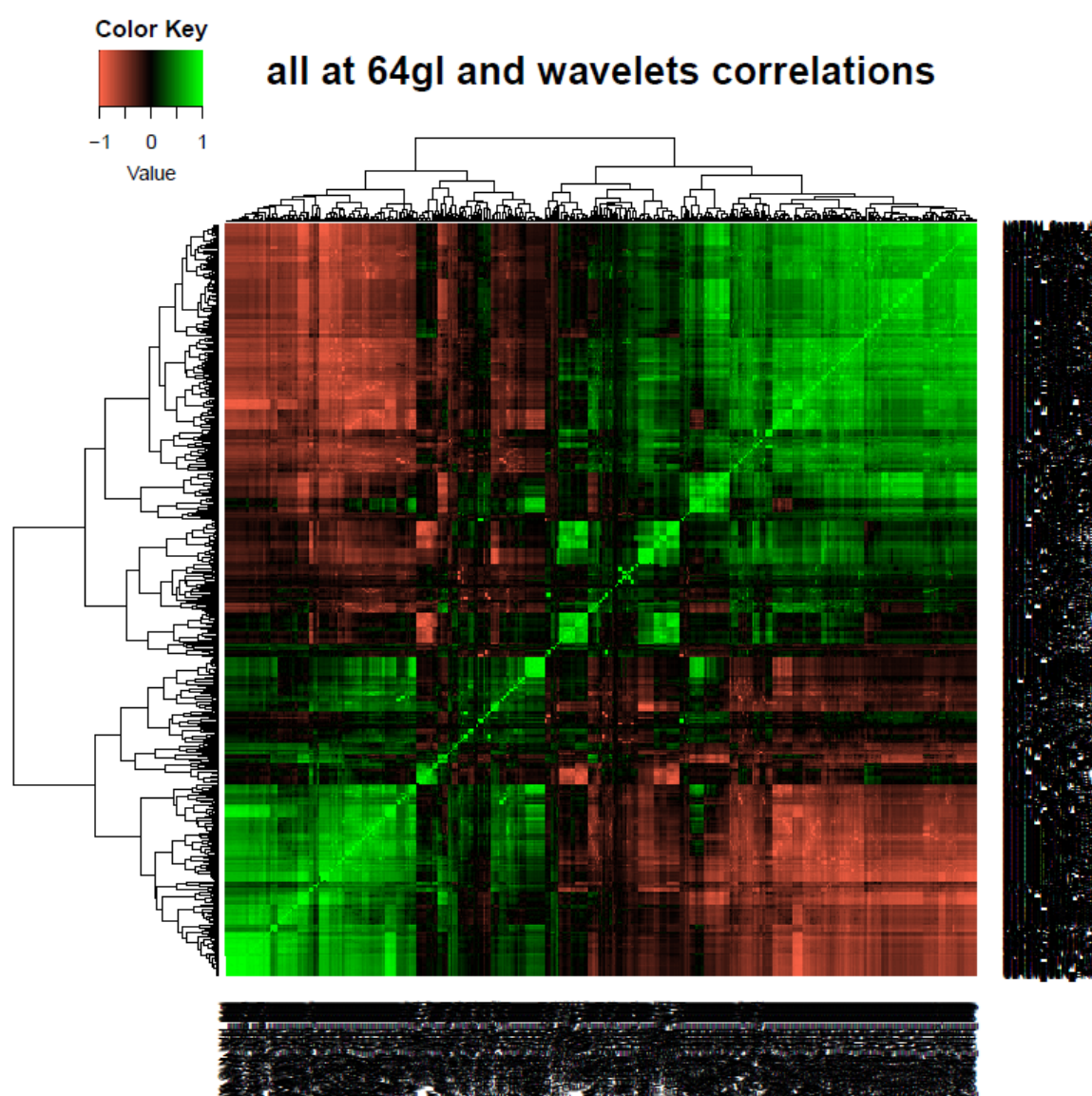
Figure 5.16. Spearman rank correlation of the radiomics features displayed as a heatmap. High levels of correlation with clustering of features are seen.

One aspect that needs to be stated is that although machine-learning methods have been developed with binomial outcomes, the methods to adapt them for continuous survival outcomes and also account for censored data have yet to be decided. In our study, two methods were used to split the data into two groups, one of which was the median survival and the other was the KM ROC method linked to survival.

In this thesis, a form of penalized regression, named LASSO has been implemented (Figure 5.17).

a)

b)
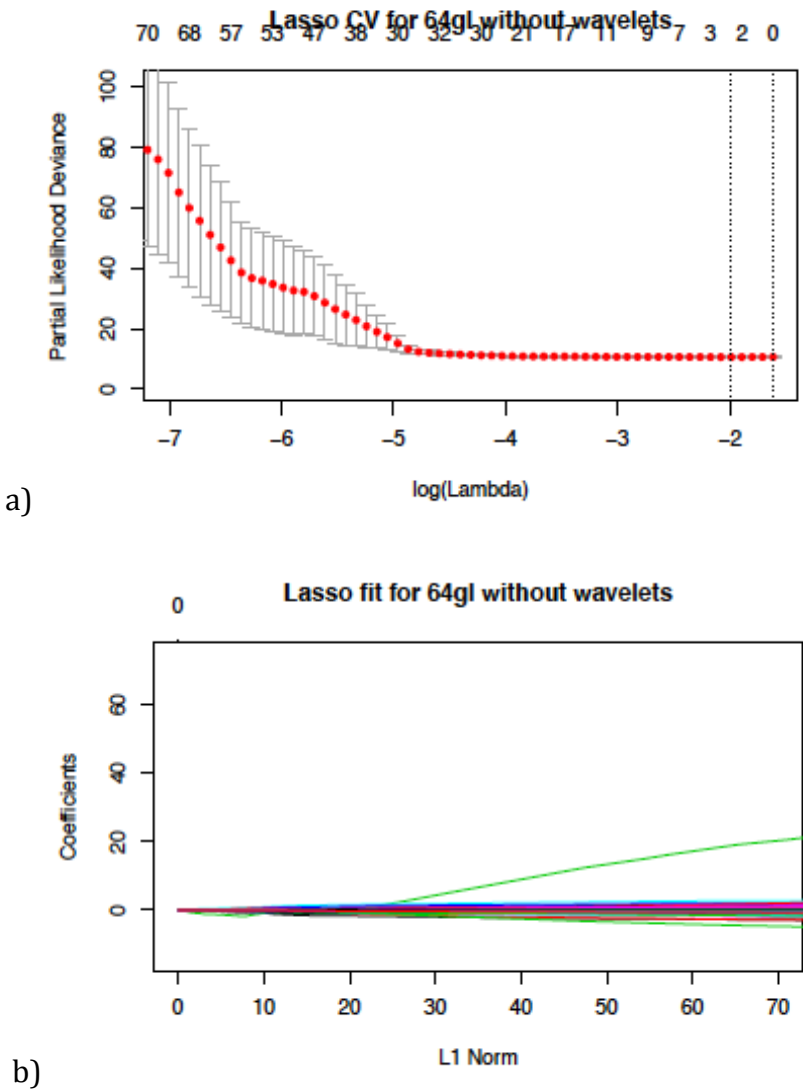
Figure 5.17 Radiomics feature selection using the binary logistic regression model, LASSO.

(a) The area under the receiver operating curve (AUC) was plotted against the logarithm of the tuning parameter (λ) by determining the minima from cross validation. The optimal values representing x1 SE of the minima are plotted as vertical lines. (b) A LASSO coefficient profile plot showing LASSO coefficients plotted against Normalised values. The vertical line represents the optimal number of non-zero coefficients obtained through cross-validation.
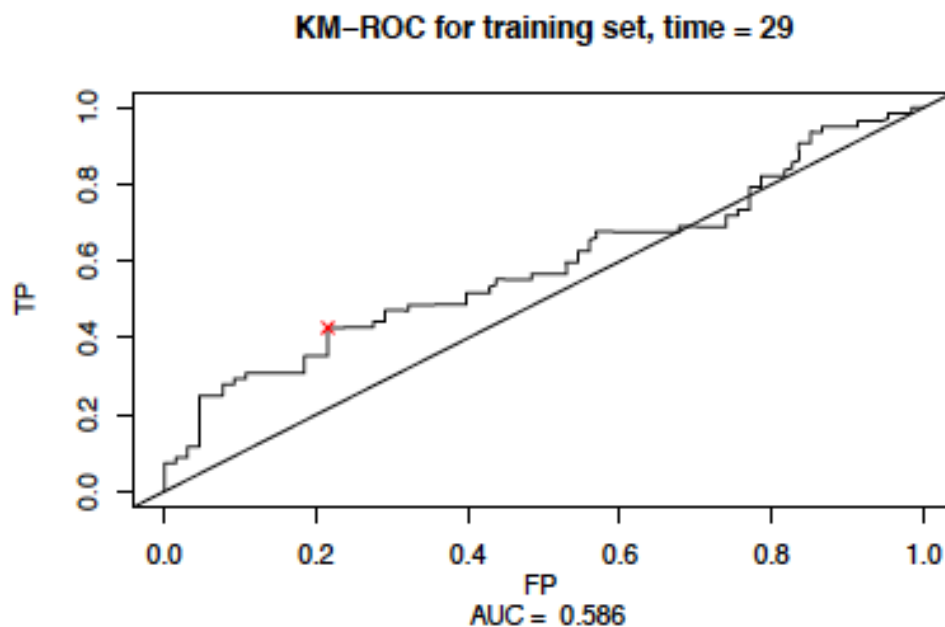
Figure 5.18 Using Youden's J method to determine the optimal cutoff from the ROC curve at optimal survival at 29 months. The cutoff was used to split the subsequent data.

The LASSO method helped develop a two feature composite, termed FVX (optimal feature vector), which was scrutinized on the three different sets (Figures 5.19 and 5.20).
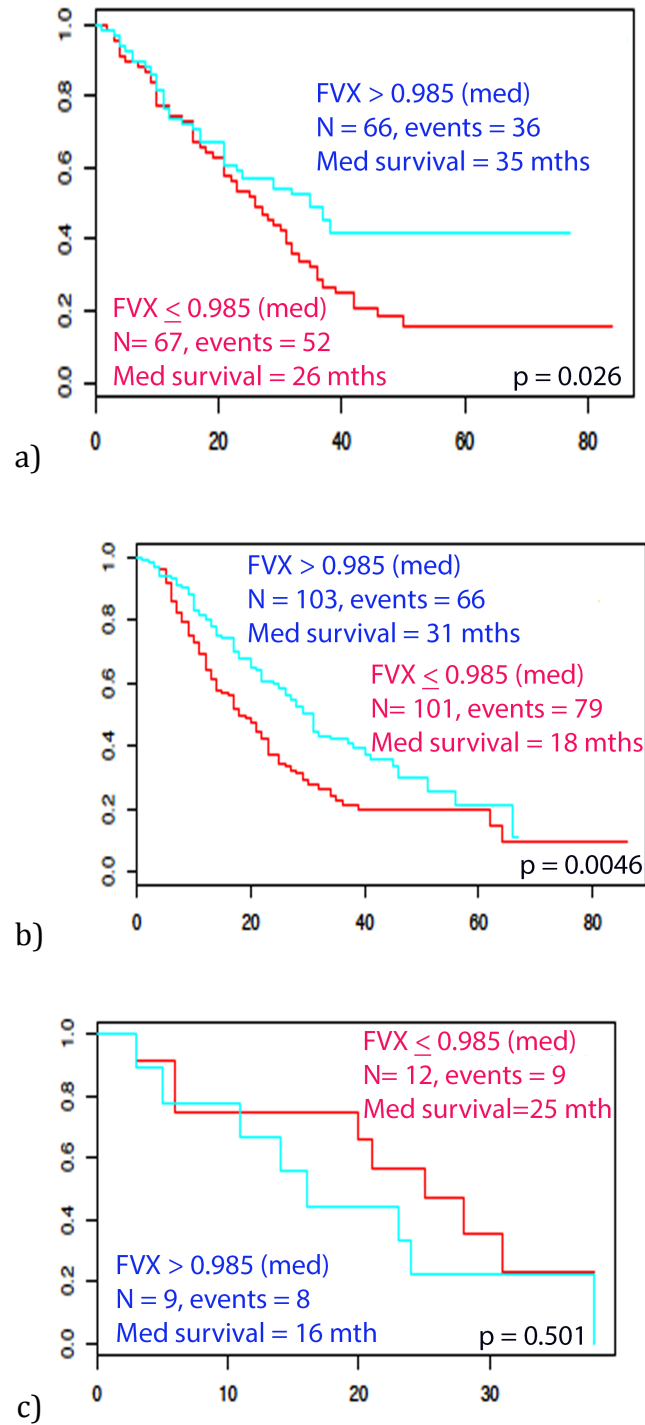
Figure 5.19 Testing the FVX composite texture feature on the datasets a) training, b) validation, c) TEST1. The FVX as split by the median gave significant results in the first two sets.
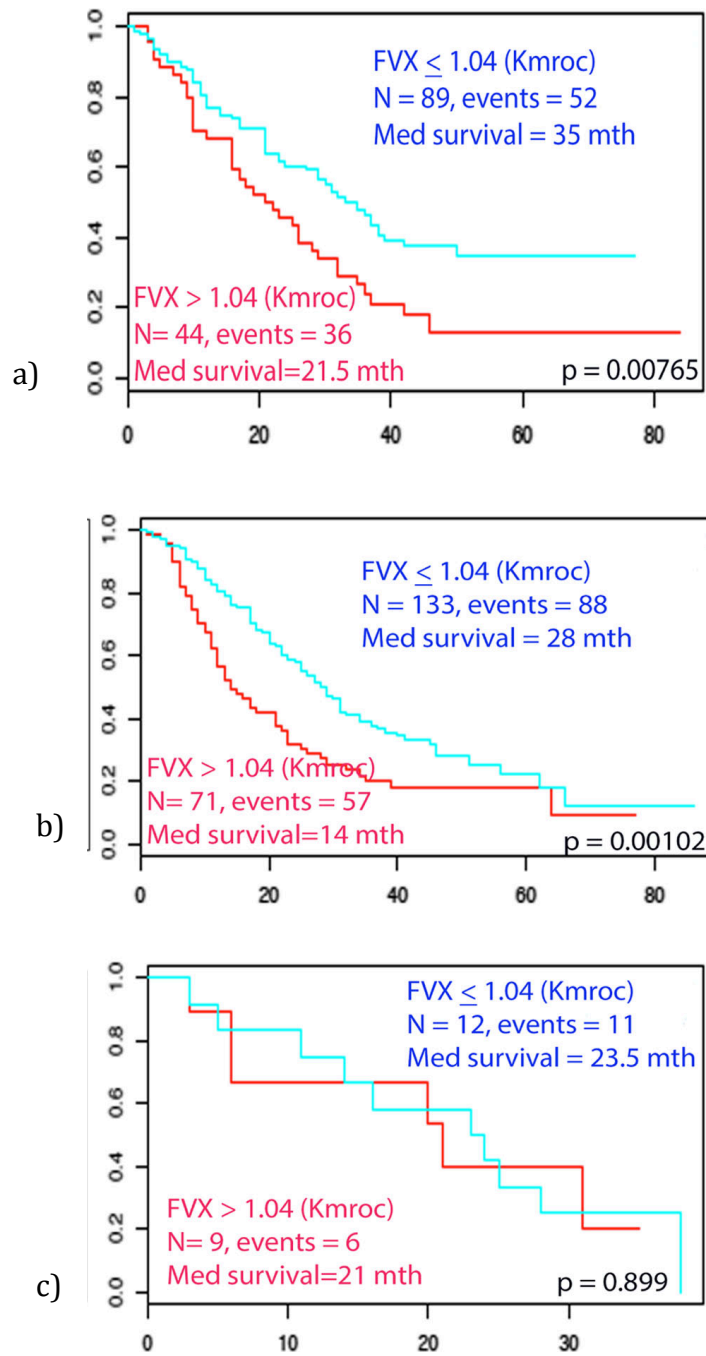
Figure 5.20 Testing the FVX composite texture feature on the datasets a) training, b) validation, c) TEST1. The FVX as split by the Kmroc gave significant results in the first two sets.

Thus in both Figures 5.19 and 5.20, there were significant results in the training and validation sets using two different methods to split the data. The FVX comprised of two texture features, weighted as given in Table 5.12.

Table 5.12 LASSO selected composite variable with weightings given for the contribution of each texture feature

Texture Variable	Weights
GLSZM Size Zone Variance	0.128
NGTDM Complexity	-0.018

The FVX was then tested in a Cox regression analysis with volume and stage. Volume was not significant and was removed. Table 5.13 demonstrated the findings.

Table 5.13 Cox regression analysis including the stage and the composite texture feature. Cox Regression analysis overall: p = 0.000196. N = 204, events = 145. Concordance between stage and FVX is -0.283.

Factor	p	HR	95 % Confidence Interval
Stage	0.352	1.12	0.88-1.43
Texture	0.0024	9.62	1.35-68.7

5.9 Differences between the Test1 set and the remainder of the sets

For all the factors that were significant, there was a constant finding that none were significant on the TEST1 set. The reasons for this were be explored in Figures 5.21, 5.22 and 5.23.
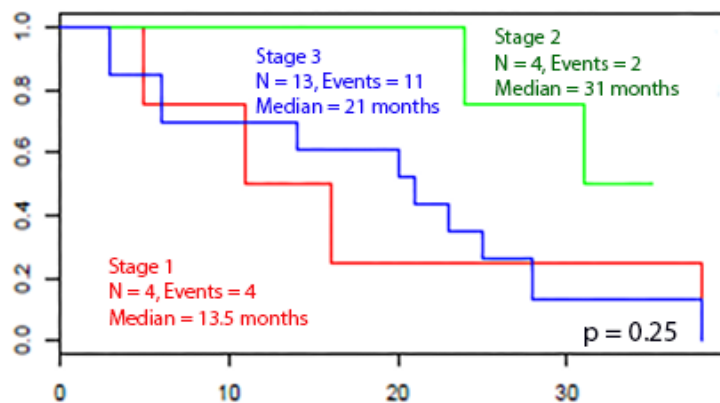
Figure 5.21 KM plot for stage for TEST1. The stage 1 group had the poorest survival.
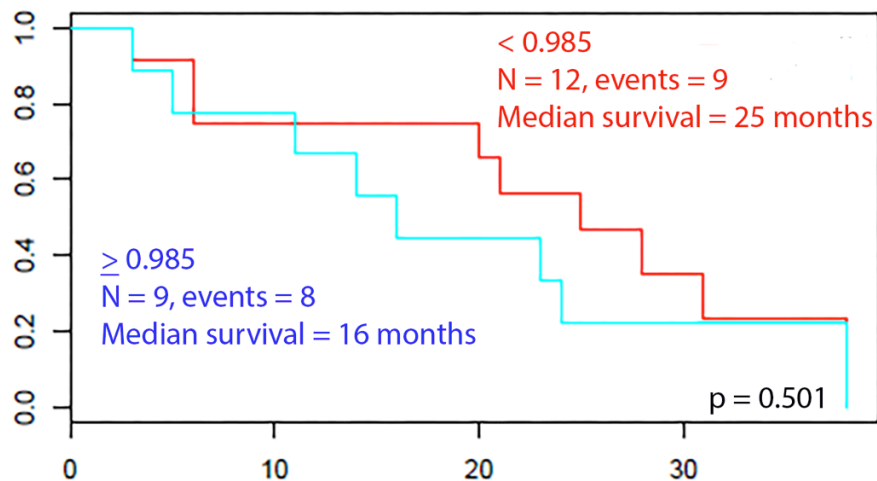
Figure 5.22 KM plot for stage split with median for TEST1. No significance.
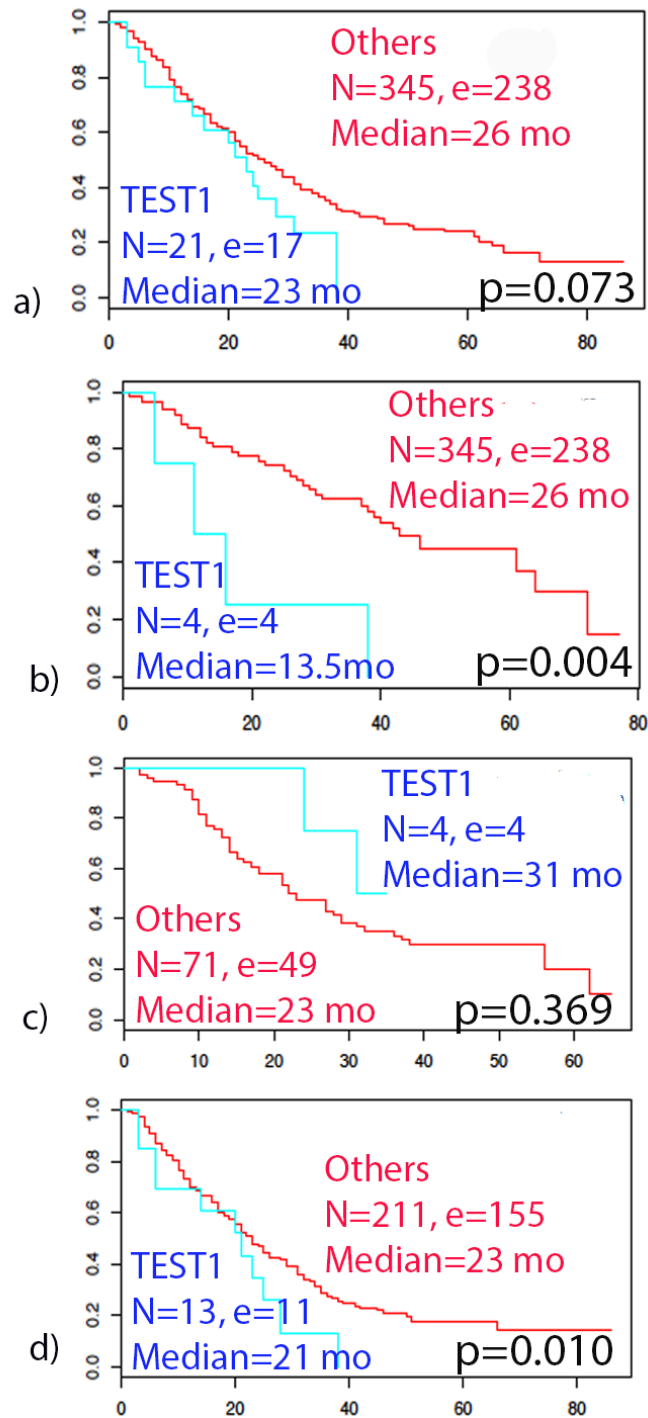
Figure 5.23. KM plots comparing TEST1 with others a) overall, b) Stage 1, c) Stage 2 and d) stage 3. The stage 1 group (number of patients = 4) was significantly different from the rest, which caused a deviation of the results. Interestingly, the stage 2 subset (number of patients = 4) performed better than the rest but this was not significant.

5.10 Test previous papers

Having developed our own texture composite feature, we wanted to test other texture features in the literature. We had collected a large number of datasets that facilitated the use of this (Table 5.14).

Table 5.14 Five previous papers that used texture analysis with survival. The cancer types from which they were derived are also provided as well as the texture features.

Author and year	Tumour type	Positive texture feature	Gray Level used
a) Cook (Cook et al., 2013)	NSCLC	Contrast Busyness Coarseness	Not stated
b) Cheng 2015 (Cheng et al., 2015)	Oropharyngeal Cancer	Zone-size non uniformity GL non-uniformity	16,32,64 16,32
c) Cheng 2013 (Cheng et al., 2013)	Oropharyngeal Cancer	Uniformity	4
d) Hatt 2015 (Hatt et al., 2015)	Oesophageal NSCLC	Combined: entropy, dissimilarity, high intensity large area emphasis, Zone percentage	64
e) Aerts (Aerts et	NSCLC	SNS_{com2}	NA

		FOS$_{Energy}$ GLRLM$_{GLN}$ GLRLM$_{GLN\ HLH}$	
al., 2014a)			

a) Cook and co-workers examined NSCLC and noted that NGTDM Contrast, Busyness and Coarseness were positively associated with survival - No composite function was given. However, they did not give a gray-level. Furthermore, this paper used the Kings dataset that we also used in our training and validation sets, so in order to test the findings of this paper we excluded this subset from our analysis.

In order to test their paper, we looked at each of their positive texture features in turn for each gray-level.

b) Cheng 2015

The Cheng 2015 paper looked at oropharyngeal cancer and found GLSZM Zone-size non-uniformity positively associated with survival at 16, 32 and 64 gray-levels, and GL non-uniformity at 16 and 32 gray-levels.

In order to test their paper we looked at each variable in turn at each gray-level.

c) Cheng 2013

The Cheng 2013 paper also looked at oropharyngeal cancer but found GLCM

Uniformity at 4 gray-level to be positively associated with survival.

They looked at GLCM_Angsmo_4gl > 0.138, however, when we looked at this variable, it was not significant in our dataset. GLCM_Angsmo_4gl > 0.138 was not significant with p= 0.0984.

d) Hatt 2015

Hatt looked at oesophageal cancer and NSCLC. They combined $GLCM_{Entropy}$ (E13 & E1), $GLCM_{Dissimilarity}$ (D13 & D1), GLSZM high intensity large area emphasis (HILAE), and $GLSZM_{Zone\ percentage}$ (Zone percentage) at 64 gray-levels.

e) Aerts

Aerts worked on CT rather than PET and had a fixed gray-level scheme that does not really make sense for PET. We looked at all gray-levels across all our data except the TEST1 dataset (Table 4.15).

Table 5.15 Summary of the texture features obtained from our lung dataset and those previously defined variables from 5 publications.

File name	Gray levels	Wavelets	No of features	Features	Significance with validation set (texture and stage model)	Stage and group model: texture sig	Stage and group model: texture hazard	Stage and group model: stage sig	Stage and group model: hazard ratio
Texture feature develop in this study	64	No	2	GLSZMSzVariance	0.000196		0.573	0.011	1.36
				NGTDMComplex	0.00116				
a) Cook paper	All non-significant in Cox regression								

b) Cheng 2015	16gl	No	1	GLSZMSzNo nUnif	0.000163	0.0084	1.48	0.011	1.28
	64gl	No	1	GLSZMSzNo nUnif	0.000115	0.0058	1.15	0.025	1.25
	32gl	No	1	GLSZMGINo nUnif	0.000421	0.0245	1.41	0.0241	1.256
c) Cheng 2013	Not significant in Cox regression								
d) Hatt 2015	Not significant in Cox regression								
e) Aerts 2015	4gl	Yes	4	SNScom2 FOSEnergy GLRLMGLN GLRLM GLN	0.000213	0.011	1.46	0.018	1.27

				HLH				
8gl	As above	4	As above	8.31x10-5	0.0035	1.57	0.033	1.24
16gl	As above	4	As above	0.000179	0.0082	1.51	0.0285	1.25
32gl	As above	4	As above	0.000148	0.0066	1.54	0.0342	1.24
64gl	As above	4	As above	0.000232	0.011	1.49	0.0228	1.26
128gl	As above	4	As above	0.000357	0.0182	1.44	0.0147	1.27
256gl	As above	4	As above	0.000183	0.00827	1.5	0.0164	1.27

Using the methods described in each paper, we used our dataset to test out these methods as an independent dataset. The highlighted areas denote significance. There is a complete table of results in Appendix D.

It is very interesting to compare our dataset with the previous papers (Table 5.15). Although, all bar the Aerts paper developed their model with much smaller numbers, it was interesting to see that there appeared to be some ability to predict prognosis. There may be an intrinsic phenotype independent of the method of acquisition (as CT is x-ray attenuation and FDG is glucose uptake) or voxel dimensions that accounts for this.

Chapter 6 Discussion NSCLC study

6.1 Summary of results

- 358 scans from 535 originals were selected. 133 were used for the training set, 204 in the validation set and 21 in TEST1

- The majority of patients were segmented at 40% (47.8%) and 30% (27.7%) thresholds. There was 9.5% manual adjustment made.

- There was excellent intra (0.9 ICC) and inter-observer (0.86) agreement for segmentation.

- The manufacturer, model type, slice thickness and matrix size did not affect the variability of the texture.

- SUV_{mean} was significant for the training set only on the KM curves. The SUV_{max}, SUV_{mean} and SUV_{peak} for the reminder of the datasets were not significant.

- The TLG was significant on the training and validation set on the KM curves. However, on a multivariate Cox regression including stage was found to be not significant.

- The composite feature vector (FVX) demonstrated significance for the training and validation sets. This was also significant on Cox regression with volume and stage included.

- The TEST1 dataset was not significant for the FVX feature but differed significantly from the remainder of the datasets with a disparate high number of Stage 1 patients who underwent chemo-radiotherapy and had a poor overall survival.

- When testing the texture features of other published data on our dataset, the Aerts (Aerts et al., 2014a, Cheng et al., 2015)feature vector developed on CT was highly significant on our dataset. The Cheng texture feature (Cheng et al., 2015) developed on oropharyngeal cancer and PET was also significant on our dataset.

The NSCLC multi-institution retrospective study demonstrated that the radiomics feature, FVX, derived from FDG-PET primary lung cancer was invariant to the PET scanner properties and predicted OS. The texture features in the radiomics feature contained GLSZM size variance and NGTDM Complexity. Although there is no direct histological correlation or immediate physiological correlation, the GLSZM is a regional 'homogeneity' texture that calculates the zones of homogeneous voxels. In particular, the size variance examines the variance in number blocks by size and is negatively correlated with survivl, perhaps consistent with hypoxic or necrotic regions. NGTDM examines the contrast difference or information content between the voxel and 26 available directions in the 3D space. The complexity of the NGTDM refers to the average visual complexity and positively correlated with survival, although with less of a magnitude that SzVariance, and possibly acting as a counterbalance to the latter.

Although PET parameters, in particular the SUV_{max}, have been shown to correlate with survival with a meta-analysis of 13 studies demonstrating that the primary tumour SUV_{max} correlated with survival (Berghmans et al., 2008) and that higher SUV_{max} was associated with poorer overall survival in patients receiving

stereotactic body radiotherapy (SBRT) (Na et al., 2014), this was not borne out in our study, where the SUV features were not prognostically significant. The TLG was also not significant in this study on multivariate analysis. In the literature, in a study of 196 inoperable Stage IIb/III NSCLC the TLG (and the MTV, which we found in our study to be highly correlated with the TLG) to be strongly prognostic for OS, while SUV_{max} was not (Salavati et al., 2017). The reasons why the MTV/TLG and the SUV_{max} have been significant in prognosis is some studies but not others, including this one, are not entirely clear.

6.1 Study strengths

The biggest strength of this study was that it utilised a multi-centre approach. This helped in many ways to increase the number of subjects as well as introduce real-life complexities in the data. Questions perpetually asked in the literature that have been explored in this study include the variation in acquisition parameters and different scanners.

The other strength of this study was the ability to build on the work of others to have a large number of texture features available. Utilising fractals, wavelets and other features for possibly the first time in PET studies has been advancement from what was utilised previously.

Another big advantage, closely allied to having a large dataset was the ability to develop independent training, validation and test sets. This has increased the statistical power of the study and allowed firm conclusions to be drawn.

Using the compiled dataset to test the findings of previously published studies in the literature was an interesting revelation. This demonstrated that some findings were indeed type 1 errors(Chalkidou et al., 2015) and others have yielded promising results.

6.2 Study limitations

6.2.1 Sample size

The large number of centres also has its disadvantages regarding information loss, particularly as it was retrospective, but these are common to all large retrospective studies and are not unique to this study.

The study was developed utilising as many patients that could be obtained. The majority of these were obtained from large centres around London and the UK. A freely available dataset from centres from the United States was also obtained, albeit this accounted for less than 10% of the final numbers. The end result was that this cohort in combination is the largest PET study to date for texture analysis.

The positive results do show that the methods work. There is a concern that utilising a large number of texture features on a limited number of studies can result in over-fitting of the data and type 1 errors (Chalkidou et al., 2015). In addition, the utilisation of machine learning, although novel, does require a very large number of subjects, somewhere in the order of thousands as opposed to hundreds. It remains to be seen if a large enough cohort will ever be tested.

Another issue, which is allied to the previous point, is the number of hours required to curate the data, to acquire the segmentations and all the appropriate clinical end-points. These are hours rarely mentioned in the literature, except by Gillies and co-workers (Gillies et al., 2016) where many hours of man-hours have been mentioned. This is a serious hindrance to this technology, which no doubt will be streamlined in the future as most studies have been testing their theories in a retrospective fashion.

The minimal volume threshold was set at 5ml, in keeping with previous work by Soussan in breast cancer PET scans (Soussan et al., 2014), however this may still be too small for PET given the fact that there is a statistical limitation in applying radiomics to PET data in small tumours (less data-points in the VOI compared with CT, 14% of the initial 535 patients initially screened had tumour volumes <5ml). In view of this, the results should be treated with caution when trying to apply it to small sized tumours.

To move forwards, a large prospective study, utilising data from many centres in order to increase numbers is warranted. A clear advantage identified is that the texture features from the primary lesion are able to recognise poor prognosis earlier than even stage.

6.2.2 Effects of PET imaging parameters on the results

6.2.2.1 Effects of image resolution and PVEs

In order to truly spread this technique into the clinical domain, some uniformity of acquisition and elimination of PVEs needs to be made. This study has gone quite a way to compare the differences between the different centres and machines, an assessment that has never been performed before. The maximum number of different scanners in previous studies was limited to two.

6.2.2.2 Effects of image reconstruction parameters (texture analysis)

The effects of image reconstruction have been shown to affect the texture analysis (Galavis et al., 2010). The fact that different scanners from different centres were utilised and the initial PCA analysis demonstrated no significant difference between the different groups is very beneficial for the future of this technique in PET. Certainly the findings in PET studies in earlier work will remain pertinent in the future even as PET technology advances.

6.2.2.3 Target segmentation

In order to utilise the best methods to segment tumours, a 40% semi-quantitative method was used to segment the tumours. Whether adaptive 40% is better than manual segmentation or another automated or semi-automated method remains to be seen. This area was not explored thoroughly in this thesis and although some data in the literature exists that states that the method of segmentation does not significantly alter the texture analysis results (Galavis et al., 2010), the previous number of studies were performed on small samples and whether methods of segmentation affect texture analysis in larger studies remains to be seen.

6.2.3 VOI size and target to background ratio

A lot of work has been performed in this area and the literature(Hatt et al., 2010a, Hatt et al., 2011a, Orlhac et al., 2014, Brooks and Grigsby, 2014) states that the limitation of PET is its poor full-width half maximum, permitting voxels which are much larger than those in MRI and CT. The minimal volume of 5mls utilised in this study was based on a study on FDG in breast cancer (Soussan et al., 2014). Further research has emerged since then that the minimum is actually 10mls (Hatt et al., 2015), although the authors add a caveat that the minimum volume could be less than this when the volume's relationship with other features are taken into account. The methodology used in this paper utilised a variety of tumour sites therefore it is unclear if this could definitely be applied to lung cancer in this these.

6.2.4 Statistical methods and machine learning

There have been a variety of statistical techniques utilised to assess texture analysis, from the simple linear regression to the more complex recently utilised machine learning methods. In this study, LASSO, a penalised regression method was utilised. Supervised PCA was also attempted but this method had to be abandoned, as we could not overcome some serious hurdles including how to use dichotomous outcomes on continuous data. There are however over 100 different techniques for machine learning, such as random forests, etc. and favouring one method over the other may be a potential fallacy. However, due to the interests of time, only a limited number of methods could be explored.

6.3 Comparison with existing Literature

6.3.1 Texture analysis

A very important outcome of this work was the exploration of radiomics variables optimised from the literature on our lung dataset. Some of the interesting conclusions, which could be explored in future, demonstrate that the texture features that are associated with prognosis are not linked to a specific tumour type, or more accurately a radiomics feature optimised from one tumour types may be applicable in another tumour type. Indeed the Cheng et al study (Cheng et al., 2015) carried out their work on oesophageal cancer, but the same variables were strongly significant in our lung dataset (see Table 5.15).

The other very interesting point was that a 4 feature set developed on CT worked on our lung dataset (Table 5.15). The work of Aerts (Aerts et al., 2014a), for instance was developed on the same tumour type (NSCLC) and also tested on head and neck cancer. However, the fact that a completely different modality was used in optimisation is very interesting and exciting. What this means in this field is difficult to ascertain at this moment but could indicate that there are global texture features, which quantify heterogeneity and that are independent of modality and tumour type.

6.4 Perspectives and Future work

6.4.1 PET intra-tumour heterogeneity and underlying biology

A large question mark hangs over texture analysis in PET due to the inability to explain the texture features and findings with their underlying biology. Indeed, in this study, where the tumours were irradiated, and only a small sample of tumour biopsy was obtained to confirm the diagnosis for treatment purposes, no corresponding underlying tumour biology could be compared. The inevitable question that hangs over texture analysis for the foreseeable future remains, "What do the numbers actually mean?" This also formed part of the thought proocesses behind the ovarian cancer texture analysis, where tissue and the ability to obtain from other 'omics' data were available and will be discussed shortly.

6.4.2 Informative biomarkers of intra-tumour biological heterogeneity in the clinical context

The difficult process of translating laboratory derived data and clinical studies are well documented. The future for texture analysis is promising but a few hurdles have to be overcome. Some of these have been mentioned already such as the long time it takes to processing the images and marrying them up with their clinical data, etc. Issues of consent and data protection abound and in addition to this, the time taken by the radiation oncologist or radiologist to acquire or perform this is a severe limitation on a background of their ever-increasing workloads.

However, an improved workflow, working with data scientists, software developers and biostatisticians could make this technique ever more useful.

6.4.2 Effects of patient motion and positioning

Ongoing studies demonstrating the changes of FDG in relation to patient positioning and location of the tumour should be pursued in the future. Whether, this affects the final texture analysis remains to be seen. The retrospective nature of the current study did not allow the exploration of the effect of breathing especially in the lower lobes.

6.4.3 Optimal parameters to image heterogeneity with PET

The optimal parameters for PET imaging in texture analysis are again something that is still being explored in the literature(Galavis et al., 2010, Willaime et al.,

2013). Given the fact that PCA clusters were not significant does demonstrate that intrinsic variances in texture may be due to underlying biological differences. However, much like batching of data in multi array RNA accounts (Guo et al., 2014) for these differences, a similar method can be applied to PET and other modalities.

6.4.4 Monitoring dynamic and adaptive biological complexity with PET

Questions often arise as to what are the advantages of dynamic PET over static images that are used in clinical practice. Although, this is currently something that is not routinely used on the clinical side, questions do arise if a dynamic data set reflects heterogeneity more accurately and whether the texture analysis can tease this out. The retrospective clinically derived images from this study have not made it possible to explore this avenue.

Chapter 7 – Results Ovarian study

The lung texture analysis provided useful information on prognosis, however, the underlying biology was unknown. Thus, in order to investigate this further, we obtained further information on contrast enhanced CT for patients who had fresh frozen material banked at the time of surgery to enable correlation with other radiomic and –omic factors. This was known as the Hammersmith cohort.

7.1 Results

7.1.1 Final numbers

7.1.1.1 Training and validation set

The final number obtained from the Hammersmith cohort was 363 CT scans. This was taken from an original number of 546 consecutive scans, of which 187 were excluded. The reasons for the exclusions are as follows with percentage in brackets:

CT for recurrent disease	28 (15.0%)
Pre-operative scan not available	106 (56.7%)
Non-ovarian histology	21 (11.1%)
Non-contrast scan	10 (5.4%)
No primary seen	22 (11.2%)
Could not be anonymised	1 (0.5%)

Non-ovarian Histology

Endometrial	13 (11.1%)
Primary peritoneal	5 (4.3%)
Breast cancer	1 (0.9%)
Uterine	1 (0.9%)
Other	2 (1.7%)
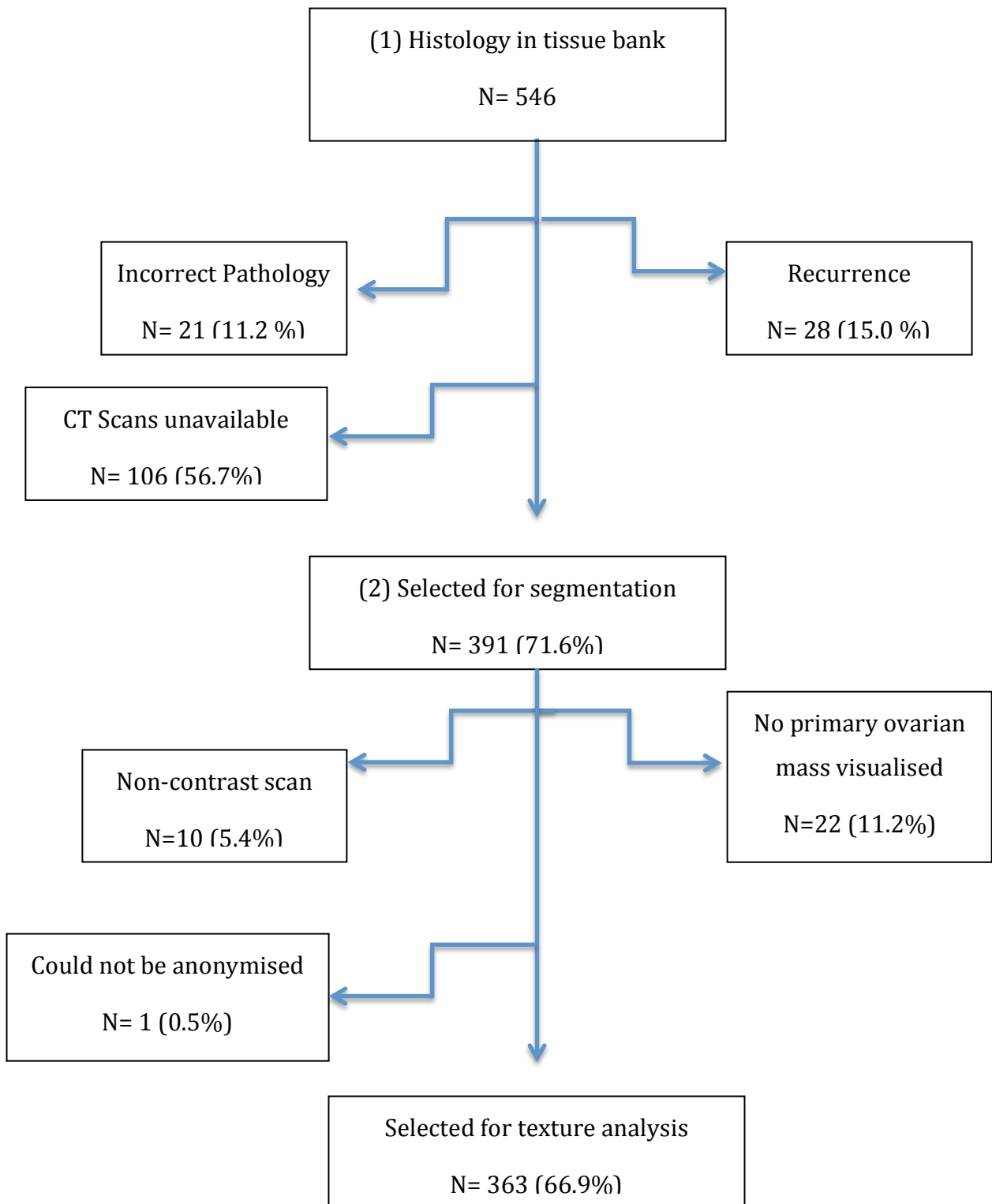
Figure 7.1 Overview demonstrating the selection of the final number of scans in the Hammersmith cohort. This was further split into a training dataset of 136 patients and validation set of 77 patients.

7.1.1.2 Validation set 2

The acquisition of the scans for the validation set has been detailed in the previous section.

For the Validation set the final numbers were as follows:

From 143 scans available, 70 were selected. The reasons for the exclusions were as follows (percentage of excluded from 73 total given in brackets):

Corrupted file	1 (1.4%)
Incomplete or no pre-op scan:	62 (85.0%)
No primary ovarian mass visualised	9 (12.3%)
Metallic artefact	1 (1.4%)
Non-contrast scans	1 (1.4%)
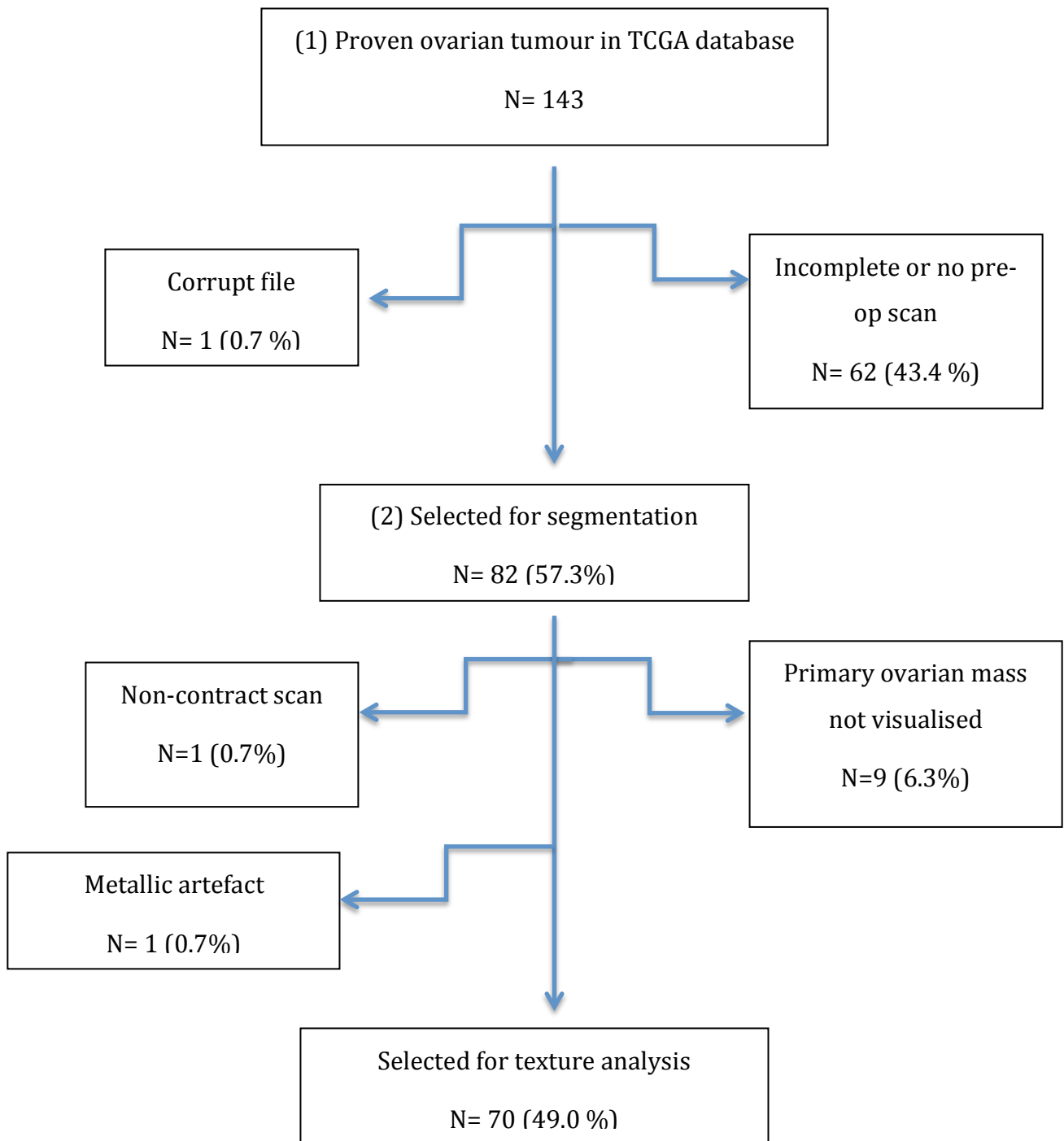
Figure 7.2 Overview demonstrating the selection of the final number of scans in the second validation set.

Table 7.1 Breakdown of the scanner types, matrix, slice thickness, etc. used in the study

Institution	Number (%)	Vendor	Number (%)	Scanner model	Number (%)	Matrix	Number (%)	Slice Thickness
Alliance	1 (0.3)	GE	150 (41.3)	Aquilion	48 (13.3)	512 x 512	349 (96.1)	0.63
BMI Chiltern	1 (0.3)	Phillips	69 (19)	Brilliance 16	2 (0.6)	768 x 768	14 (3.9)	1
Central Mid	11 (3.1)	Siemens	94 (25.9)	Brilliance 64	28 (7.7)			1.25
Northwest (Northwick Park)	29 (8)	Toshiba	48 (13.2)	Definition AS +	26 (7.2)			1.5
Charing Cross	14 (3.9)	Picker	2 (0.6)	Discovery CT750H	1 (0.3)			2

Hammersmith	147 (40.5)	Emotion	1 (0.3)	2.5
St Marys	33 (9.1)	Emotion Duo	3 (0.8)	3
UHW Cardiff	1 (0.3)	iCT 256	6 (1.7)	3.2
Derriford	1 (0.3)	Ingenuity Core 12	1 (0.3)	5
Ealing	19 (5.2)	Ingenuity CT	11 (3)	5.5
Heatherwood	7 (1.9)	LightSpeed 16	1 (0.3)	7.5
Hillingdon	34 (9.3)	LightSpeed Plus	12 (3.3)	8
Lancashire	1 (0.3)	LightSpeed Ultra	96 (26.5)	10

Lister	1 (0.3)		LightSpeed VCT	26 (7.2)
Mount Vernon	4 (1.2)		Mx8000	13 (3.6)
Royal Shrewsbury	1 (0.3)		Mx8000 IDT 16	8 (2.2)
St Georges	1 (0.3)		MxTwin	2 (0.6)
Wellington	1 (0.3)		Optima CT600	14 (3.6)
West Mid	30 (8.3)		Sensation 10	1 (0.3)
Wexham Park	19 (5.2)		Sensation 16	15 (4.1)

		Total	363		363			363		363	
Somatom Definition								47 (12.9)			
	Somatom Plus 4							1 (0.3)			

7.2 The Patient Demographics

Table 7.2 Demographics of the patients involved in the study

These are split between the patients that went straight to surgery and those that had neo-adjuvant chemotherapy. A total of 293 patients underwent primary debulking and 48 delayed primary debulking.

Feature		Primary debulking (N=293)	Delayed primary debulking (N=48)
Age at diagnosis	Median	60	67.5
	Range	19-91	43-83
Type	Benign	6	0
	Borderline	35	0
	Malignant	247	47
	Unknown	5	1
Stage	I-II	70	0
	III-IV	190	45
	Unknown	33	3
Grade	Low	15	3
	High	218	44

	Unknown	60	1
Histology	Serous	193	42
	Non-serous	100	6
	Unknown	0	0
Residual disease	No	199	37
	Yes	52	7
	Unknown	42	4
Relapsed	No	141	21
	Yes	117	22
	Unknown	62	5
Deceased	No	102	30
	Yes	187	17
	Unknown	4	1
Molecular data	Texture	268	42
	RPPA	236	35
	tRNA-seq	173	
	WGS	84	

7.3 Comparing patients: Scanners

(a)

(b)

Figure 7.3 PCA and Multidimensional scaling to elucidate any batch effect

a) Principal component analysis and b) multidimensional scaling by Euclidean distance used to identify batch effect on thickness, matrix and vendor (manufacturer). There is strong clustering in the matrix and different vendor plots indicating that any variance observed are not due to intrinsic differences between the scanners. The thickness demonstrated the strongest effect and as such the 3.2 and 1.25 mm thickness scans were removed from further analysis.

7.3.1 The effect of the different scanners

Using two separate methods (Figure 7.3a) to assess the texture variability of the different CT scanners demonstrates that there is strong clustering, with minimal variation. This can be seen with the principal component analysis of the first two components accounted for 20% of the variance of the dataset. The groupings for thickness, matrix and vendor were very tight. The thickness data will be explored in more detail later on.

In addition, in   Figure 7.3b another method to demonstrate the variability of the data, namely the Euclidean distance, again demonstrates a very similar result to the PCA, with clustering of the data points. In short, this suggests that the three factors explored, namely the slice thickness, matrix and vendor and invariant between the different scanners and thus any difference that we do see can be inferred to as belonging to the inherent tumour biological differences.

7.4 Unsupervised clustering

Two separate methods were used to dichotomise the data so that underlying differences could be delineated. First, an unsupervised clustering method (in our case, hierarchical clustering) was used. An unsupervised clustering method is when the data is arranged or sorted without any priori data used. For instance, no stage or outcome data is added to the algorithm. This allows the mining tool to develop clusters of similarity, which can be tested against survival and prediction of response. The unsupervised clustering method was able to separate out the data, and when linked to progression free survival (p=0.009), this was significant, but overall survival was not significant OS (p=0.09). (Figures 7.4 and 7.5).

The findings from this indicate that there are some inherent differences in the data that can be determined and linked to survival. The model was further refined to strengthen these early promising findings.

In addition, the different cluster groups were also very similar when compared using Pearson and Euclidean distances (Figure 7.3).

Hierarchical clustering by Pearson correlation
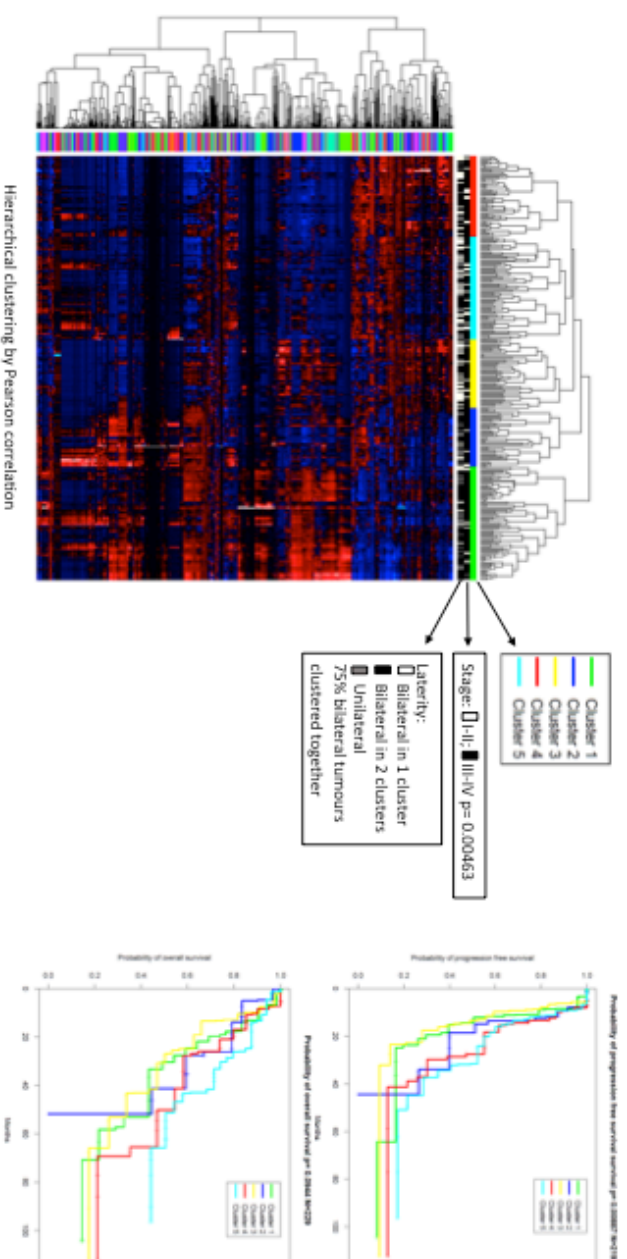
Figure 7.4 Unsupervised clustering of all the serous cases demonstrate diagnostic and prognostic potential.

Using hierarchical clustering by Pearson correlation, the scans were separated by low (stage 1-2) and high (stage 3-4) and whether the primary tumour involved one ovary or both. Interestingly, 75% of bilateral tumours clustered together using this method. Also using a KM curve of these five groups, there was a significant association with the PFS (p=0.009) but not with OS (p=0.09). Acknowledgments go to Haonan Lu and Edward Curry for their help in generating this data.

By Pearson correlation

Dendrogram of serous cases
N= 296, hcluster by Pearson correlation

By Euclidean distance

Dendrogram of serous cases
N= 296, hcluster by Euclidean distance
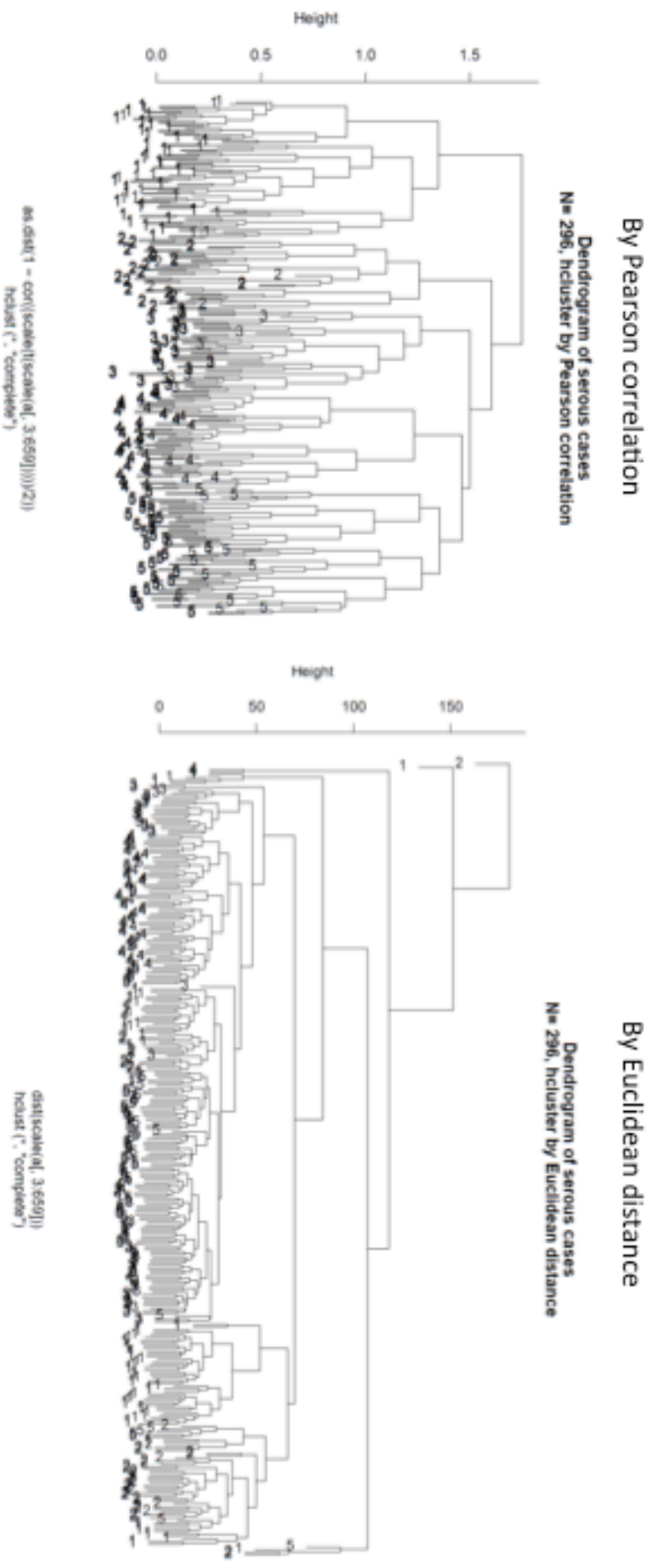
Figure 7.5 Using two different methods to show the clusters obtained from the hierarchical clustering.

There was a very close relationship and similarity of the clustering using the Pearson correlation and that obtained from Euclidean distance.

7.5 Slice Thickness and survival

In order to further improve the model, an examination of the slice thickness and the group clustering was made. In Figure 7.6, a comparison was made between the slice thickness and clusters. Cluster 2 was significantly different from all the other clusters (p=3.4 x10^{-10}). This cluster was also significantly biased due to small numbers and the decision was made to remove this variable (Figure 7.6). Removing this outlier group, demonstrated that the remainder of the groups were stable (Figure 7.7). This was further confirmed when a box-, mosaic plots and Kaplan Meier curves of all the different slice thickness demonstrated a better survival with a slice thickness of 1.5 and 10 mm (Figures 7.8 and 7.9). As the scan type should not contribute to survival, these are clearly outliers and should be removed from further analysis.
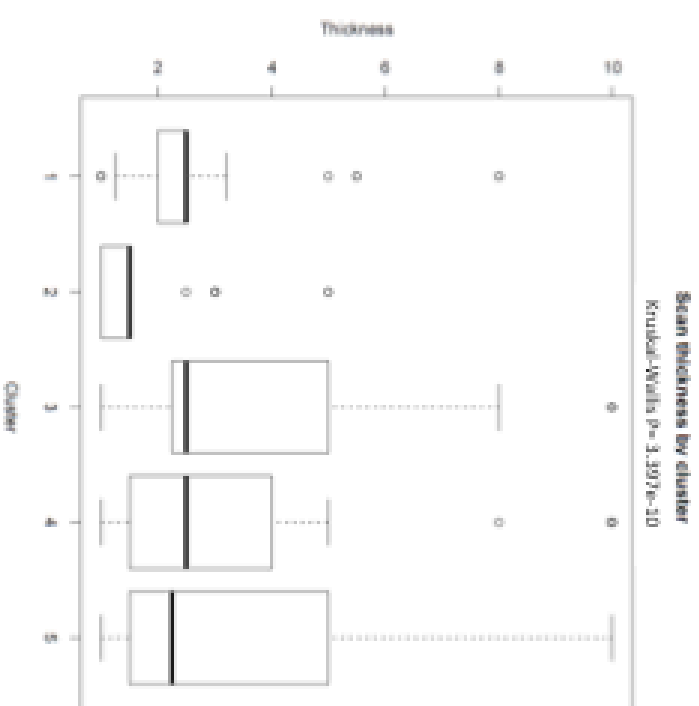
Figure 7.6 Comparing the 5 clusters with scan thickness.

There is a significant difference between cluster 2 and the others when compared with slice thickness. Cluster 2 was associated with thinner slices (p = 3.4 x10^{-10} by Kruskall Wallis). The decision was made to remove this cluster and re-run the clustering.

Supplementary 3. Unsupervised clustering without the outlier shows stable subgroups
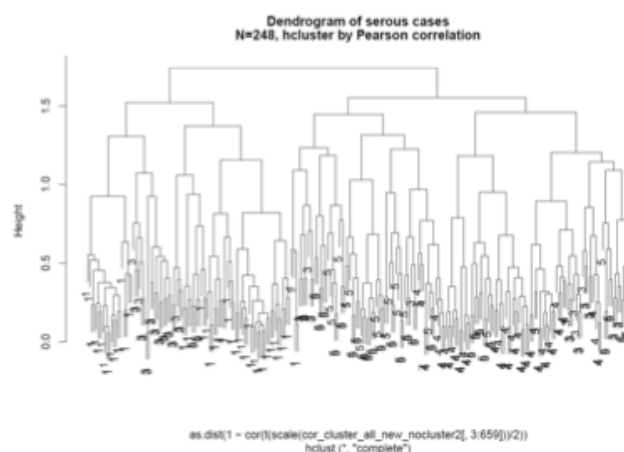
Dendrogram of serous cases
N=248, hcluster by Pearson correlation

as.dist(1 – cor(t(scale(cor_cluster_all_new_nocluster2[, 3:659]))/2))
hclust (*, "complete")

Supplementary 3. Unsupervised clustering without the outlier shows stable subgroups

By Pearson correlation   By Euclidean distance

Dendrogram of serous cases   Dendrogram of serous cases
N=248, hcluster by Pearson correlation   N=248, hcluster by Euclidean distance

as.dist(1 – cor(t(scale(cor_cluster_all_new_nocluster2[, 3:659]))/2))   dist()(scale(cor_cluster_all_new_nocluster2[, 3:659]))
hclust (*, "complete")   hclust (*, "complete")
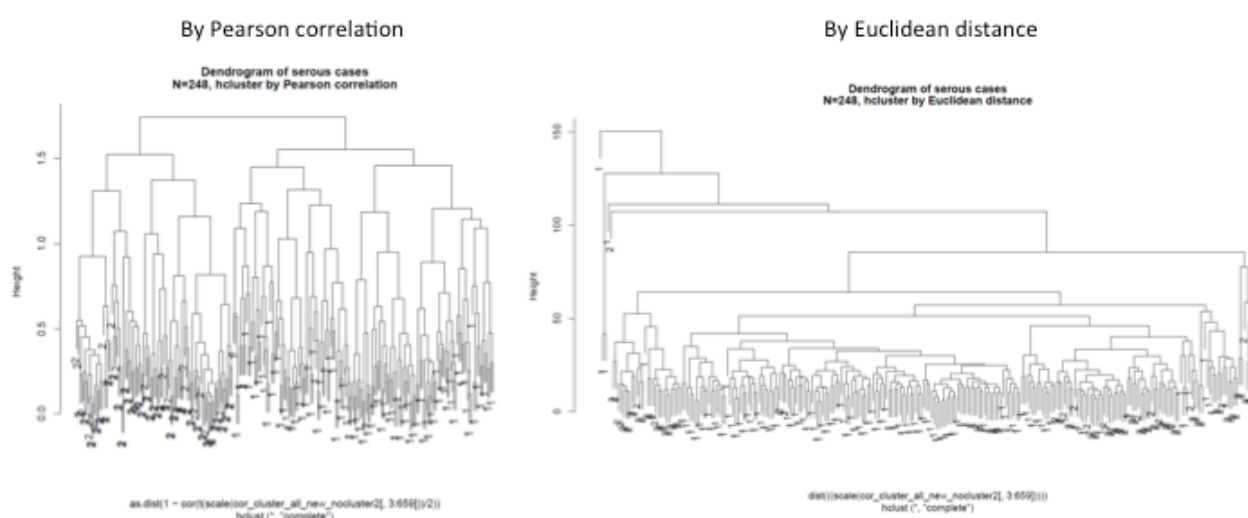
Figure 7.7 Re-run of the clustering without the outlier demonstrates stable subgroups

The unsupervised clustering of the groups has resulted in stable sub-groups that were again checked with Pearson correlation and Euclidean distance.
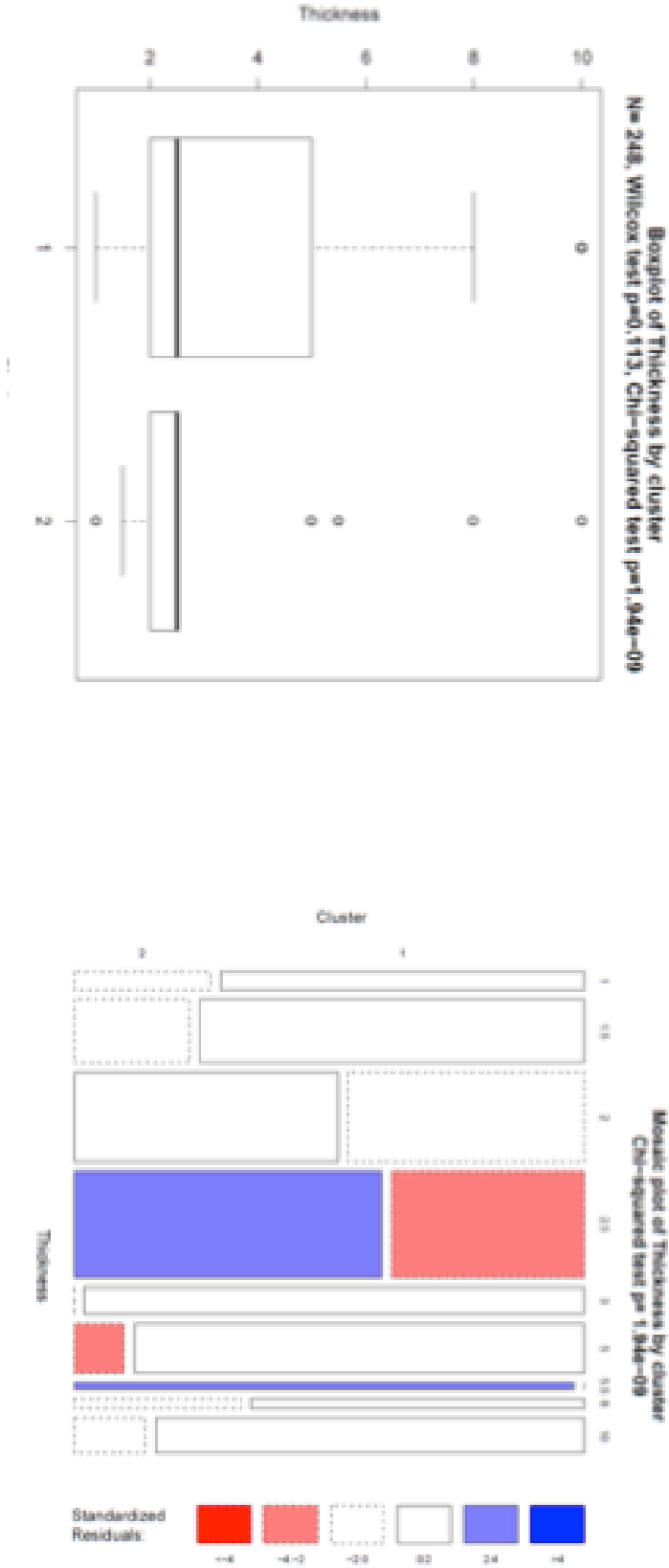
Figure 7.8 The association between the clusters and the slice thickness

There is an association between the clusters and the size thickness as demonstrated by the boxplot (p=9.4 x10⁻⁹ and the mosaic plot (p = 1.9 x10⁻⁹).
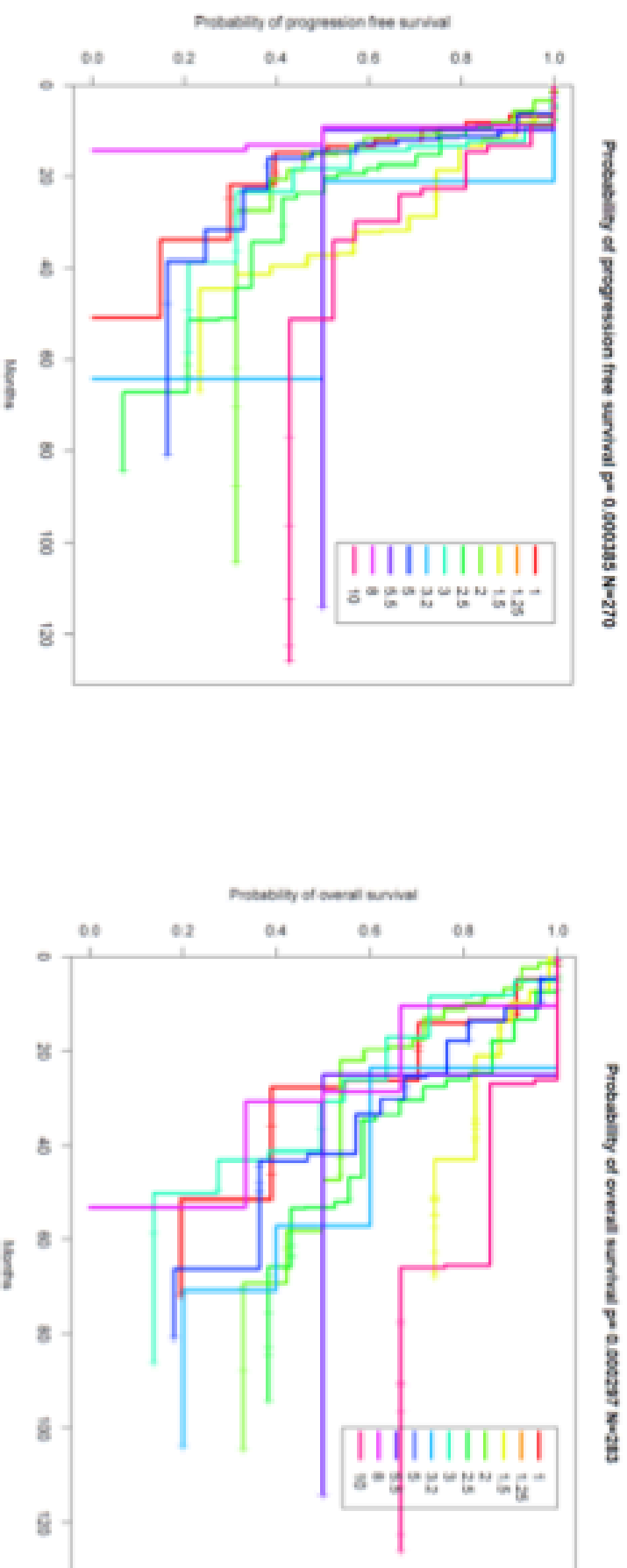
Figure 7.9 Kaplan-Meier plots of the slice thickness and survival a) PFS, b) OS.

Interestingly, there was a survival difference depending on the slice thickness on both the PFS (p=0.0004) and OS (p=0.0003). The scans with thickness of 1.5 and 10 mm had better survival

7.6 Rerun of the unsupervised clustering

The analysis was re-run with the removal of the small number of outliers. The new data generated demonstrated an improved model (Figure 7.10 and 7.11). These show that the OS KM curve is significant with p=0.0214, HR 1.59. Interestingly, when there were bilateral tumours, there was a strong clustering of these (96%). There was also no significant association of the unsupervised clustering method outcome with stage.

7.6 Rerun of the unsupervised clustering
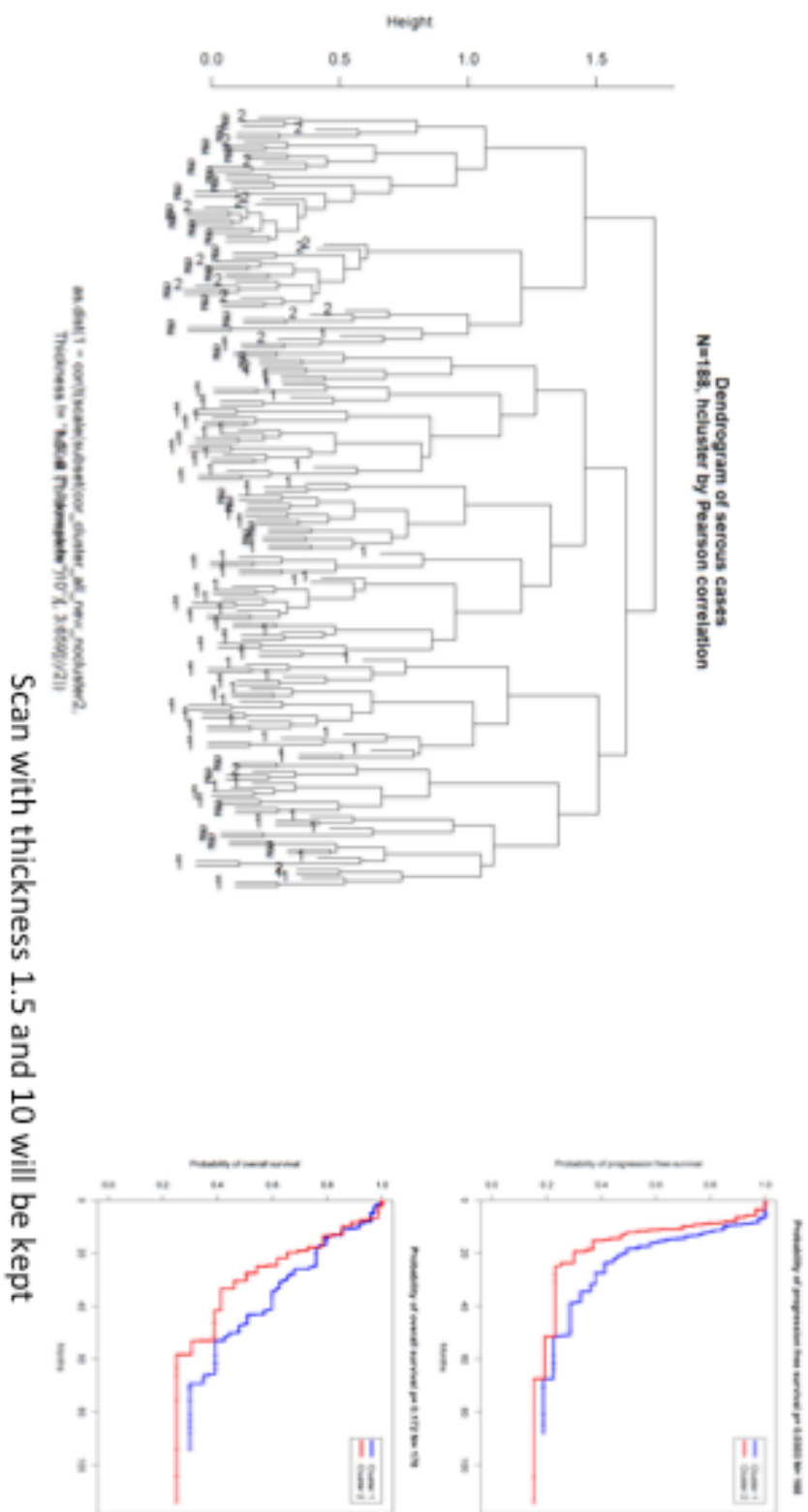
Figure 7.10 Removing the slice thickness of 1.5 and 10mm demonstrated stable clusters and prognostic power.

Given these results, the decision was made to keep these factors in the study.
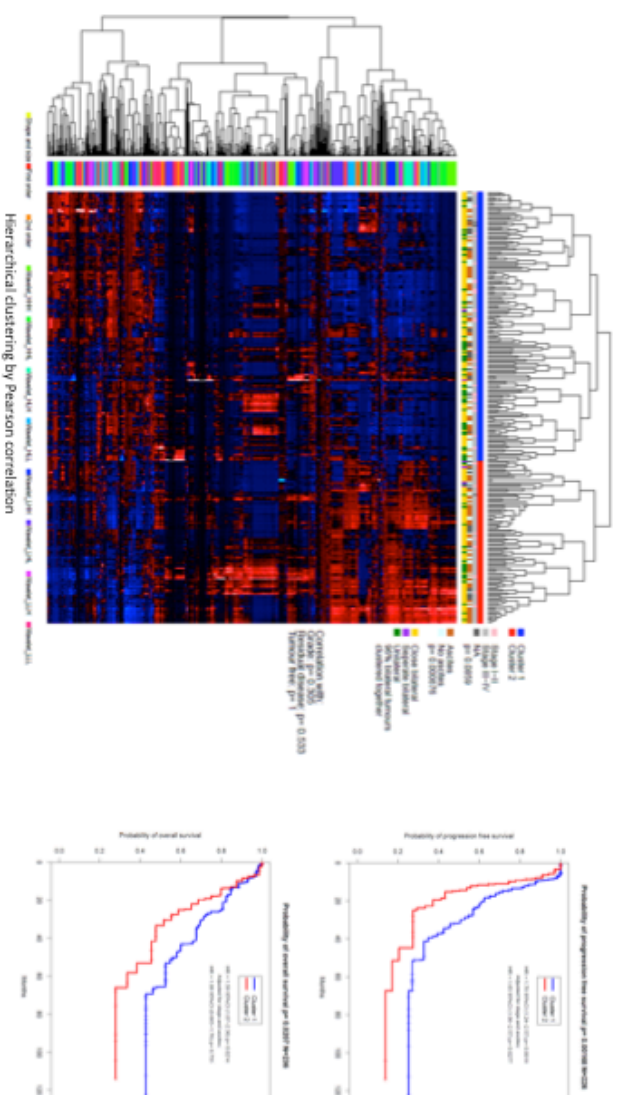
Figure 7.11 Unsupervised clustering of corrected clusters demonstrates diagnostic and prognostic potential

Two clusters were then generated. The clusters were not associated with Stage (p=0.089) but were significantly associated with the presence of ascites (p=0.0009). 96% of bilateral tumours clustered together. The Kaplan-Meier curves demonstrate significant associations with the PFS (p=0.002, HR 1.78, 95% CI 1.24-2.57). When this was adjusted for stage and ascites p=0.0277, HR 1.65, 95% CI (1.06-2.57). There was also an association with OS with p=0.0214, HR 1.59 (95% CI 1.07-2.36), although the significance disappeared when stage and ascites were taken into account (p=0.755).

7.7 Association between different clinical and CT factors.

To ensure that certain known contributing factors (Ahmed and Stenvers, 2013, Gupta and Lis, 2009) were accounted for, boxplots for stage and slice thickness and slice thickness and ascites were performed (Figure 7.12). Neither of these were significant. The CA125 and cluster were also plotted on a boxplot (Figure 7.13) but this was again not significant.
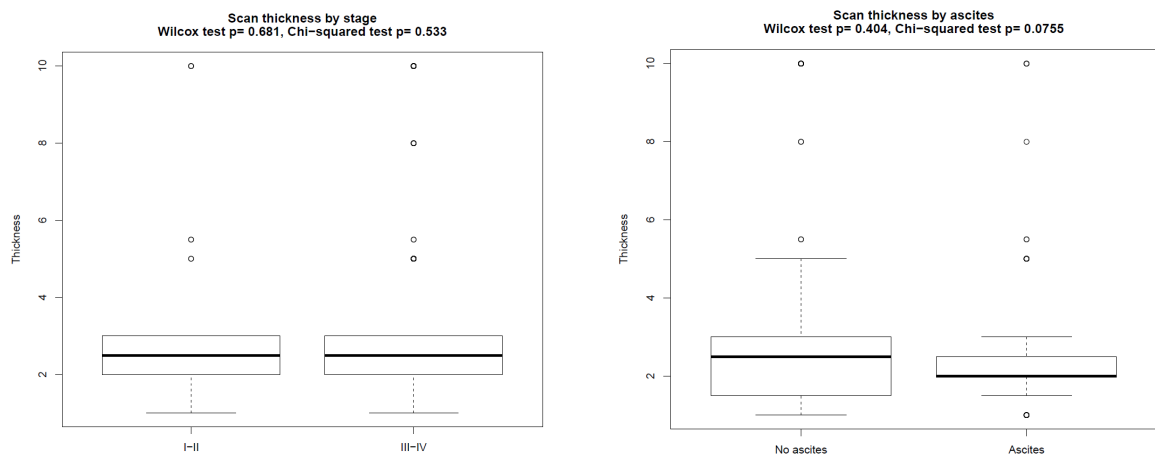
Figure 7.12 The scan thickness was not associated with stage or the presence of ascites

Using a boxplot, stage did not have an association with thickness (Wilcoxin test p=0.681, Chi-squared test p=0.533). Using the same methodology, the presence of ascites also did not have an association with slice thickness (Wilcoxin p=0.404, Chi-squared test p=0.08).
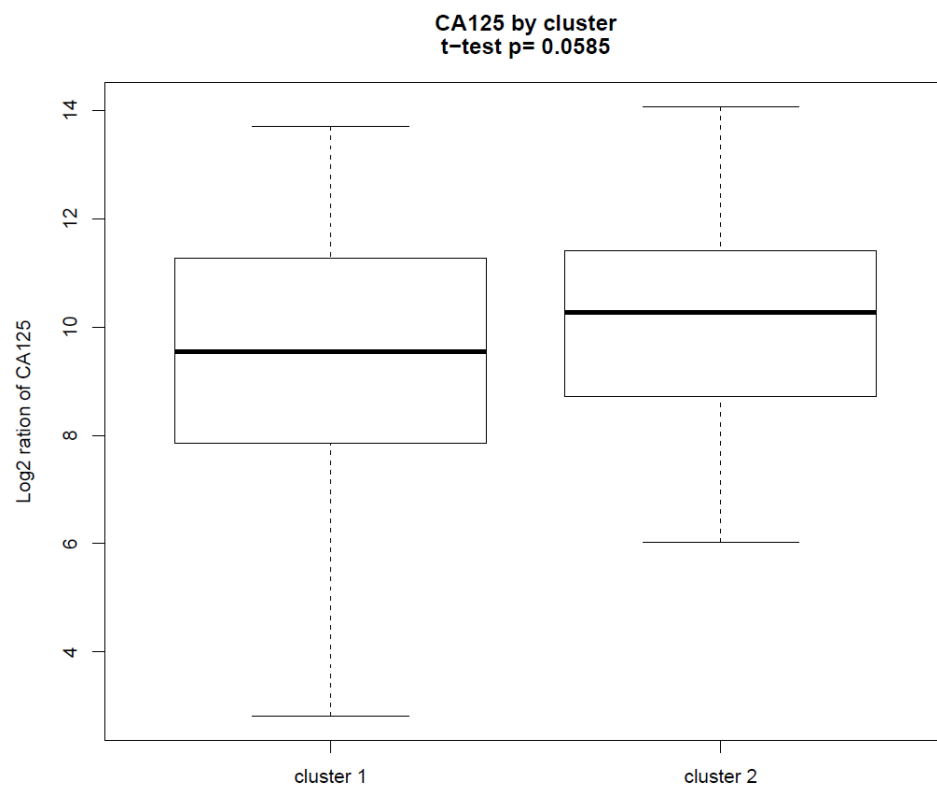
Figure 7.13 CA-125 was not significantly different between the two clusters

The CA-125, a blood-based tumour marker, used for the detecting the presence of ovarian malignancy, was not significantly different between the two clusters (p= 0.06.)

The two clusters that we have developed seem to be independent of known factors such as ascites and CA-125.

7.8 Univariate method adjusted for clinical factors

Using single univariate analysis, KM curves were generated which accounted for known associations like stage, presence of ascites, etc. (Figure 7.14). This was used to further strengthen our model. The results (Figure 7.14a) demonstrate that there was an association with the adjusted texture factor, NGTDM$_{contrast}$ and PFS, p=0.055, HR 1.34, 95% CI 1.1-1.64, and in the validation set on the KM curve p=0.003 and on the multivariate analysis, p=0.012, HR 1.5, 95% CI (1.11-2.26).

In Figure 7.14, using the texture feature, GLRLM$_{SRLGLE}$_LLL, split at the median, the KM curve had a p=0.023 in the training set and on multivariate analysis (adjusted for stage, grade, residual disease, ascites and thickness): p=0.001, HR 1.34, 95% CI 1.2-1.54, and in the test set on the KM curve p=0.037 and on the multivariate analysis, p=0.0004, HR 1.71, 95% CI (1.27-2.29).

a)

b)
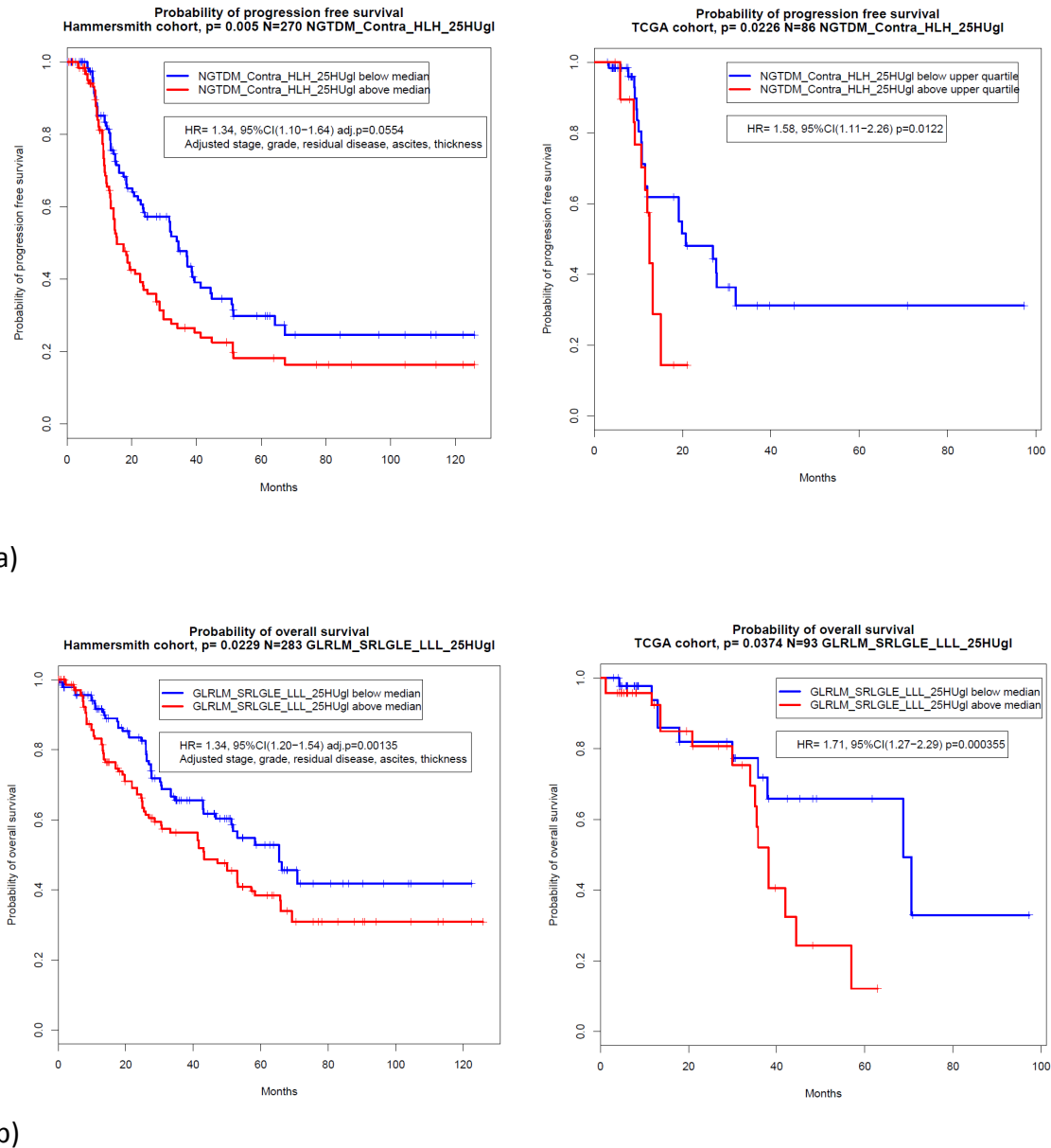
Figure 7.14 Texture Features may predict survival

(a) For PFS: using the texture feature NGTDM$_{contrast}$ split at the median, the KM curve had a p=0.005 in the training set and on multivariate analysis (adjusted for stage, grade, residual disease, ascites and thickness): p=0.055, HR 1.34, 95% CI 1.1-1.64, and in the test set (TCGA) on the KM curve p=0.003 and on the multivariate analysis, p=0.012, HR 1.5, 95% CI (1.11-2.26).

(b) For OS: using the texture feature GLRLM$_{SRLGLE}$_LLL split at the median, the KM curve had a p=0.023 in the training set and on multivariate analysis

(adjusted for stage, grade, residual disease, ascites and thickness): p=0.001, HR 1.34, 95% CI 1.2-1.54, and in the test set (TCGA) on the KM curve p=0.037 and on the multivariate analysis, p=0.0004, HR 1.71, 95% CI (1.27-2.29).

7.9 Using supervised clustering using LASSO

The fact that unsupervised clustering was able to split the data into two significant groups was very promising. Thus a supervised method, which incorporated outcome data, was implemented to develop this further. We first performed Cox regression with overall survival examining each feature in turn, using the texture features from the primary tumours of the HH training dataset. To develop a more powerful model, LASSO, a penalised regression method, which was used for the NSCLC, was again implemented. Forty-two radiomic features were found to be significantly associated with OS (false discovery rate <0.05). The LASSO model utilised the 657 generated texture features and created a new texture factor, which included 4 weighted texture variables. This was tested on the training set (Figure 7.21). The newly discovered feature vector was able to give a Radiomic Prognostic Vector (RPV) for each tumour.

The RPV consists of four radiomic features: a) FD_max_25HUgl (coefficient: -0.0876), b) GLRLM_SRLGLE_LLL_25HUgl (coefficient: 0.0869), c) NGTDM_Contra_HLL_25HUgl (coefficient: 0.165), and d) FOS_Imedian_LHH (coefficient: 0.250). All the features have approximately even weighting and relate to tumor macro-architecture at the 25 Hounsfield Unit gray level (and discrete wavelet filters).

An unsupervised k-means clustering approach was used to split the patients from the three cohorts based on their RPV into three subgroups (low-, medium, and high-risk). The patient groups stratified by RPV differed significantly on overall survival on the training dataset (N=136, p< 0.0001; log-rank test). Using the same boundaries as the testing set, the independent datasets were also tested and this confirmed the findings (HH validation set, N = 77, p = 0.0274, log rank and TCGA validation dataset, N=70, p = 0.000105, log rank, Fig 7.21)

In a Cox regression model including age, stage, post-operative residual disease, neo-adjuvant chemotherapy and the scan thickness, RPV remained significant and remained associated with OS in the training dataset (Hazard ratio: 3.83, 95% Confidence Interval (2.27-6.46), p= 5.11x10-7; RPV range: -0.322 to 3.16), as well as the TCGA validation dataset (HR: 4.87, 95% CI (1.67-14.2), p= 0.00380) and the HH validation dataset (HR: 7.36, 95% CI (1.29-41.9), p= 0.0245). Age, stage and post-operative residual disease were significantly associated with OS in either uni- or multivariable analysis in the combined HH cohort while RPV remained the strongest prognostic factor, suggesting RPV is prognostic in a representative HGSOC cohort.

Notably, RPV possessed a better prognostic power when compared to the existing prognostic markers including CA125 and the transcriptome-based molecular subtype and potentially synergises with existing CT-based morphological approaches. Apart from prognosis, high RPV was found significantly associated with primary chemotherapy resistance, shorter PFS and

poor surgical outcome (Fig. 7.21) suggesting RPV as a potential predictive marker in HGSOC.
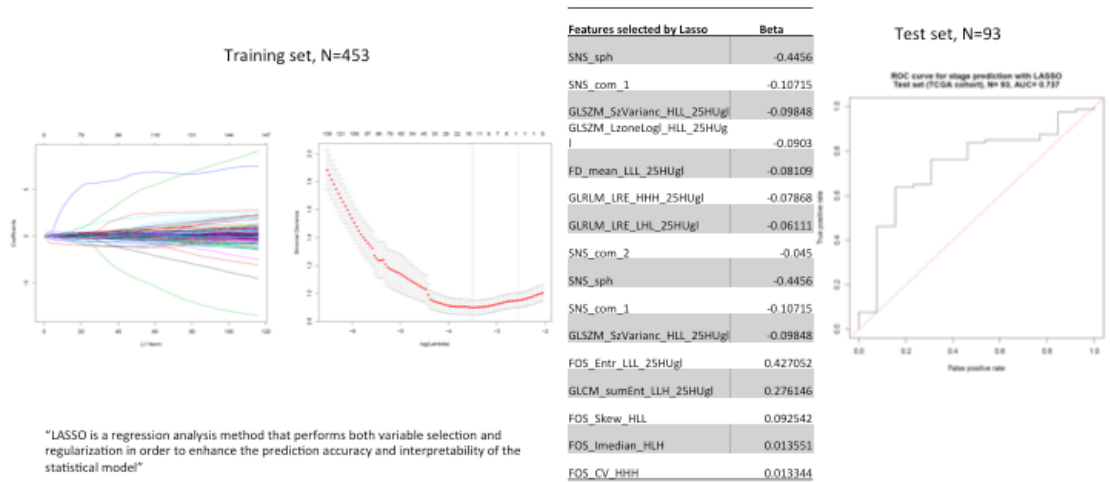
Figure 7.15 Texture features predict stage using LASSO regression. Using the LASSO method for penalised regression, feature reduction was utilised and the 16 features signature (feature value) was tested on the training set with very significant results. There was an AUC (area under the curve) of 0.737. The reason why N is 453 is because all the right and left sided ovarian masses were treated as separate.

7.10 Comparing the radiomic data between the datasets

Although, the validation dataset has been incorporated in testing the unsupervised models, further details and analysis were needed to ensure that the validation dataset was similar to the training set. This was to ensure that neither dataset has unique features that would preclude to making any general conclusions and also to reduce the possibility of false discovery. Using different

methods to compare the radiomic datasets for the two cohorts, it was seen that there were no significant differences between the two groups (Figure 7.16). The Pearson rank correlation between the two groups was very high (r = 0.817, p < 2.2 x10^{-16}). The same methods were adopted for comparing the molecular portions of both datasets, which is expanded in the next section.

a)

b)
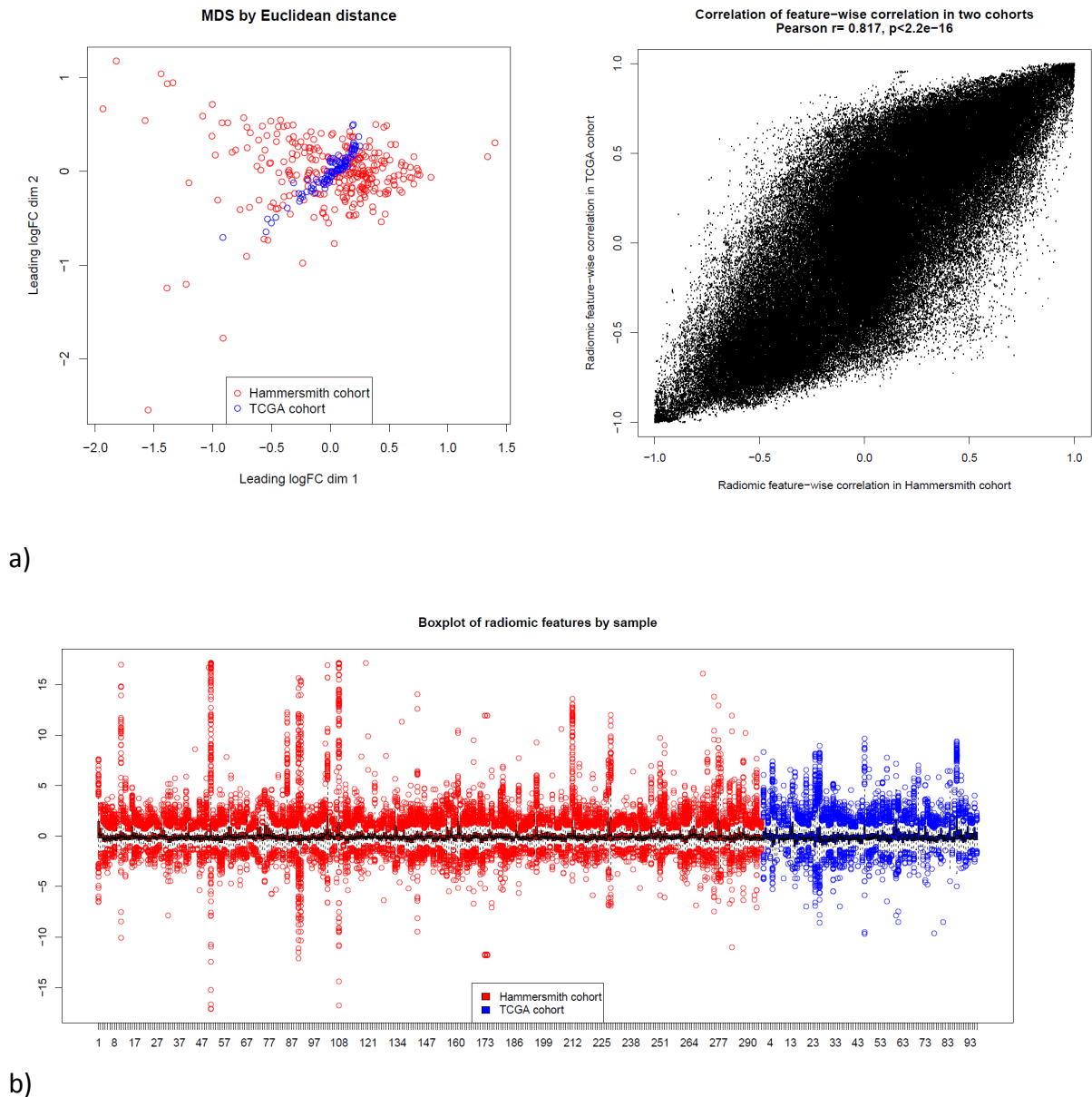
Figure 7.16 The Radiomic data from the two cohorts is similar

Using the multidimensional scaling (MDS) by a) Euclidean method and Pearson

Correlation, there was a very strong correlation between the two datasets (in the

latter r = 0.817, p < 2.2 x10^{-16}). (b)The boxplot of the features also demonstrates

this.

Using the gene expression profiles collected in parallel with radiomic profiles, a surrogate marker of RPV was constructed based on a weighted list of mRNA expressions in the TCGA validation dataset where both CT scans and gene expression profiles were available (eRPV). The eRPV strongly correlated with RPV (r= 0.720) in the TCGA validation dataset and significantly interacted with RPV in the Cox regression model. It showed a similar prognostic potential as RPV in two additional cohorts (TCGA dataset without publicly available CT scans: n=448, HR= 2.19, 95% CI (1.23 - 4.25), p= 0.0208; Tothill dataset: n= 228, HR= 7.94, 95% CI (2.02 - 31.3), p= 0.00303; adjusted for stage, grade, residual disease, age and neo-adjuvant chemotherapy). The eRPV was thus considered as a surrogate of RPV and subsequently used eRPV in a subset of the TCGA dataset without publicly available CT scans, as an extension of RPV (Noted as 'eRPV' in Fig 7.21).

Overall, we observed RPV to be associated with overall survival, independent of known clinical prognostic factors, suggesting that it may reflect distinct aspects of clinically relevant variation across HGSOC.

7.11 The molecular pathways

The underlying molecular pathways and their associations with texture analysis were further explored. First, the RPPA (reverse phase protein arrays) quality in the training set (Figure 7.17) was demonstrated in an MDS plot. This permitted

the removal of low tumour content samples (<30%). This was important for further analysis as the validation set included high tumour content samples only.

Secondly, further comparisons of the protein data between the two datasets were examined (Figures 7.17 and 7.18). It was demonstrated that there was no significant difference between the two groups, which was very important when using the sets to test the findings on the training set. Again two different methods were used to demonstrate this, the Euclidean and Pearson methods and the boxplot method (Figure 7.19a and b). The Pearson rank correlation was (r=0.48, p <2.2 x10^{-16}). Although this was not as high as the texture features, it was still significant.

A different method to demonstrate the similarities between the two groups was used to quantify the relative amount of specific proteins in both groups (Figure 7.14). This demonstrated the levels of transglutaminase and Cyclin B1 proteins were similar in the training and validation sets. The levels of transglutaminase compared with stage in the training and validation sets were similar and significant. Comparing Cyclin B and grade demonstrated a Spearman Rank of r=0.447, p= 5.7 x10-6 in the training set and r=0.190, p= 0.00015 in the validation set.
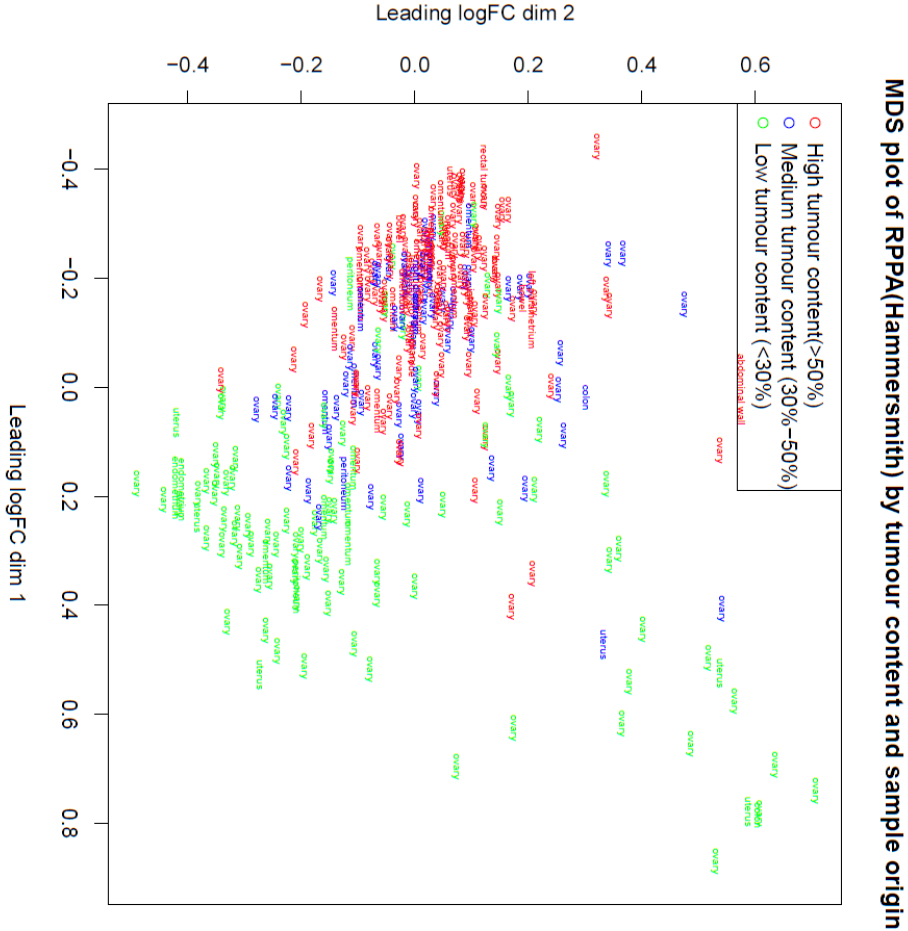
Figure 7.17 RPPA quality control

Using the MDS plot of the protein by tumour content and sample origin, the low tumour-content samples were removed.

MDS plot of RPPA(Hammersmith) by tumour content and sample origin

a)

b)
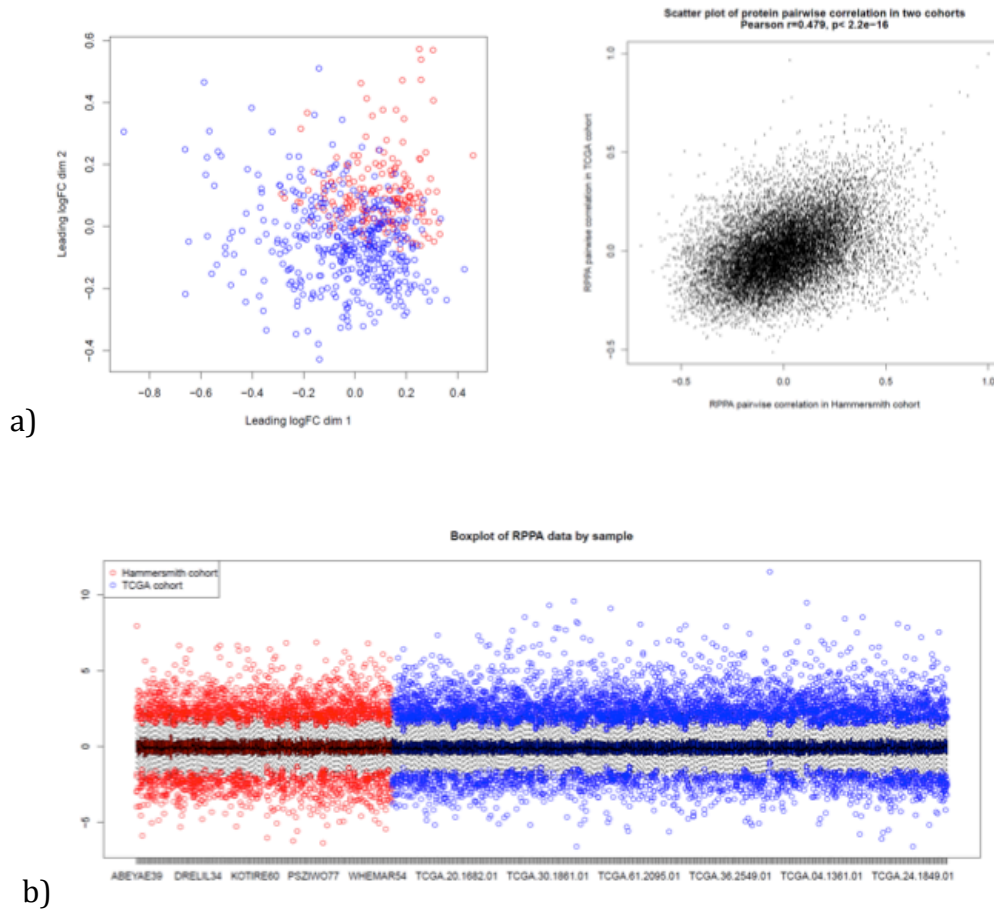
Figure 7.18 The RPPA data in the two samples is similar

The MDS by (a) Euclidian, Pearson correlation (r=0.48, p <2.2 x10^{-16}) and (b) boxplot demonstrate the similarity of the two datasets. There is, however, less correlation than that compared with the texture features.

a)

b)
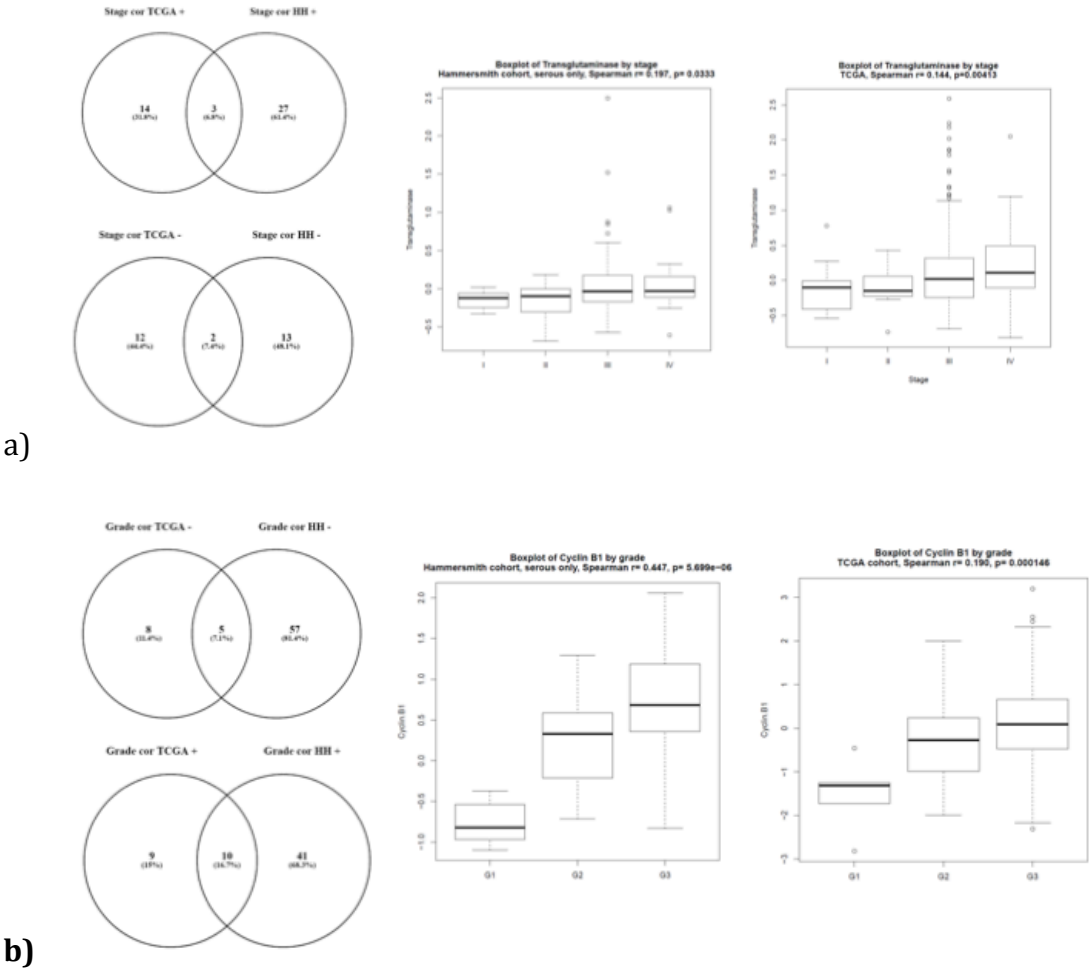
Figure 7.19 The RPPA data in the two samples is similar part 2

Boxplots of (a) transglutaminase and (b) Cyclin B1 show very similar levels across stage in the training and validation sets.
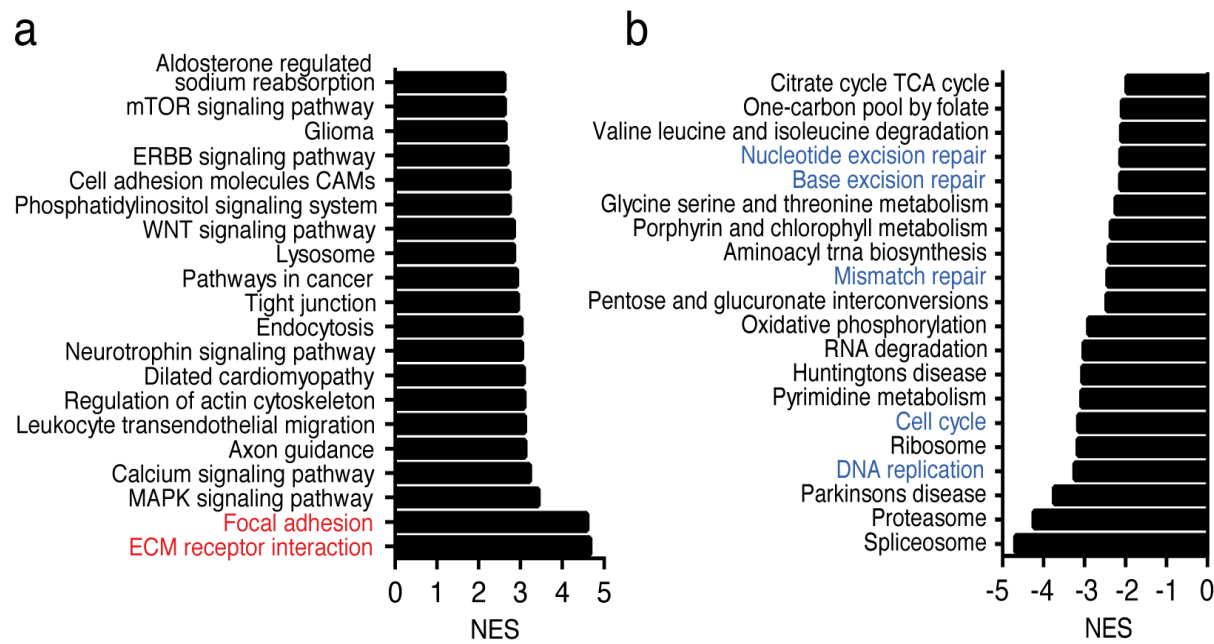
Fig 7.20 Molecular characteristics associated with RPV in HGSOC. The Gene enrichment analysis identified a) RPV-positively correlated biological pathways and b) RPV-negatively correlated pathways (FDR <0.05).
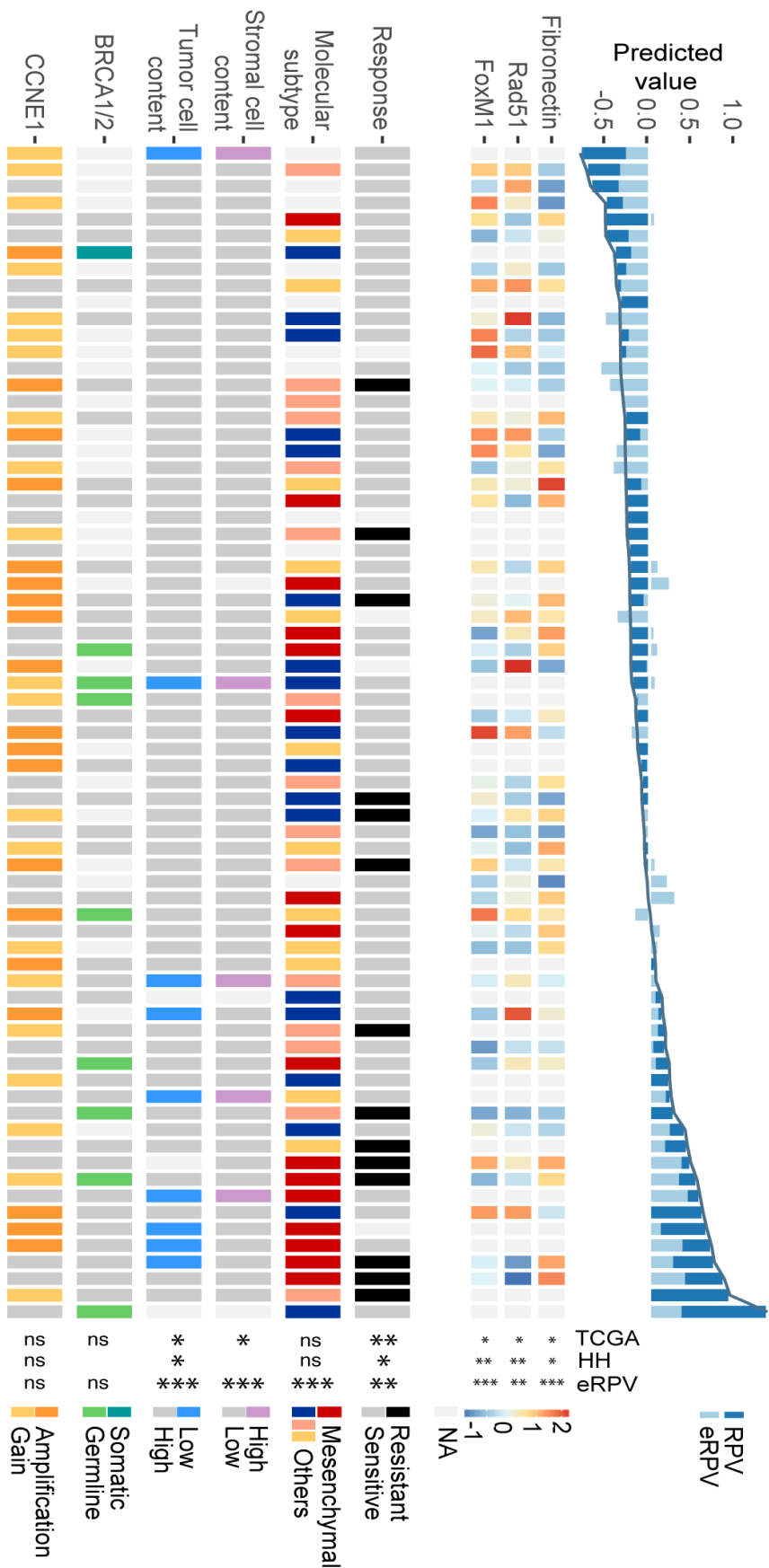
Fig 7.21 Heatmap and associations of RPV in HGSOC Heatmap demonstrating protein expression correlation (Fibronectin, Rad51 and FoxM1) with RPV for 47 cases in the TCGA validation dataset. These protein features were compared with RPV with significance shown for 119 cases in the HH cohort and the eRPV from 353 additional TCGA cases. In the top panel the RPV is ranked from low to high (left to

right) and their corresponding eRPV (light blue). In the lower panel, the protein expression level of the Fibronectin, Rad51 and FoxM1. The p-values are given by one-sided Spearman's correlation test as validation of the transcriptomic analyses. The clinical, histological and genetic characteristics associated with the RPV in the TCGA and HH cohorts are shown. A single rectangle block represents a patient in the TCGA validation dataset. The significance of the association between these characteristics with the RPV in the two validation datasets and the additional TCGA dataset is shown on the right. The significance was calculated from Krukal-Wallis (molecular subtype) or two-tailed Wilcoxon rank-sum test. The association between RPV and proliferation or DNA damage is highlighted. *** p<0.001, **p<0.01, *p<0.05, ns p>0.1.

7.2 Peritoneal Cancer Index on preoperative CT in ovarian cancer

297 patients were eligible and the surgical outcome was available in 277 (93%) patients. The tumour grading was available in 187 (63%) (Table 7.3).

The surgical procedure included a total hysterectomy and bilateral salpingo-oophorectomy, omentectomy, lymphadenectomy in the case of bulky lymph nodes. There was also careful evaluation of all peritoneal surface with resection of any suspicious peritoneal lesions or other tumour involved organs or structures. A diaphragmatic stripping or full thickness resection was performed in 61% of patients,

splenectomy in 18%, liver capsule resection in 37%, a bowel resection in 63%, a pelvic lymphadenectomy in 43% and a para-aortic lymphadenectomy in 43%. Total macroscopic tumour clearance was achieved in 77% of the patients.

Serous ovarian cancer (both high- and low- grade subtypes) had significantly higher CT-PCI in comparison with all the other histological subtypes (p<0.001) and was associated with a significantly more frequent upper abdominal and bowel involvement (p<0.001).

Table 7.3 Descriptive Analysis of those included in the Ovarian CT-PCI with clinical and PCI details

AGE		MEDIAN	RANGE		
		61	19-91		
HISTOLOGY	SEROUS	ENDOMETRIAL	MUCINOUS	CLEAR CELLS	OTHER
	237/297 (79.8%)	17/297 (5.7%)	12/297 (4.0%)	12/297 (4%)	19/297 (6.5%)

FIGO STAGE	I	II	III	IV
	45/297 (15.2%)	26/297 (8.7%)	155/297 (52.2%)	71/297 (23.9%)

BOWEL INVOLVEMENT AT CT				
	54/297 (18.2%)			

HISTOLOGICAL GRADE	1	2	3	
	13/187 (7.0%)	42/187 (22.4%)	132/187 (70.6%)	

Surgical outcome	COMPLETE CYTOREDUCTIVE	INCOMPLETE CYTOREDUCTIVE		
	214/277 (77.3%)	63/277 (22.7%)		

CT PCI VALUE	MEDIAN	MEAN	RANGE		
	9	10.9	0-39		

CT-PCI groups	CT-PCI=0	0<CT-PCI≤5	5<CT-PCI≤10	10<CT-PCI≤20	CT-PCI>20
	85/297 (28.6%)	41/297 (13.8%)	27/297 (9.1%)	81/297 (27.3%)	63/297 (21.2%)

DISEASE FREE SURVIVAL (MONTHS)	MEDIAN	MEAN	RANGE
	14.6	25.5	0-126

OVERALL SURVIVAL (MONTH)	MEDIAN	MEAN	RANGE
	24.7	34.4	0-126

7.2.1 Correlation between CT-PCI and postoperative residual disease

CT-PCI was positively associated with the risk of postoperative residual disease of any size (OR=1.04, 95%CI: 1.01-1.07; p=0.003). The median CT-PCI of patients with residual disease after surgery was 16 (interquartile range, IQR: 6-20), while amongst patients with complete tumour cytoreductive surgery, it was 7 (IQR: 0-19). The proportion of patients with suboptimal cytoreduction was 32% (31/97) among those with PCI>16 and 18% (32/180), amongst those with PCI<=16 (p=0.001). The ROC curve analysis returned an area under the curve (AUC) of 0.63 (95%CI: 0.56-0.71). The best positive predictive value for complete resection was 0.36, corresponding to a PCI threshold of 9, while the best negative predictive value was 0.90, corresponding to a PCI threshold of 16.

7.2.2 Correlation between CT-PCI interval categories and survival

Considering the value of CT-PCI, the Kaplan-Meier curves show differences between the five groups based solely on the PCI intervals, especially for patients with no visible disease at the completion of surgery (Fig 7.21). Survival analysis according to Cox-regression model showed that FIGO stage (p<0.001), histological sub-type (p<0.01) and the CT-PCI value (HR=1.03, 95%CI: 1.00-1.05; p<0.05) were significant predictors of PFS. Stage and histological subtype were time-dependent variables; whilst their estimate as prognostic markers of recurrence was maximal at the time of surgery, and declined with time. This

means their strength as prognostic factor for recurrence is very high at the time of surgery but declines afterwards; in contrast, a relatively higher CT-PCI maintained the risk of recurrence stable across time, compared to a lower CT-PCI.

a)

b)
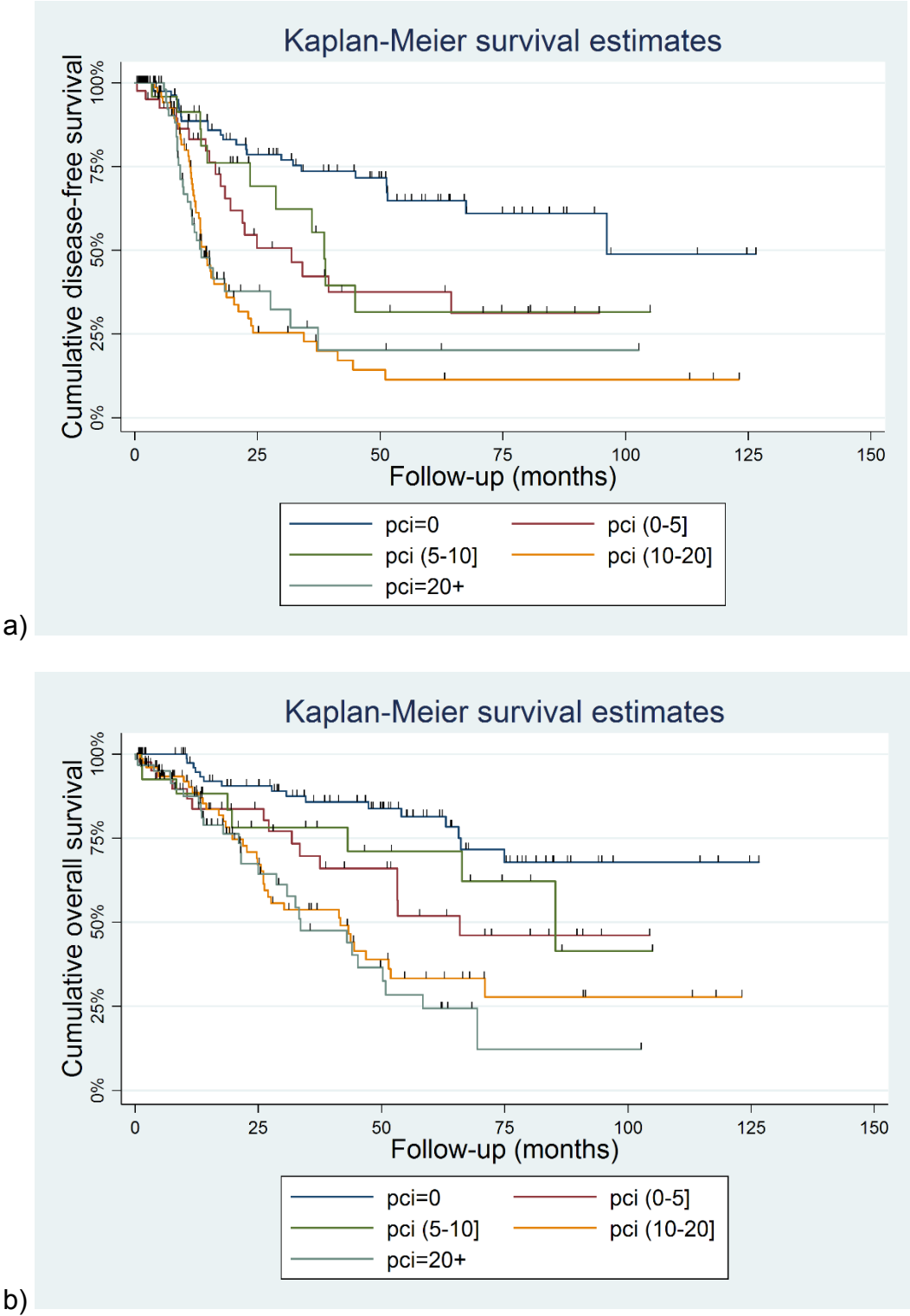
Fig 7.22: Kaplan-Meier survival curves of patients, across 5 ordered PCI interval categories. Panel a) Progression Free Survival. Panel b) Overall Survival.

Tick marks on each line represent censored data.

Similar findings were obtained for OS, where in multivariate Cox-regression analysis, FIGO stage (p<0.001) and histology (p<0.01) and CT-PCI (HR=1.03, 95%CI: 1.01-1.06; p<0.05) were independent predictors of OS (all of them satisfy the proportional hazard assumption with time). Kaplan-Meier curves for OS of the five groups are shown in Fig7.21b).

7.2.3 Correlation between CT-PCI anatomical categories and survival

The population was divided into four groups based on the abdominal regions involved on CT, independently of the PCI value: 1. No peritoneal disease, 2. Peritoneal disease in the lower abdomen only, 3. Presence of peritoneal disease in the upper abdomen, 4. Peritoneal disease in the upper abdomen with small bowel involvement.

In multivariable Cox-regression analysis on 259 patients (excluding patients with mucinous carcinoma for small numbers and with missing data), those with peritoneal involvement of the lower abdominal regions (HR=2.04, 95%CI: 0.94-4.45; p=0.07) and with bowel involvement (HR=2.28, 95%CI: 1.25-4.16; p=0.007) had a shorter PFS in comparison with patients without peritoneal spread.

With respect to the OS, patients with tumour involvement of the small bowel were estimated at higher hazard of death or recurrence compared to patients with no region involvement (HR=3.01, 95%CI: 1.49-6.08; p=0.002).

Chapter 8 Discussion – Ovarian cancer

8.1 Overview

The results of the ovarian masses data will be outlined and then the study strengths and weaknesses are discussed. The role of this data in the larger context and applicability to the clinical context will also be discussed.

8.1.1 Results summarised

The results demonstrated that both an unsupervised clustering method (using hierarchical clustering) and supervised clustering (using LASSO) were linked with survival. After modifications, the unsupervised clustering method had an association with PFS ((p=0.002, HR 1.78, 95% CI 1.24-2.57) and when adjusted for stage and ascites p=0.0277, HR 1.65, 95% CI (1.06-2.57). There unsupervised clustering also had an association with OS, p=0.0214, HR 1.59 (95% CI 1.07-2.36), although the significance disappeared when stage and ascites were taken into account (p=0.755).

Using the supervised clustering method, a feature vector was developed that had an area under the curve of 0.737 for stage prediction in the validation dataset. A four feature weighted texture feature developed on the training set, called the Radiomic prognostic vector (RPV) was able to differentiate the two validation

cohorts into low-, medium and high-risk groups (TCGA: N=70, p= 0.000105, log-rank test; Fig 7.20e) and the HH validation dataset: N= 77, p= 0.0274).

The similarities between our dataset and the publically available TCGA (validation) dataset were explored and the texture features (Pearson rank, (r = 0.817, p < 2.2 x10^{-16}) and RPPA (Pearson correlation (r=0.48, p <2.2 x10^{-16})) were significantly similar and correlated, although the association was stronger with the texture features.

The CT-PCI positively correlated with the probability of postoperative residual disease (p=0.0001). The performance of CT-PCI on predicting resectabilty in ovarian cancer was suboptimal (AUC = 0.64). Using the CT-PCI as a surrogate marker of disease burden correlated with progression free and overall survival. The involvement of specific anatomic regions on the pre-operative CT, such as the small bowel, was a negative prognostic factor for the overall survival (p=0.001).

8.2 Biological interpretation of the radiomic prognostic vector

The ECM-receptor interaction and focal adhesion were the two pathways thatwere most significantly enriched for associations with high RPV. These two pathways contained ECM components (*TIMP3* (*r*= 0.530), *COL11A1* (*r*= 0.460)) and focal adhesion receptors (*ITGA5* (*r*= 0.368), *ITGB5* (*r*= 0.387)), and from previous studies both pathways were enriched in stroma (Finak et al., 2008,

Malhotra et al., 2012). Accordingly, gene expression that correlated with high RPV were significantly enriched for genes expressed in the stromal component (Fig 7.20c, chi-squared test p<0.0001). Additionally, RPV was positively correlated with a stromal marker, fibronectin, at the protein level in both the TCGA and the HH cohorts (Fig 7.20d). High RPV was also associated with a high proportion of tumor-associated stromal cells, evidenced from both histological data (Fig 7.20e) and stroma score estimated from transcriptomic data (Yoshihara et al., 2013). Lower tumor cell content is inversely related to high stromal content. Consistent with previous results, we noted that higher RPV was associated with lower tumor cellularity (Fig 7.20e) in the TCGA cohort and in the HH cohort.

Besides stroma-related pathways, a number of proliferation and DNA damage response (DDR) pathways, including DNA replication, cell cycle, mismatch repair, base excision repair, nucleotide excision repair and homologous recombination, were among the top pathways activated in the RPV-low tumors (Fig 7.20b). To verify the validity of the pathway analysis, an analysis was performed on reverse phase protein array (RPPA) data from both HH and TCGA cohorts and the expression proliferation and DDR pathway marker proteins including Stathmin 1, FoxM1 and Rad51 was found to be higher at the protein level in tumors with low RPV in the two independent datasets (Fig 7.20d), which was consistent with the transcriptomic and pathway analysis. Existence of highly proliferative cancer cells with impaired DNA damage response mechanism (e.g. *TP53* mutation) could elicit accumulation of DNA damage (Gaillard et al., 2015). Accordingly,

higher tumour mutational burden and CNA burden were observed in RPV-low tumours. Collectively, these molecular features suggest that RPV-low patients could benefit from therapeutic DDR inhibitors (PARPi) and immunotherapy (anti-PD1/PD-L1) (Rizvi et al., 2015).

Molecular subtype, *BRCA1/2* mutations and *CCNE1* amplification are well-established molecular characteristics contributing to primary chemotherapy response and prognosis. However, they were not found correlated with RPV, highlighting the independent disease mechanisms associated with RPV (Fig 7.20e).

Overall, the stromal phenotype seem sot be activated in the high RPV tumours and the proliferation and DNA damage response pathways in low RPV tumours, all of which are potential therapeutic targets in HGSOC.

8.2.1 Texture RPV and underlying biological significance

In biological terms, the individual components of RPV combine to define tumour mesoscopic structure: a) maximal fractal dimension of the tumour and its microenvironment, which was negatively correlated with survival, together with the following positively correlated features; b) proportions of runs that have short lengths in the low pass filtered image; a function which gives coarse low-density textures, e.g. intermixed fibrotic stroma and tumour cells; c) the average visual contrast across the tumour weighted by sharpening in the X-axis, and

blurring in the Y and Z axes reflecting local heterogeneity, and d) the median of the distribution of voxel intensities across the entire tumour weighted by blurring in the X-axis and sharpening in the Y and Z axes, reflecting global heterogeneity, respectively.

In addition to building a prognostic model, we further demonstrated that the radiomics-derived signature is closely linked to a stromal phenotype and DNA-damage response through genetic, transcriptomic, proteomic and histological analysis. This finding is consistent with the poor prognostic value of stromal phenotype identified in many cancers including ovarian (Busuttil et al., 2014, Zhang et al., 2015, Chen et al., 2015, Cheung et al., 2013, Bonome et al., 2005), pancreatic (Wang et al., 2016a), prostate (Mo et al., 2017), colorectal (Isella et al., 2015, Cheung et al., 2013), gastric (Busuttil et al., 2014), lung (Cheung et al., 2013) and breast cancer(Frings et al., 2013).

The tumour stroma consists of immune cells, endothelial cells, fibroblasts and extracellular-matrix (ECM)(Kalluri and Zeisberg, 2006) all of which could directly contribute to outcome via distinct mechanisms in EOC (Mhawech-Fauceglia et al., 2015, Davidson et al., 2014, Sherman-Baust et al., 2003, Wang et al., 2016b). There has been a demonstration, based on the strong association between RPV and response to primary chemotherapy or surgery, that patients with high RPV have a significantly high risk of failing quality surgery or systemic strategies and suggest that they possibly need to be directed towards alternative therapeutic approaches including stroma modifying therapies

Interestingly, in our HGSOC cohort we did not observe a strong association between RPV and any single cancer driver events including ovarian cancer 'molecular subtypes', specific gene mutations or CNA, suggesting that the RPV phenotype and related poor prognosis may be shaped by non-canonical genetic alterations or pathways.

8.2.2 CT-PCI

The radiological PCI correlated with the probability of post-operative residual disease in patients with epithelial ovarian cancer undergoing primary cytoreductive surgery. In addition, consistent with other studies, serous histology was significantly associated with higher CT-PCI scores. CT-PCI was however unsuccessful as a tool to identify patients who would not benefit from cytoreductive surgery in terms of total macroscopic tumour clearance.

These results partly correspond with data from previous studies carried out with other tumour groups. Some studies in peritoneal carcinomatosis in gastrointestinal cancer found that CT PCI is not a good indicator of subsequent surgical findings (Rivard et al., 2014), the sensitivity is dependent on lesion size (Koh et al., 2009) and the CT-PCI has a lower sensitivity and underestimates the volume of disease (Chua et al., 2011) compared with surgical PCI. On the other hand, even if CT-PCI is lower than surgical PCI, the clinical significance of this

difference is limited (Esquivel et al., 2010) and that the CT-PCI in gastrointestinal cancer is helpful to predict resectability and survival (Suzuki et al., 2018).

In ovarian cancer, there has been limited application of preoperative radiological assessment of PCI. In a small study (Schmidt et al., 2015) of 15 patients, ten of whom had peritoneal carcinomatosis, CT, MRI and PET-CT had good accuracy in estimating the intraoperative PCI. A further study (Mazzei et al., 2013) found that CT could be used as single technique to select patients for either primary cytoreductive surgery or neoadjuvant chemotherapy, if performed with a dedicated protocol and read by an expert radiologist.

Our results contrast with the previously reported experience in ovarian cancer, as the performance of CT-PCI in predicting total macroscopic tumour clearance was unsatisfactory. However, even if it cannot be used alone as a reliable triage tool to avoid suboptimal surgery, the positive association between CT-PCI and postoperative residual disease supports further investigation, as it may be an auxiliary predictor of patients at high-risk for residual disease. The value of the odds ratio of CT-PCI for incomplete cytoreduction may seem low but, considering that it indicates the relative risk for each point of CT-PCI and that CT-PCI has a wide spectrum of values (0-39), the difference of relative risk may differ consistently.

The prognostic value of surgical PCI is helpful and recent studies have indicated that this may also be so in ovarian cancer (Gasimli et al., 2015, Lampe et al., 2015). Very little data has been published about the prognostic value of

radiological PCI: in gastrointestinal cancer (Suzuki et al., 2018) CT-PCI correlated with survival; in ovarian cancer, CT-PCI is correlated with CA-125 level and 5-years survival rates (Diaz-Gil et al., 2016).

Our study found that CT-PCI was an independent prognostic factor of both PFS and OS. Thus, it may aid in stratifying patient risk, especially when the peritoneal involvement differs widely (it may happen in FIGO stage III and IV). In particular, patients with serosal bowel involvement, which is a challenging site for complete tumour resection especially when the mesentery is involved in a diffuse fashion, are at increased risk, with a demonstrated shorter OS (Rosendahl et al., 2018). The negative impact of high radiological tumour burden on patients' clinical outcome was expected since a high initial tumour burden indicates more aggressive and more advanced disease that is inherently associated with poorer outcome. Horowitz et al. examined data from 2,655 patients with EOC and concluded that initial tumour burden was a significant prognostic factor even after adjusting for residual disease (Horowitz et al., 2015).

8.3.1 Study strengths

The biggest strength of this study was that it was large and used the expertise and resources available to greatly expand our knowledge and understanding of texture analysis, on a tumour group that has not been previously extensively studied with texture analysis. In this way, a comprehensive database has been developed which may provide valuable information in the foreseeable future. Although at the time of writing, complete results were unavailable.

The other strength of this study is that it has built on the previous knowledge of texture analysis, and built a system to extract many texture features easily from available images.

Another major advantage was the ability to test out our findings on a completely independent dataset which had both molecular and CT outcome data available (TCGA). The ability to utilise publically available data will only enhance future studies and make the findings more robust.

8.4 Study limitations

8.4.1 Sample Size

All data were obtained from one super-centre, although it can be seen that (See methods Section 7.1 and Table 7.1) there were a wide number of affiliated centres, which supplied CT scans. Interestingly, despite the extensive network, more than half of the scans (53.4%) were imaged in Imperial College NHS Trust. The end-result was that this cohort was the largest and at the time of writing one of two texture analysis studies(Vargas et al., 2017) for ovarian cancer on CT.

The ability of texture to correlate with prognosis demonstrates promising findings. There is a concern that utilising a large number of texture features on a limited number of studies can result in over-fitting of the data and type 1 errors

(Chalkidou et al., 2015), so the study design is important to ensure that the results can be generalised to other datasets. Testing the texture discovery features on independent validation set has strengthened the statistical findings. In addition, the utilisation of machine learning, although novel, does require a very large number of subjects, somewhere in the order of thousands as opposed to hundreds. This was similar to the conclusions drawn from the Lung texture study (Chapter 3). It remains to be seen if a large enough cohort will ever be tested.

Another issue, which is allied to the previous point, is the number of hours required to curate the data. This has been discussed previously.

To advance the field, a large prospective study, utilising many centres in order to increase numbers is warranted. A clear advantage identified is that the texture features are able to recognise poor prognosis earlier than even stage.

Another limitation is that the stromal component contains a mixture of cells of different origin and extracellular matrix composition, thus the exact elements in the stroma measured by RPV are unclear. A study to associate the RPV with each stromal component including but not limited to the fibroblast activation, immune cell infiltration and ECM density is necessary to better understand na delineate the prognostic power of RPV. In addition, as discussed earlier, the EOC patients often had bilateral disease and one tumour was chosen to represent the

patient. Further investigation into heterogeneity in RPV for bilateral tumours may help elucidate this further.

In the PCI-CT part, the retrospective nature of the study meant that we could not directly correlate the CT-PCI with surgical PCI as the surgeons did not systematically report it.

8.4.2 Effects of CT imaging parameters on the results

8.4.2.1 Effects of image resolution

The scans for this study were taken over a period of over 15 years starting from the beginning of the last decade. Over the course of this period, there has been extensive development in the resolution of CT technology with slice thickness reducing from 10mm to less than 1mm over a decade. Imaging resolution, as a result, appears to have had an effect on the texture analysis. As there has been such a mixture of CT scans entered into this study, from different manufactures, we have been able to study the effects of different scanners, manufacturers and matrices on CT for the first time. Some of this has yielded results showing that these parameters have no effect on the texture parameters (Manufacturer and model type), whereas others have shown that these parameters have an effect (slice thickness). As these changes are a more accurate reflection of the real-life situation, this has potentially proven that texture analysis is resilient and useful in the real-life setting. The work is currently on going but will hopefully pave the

way for more studies so as to implement a more realistic model for texture analysis.

8.4.2.2 Effects of patient motion, positioning and contrast

The effects of patient motion and positioning had a minimal impact on the scans and subsequent texture analysis. CT scan acquisition is rapid, often lasting less than a minute. Although there is movement artefact from bowel peristalsis, etc. we did not encounter this in our experience. The only issues that we did encounter were metallic artefact (for instance, from metallic hip work) causing photon starvation and potential abnormal texture features, which we had to exclude. This only affected less than 0.5% of the scans used.

Unfortunately, no studies exist that compare the differences in the texture features obtained between contrast and non-contrast studies. Presuming that there will be an effect with contrast on the texture features, all non-contrast scans were excluded.

8.4.2.3 Effects of image reconstruction parameters

The effects of image reconstruction have been shown to affect the texture analysis in PET studies (Galavis et al., 2010). However, the equivalent studies do not exist in CT studies. We did demonstrate a modest effect of slice thickness on the variability of the texture features although more work is required in this area.

8.4.2.4 Target segmentation

Manual segmentation was incorporated only on this study. The effect of segmentation on texture stability is currently being explored.

Of note, this is the rate-limiting step in terms of time, which could be a potential problem in the clinical implantation of the technology in the future. Many man-hours are required to segment tumours and check them. An easier way to semi-automatically segment the tumours was not explored, as in the PET study performed in the first part of this research. The contrast differences on CT are much less than on PET.

8.4.3 VOI size

There has been a lot of work in the PET literature of the effect of VOI size(Hatt et al., 2010b, Hatt et al., 2011a, Hatt et al., 2009, Orlhac et al., 2014) on texture analysis, especially the lowest volume permissible. The equivalent studies on CT scans have not been established (however a recent study on this subject on CT phantoms has revealed some interesting results (Shafiq-ul-Hassan et al., 2017)). Certainly, the size of the voxels on CT is much smaller than those seen on PET (approximately $1 \times 1 \times 2$ mm^3 in CT compared with $2.6 \times 2.6 \times 2.4$ mm^3 or $5.2 \times 5.2 \times 2.4$ mm^3 in PET (Jentzen et al., 2007)) and unlike the latter, the size of the voxels and hence the resolution has decreased significantly in CT over the last decade. In the only CT study to date 150/213 texture features were reproducible across voxel size (Shafiq-ul-Hassan et al., 2017).

8.4.4 Statistical methods and machine learning

A variety of techniques were used in this study, some of which were utilised for the first time having been initially developed in the field of multi-array RNA studies (E Bair, 2006, Tibshirani, 1996, Tibshirani, 1997). The number of texture features easily outnumbers the number of cases and the potential for type 1 statistical errors is large as a result, but these problems were also encountered in other areas, such as the aforementioned multi-array RNA, and penalised regression (like the use of LASSO in our study) and using independent datasets (from the TCGA dataset) reduced these types of errors. The use of independent datasets has shown the potential for this technique and potentially validates the results obtained.

8.5 Perspectives and future work

This study has generated a lot of data, some of which may form the basis of future work and some of which may take many years to analyse in the context of matching whole genomic sequences with CT, RPPA and RNA sequencing(Leong et al., 2015). The omental and peritoneal disease texture has not been analysed as of the time of writing.

The semantic data that was gathered has the potential to identify disease patterns that correlate with outcome. The findings from the Peritoneal Cancer Index (not presented) are due to be published. With the continuation of building up the dataset with more imaging and biological samples, there is scope for interesting developments in data mining allowing hypothesis generation.

8.6.1 Tumour heterogeneity and underlying biology

There was addition information regarding how texture is correlated to the biological pathways of tumours and underlying molecular data. Thus, this study may go some way to try to explain what the underlying texture features mean biologically but at the time of writing this has not been completely finalised.

8.6.2 Optimal parameters to image heterogeneity with CT

If the underlying molecular and proteomic data were reflected in the appearances on the CT, this would provide a novel non-invasive approach to guide treatment. However, there is still some way to go before this is realised.

One of the parameters explored in this study was the effect of slice thickness on texture analysis. The slice thickness is intrinsic to the type of scanner, and although in post processing it is possible to make the slices thicker than when they were acquired, the converse is not true. We have witnessed a revolution in the CT scanning equipment in this study where thicker slices with higher doses have been replaced with thinner slice acquisition (Shafiq-ul-Hassan et al., 2017, Goldman, 2008). More work is required on this, particularly the effect of mAs and kVp on the texture parameters(Kim et al., 2015, Miles et al., 2009). Until this time, we cannot recommend any definite effects on the texture parameters except the slice thickness (See Section 7.3).

Chapter 9 Conclusion

Texture analysis was employed in two studies, which were different from each other. The first study focused on 358 multi-centre pre-therapy FDG-PET and NSCLC. A radiomics feature vector (FV) containing two texture features was developed on 133 scan training set using LASSO regression analysis and feature discovery. In an independent validation set of 204 scans, this FV was significant (p = 0.00465; HR = 1.61, 95% CI 1.16–2.24), and independent of known prognostic factors, such as stage and tumour volume. With a median cut-off there was a fourteen-month survival difference between the two groups. In a separate smaller TEST1 cohort, who had an unusual poor survival for Stage 1 cancers, FVX indicated a lack of survival difference (N = 21, p = 0.501). PET variables such as the SUV_{max}, SUV_{mean} and SUV_{peak} lacked any prognostic information.

In the second study, 657 texture features were extracted from the preoperative CT scans of 364 epithelial ovarian cancer (EOC) patients. Using machine learning, a radiomic prognostic vector (RPV) using 4 texture features was developed. This was able to reliably identify the 5% of patients with a median overall survival of less than 2 years,, and this was validated in two independent multi-centre cohorts. In addition, genetic, transcriptomic and proteomic analysis from two independent datasets elucidated that the stromal phenotype and DNA damage reponse pathways were activated in RPV-stratified tumours.

In the same cohort, 297 patients with pre-operative EOC, had their pre-operative CTs reviewed by an expert radiologist according to the peritoneal cancer index (PCI) to assess and evaluate the extent of their peritoneal metastases, if they were present. This was compared with their operative PCI findings. Although there was a positive correlation with postoperative residual disease (Odds ratio (OR=1.04, 95% CI: 1.01-1.07, p=0.003). ROC curve analysis returned AUC=0.64 for prediction of total macroscopic tumour clearance. In multivariable analysis, patients with no peritoneal disease seen on CT had a significantly longer PFS (Hazard ratio, HR=2.28, p=0.007). Radiological serosal small bowel involvement was also an independent predictor for shorter OS (HR=3.01, p=0.002).

This work contributes to the demonstration that texture features based on image analysis can supplement existing methods in imaging to provide information that has hitherto been unavailable.

Bibliography

2017. *Ovarian Cancer Statistics* [Online]. http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer - heading-Zero. [Accessed 31/8/2017 2017].

ABHISHEK, K. & KHUNGER, N. 2015. Complications of skin biopsy. *Journal of Cutaneous and Aesthetic Surgery,* 8**,** 239-241.

AERTS, H. J., VELAZQUEZ, E. R., LEIJENAAR, R. T., PARMAR, C., GROSSMANN, P., CAVALHO, S., BUSSINK, J., MONSHOUWER, R., HAIBE-KAINS, B., RIETVELD, D., HOEBERS, F., RIETBERGEN, M. M., LEEMANS, C. R., DEKKER, A., QUACKENBUSH, J., GILLIES, R. J. & LAMBIN, P. 2014a. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun,* 5**,** 4006.

AERTS, H. J. W. L., VELAZQUEZ, E. R., LEIJENAAR, R. T. H., PARMAR, C., GROSSMANN, P., CARVALHO, S., BUSSINK, J., MONSHOUWER, R., HAIBE-KAINS, B., RIETVELD, D., HOEBERS, F., RIETBERGEN, M. M., LEEMANS, C. R., DEKKER, A., QUACKENBUSH, J., GILLIES, R. J. & LAMBIN, P. 2014b. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. 5**,** 4006.

AHMAD, A. S., ORMISTON-SMITH, N. & SASIENI, P. D. 2015. Trends in the lifetime risk of developing cancer in Great Britain: comparison of risk for those born from 1930 to 1960. *Br J Cancer,* 112**,** 943-947.

AHMED, N. & STENVERS, K. L. 2013. Getting to Know Ovarian Cancer Ascites: Opportunities for Targeted Therapy-Based Translational Research. *Frontiers in Oncology,* 3**,** 256.

AHN, S. J., KIM, J. H., PARK, S. J. & HAN, J. K. 2016. Prediction of the therapeutic response after FOLFOX and FOLFIRI treatment for patients with liver metastasis from colorectal cancer using computerized CT texture analysis. *Eur J Radiol,* 85**,** 1867-1874.

AL-KADI, O. S. & WATSON, D. 2008. Texture analysis of aggressive and nonaggressive lung tumor CE CT images. *IEEE Trans Biomed Eng,* 55**,** 1822-30.

ALAM, I. S., ARSHAD, M. A., NGUYEN, Q. D. & ABOAGYE, E. O. 2015. Radiopharmaceuticals as probes to characterize tumour tissue. *Eur J Nucl Med Mol Imaging,* 42**,** 537-61.

ALBERG, A. J., BROCK, M. V., FORD, J. G., SAMET, J. M. & SPIVACK, S. D. 2013. Epidemiology of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest,* 143**,** e1S-29S.

ALESSIO, A. & KIDNAHAN, P. 2006. PET Image Reconstruction. *In:* HENKIN, R. E., BOVA, D., DILLEHAY, G., KARESH, S. M., HALAMA, J. R. & WAGNER, R. H. (eds.) *Nuclear Medicine.* 2nd ed. Philadelphia: Elsevier Mosby.

ALIZADEH, A. A., ARANDA, V., BARDELLI, A., BLANPAIN, C., BOCK, C., BOROWSKI, C., CALDAS, C., CALIFANO, A., DOHERTY, M., ELSNER, M., ESTELLER, M., FITZGERALD, R., KORBEL, J. O., LICHTER, P., MASON, C. E., NAVIN, N., PE'ER, D., POLYAK, K., ROBERTS, C. W. M., SIU, L., SNYDER, A., STOWER, H., SWANTON, C., VERHAAK, R. G. W., ZENKLUSEN, J. C., ZUBER, J. & ZUCMAN-ROSSI, J. 2015. Toward understanding and exploiting tumor heterogeneity. *Nature medicine,* 21**,** 846-853.

AMADASUN, M. & R, K. 1989. Texural Features Corresponding to Texural Properties. *1989,* 19**,** 1264-1274.

ASSELIN, M. C., O'CONNOR, J. P., BOELLAARD, R., THACKER, N. A. & JACKSON, A. 2012. Quantifying heterogeneity in human tumours using MRI and PET. *Eur J Cancer,* 48**,** 447-55.

AVRIL, N., SASSEN, S. & ROYLANCE, R. 2009. Response to therapy in breast cancer. *J Nucl Med,* 50 Suppl 1**,** 55s-63s.

BABYAK, M. A. 2004. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosom Med,* 66**,** 411-21.

BAGCI, U., YAO, J., MILLER-JASTER, K., CHEN, X. & MOLLURA, D. J. 2013. Predicting future morphological changes of lesions from radiotracer uptake in 18F-FDG-PET images. *PLoS One,* 8**,** e57105.

BALAGURUNATHAN, Y., GU, Y., WANG, H., KUMAR, V., GROVE, O., HAWKINS, S., KIM, J., GOLDGOF, D. B., HALL, L. O., GATENBY, R. A. & GILLIES, R. J. 2014. Reproducibility and Prognosis of Quantitative Features Extracted from CT Images. *Transl Oncol,* 7**,** 72-87.

BARRINGTON, S. F., MIKHAEEL, N. G., KOSTAKOGLU, L., MEIGNAN, M., HUTCHINGS, M., MUELLER, S. P., SCHWARTZ, L. H., ZUCCA, E., FISHER, R. I., TROTMAN, J., HOEKSTRA, O. S., HICKS, R. J., O'DOHERTY, M. J., HUSTINX, R., BIGGI, A. & CHESON, B. D. 2014. Role of imaging in the staging and response assessment of lymphoma: consensus of the International Conference on Malignant Lymphomas Imaging Working Group. *J Clin Oncol,* 32**,** 3048-58.

BASU, S., KWEE, T. C., GATENBY, R., SABOURY, B., TORIGIAN, D. A. & ALAVI, A. 2011. Evolving role of molecular imaging with PET in detecting and characterizing heterogeneity of cancer tissue at the primary and metastatic sites, a plausible explanation for failed attempts to cure malignant disorders. *Eur J Nucl Med Mol Imaging,* 38**,** 987-91.

BERGHMANS, T., DUSART, M., PAESMANS, M., HOSSEIN-FOUCHER, C., BUVAT, I., CASTAIGNE, C., SCHERPEREEL, A., MASCAUX, C., MOREAU, M., ROELANDTS, M., ALARD, S., MEERT, A. P., PATZ, E. F., JR., LAFITTE, J. J., SCULIER, J. P. & EUROPEAN LUNG CANCER WORKING PARTY FOR THE, I. L. C. S. P. 2008. Primary tumor standardized uptake value (SUVmax) measured on fluorodeoxyglucose positron emission tomography (FDG-PET) is of prognostic value for survival in non-small cell lung cancer (NSCLC): a systematic review and meta-analysis (MA) by the European Lung Cancer Working Party for the IASLC Lung Cancer Staging Project. *J Thorac Oncol,* 3**,** 6-12.

BIEHL, K. J., KONG, F. M., DEHDASHTI, F., JIN, J. Y., MUTIC, S., EL NAQA, I., SIEGEL, B. A. & BRADLEY, J. D. 2006. 18F-FDG PET definition of gross tumor volume for radiotherapy of non-small cell lung cancer: is a single standardized uptake value threshold approach appropriate? *J Nucl Med,* 47**,** 1808-12.

BOELLAARD, R. 2009. Standards for PET image acquisition and quantitative data analysis. *J Nucl Med,* 50 Suppl 1**,** 11s-20s.

BONOME, T., LEE, J. Y., PARK, D. C., RADONOVICH, M., PISE-MASISON, C., BRADY, J., GARDNER, G. J., HAO, K., WONG, W. H., BARRETT, J. C., LU, K. H., SOOD, A. K., GERSHENSON, D. M., MOK, S. C. & BIRRER, M. J. 2005. Expression profiling of serous low malignant potential, low-grade, and high-grade tumors of the ovary. *Cancer Res,* 65**,** 10602-12.

BOUSSION, N., HATT, M., LAMARE, F., BIZAIS, Y., TURZO, A., CHEZE-LE REST, C. & VISVIKIS, D. 2006. A multiresolution image based approach for correction of partial volume effects in emission tomography. *Phys Med Biol,* 51**,** 1857-76.

BRINDLE, K. 2008. New approaches for imaging tumour responses to treatment. *Nat Rev Cancer,* 8**,** 94-107.

BROOKS, F. J. & GRIGSBY, P. W. 2011. Current measures of metabolic heterogeneity within cervical cancer do not predict disease outcome. *Radiat Oncol,* 6**,** 69.

BROOKS, F. J. & GRIGSBY, P. W. 2013a. FDG uptake heterogeneity in FIGO IIb cervical carcinoma does not predict pelvic lymph node involvement. *Radiat Oncol,* 8**,** 294.

BROOKS, F. J. & GRIGSBY, P. W. 2013b. Quantification of heterogeneity observed in medical images. *BMC Med Imaging,* 13**,** 7.

BROOKS, F. J. & GRIGSBY, P. W. 2014. The effect of small tumor volumes on studies of intratumoral heterogeneity of tracer uptake. *J Nucl Med,* 55**,** 37-42.

BUSUTTIL, R. A., GEORGE, J., TOTHILL, R. W., IOCULANO, K., KOWALCZYK, A., MITCHELL, C., LADE, S., TAN, P., HAVIV, I. & BOUSSIOUTAS, A. 2014. A signature predicting poor prognosis in gastric and ovarian cancer represents a coordinated macrophage and stromal response. *Clin Cancer Res,* 20**,** 2761-72.

CAMPBELL, L. L. & POLYAK, K. 2007. Breast tumor heterogeneity: cancer stem cells or clonal evolution? *Cell Cycle,* 6**,** 2332-8.

CASTELLANO, G., BONILHA, L., LI, L. M. & CENDES, F. 2004. Texture analysis of medical images. *Clin Radiol,* 59**,** 1061-9.

CHALKIDOU, A., O'DOHERTY, M. J. & MARSDEN, P. K. 2015. False Discovery Rates in PET and CT Studies with Texture Features: A Systematic Review. *PLoS One,* 10**,** e0124165.

CHEN, P., HUHTINEN, K., KAIPIO, K., MIKKONEN, P., AITTOMAKI, V., LINDELL, R., HYNNINEN, J., AURANEN, A., GRENMAN, S., LEHTONEN, R., CARPEN, O. & HAUTANIEMI, S. 2015. Identification of Prognostic Groups in High-Grade Serous Ovarian Cancer Treated with Platinum-Taxane Chemotherapy. *Cancer Res,* 75**,** 2987-98.

CHENG, J. C., LAFOREST, R. & O'SULLIVAN, J. A. 2011. A scatter and randoms weighted (SRW) iterative PET reconstruction. *Med Phys,* 38**,** 3186-92.

CHENG, N. M., FANG, Y. H., CHANG, J. T., HUANG, C. G., TSAN, D. L., NG, S. H., WANG, H. M., LIN, C. Y., LIAO, C. T. & YEN, T. C. 2013. Textural features of pretreatment 18F-FDG PET/CT images: prognostic significance in patients with advanced T-stage oropharyngeal squamous cell carcinoma. *J Nucl Med,* 54**,** 1703-9.

CHENG, N. M., FANG, Y. H., LEE, L. Y., CHANG, J. T., TSAN, D. L., NG, S. H., WANG, H. M., LIAO, C. T., YANG, L. Y., HSU, C. H. & YEN, T. C. 2015. Zone-size nonuniformity of 18F-FDG PET regional textural features predicts survival in patients with oropharyngeal cancer. *Eur J Nucl Med Mol Imaging,* 42**,** 419-28.

CHERRY, S. R. 2006. The 2006 Henry N. Wagner Lecture: Of mice and men (and positrons)--advances in PET imaging technology. *J Nucl Med,* 47**,** 1735-45.

CHEUNG, T. T., HO, C. L., LO, C. M., CHEN, S., CHAN, S. C., CHOK, K. S., FUNG, J. Y., YAN CHAN, A. C., SHARR, W., YAU, T., POON, R. T. & FAN, S. T. 2013. 11C-acetate and 18F-FDG PET/CT for clinical staging and selection of patients with hepatocellular carcinoma for liver transplantation on the basis of Milan criteria: surgeon's perspective. *J Nucl Med,* 54**,** 192-200.

CHICKLORE, S., GOH, V., SIDDIQUE, M., ROY, A., MARSDEN, P. K. & COOK, G. J. 2013. Quantifying tumour heterogeneity in 18F-FDG PET/CT imaging by texture analysis. *Eur J Nucl Med Mol Imaging,* 40**,** 133-40.

CHUA, T. C., AL-ZAHRANI, A., SAXENA, A., GLENN, D., LIAUW, W., ZHAO, J. & MORRIS, D. L. 2011. Determining the association between preoperative computed tomography findings and postoperative outcomes after cytoreductive surgery and perioperative intraperitoneal chemotherapy for pseudomyxoma peritonei. *Ann Surg Oncol,* 18**,** 1582-9.

CIDON, E. U. 2017. New therapeutic approaches to metastatic gastroenteropancreatic neuroendocrine tumors: A glimpse into the future. *World Journal of Gastrointestinal Oncology,* 9**,** 4-20.

CLAUSI, D. A. 2002. An analysis of co-occurrence texture statistics as a function of grey level quantization. *Can. J. Remote Sensing,* 28**,** 45-62.

CONTRACTOR, K. B. & ABOAGYE, E. O. 2009. Monitoring predominantly cytostatic treatment response with 18F-FDG PET. *J Nucl Med,* 50 Suppl 1**,** 97s-105s.

COOK, G. J., 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? *J Nucl Med,* 54**,** 19-26.

CRUK, C. R. 2017. *Lifetime risk of cancer* [Online]. Cancer Research UK. Available: http://www.cancerresearchuk.org/health-professional/cancer-statistics/risk/lifetime-risk - heading-Zero [Accessed 17/10/2017].

CURTIS, C., SHAH, S. P., CHIN, S. F., TURASHVILI, G., RUEDA, O. M., DUNNING, M. J., SPEED, D., LYNCH, A. G., SAMARAJIWA, S., YUAN, Y., GRAF, S., HA, G., HAFFARI, G., BASHASHATI, A., RUSSELL, R., MCKINNEY, S., LANGEROD, A., GREEN, A., PROVENZANO, E., WISHART, G., PINDER, S., WATSON, P., MARKOWETZ, F., MURPHY, L., ELLIS, I., PURUSHOTHAM, A., BORRESEN-DALE, A. L., BRENTON, J. D., TAVARE, S., CALDAS, C. & APARICIO, S. 2012. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature,* 486**,** 346-52.

DAVIDSON, B., TROPE, C. G. & REICH, R. 2014. The role of the tumor stroma in ovarian cancer. *Front Oncol,* 4**,** 104.

DE UNAMUNO BUSTOS, B., MURRIA ESTAL, R., PÉREZ SIMÓ, G., DE JUAN JIMENEZ, I., ESCUTIA MUÑOZ, B., RODRÍGUEZ SERNA, M., ALEGRE DE MIQUEL, V., LLAVADOR ROS, M., BALLESTER SÁNCHEZ, R., NAGORE ENGUÍDANOS, E., PALANCA SUELA, S. & BOTELLA ESTRADA, R. 2017. Towards Personalized Medicine in Melanoma: Implementation of a Clinical Next-Generation Sequencing Panel. *Scientific Reports,* 7**,** 495.

DETTERBECK, F. C., POSTMUS, P. E. & TANOUE, L. T. 2013. The stage classification of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest,* 143**,** e191S-210S.

DIAZ-CANO, S. J. 2012. Tumor heterogeneity: mechanisms and bases for a reliable application of molecular marker design. *Int J Mol Sci,* 13**,** 1951-2011.

DIAZ-GIL, D., FINTELMANN, F. J., MOLAEI, S., ELMI, A., HEDGIRE, S. S. & HARISINGHANI, M. G. 2016. Prediction of 5-year survival in advanced-stage ovarian cancer patients based on computed tomography peritoneal carcinomatosis index. *Abdom Radiol (NY),* 41**,** 2196-2202.

DIEHN, M., NARDINI, C., WANG, D. S., MCGOVERN, S., JAYARAMAN, M., LIANG, Y., ALDAPE, K., CHA, S. & KUO, M. D. 2008. Identification of noninvasive imaging surrogates for brain tumor gene-expression modules. *Proc Natl Acad Sci U S A,* 105**,** 5213-8.

DIMITRAKOPOULOU-STRAUSS, A., HOFFMANN, M., BERGNER, R., UPPENKAMP, M., EISENHUT, M., PAN, L., HABERKORN, U. & STRAUSS, L. G. 2007. Prediction of short-term survival in patients with advanced nonsmall cell lung cancer following chemotherapy based on 2-deoxy-2-[F-18]fluoro-D-glucose-positron emission tomography: a feasibility study. *Mol Imaging Biol,* 9**,** 308-17.

DIMITRAKOPOULOU-STRAUSS, A., STRAUSS, L. G., BURGER, C., RUHL, A., IRNGARTINGER, G., STREMMEL, W. & RUDI, J. 2004. Prognostic aspects of 18F-FDG PET kinetics in patients with metastatic colorectal carcinoma receiving FOLFOX chemotherapy. *J Nucl Med,* 45**,** 1480-7.

DIMITRAKOPOULOU-STRAUSS, A., STRAUSS, L. G., HEICHEL, T., WU, H., BURGER, C., BERND, L. & EWERBECK, V. 2002. The role of quantitative (18)F-FDG PET studies for the differentiation of malignant and benign bone lesions. *J Nucl Med,* 43**,** 510-8.

DONG, X., XING, L., WU, P., FU, Z., WAN, H., LI, D., YIN, Y., SUN, X. & YU, J. 2013. Three-dimensional positron emission tomography image texture analysis of esophageal squamous cell carcinoma: relationship between tumor 18F-fluorodeoxyglucose uptake heterogeneity, maximum standardized uptake value, and tumor stage. *Nucl Med Commun,* 34**,** 40-6.

E BAIR, T. H., D PAUL, R TIBSHIRANI 2006. Prediction by Supervised Principal Components. *Journal of the American Statistical Association,* 101**,** 119-137.

EARY, J. F., O'SULLIVAN, F., O'SULLIVAN, J. & CONRAD, E. U. 2008. Spatial heterogeneity in sarcoma 18F-FDG uptake as a predictor of patient outcome. *J Nucl Med,* 49**,** 1973-9.

EISENHAUER, E. A., THERASSE, P., BOGAERTS, J., SCHWARTZ, L. H., SARGENT, D., FORD, R., DANCEY, J., ARBUCK, S., GWYTHER, S., MOONEY, M., RUBINSTEIN, L., SHANKAR, L., DODD, L., KAPLAN, R., LACOMBE, D. & VERWEIJ, J. 2009. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer,* 45**,** 228-47.

EL NAQA, I., GRIGSBY, P., APTE, A., KIDD, E., DONNELLY, E., KHULLAR, D., CHAUDHARI, S., YANG, D., SCHMITT, M., LAFOREST, R., THORSTAD, W. & DEASY, J. O. 2009. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recognit,* 42**,** 1162-1171.

ERLANDSSON, K., BUVAT, I., PRETORIUS, P. H., THOMAS, B. A. & HUTTON, B. F. 2012. A review of partial volume correction techniques for emission tomography and their applications in neurology, cardiology and oncology. *Phys Med Biol,* 57**,** R119-59.

ERTAY, T., SENCAN EREN, M., KARAMAN, M., OKTAY, G. & DURAK, H. 2017. 18F-FDG-PET/CT in Initiation and Progression of Inflammation and Infection. *Molecular Imaging and Radionuclide Therapy,* 26**,** 47-52.

ESQUIVEL, J., CHUA, T. C., STOJADINOVIC, A., MELERO, J. T., LEVINE, E. A., GUTMAN, M., HOWARD, R., PISO, P., NISSAN, A., GOMEZ-PORTILLA, A., GONZALEZ-BAYON, L., GONZALEZ-MORENO, S., SHEN, P., STEWART, J. H., SUGARBAKER, P. H., BARONE, R. M., HOEFER, R., MORRIS, D. L., SARDI, A. & STICCA, R. P. 2010. Accuracy and clinical relevance of computed tomography scan interpretation of peritoneal cancer index in colorectal cancer peritoneal carcinomatosis: a multi-institutional study. *J Surg Oncol,* 102**,** 565-70.

FINAK, G., BERTOS, N., PEPIN, F., SADEKOVA, S., SOULEIMANOVA, M., ZHAO, H., CHEN, H., OMEROGLU, G., METERISSIAN, S., OMEROGLU, A., HALLETT, M. & PARK, M. 2008. Stromal gene expression predicts clinical outcome in breast cancer. *Nat Med,* 14**,** 518-27.

FISHER, R., PUSZTAI, L. & SWANTON, C. 2013. Cancer heterogeneity: implications for targeted therapeutics. *Br J Cancer,* 108**,** 479-85.

FOWLER, J. S. & IDO, T. 2002. Initial and subsequent approach for the synthesis of ¹⁸FDG. *Seminars in Nuclear Medicine,* 32**,** 6-12.

FRINGS, O., AUGSTEN, M., TOBIN, N. P., CARLSON, J., PAULSSON, J., PENA, C., OLSSON, E., VEERLA, S., BERGH, J., OSTMAN, A. & SONNHAMMER, E. L. 2013. Prognostic significance in breast cancer of a gene signature capturing stromal PDGF signaling. *Am J Pathol,* 182**,** 2037-47.

FUKUOKA, M., YANO, S., GIACCONE, G., TAMURA, T., NAKAGAWA, K., DOUILLARD, J. Y., NISHIWAKI, Y., VANSTEENKISTE, J., KUDOH, S., RISCHIN, D., EEK, R., HORAI, T., NODA, K., TAKATA, I., SMIT, E., AVERBUCH, S., MACLEOD, A., FEYEREISLOVA, A., DONG, R. P. & BASELGA, J. 2003. Multi-institutional randomized phase II trial of gefitinib for previously treated patients with advanced non-small-cell lung cancer (The IDEAL 1 Trial) [corrected]. *J Clin Oncol,* 21**,** 2237-46.

GAILLARD, H., GARCIA-MUSE, T. & AGUILERA, A. 2015. Replication stress and cancer. *Nat Rev Cancer,* 15**,** 276-89.

GALAVIS, P. E., HOLLENSEN, C., JALLOW, N., PALIWAL, B. & JERAJ, R. 2010. Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters. *Acta Oncol,* 49**,** 1012-6.

GALLAMINI, A., BARRINGTON, S. F., BIGGI, A., CHAUVIE, S., KOSTAKOGLU, L., GREGIANIN, M., MEIGNAN, M., MIKHAEEL, G. N., LOFT, A., ZAUCHA, J. M., SEYMOUR, J. F., HOFMAN, M. S., RIGACCI, L., PULSONI, A., COLEMAN, M., DANN, E. J., TRENTIN, L., CASASNOVAS, O., RUSCONI, C., BRICE, P., BOLIS, S., VIVIANI, S., SALVI, F., LUMINARI, S. & HUTCHINGS, M. 2014. The predictive role of interim positron emission tomography for Hodgkin lymphoma treatment outcome is confirmed using the

interpretation criteria of the Deauville five-point scale. *Haematologica,* 99, 1107-1113.

GALLOWAY, M. 1975. Texture Analysis Using Grey Level Run Lengths. *Computer Graphics and Image Processing,* 4, 172-179.

GALUN, D., SRDIC-RAJIC, T., BOGDANOVIC, A., LONCAR, Z. & ZUVELA, M. 2017. Targeted therapy and personalized medicine in hepatocellular carcinoma: drug resistance, mechanisms, and treatment strategies. *J Hepatocell Carcinoma,* 4, 93-103.

GANESHAN, B., ABALEKE, S., YOUNG, R. C., CHATWIN, C. R. & MILES, K. A. 2010a. 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, 137-43.

GANESHAN, B., ABALEKE, S., YOUNG, R. C. D., CHATWIN, C. R. & MILES, K. A. 2010b. 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 : the official publication of the International Cancer Imaging Society,* 10, 137-143.

GANESHAN, B., SKOGEN, K., PRESSNEY, I., COUTROUBIS, D. & MILES, K. 2012. Tumour heterogeneity in oesophageal cancer assessed by CT texture analysis: preliminary evidence of an association with tumour metabolism, stage, and survival. *Clin Radiol,* 67, 157-64.

GASIMLI, K., BRAICU, E. I., RICHTER, R., CHEKEROV, R. & SEHOULI, J. 2015. Prognostic and Predictive Value of the Peritoneal Cancer Index in Primary Advanced Epithelial Ovarian Cancer Patients After Complete Cytoreductive Surgery: Study of Tumor Bank Ovarian Cancer. *Ann Surg Oncol,* 22, 2729-37.

GATENBY, R. A. & GILLIES, R. J. 2004. Why do cancers have high aerobic glycolysis? *Nat Rev Cancer,* 4, 891-9.

GATENBY, R. A. & VINCENT, T. L. 2003. Application of quantitative models from population biology and evolutionary game theory to tumor therapeutic strategies. *Mol Cancer Ther,* 2, 919-27.

GEORGE, J., CLAES, P., VUNCKX, K., TEJPAR, S., DEROOSE, C. M., NUYTS, J., LOECKX, D., SUETENS, P. & IEEE 2012. A TEXTURAL FEATURE BASED TUMOR THERAPY RESPONSE PREDICTION MODEL FOR LONGITUDINAL EVALUATION WITH PET IMAGING. *2012 9th Ieee International Symposium on Biomedical Imaging (Isbi)*, 1048-1051.

GERLINGER, M., ROWAN, A. J., HORSWELL, S., LARKIN, J., ENDESFELDER, D., GRONROOS, E., MARTINEZ, P., MATTHEWS, N., STEWART, A., TARPEY, P., VARELA, I., PHILLIMORE, B., BEGUM, S., MCDONALD, N. Q., BUTLER, A., JONES, D., RAINE, K., LATIMER, C., SANTOS, C. R., NOHADANI, M., EKLUND, A. C., SPENCER-DENE, B., CLARK, G., PICKERING, L., STAMP, G., GORE, M., SZALLASI, Z., DOWNWARD, J., FUTREAL, P. A. & SWANTON, C. 2012. Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *New England Journal of Medicine,* 366, 883-892.

GERLINGER, M. & SWANTON, C. 2010. How Darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine. *Br J Cancer,* 103, 1139-43.

GIARDINO, A., GUPTA, S., OLSON, E., SEPULVEDA, K., LENCHIK, L., IVANIDZE, J., RAKOW-PENNER, R., PATEL, M. J., SUBRAMANIAM, R. M. & GANESHAN, D. 2017. Role of Imaging in the Era of Precision Medicine. *Acad Radiol,* 24, 639-649.

GILLIES, R. J., ANDERSON, A. R., GATENBY, R. A. & MORSE, D. L. 2010. The biology underlying molecular imaging in oncology: from genome to anatome and back again. *Clin Radiol,* 65, 517-21.

GILLIES, R. J., KINAHAN, P. E. & HRICAK, H. 2016. Radiomics: Images Are More than Pictures, They Are Data. *Radiology,* 278, 563-77.

GILLY, F. N., COTTE, E., BRIGAND, C., MONNEUSE, O., BEAUJARD, A. C., FREYER, G. & GLEHEN, O. 2006. Quantitative prognostic indices in peritoneal carcinomatosis. *Eur J Surg Oncol,* 32**,** 597-601.

GOH, V., GANESHAN, B., NATHAN, P., JUTTLA, J. K., VINAYAN, A. & MILES, K. A. 2011. Assessment of response to tyrosine kinase inhibitors in metastatic renal cell cancer: CT texture as a predictive biomarker. *Radiology,* 261**,** 165-71.

GOLDMAN, L. W. 2007. Principles of CT and CT technology. *J Nucl Med Technol,* 35**,** 115-28; quiz 129-30.

GOLDMAN, L. W. 2008. Principles of CT: multislice CT. *J Nucl Med Technol,* 36**,** 57-68; quiz 75-6.

GOLDSTRAW, P., CHANSKY, K., CROWLEY, J., RAMI-PORTA, R., ASAMURA, H., EBERHARDT, W. E., NICHOLSON, A. G., GROOME, P., MITCHELL, A. & BOLEJACK, V. 2016. The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. *J Thorac Oncol,* 11**,** 39-51.

GROHEUX, D., GIACCHETTI, S., HATT, M., MARTY, M., VERCELLINO, L., DE ROQUANCOURT, A., CUVIER, C., COUSSY, F., ESPIE, M. & HINDIE, E. 2013. HER2-overexpressing breast cancer: FDG uptake after two cycles of chemotherapy predicts the outcome of neoadjuvant treatment. *Br J Cancer,* 109**,** 1157-64.

GUO, Y., ZHAO, S., SU, P.-F., LI, C.-I., YE, F., FLYNN, C. R. & SHYR, Y. 2014. Statistical strategies for microRNAseq batch effect reduction. *Translational cancer research,* 3**,** 260-265.

GUPTA, D. & LIS, C. G. 2009. Role of CA125 in predicting ovarian cancer survival - a review of the epidemiological literature. *Journal of Ovarian Research,* 2**,** 13-13.

HANAHAN, D. & WEINBERG, R. A. 2000. The Hallmarks of Cancer. *Cell,* 100**,** 57-70.

HANAHAN, D. & WEINBERG, R. A. 2011. Hallmarks of cancer: the next generation. *Cell,* 144**,** 646-74.

HARALICK, R. 1979. Statistical and structural approaches to texture. *Proceedings of the IEEE* 67 786-804.

HARALICK, R., SHANMUGAN, K. & DINSTEIN, I. 1973. Textural Features for Image Classification. *IEEE Transactions on Systems, Man, And Cybernetics,* SMC 3**,** 610-621.

HATT, M. 2012. *Multi modal images analysis and processing in oncology.* PhD, Brest - France.

HATT, M., CHEZE LE REST, C., DESCOURT, P., DEKKER, A., DE RUYSSCHER, D., OELLERS, M., LAMBIN, P., PRADIER, O. & VISVIKIS, D. 2010a. Accurate automatic delineation of heterogeneous functional volumes in positron emission tomography for oncology applications. *Int J Radiat Oncol Biol Phys,* 77**,** 301-8.

HATT, M., CHEZE LE REST, C., TURZO, A., ROUX, C. & VISVIKIS, D. 2009. A fuzzy locally adaptive Bayesian segmentation approach for volume determination in PET. *IEEE Trans Med Imaging,* 28**,** 881-93.

HATT, M., CHEZE-LE REST, C., VAN BAARDWIJK, A., LAMBIN, P., PRADIER, O. & VISVIKIS, D. 2011a. Impact of tumor size and tracer uptake heterogeneity in (18)F-FDG PET and CT non-small cell lung cancer tumor delineation. *J Nucl Med,* 52**,** 1690-7.

HATT, M., GROHEUX, D., MARTINEAU, A., ESPIE, M., HINDIE, E., GIACCHETTI, S., DE ROQUANCOURT, A., VISVIKIS, D. & CHEZE-LE REST, C. 2013a. Comparison between 18F-FDG PET image-derived indices for early prediction of response to neoadjuvant chemotherapy in breast cancer. *J Nucl Med,* 54**,** 341-9.

HATT, M., MAJDOUB, M., VALLIERES, M., TIXIER, F., LE REST, C. C., GROHEUX, D., HINDIE, E., MARTINEAU, A., PRADIER, O., HUSTINX, R., PERDRISOT, R., GUILLEVIN, R., EL NAQA, I. & VISVIKIS, D. 2015. 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,* 56**,** 38-44.

HATT, M., REST, C. C.-L., ABOAGYE, E. O., KENNY, L. M., ROSSO, L., TURKHEIMER, F. E., ALBARGHACH, N. M., METGES, J.-P., PRADIER, O. & VISVIKIS, D. 2010b. Reproducibility of 18F-FDG and 39-Deoxy-39-18F-Fluorothymidine PET Tumor Volume Measurements. *THE JOURNAL OF NUCLEAR MEDICINE,* 51**,** 1368-1376.

HATT, M., TIXIER, F., CHEZE LE REST, C., PRADIER, O. & VISVIKIS, D. 2013b. Robustness of intratumour (1)(8)F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma. *Eur J Nucl Med Mol Imaging,* 40**,** 1662-71.

HATT, M., VAN STIPHOUT, R., LE POGAM, A., LAMMERING, G., VISVIKIS, D. & LAMBIN, P. 2013c. Early prediction of pathological response in locally advanced rectal cancer based on sequential 18F-FDG PET. *Acta Oncol,* 52**,** 619-26.

HATT, M., VISVIKIS, D., ALBARGHACH, N. M., TIXIER, F., PRADIER, O. & CHEZE-LE REST, C. 2011b. Prognostic value of 18F-FDG PET image-based parameters in oesophageal cancer and impact of tumour delineation methodology. *Eur J Nucl Med Mol Imaging,* 38**,** 1191-202.

HAYANO, K., TIAN, F., KAMBADAKONE, A. R., YOON, S. S., DUDA, D. G., GANESHAN, B. & SAHANI, D. V. 2015. Texture Analysis of Non-Contrast-Enhanced Computed Tomography for Assessing Angiogenesis and Survival of Soft Tissue Sarcoma. *J Comput Assist Tomogr,* 39**,** 607-12.

HAYES, S. A., PIETANZA, M. C., O'DRISCOLL, D., ZHENG, J., MOSKOWITZ, C. S., KRIS, M. G. & GINSBERG, M. S. 2016. Comparison of CT volumetric measurement with RECIST response in patients with lung cancer. *European journal of radiology,* 85**,** 524-533.

HE, D. C. & LI, W. 1991. Texture Features Based on Texture Spectrum. *Pattern Recognition,* 24**,** 391-399.

HE, Z., SUN, M., KE, Y., LIN, R., XIAO, Y., ZHOU, S., ZHAO, H., WANG, Y., ZHOU, F. & ZHOU, Y. 2017. Identifying biomarkers of papillary renal cell carcinoma associated with pathological stage by weighted gene co-expression network analysis. *Oncotarget,* 8**,** 27904-27914.

HECTORS, S. J., WAGNER, M., BANE, O., BESA, C., LEWIS, S., REMARK, R., CHEN, N., FIEL, M. I., ZHU, H., GNJATIC, S., MERAD, M., HOSHIDA, Y. & TAOULI, B. 2017. Quantification of hepatocellular carcinoma heterogeneity with multiparametric magnetic resonance imaging. *Scientific Reports,* 7**,** 2452.

HEINZMANN, K., CARTER, L. M., LEWIS, J. S. & ABOAGYE, E. O. 2017. Multiplexed imaging for diagnosis and therapy. *Nature Biomedical Engineering,* 1**,** 697-713.

HENNESSY, B. T., LU, Y., GONZALEZ-ANGULO, A. M., CAREY, M. S., MYHRE, S., JU, Z., DAVIES, M. A., LIU, W., COOMBES, K., MERIC-BERNSTAM, F., BEDROSIAN, I., MCGAHREN, M., AGARWAL, R., ZHANG, F., OVERGAARD, J., ALSNER, J., NEVE, R. M., KUO, W.-L., GRAY, J. W., BORRESEN-DALE, A.-L. & MILLS, G. B. 2010. A Technical Assessment of the Utility of Reverse Phase Protein Arrays for the Study of the Functional Proteome in Non-microdissected Human Breast Cancers. *Clinical proteomics,* 6**,** 129-151.

HOANG, J. K., DAS, S. K., CHOUDHURY, K. R., YOO, D. S. & BRIZEL, D. M. 2013. Using FDG-PET to measure early treatment response in head and neck squamous cell carcinoma: quantifying intrinsic variability in order to understand treatment-induced change. *AJNR Am J Neuroradiol,* 34**,** 1428-33.

HOFFMAN, E. J., HUANG, S. C. & PHELPS, M. E. 1979. Quantitation in positron emission computed tomography: 1. Effect of object size. *J Comput Assist Tomogr,* 3**,** 299-308.

HOROWITZ, N. S., MILLER, A., RUNGRUANG, B., RICHARD, S. D., RODRIGUEZ, N., BOOKMAN, M. A., HAMILTON, C. A., KRIVAK, T. C. & MAXWELL, G. L. 2015. Does aggressive surgery improve outcomes? Interaction between preoperative disease burden and complex surgery in patients with advanced-stage ovarian cancer: an analysis of GOG 182. *J Clin Oncol,* 33**,** 937-43.

HUNTER, L., CHEN, Y., ZHANG, L., LIAO, Z., DONG, L. & COURT, L. 2012. SU-E-J-186: CT Textures Can be Predictive for Tumor Shrinkage. *Med Phys,* 39**,** 3695.

ISELLA, C., TERRASI, A., BELLOMO, S. E., PETTI, C., GALATOLA, G., MURATORE, A., MELLANO, A., SENETTA, R., CASSENTI, A., SONETTO, C., INGHIRAMI, G., TRUSOLINO, L., FEKETE, Z., DE RIDDER, M., CASSONI, P., STORME, G., BERTOTTI, A. & MEDICO, E. 2015. Stromal contribution to the colorectal cancer transcriptome. *Nat Genet,* 47**,** 312-9.

JACQUET, P. & SUGARBAKER, P. H. 1996. Clinical research methodologies in diagnosis and staging of patients with peritoneal carcinomatosis. *Cancer Treat Res,* 82**,** 359-74.

JAMES, K., EISENHAUER, E., CHRISTIAN, M., TERENZIANI, M., VENA, D., MULDAL, A. & THERASSE, P. 1999. Measuring response in solid tumors: unidimensional versus bidimensional measurement. *J Natl Cancer Inst,* 91**,** 523-8.

JANKU, F., STEWART, D. J. & KURZROCK, R. 2010. Targeted therapy in non-small-cell lung cancer--is it becoming a reality? *Nat Rev Clin Oncol,* 7**,** 401-14.

JENTZEN, W., FREUDENBERG, L., EISING, E. G., HEINZE, M., BRANDAU, W. & BOCKISCH, A. 2007. Segmentation of PET volumes by iterative image thresholding. *J Nucl Med,* 48**,** 108-14.

JULESZ, B. 1975. Experiments in the visual perception of texture. *Sci Am,* 232**,** 34-43.

KALLURI, R. & ZEISBERG, M. 2006. Fibroblasts in cancer. *Nat Rev Cancer,* 6**,** 392-401.

KARAKATSANIS, N. A., FOKOU, E. & TSOUMPAS, C. 2015. Dosage optimization in positron emission tomography: state-of-the-art methods and future prospects. *American Journal of Nuclear Medicine and Molecular Imaging,* 5**,** 527-547.

KEATING, G. M. 2014. Afatinib: a review of its use in the treatment of advanced non-small cell lung cancer. *Drugs,* 74**,** 207-21.

KENNY, L., COOMBES, R. C., VIGUSHIN, D. M., AL-NAHHAS, A., SHOUSHA, S. & ABOAGYE, E. O. 2007. Imaging early changes in proliferation at 1 week post chemotherapy: a pilot study in breast cancer patients with 3'-deoxy-3'-[18F]fluorothymidine positron emission tomography. *Eur J Nucl Med Mol Imaging,* 34**,** 1339-47.

KESSLER, R. M., ELLIS, J. R., JR. & EDEN, M. 1984. Analysis of emission tomographic scan data: limitations imposed by resolution and background. *J Comput Assist Tomogr,* 8**,** 514-22.

KIDD, E. A. & GRIGSBY, P. W. 2008. Intratumoral metabolic heterogeneity of cervical cancer. *Clin Cancer Res,* 14**,** 5236-41.

KIDD, E. A., THOMAS, M., SIEGEL, B. A., DEHDASHTI, F. & GRIGSBY, P. W. 2013. Changes in cervical cancer FDG uptake during chemoradiation and association with response. *Int J Radiat Oncol Biol Phys,* 85**,** 116-22.

KIM, H. G., CHUNG, Y. E., LEE, Y. H., CHOI, J.-Y., PARK, M.-S., KIM, M.-J. & KIM, K. W. 2015. Quantitative Analysis of the Effect of Iterative Reconstruction Using a Phantom: Determining the Appropriate Blending Percentage. *Yonsei Medical Journal,* 56**,** 253-261.

KJELL, E., IRÈNE, B., PRETORIUS, P. H., BENJAMIN, A. T. & BRIAN, F. H. 2012. A review of partial volume correction techniques for emission tomography and their applications in neurology, cardiology and oncology. *Physics in Medicine and Biology,* 57**,** R119.

KNOGLER, T., EL-RABADI, K., WEBER, M., KARANIKAS, G. & MAYERHOEFER, M. E. 2014. Three-dimensional texture analysis of contrast enhanced CT images for treatment response assessment in Hodgkin lymphoma: comparison with F-18-FDG PET. *Med Phys,* 41**,** 121904.

KOH, J. L., YAN, T. D., GLENN, D. & MORRIS, D. L. 2009. Evaluation of preoperative computed tomography in estimating peritoneal cancer index in colorectal peritoneal carcinomatosis. *Ann Surg Oncol,* 16**,** 327-33.

KUMAR, V., GU, Y., BASU, S., BERGLUND, A., ESCHRICH, S. A., SCHABATH, M. B., FORSTER, K., AERTS, H. J., DEKKER, A., FENSTERMACHER, D., GOLDGOF, D. B., HALL, L. O., LAMBIN,

P., BALAGURUNATHAN, Y., GATENBY, R. A. & GILLIES, R. J. 2012. Radiomics: the process and the challenges. *Magn Reson Imaging,* 30**,** 1234-48.

KWEE, T. C., TAKAHARA, T., KLOMP, D. W. J. & LUIJTEN, P. R. 2010. Cancer imaging: novel concepts in clinical magnetic resonance imaging. *Journal of Internal Medicine,* 268**,** 120-132.

LAMBIN, P., RIOS-VELAZQUEZ, E., LEIJENAAR, R., CARVALHO, S., VAN STIPHOUT, R. G., GRANTON, P., ZEGERS, C. M., GILLIES, R., BOELLARD, R., DEKKER, A. & AERTS, H. J. 2012. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer,* 48**,** 441-6.

LAMPE, B., KROLL, N., PISO, P., FORNER, D. M. & MALLMANN, P. 2015. Prognostic significance of Sugarbaker's peritoneal cancer index for the operability of ovarian carcinoma. *Int J Gynecol Cancer,* 25**,** 135-44.

LANDI, L. & CAPPUZZO, F. 2011. Targeted therapies: Front-line therapy in lung cancer with mutations in EGFR. *Nat Rev Clin Oncol,* 8**,** 571-3.

LARSON, S. M., ERDI, Y., AKHURST, T., MAZUMDAR, M., MACAPINLAC, H. A., FINN, R. D., CASILLA, C., FAZZARI, M., SRIVASTAVA, N., YEUNG, H. W., HUMM, J. L., GUILLEM, J., DOWNEY, R., KARPEH, M., COHEN, A. E. & GINSBERG, R. 1999. Tumor Treatment Response Based on Visual and Quantitative Changes in Global Tumor Glycolysis Using PET-FDG Imaging. The Visual Response Score and the Change in Total Lesion Glycolysis. *Clin Positron Imaging,* 2**,** 159-171.

LARSON, S. M. & SCHODER, H. 2009. New PET tracers for evaluation of solid tumor response to therapy. *Q J Nucl Med Mol Imaging,* 53**,** 158-66.

LASNON, C., QUAK, E., LE ROUX, P.-Y., ROBIN, P., HOFMAN, M. S., BOURHIS, D., CALLAHAN, J., BINNS, D. S., DESMONTS, C., SALAUN, P.-Y., HICKS, R. J. & AIDE, N. 2017. EORTC PET response criteria are more influenced by reconstruction inconsistencies than PERCIST but both benefit from the EARL harmonization program. *EJNMMI Physics,* 4**,** 17.

LAW, W. P. & MILES, K. A. 2013. Incorporating prognostic imaging biomarkers into clinical practice. *Cancer Imaging,* 13**,** 332-341.

LEE, E. S., SON, D. S., KIM, S. H., LEE, J., JO, J., HAN, J., KIM, H., LEE, H. J., CHOI, H. Y., JUNG, Y., PARK, M., LIM, Y. S., KIM, K., SHIM, Y., KIM, B. C., LEE, K., HUH, N., KO, C., PARK, K., LEE, J. W., CHOI, Y. S. & KIM, J. 2008. Prediction of recurrence-free survival in postoperative non-small cell lung cancer patients by using an integrated model of clinical information and gene expression. *Clin Cancer Res,* 14**,** 7397-404.

LEIJENAAR, R. T., NALBANTOV, G., CARVALHO, S., VAN ELMPT, W. J., TROOST, E. G., BOELLAARD, R., AERTS, H. J., GILLIES, R. J. & LAMBIN, P. 2015. The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis. *Sci Rep,* 5**,** 11075.

LEONG, H. S., GALLETTA, L., ETEMADMOGHADAM, D., GEORGE, J., AUSTRALIAN OVARIAN CANCER, S., KOBEL, M., RAMUS, S. J. & BOWTELL, D. 2015. Efficient molecular subtype classification of high-grade serous ovarian cancer. *J Pathol,* 236**,** 272-7.

LIM, C., SEKHON, H. S., CUTZ, J. C., HWANG, D. M., KAMEL-REID, S., CARTER, R. F., SANTOS, G. D. C., WADDELL, T., BINNIE, M., PATEL, M., PAUL, N., CHUNG, T., BRADE, A., EL-MARAGHI, R., SIT, C., TSAO, M. S. & LEIGHL, N. B. 2017. Improving molecular testing and personalized medicine in non-small-cell lung cancer in Ontario. *Current Oncology,* 24**,** 103-110.

LINARDOU, H., DAHABREH, I. J., BAFALOUKOS, D., KOSMIDIS, P. & MURRAY, S. 2009. Somatic EGFR mutations and efficacy of tyrosine kinase inhibitors in NSCLC. *Nat Rev Clin Oncol,* 6**,** 352-66.

LONGO, D. L. 2012. Tumor heterogeneity and personalized medicine. *N Engl J Med,* 366**,** 956-7.

LOPES, R. & BETROUNI, N. 2009. Fractal and multifractal analysis: a review. *Med Image Anal,* 13**,** 634-49.

LOVINFOSSE, P., JANVARY, Z. L., COUCKE, P., JODOGNE, S., BERNARD, C., HATT, M., VISVIKIS, D., JANSEN, N., DUYSINX, B. & HUSTINX, R. 2016. FDG PET/CT texture analysis for predicting the outcome of lung cancer treated by stereotactic body radiation therapy. *Eur J Nucl Med Mol Imaging,* 43**,** 1453-60.

LUBNER, M. G., STABO, N., ABEL, E. J., DEL RIO, A. M. & PICKHARDT, P. J. 2016. CT Textural Analysis of Large Primary Renal Cell Carcinomas: Pretreatment Tumor Heterogeneity Correlates With Histologic Findings and Clinical Outcomes. *AJR Am J Roentgenol,* 207**,** 96-105.

LUBNER, M. G., STABO, N., LUBNER, S. J., DEL RIO, A. M., SONG, C., HALBERG, R. B. & PICKHARDT, P. J. 2015. CT textural analysis of hepatic metastatic colorectal cancer: pre-treatment tumor heterogeneity correlates with pathology and clinical outcomes. *Abdom Imaging,* 40**,** 2331-7.

LYNCH, T. J., PATEL, T., DREISBACH, L., MCCLEOD, M., HEIM, W. J., HERMANN, R. C., PASCHOLD, E., IANNOTTI, N. O., DAKHIL, S., GORTON, S., PAUTRET, V., WEBER, M. R. & WOYTOWITZ, D. 2010. 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**,** 911-7.

MACHTAY, M., DUAN, F., SIEGEL, B. A., SNYDER, B. S., GORELICK, J. J., REDDIN, J. S., MUNDEN, R., JOHNSON, D. W., WILF, L. H., DENITTIS, A., SHERWIN, N., CHO, K. H., KIM, S.-K., VIDETIC, G., NEUMANN, D. R., KOMAKI, R., MACAPINLAC, H., BRADLEY, J. D. & ALAVI, A. 2013. Prediction of Survival by [18F]Fluorodeoxyglucose Positron Emission Tomography in Patients With Locally Advanced Non–Small-Cell Lung Cancer Undergoing Definitive Chemoradiation Therapy: Results of the ACRIN 6668/RTOG 0235 Trial. *Journal of Clinical Oncology,* 31**,** 3823-3830.

MALHOTRA, D., FLETCHER, A. L., ASTARITA, J., LUKACS-KORNEK, V., TAYALIA, P., GONZALEZ, S. F., ELPEK, K. G., CHANG, S. K., KNOBLICH, K., HEMLER, M. E., BRENNER, M. B., CARROLL, M. C., MOONEY, D. J., TURLEY, S. J. & IMMUNOLOGICAL GENOME PROJECT, C. 2012. Transcriptional profiling of stroma from inflamed and resting lymph nodes defines immunological hallmarks. *Nat Immunol,* 13**,** 499-510.

MANDELBROT, B. 1982. *The fractal geometry of nature,* San Francisco, W.H. Freeman.

MARKEL, D., CALDWELL, C., ALASTI, H., SOLIMAN, H., UNG, Y., LEE, J. & SUN, A. 2013. Automatic Segmentation of Lung Carcinoma Using 3D Texture Features in 18-FDG PET/CT. *Int J Mol Imaging,* 2013**,** 980769.

MARUSYK, A., ALMENDRO, V. & POLYAK, K. 2012. Intra-tumour heterogeneity: a looking glass for cancer? *Nat Rev Cancer,* 12**,** 323-34.

MARUSYK, A. & POLYAK, K. 2010. Tumor heterogeneity: causes and consequences. *Biochim Biophys Acta,* 1805**,** 105-17.

MATTHEWS, P. A. G. B. J. 2002. *Statistical Methods In Medical Research*, Blackwell Science.

MATTONEN, S. A., TETAR, S., PALMA, D. A., LOUIE, A. V., SENAN, S. & WARD, A. D. 2015. Imaging texture analysis for automated prediction of lung cancer recurrence after stereotactic radiotherapy. *J Med Imaging (Bellingham),* 2**,** 041010.

MAUGHAN, T. 2017. The Promise and the Hype of 'Personalised Medicine'. *The New Bioethics,* 23**,** 13-20.

MAZZEI, M. A., KHADER, L., CIRIGLIANO, A., CIOFFI SQUITIERI, N., GUERRINI, S., FORZONI, B., MARRELLI, D., ROVIELLO, F., MAZZEI, F. G. & VOLTERRANI, L. 2013. Accuracy of MDCT in the preoperative definition of Peritoneal Cancer Index (PCI) in patients with advanced ovarian cancer who underwent peritonectomy and hyperthermic intraperitoneal chemotherapy (HIPEC). *Abdom Imaging,* 38**,** 1422-30.

MERTENS, J., DE BRUYNE, S., VAN DAMME, N., SMEETS, P., CEELEN, W., TROISI, R., LAURENT, S., GEBOES, K., PEETERS, M., GOETHALS, I. & VAN DE WIELE, C. 2013. Standardized added metabolic activity (SAM) IN (1)(8)F-FDG PET assessment of treatment response in colorectal liver metastases. *Eur J Nucl Med Mol Imaging,* 40**,** 1214-22.

MERTENS, J., DOBBELEIR, A., HAM, H., D'ASSELER, Y., GOETHALS, I. & VAN DE WIELE, C. 2012. Standardized added metabolic activity (SAM): a partial volume independent marker of total lesion glycolysis in liver metastases. *Eur J Nucl Med Mol Imaging,* 39**,** 1441-8.

MHAWECH-FAUCEGLIA, P., YAN, L., SHARIFIAN, M., REN, X., LIU, S., KIM, G., GAYTHER, S. A., PEJOVIC, T. & LAWRENSON, K. 2015. Stromal Expression of Fibroblast Activation Protein Alpha (FAP) Predicts Platinum Resistance and Shorter Recurrence in patients with Epithelial Ovarian Cancer. *Cancer Microenviron,* 8**,** 23-31.

MICHAELIS, L. C. & RATAIN, M. J. 2006. Measuring response in a post-RECIST world: from black and white to shades of grey. *Nat Rev Cancer,* 6**,** 409-14.

MIDHA, A., DEARDEN, S. & MCCORMACK, R. 2015. EGFR mutation incidence in non-small-cell lung cancer of adenocarcinoma histology: a systematic review and global map by ethnicity (mutMapII). *Am J Cancer Res,* 5**,** 2892-911.

MILES, K. A., GANESHAN, B., GRIFFITHS, M. R., YOUNG, R. C. & CHATWIN, C. R. 2009. Colorectal cancer: texture analysis of portal phase hepatic CT images as a potential marker of survival. *Radiology,* 250**,** 444-52.

MILLER, T. R., PINKUS, E., DEHDASHTI, F. & GRIGSBY, P. W. 2003. Improved prognostic value of 18F-FDG PET using a simple visual analysis of tumor characteristics in patients with cervical cancer. *J Nucl Med,* 44**,** 192-7.

MINCHINTON, A. I. & TANNOCK, I. F. 2006. Drug penetration in solid tumours. *Nat Rev Cancer,* 6**,** 583-92.

MO, F., LIN, D., TAKHAR, M., RAMNARINE, V. R., DONG, X., BELL, R. H., VOLIK, S. V., WANG, K., XUE, H., WANG, Y., HAEGERT, A., ANDERSON, S., BRAHMBHATT, S., ERHO, N., WANG, X., GOUT, P. W., MORRIS, J., KARNES, R. J., DEN, R. B., KLEIN, E. A., SCHAEFFER, E. M., ROSS, A., REN, S., SAHINALP, S. C., LI, Y., XU, X., WANG, J., WANG, J., GLEAVE, M. E., DAVICIONI, E., SUN, Y., WANG, Y. & COLLINS, C. C. 2017. Stromal Gene Expression is Predictive for Metastatic Primary Prostate Cancer. *Eur Urol*.

MOISEENKO, F., VOLKOV, N., BOGDANOV, A., DUBINA, M. & MOISEYENKO, V. 2017. Resistance mechanisms to drug therapy in breast cancer and other solid tumors: An opinion. *F1000Research,* 6**,** 288.

MUZIC, R. F. & DIFILIPPO, F. P. 2014. PET/MRI – Technical Review. *Seminars in roentgenology,* 49**,** 242-254.

NA, F., WANG, J., LI, C., DENG, L., XUE, J. & LU, Y. 2014. Primary tumor standardized uptake value measured on F18-Fluorodeoxyglucose positron emission tomography is of prediction value for survival and local control in non-small-cell lung cancer receiving radiotherapy: meta-analysis. *J Thorac Oncol,* 9**,** 834-42.

NAKAJO, M., JINGUJI, M., TANI, A., NAGANO, H., NAKABEPPU, Y., NAKAJO, M. & YOSHIURA, T. 2017. Texture analysis of 18F-FDG PET/CT for predicting the malignant nature in thymic epithelial tumors. *Journal of Nuclear Medicine,* 58**,** 1048.

NECIB, H., DUSART, M., VANDERLINDEN, B. & BUVAT, I. 2008. Detection and characterization of the tumor change between two FDG PET scans using parametric imaging. 21-24.

NECIB, H., GARCIA, C., WAGNER, A., VANDERLINDEN, B., EMONTS, P., HENDLISZ, A., FLAMEN, P. & BUVAT, I. 2011. Detection and characterization of tumor changes in 18F-FDG PET patient monitoring using parametric imaging. *J Nucl Med,* 52**,** 354-61.

NEHMEH, S. A. & ERDI, Y. E. 2008. Respiratory motion in positron emission tomography/computed tomography: a review. *Semin Nucl Med,* 38**,** 167-76.

NG, F., GANESHAN, B., KOZARSKI, R., MILES, K. A. & GOH, V. 2013a. Assessment of primary colorectal cancer heterogeneity by using whole-tumor texture analysis: contrast-enhanced CT texture as a biomarker of 5-year survival. *Radiology,* 266**,** 177-84.

NG, F., KOZARSKI, R., GANESHAN, B. & GOH, V. 2013b. Assessment of tumor heterogeneity by CT texture analysis: can the largest cross-sectional area be used as an alternative to whole tumor analysis? *Eur J Radiol,* 82**,** 342-8.

NISHINO, M., DAHLBERG, S. E., CARDARELLA, S., JACKMAN, D. M., RABIN, M. S., RAMAIYA, N. H., HATABU, H., JANNE, P. A. & JOHNSON, B. E. 2013. Volumetric tumor growth in advanced non-small cell lung cancer patients with EGFR mutations during EGFR-tyrosine kinase inhibitor therapy: developing criteria to continue therapy beyond RECIST progression. *Cancer,* 119**,** 3761-8.

NOER, M. C., SPERLING, C. D., OTTESEN, B., ANTONSEN, S. L., CHRISTENSEN, I. J. & HOGDALL, C. 2017. Ovarian Cancer and Comorbidity: Is Poor Survival Explained by Choice of Primary Treatment or System Delay? *Int J Gynecol Cancer,* 27**,** 1123-1133.

NOOR, A. M., HOLMBERG, L., GILLETT, C. & GRIGORIADIS, A. 2015. Big Data: the challenge for small research groups in the era of cancer genomics. *Br J Cancer,* 113**,** 1405-1412.

NOWELL, P. C. 1976. The clonal evolution of tumor cell populations. *Science,* 194**,** 23-8.

O'CONNOR, J. P. B., ABOAGYE, E. O., ADAMS, J. E., AERTS, H. J. W. L., BARRINGTON, S. F., BEER, A. J., BOELLAARD, R., BOHNDIEK, S. E., BRADY, M., BROWN, G., BUCKLEY, D. L., CHENEVERT, T. L., CLARKE, L. P., COLLETTE, S., COOK, G. J., DESOUZA, N. M., DICKSON, J. C., DIVE, C., EVELHOCH, J. L., FAIVRE-FINN, C., GALLAGHER, F. A., GILBERT, F. J., GILLIES, R. J., GOH, V., GRIFFITHS, J. R., GROVES, A. M., HALLIGAN, S., HARRIS, A. L., HAWKES, D. J., HOEKSTRA, O. S., HUANG, E. P., HUTTON, B. F., JACKSON, E. F., JAYSON, G. C., JONES, A., KOH, D.-M., LACOMBE, D., LAMBIN, P., LASSAU, N., LEACH, M. O., LEE, T.-Y., LEEN, E. L., LEWIS, J. S., LIU, Y., LYTHGOE, M. F., MANOHARAN, P., MAXWELL, R. J., MILES, K. A., MORGAN, B., MORRIS, S., NG, T., PADHANI, A. R., PARKER, G. J. M., PARTRIDGE, M., PATHAK, A. P., PEET, A. C., PUNWANI, S., REYNOLDS, A. R., ROBINSON, S. P., SHANKAR, L. K., SHARMA, R. A., SOLOVIEV, D., STROOBANTS, S., SULLIVAN, D. C., TAYLOR, S. A., TOFTS, P. S., TOZER, G. M., VAN HERK, M., WALKER-SAMUEL, S., WASON, J., WILLIAMS, K. J., WORKMAN, P., YANKEELOV, T. E., BRINDLE, K. M., MCSHANE, L. M., JACKSON, A. & WATERTON, J. C. 2016. Imaging biomarker roadmap for cancer studies. *Nat Rev Clin Oncol,* advance online publication.

O'SULLIVAN, F., ROY, S. & EARY, J. 2003. A statistical measure of tissue heterogeneity with application to 3D PET sarcoma data. *Biostatistics,* 4**,** 433-48.

O'SULLIVAN, F., WOLSZTYNSKI, E., O'SULLIVAN, J., RICHARDS, T., CONRAD, E. U. & EARY, J. F. 2011. A statistical modeling approach to the analysis of spatial patterns of FDG-PET uptake in human sarcoma. *IEEE Trans Med Imaging,* 30**,** 2059-71.

OIKONOMOPOULOU, K., LI, L., ZHENG, Y., SIMON, I., WOLFERT, R. L., VALIK, D., NEKULOVA, M., SIMICKOVA, M., FRGALA, T. & DIAMANDIS, E. P. 2008. Prediction of ovarian cancer prognosis and response to chemotherapy by a serum-based multiparametric biomarker panel. *British Journal of Cancer,* 99**,** 1103-1113.

OKEREKE, I. C., GANGADHARAN, S. P., KENT, M. S., NICOTERA, S. P., SHEN, C. & DECAMP, M. M. 2009. Standard uptake value predicts survival in non-small cell lung cancer. *Ann Thorac Surg,* 88**,** 911-5; discussion 915-6.

ORLHAC, F., SOUSSAN, M., CHOUAHNIA, K., MARTINOD, E. & BUVAT, I. 2015. 18F-FDG PET-Derived Textural Indices Reflect Tissue-Specific Uptake Pattern in Non-Small Cell Lung Cancer. *PLoS One,* 10**,** e0145063.

ORLHAC, F., SOUSSAN, M., MAISONOBE, J. A., GARCIA, C. A., VANDERLINDEN, B. & BUVAT, I. 2014. Tumor texture analysis in 18F-FDG PET: relationships between texture

parameters, histogram indices, standardized uptake values, metabolic volumes, and total lesion glycolysis. *J Nucl Med,* 55**,** 414-22.

PAANS, A. & VAN WAARDE, A. 2002. Positron emission tomography. *Methods,* 27**,** 193-4.

PAPATHANASSIOU, D., BRUNA-MURAILLE, C., LIEHN, J. C., NGUYEN, T. D. & CURE, H. 2009. Positron Emission Tomography in oncology: present and future of PET and PET/CT. *Crit Rev Oncol Hematol,* 72**,** 239-54.

PARK, C. C., BISSELL, M. J. & BARCELLOS-HOFF, M. H. 2000. The influence of the microenvironment on the malignant phenotype. *Mol Med Today,* 6**,** 324-9.

PATANI, N., MARTIN, L. A. & DOWSETT, M. 2013. Biomarkers for the clinical management of breast cancer: international perspective. *Int J Cancer,* 133**,** 1-13.

PATEL, G. S., KIUCHI, T., LAWLER, K., OFO, E., FRUHWIRTH, G. O., KELLEHER, M., SHAMIL, E., ZHANG, R., SELVIN, P. R., SANTIS, G., SPICER, J., WOODMAN, N., GILLETT, C. E., BARBER, P. R., VOJNOVIC, B., KERI, G., SCHAEFFTER, T., GOH, V., O'DOHERTY, M. J., ELLIS, P. A. & NG, T. 2011. The challenges of integrating molecular imaging into the optimization of cancer therapy. *Integr Biol (Camb),* 3**,** 603-31.

PERDIGONES, N. & MURTAZA, M. 2017. Capturing tumor heterogeneity and clonal evolution in solid cancers using circulating tumor DNA analysis. *Pharmacol Ther,* 174**,** 22-26.

PETROU, M. & GARCÍA SEVILLA, P. 2006. Image Processing: Dealing with Texture. *Image Processing.* John Wiley & Sons, Ltd.

PHELPS, M. E. 2000. Positron emission tomography provides molecular imaging of biological processes. *Proc Natl Acad Sci U S A,* 97**,** 9226-33.

PIRKER, R., PEREIRA, J. R., SZCZESNA, A., VON PAWEL, J., KRZAKOWSKI, M., RAMLAU, R., VYNNYCHENKO, I., PARK, K., YU, C. T., GANUL, V., ROH, J. K., BAJETTA, E., O'BYRNE, K., DE MARINIS, F., EBERHARDT, W., GODDEMEIER, T., EMIG, M. & GATZEMEIER, U. 2009. Cetuximab plus chemotherapy in patients with advanced non-small-cell lung cancer (FLEX): an open-label randomised phase III trial. *Lancet,* 373**,** 1525-31.

PRAT, J. & ONCOLOGY, F. C. O. G. 2015. FIGO's staging classification for cancer of the ovary, fallopian tube, and peritoneum: abridged republication. *Journal of Gynecologic Oncology,* 26**,** 87-89.

PRIETO, E., MARTI-CLIMENT, J. M., MORAN, V., SANCHO, L., BARBES, B., ARBIZU, J. & RICHTER, J. A. 2015. Brain PET imaging optimization with time of flight and point spread function modelling. *Phys Med,* 31**,** 948-55.

PYKA, T., BUNDSCHUH, R. A., ANDRATSCHKE, N., MAYER, B., SPECHT, H. M., PAPP, L., ZSOTER, N. & ESSLER, M. 2015. Textural features in pre-treatment [F18]-FDG-PET/CT are correlated with risk of local recurrence and disease-specific survival in early stage NSCLC patients receiving primary stereotactic radiation therapy. *Radiat Oncol,* 10**,** 100.

RESEARCH, U. C. 2016. *Cancer Research UK Cancer in General Treatment* [Online]. Available: http://www.cancerresearchuk.org/about-cancer/cancers-in-general/treatment/ [Accessed 29/10/2016 2016].

RIVARD, J. D., TEMPLE, W. J., MCCONNELL, Y. J., SULTAN, H. & MACK, L. A. 2014. Preoperative computed tomography does not predict resectability in peritoneal carcinomatosis. *Am J Surg,* 207**,** 760-4; discussion 764-5.

RIZVI, N. A., HELLMANN, M. D., SNYDER, A., KVISTBORG, P., MAKAROV, V., HAVEL, J. J., LEE, W., YUAN, J., WONG, P., HO, T. S., MILLER, M. L., REKHTMAN, N., MOREIRA, A. L., IBRAHIM, F., BRUGGEMAN, C., GASMI, B., ZAPPASODI, R., MAEDA, Y., SANDER, C., GARON, E. B., MERGHOUB, T., WOLCHOK, J. D., SCHUMACHER, T. N. & CHAN, T. A. 2015. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science,* 348**,** 124-8.

RODRIGUEZ, L. V. & TERRIS, M. K. 1998. Risks and complications of transrectal ultrasound guided prostate needle biopsy: a prospective study and review of the literature. *J Urol,* 160**,** 2115-20.

ROSENDAHL, M., HARTER, P., BJORN, S. F. & HOGDALL, C. 2018. Specific Regions, Rather than the Entire Peritoneal Carcinosis Index, are Predictive of Complete Resection and Survival in Advanced Epithelial Ovarian Cancer. *Int J Gynecol Cancer,* 28**,** 316-322.

ROUSSET, O., RAHMIM, A., ALAVI, A. & ZAIDI, H. 2007. Partial Volume Correction Strategies in PET. *PET Clinics,* 2**,** 235-249.

SACCONI, B., ANZIDEI, M., LEONARDI, A., BONI, F., SABA, L., SCIPIONE, R., ANILE, M., RENGO, M., LONGO, F., BEZZI, M., VENUTA, F., NAPOLI, A., LAGHI, A. & CATALANO, C. 2017. Analysis of CT features and quantitative texture analysis in patients with lung adenocarcinoma: a correlation with EGFR mutations and survival rates. *Clin Radiol,* 72**,** 443-450.

SALAMON, J., DERLIN, T., BANNAS, P., BUSCH, J. D., HERRMANN, J., BOCKHORN, M., HAGEL, C., FRIEDRICH, R. E., ADAM, G. & MAUTNER, V. F. 2013. Evaluation of intratumoural heterogeneity on (1)(8)F-FDG PET/CT for characterization of peripheral nerve sheath tumours in neurofibromatosis type 1. *Eur J Nucl Med Mol Imaging,* 40**,** 685-92.

SALAVATI, A., DUAN, F., SNYDER, B. S., WEI, B., HOUSHMAND, S., KHIEWVAN, B., OPANOWSKI, A., SIMONE, C. B., SIEGEL, B. A., MACHTAY, M. & ALAVI, A. 2017. Optimal FDG PET/CT volumetric parameters for risk stratification in patients with locally advanced non-small cell lung cancer: results from the ACRIN 6668/RTOG 0235 trial. *European Journal of Nuclear Medicine and Molecular Imaging*.

SCHMIDT, S., MEULI, R. A., ACHTARI, C. & PRIOR, J. O. 2015. Peritoneal carcinomatosis in primary ovarian cancer staging: comparison between MDCT, MRI, and 18F-FDG PET/CT. *Clin Nucl Med,* 40**,** 371-7.

SCHMITT, T., LEHNER, B., KASPER, B., BISCHOF, M., ROEDER, F., DIETRICH, S., DIMITRAKOPOULOU-STRAUSS, A., STRAUSS, L. G., MECHTERSHEIMER, G., WUCHTER, P., HO, A. D. & EGERER, G. 2011. A phase II study evaluating neo-/adjuvant EIA chemotherapy, surgical resection and radiotherapy in high-risk soft tissue sarcoma. *BMC Cancer,* 11**,** 510.

SCHNEIDER-KOLSKY, M. E., HART, S., FOX, J., MIDOLO, P., STUCKEY, J., HOFMAN, M. & GANJU, V. 2010. The role of chemotherapeutic drugs in the evaluation of breast tumour response to chemotherapy using serial FDG-PET. *Breast Cancer Res,* 12**,** R37.

SCHREIBMANN, E., WALLER, A. F., CROCKER, I., CURRAN, W. & FOX, T. 2013. Voxel clustering for quantifying PET-based treatment response assessment. *Med Phys,* 40**,** 012401.

SEGAL, E., SIRLIN, C. B., OOI, C., ADLER, A. S., GOLLUB, J., CHEN, X., CHAN, B. K., MATCUK, G. R., BARRY, C. T., CHANG, H. Y. & KUO, M. D. 2007. Decoding global gene expression programs in liver cancer by noninvasive imaging. *Nat Biotechnol,* 25**,** 675-80.

SHAFIQ-UL-HASSAN, M., ZHANG, G. G., LATIFI, K., ULLAH, G., HUNT, D. C., BALAGURUNATHAN, Y., ABDALAH, M. A., SCHABATH, M. B., GOLDGOF, D. G., MACKIN, D., COURT, L. E., GILLIES, R. J. & MOROS, E. G. 2017. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. *Medical Physics,* 44**,** 1050-1062.

SHAH, S. N. & HUANG, S. S. 2015. Hybrid PET/MR imaging: physics and technical considerations. *Abdom Imaging,* 40**,** 1358-65.

SHEN, Y. C., TSENG, G. C., TU, C. Y., CHEN, W. C., LIAO, W. C., CHEN, W. C., LI, C. H., CHEN, H. J. & HSIA, T. C. 2017. Comparing the effects of afatinib with gefitinib or Erlotinib in patients with advanced-stage lung adenocarcinoma harboring non-classical epidermal growth factor receptor mutations. *Lung Cancer,* 110**,** 56-62.

SHERMAN-BAUST, C. A., WEERARATNA, A. T., RANGEL, L. B., PIZER, E. S., CHO, K. R., SCHWARTZ, D. R., SHOCK, T. & MORIN, P. J. 2003. Remodeling of the extracellular matrix through overexpression of collagen VI contributes to cisplatin resistance in ovarian cancer cells. *Cancer Cell,* 3**,** 377-86.

SILVA, A. S. & GATENBY, R. A. 2010. A theoretical quantitative model for evolution of cancer chemotherapy resistance. *Biol Direct,* 5**,** 25.

SKOUGAARD, K., NIELSEN, D., JENSEN, B. V. & HENDEL, H. W. 2013. Comparison of EORTC criteria and PERCIST for PET/CT response evaluation of patients with metastatic colorectal cancer treated with irinotecan and cetuximab. *J Nucl Med,* 54**,** 1026-31.

SMALLEY, K. S. M., BRAFFORD, P. A. & HERLYN, M. 2005. Selective evolutionary pressure from the tissue microenvironment drives tumor progression. *Seminars in Cancer Biology,* 15**,** 451-459.

SÖLÉTORMOS, G., DUFFY, M. J., HASSAN, S. O. A., VERHEIJEN, R. H. M., THOLANDER, B., BAST, R. C., GAARENSTROOM, K. N., STURGEON, C. M., BONFRER, J. M., PETERSEN, P. H., TROONEN, H., TORRE, G. C., KULPA, J. K., TUXEN, M. K. & MOLINA, R. 2016. Clinical Use of Cancer Biomarkers in Epithelial Ovarian Cancer: Updated Guidelines from the European Group on Tumor Markers (EGTM). *International journal of gynecological cancer : official journal of the International Gynecological Cancer Society,* 26**,** 43-51.

SORET, M., BACHARACH, S. L. & BUVAT, I. 2007. Partial-volume effect in PET tumor imaging. *J Nucl Med,* 48**,** 932-45.

SOUSSAN, M., ORLHAC, F., BOUBAYA, M., ZELEK, L., ZIOL, M., EDER, V. & BUVAT, I. 2014. Relationship between tumor heterogeneity measured on FDG-PET/CT and pathological prognostic factors in invasive breast cancer. *PLoS One,* 9**,** e94017.

STRAUSS, L. G., KLIPPEL, S., PAN, L., SCHONLEBEN, K., HABERKORN, U. & DIMITRAKOPOULOU-STRAUSS, A. 2007. Assessment of quantitative FDG PET data in primary colorectal tumours: which parameters are important with respect to tumour detection? *Eur J Nucl Med Mol Imaging,* 34**,** 868-77.

STRICKER, T., CATENACCI, D. V. & SEIWERT, T. Y. 2011. Molecular profiling of cancer--the future of personalized cancer medicine: a primer on cancer biology and the tools necessary to bring molecular testing to the clinic. *Semin Oncol,* 38**,** 173-85.

SUBRAMANIAN, J. & GOVINDAN, R. 2008. Molecular genetics of lung cancer in people who have never smoked. *Lancet Oncol,* 9**,** 676-82.

SUGAWARA, Y., ZASADNY, K. R., NEUHOFF, A. W. & WAHL, R. L. 1999. Reevaluation of the standardized uptake value for FDG: variations with body weight and methods for correction. *Radiology,* 213**,** 521-5.

SUZUKI, C., WALLGREN, H., ABRAHAM-NORDLING, M. & PALMER, G. 2018. Preoperative CT-based predictive factors for resectability and medium-term overall survival in patients with peritoneal carcinomatosis from colorectal cancer. *Clin Radiol,* 73**,** 756.e11-756.e16.

SVED, P. D., GOMEZ, P., MANOHARAN, M., KIM, S. S. & SOLOWAY, M. S. 2004. Limitations of biopsy Gleason grade: implications for counseling patients with biopsy Gleason score 6 prostate cancer. *J Urol,* 172**,** 98-102.

TAN, S., KLIGERMAN, S., CHEN, W., LU, M., KIM, G., FEIGENBERG, S., D'SOUZA, W. D., SUNTHARALINGAM, M. & LU, W. 2013. Spatial-temporal [(1)(8)F]FDG-PET features for predicting pathologic response of esophageal cancer to neoadjuvant chemoradiation therapy. *Int J Radiat Oncol Biol Phys,* 85**,** 1375-82.

TANG, X. O. 1998. Texture information in run-length matrices. *Ieee Transactions on Image Processing,* 7**,** 1602-1609.

THERASSE, P., ARBUCK, S. G., EISENHAUER, E. A., WANDERS, J., KAPLAN, R. S., RUBINSTEIN, L., VERWEIJ, J., VAN GLABBEKE, M., VAN OOSTEROM, A. T., CHRISTIAN, M. C. & GWYTHER, S. G. 2000. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst,* 92**,** 205-16.

THIBAULT, G., FERTIL, B., NAVARRO, C., PEREIRA, S., CAU, P., LEVY, N., SEQUEIRA, J. & MARI, J. 2009. Texture Indexes and Gray Level Size Zone Matrix Application to Cell Nuclei Classification. *Pattern Recognition Inf Process*, 140-145.

THIE, J. A. 2004. Understanding the Standardized Uptake Value, Its Methods, and Implications for Usage. *Journal of Nuclear Medicine,* 45, 1431-1434.

TIAN, F., HAYANO, K., KAMBADAKONE, A. R. & SAHANI, D. V. 2015. Response assessment to neoadjuvant therapy in soft tissue sarcomas: using CT texture analysis in comparison to tumor size, density, and perfusion. *Abdom Imaging,* 40, 1705-12.

TIBSHIRANI, R. 1996. Regression Shrinkage and Selection via the Lasso. *J R Statist,* 58, 267-288.

TIBSHIRANI, R. 1997. The lasso method for variable selection in the Cox model. *Stat Med,* 16, 385-95.

TIXIER, F., HATT, M., VALLA, C., FLEURY, V., LAMOUR, C., EZZOUHRI, S., INGRAND, P., PERDRISOT, R., VISVIKIS, D. & LE REST, C. C. 2014. Visual versus quantitative assessment of intratumor 18F-FDG PET uptake heterogeneity: prognostic value in non-small cell lung cancer. *J Nucl Med,* 55, 1235-41.

TIXIER, F., LE REST, C. C., HATT, M., ALBARGHACH, N., PRADIER, O., METGES, J. P., CORCOS, L. & VISVIKIS, D. 2011. Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. *J Nucl Med,* 52, 369-78.

TOMASI, G., TURKHEIMER, F. & ABOAGYE, E. 2012. Importance of quantification for the analysis of PET data in oncology: review of current methods and trends for the future. *Mol Imaging Biol,* 14, 131-46.

TØNDEL, C., VIKSE, B. E., BOSTAD, L. & SVARSTAD, E. 2012. Safety and Complications of Percutaneous Kidney Biopsies in 715 Children and 8573 Adults in Norway 1988–2010. *Clinical Journal of the American Society of Nephrology : CJASN,* 7, 1591-1597.

TSOUMPAS, C., BUERGER, C., KING, A. P., MOLLET, P., KEEREMAN, V., VANDENBERGHE, S., SCHULZ, V., SCHLEYER, P., SCHAEFFTER, T. & MARSDEN, P. K. 2011. Fast generation of 4D PET-MR data from real dynamic MR acquisitions. *Physics in Medicine and Biology,* 56, 6597.

UK, C. R. 2016. *Lifetime risk of cancer* [Online]. Cancer Research UK. Available: http://www.cancerresearchuk.org/health-professional/cancer-statistics/risk/lifetime-risk - heading-Zero [Accessed 04/10/2016].

UTO, F., SHIBA, E., ONOUE, S., YOSHIMURA, H., TAKADA, M., TSUJI, Y., FUKUGAMI, S., ASAKAWA, I., TAMAMOTO, T. & HASEGAWA, M. 2010. Phantom study on radiotherapy planning using PET/CT--delineation of GTV by evaluating SUV. *J Radiat Res,* 51, 157-64.

VAIDYA, M., CREACH, K. M., FRYE, J., DEHDASHTI, F., BRADLEY, J. D. & EL NAQA, I. 2012. Combined PET/CT image characteristics for radiotherapy tumor response in lung cancer. *Radiother Oncol,* 102, 239-45.

VALLIERES, M., KAY-RIVEST, E., PERRIN, L. J., LIEM, X., FURSTOSS, C., AERTS, H., KHAOUAM, N., NGUYEN-TAN, P. F., WANG, C. S., SULTANEM, K., SEUNTJENS, J. & EL NAQA, I. 2017. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. *Sci Rep,* 7, 10117.

VAN VELDEN, F. H., CHEEBSUMON, P., YAQUB, M., SMIT, E. F., HOEKSTRA, O. S., LAMMERTSMA, A. A. & BOELLAARD, R. 2011. Evaluation of a cumulative SUV-volume histogram method for parameterizing heterogeneous intratumoural FDG uptake in non-small cell lung cancer PET studies. *Eur J Nucl Med Mol Imaging,* 38, 1636-47.

VANDERHOEK, M., PERLMAN, S. B. & JERAJ, R. 2012. Impact of the definition of peak standardized uptake value on quantification of treatment response. *J Nucl Med,* 53, 4-11.

VANSTEENKISTE, J. F., STROOBANTS, S. G., DUPONT, P. J., DE LEYN, P. R., VERBEKEN, E. K., DENEFFE, G. J., MORTELMANS, L. A. & DEMEDTS, M. G. 1999. Prognostic importance of the standardized uptake value on (18)F-fluoro-2-deoxy-glucose-positron emission tomography scan in non-small-cell lung cancer: An analysis of 125 cases. Leuven Lung Cancer Group. *J Clin Oncol,* 17**,** 3201-6.

VARGAS, H. A., VEERARAGHAVAN, H., MICCO, M., NOUGARET, S., LAKHMAN, Y., MEIER, A. A., SOSA, R., SOSLOW, R. A., LEVINE, D. A., WEIGELT, B., AGHAJANIAN, C., HRICAK, H., DEASY, J., SNYDER, A. & SALA, E. 2017. A novel representation of inter-site tumour heterogeneity from pre-treatment computed tomography textures classifies ovarian cancers by clinical outcome. *Eur Radiol,* 27**,** 3991-4001.

VISVADER, J. E. & LINDEMAN, G. J. 2008. Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. *Nat Rev Cancer,* 8**,** 755-68.

VISVIKIS, D., LAMARE, F., BRUYANT, P., BOUSSION, N. & CHEZE LE REST, C. 2006. Respiratory motion in positron emission tomography for oncology applications: Problems and solutions. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* 569**,** 453-457.

WAHL, R. L., JACENE, H., KASAMON, Y. & LODGE, M. A. 2009. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. *J Nucl Med,* 50 Suppl 1**,** 122S-50S.

WANG, L. M., SILVA, M. A., D'COSTA, Z., BOCKELMANN, R., SOONAWALLA, Z., LIU, S., O'NEILL, E., MUKHERJEE, S., MCKENNA, W. G., MUSCHEL, R. & FOKAS, E. 2016a. The prognostic role of desmoplastic stroma in pancreatic ductal adenocarcinoma. *Oncotarget,* 7**,** 4183-94.

WANG, S. J. 2015. Surveillance radiologic imaging after treatment of oropharyngeal cancer: a review. *World Journal of Surgical Oncology,* 13**,** 94.

WANG, W., KRYCZEK, I., DOSTAL, L., LIN, H., TAN, L., ZHAO, L., LU, F., WEI, S., MAJ, T., PENG, D., HE, G., VATAN, L., SZELIGA, W., KUICK, R., KOTARSKI, J., TARKOWSKI, R., DOU, Y., RATTAN, R., MUNKARAH, A., LIU, J. R. & ZOU, W. 2016b. Effector T Cells Abrogate Stroma-Mediated Chemoresistance in Ovarian Cancer. *Cell,* 165**,** 1092-1105.

WATERTON, J. C. & PYLKKANEN, L. 2012. Qualification of imaging biomarkers for oncology drug development. *Eur J Cancer,* 48**,** 409-15.

WEBER, W. A. 2006. Positron emission tomography as an imaging biomarker. *J Clin Oncol,* 24**,** 3282-92.

WILLAIME, J. M., ABOAGYE, E. O., TSOUMPAS, C. & TURKHEIMER, F. E. 2014. A multifractal approach to space-filling recovery for PET quantification. *Med Phys,* 41**,** 112505.

WILLAIME, J. M., TURKHEIMER, F. E., KENNY, L. M. & ABOAGYE, E. O. 2013. Quantification of intra-tumour cell proliferation heterogeneity using imaging descriptors of 18F fluorothymidine-positron emission tomography. *Phys Med Biol,* 58**,** 187-203.

WIN, T., MILES, K. A., JANES, S. M., GANESHAN, B., SHASTRY, M., ENDOZO, R., MEAGHER, M., SHORTMAN, R. I., WAN, S., KAYANI, I., ELL, P. J. & GROVES, A. M. 2013. Tumor heterogeneity and permeability as measured on the CT component of PET/CT predict survival in patients with non-small cell lung cancer. *Clin Cancer Res,* 19**,** 3591-9.

YANG, F., THOMAS, M. A., DEHDASHTI, F. & GRIGSBY, P. W. 2013. Temporal analysis of intratumoral metabolic heterogeneity characterized by textural features in cervical cancer. *Eur J Nucl Med Mol Imaging,* 40**,** 716-27.

YANG, Y. S., GUCCIONE, S. & BEDNARSKI, M. D. 2003. Comparing genomic and histologic correlations to radiographic changes in tumors: a murine SCC VII model study. *Acad Radiol,* 10**,** 1165-75.

YAP, T. A., GERLINGER, M., FUTREAL, P. A., PUSZTAI, L. & SWANTON, C. 2012. Intratumor heterogeneity: seeing the wood for the trees. *Sci Transl Med,* 4**,** 127ps10.

YIP, C., DAVNALL, F., KOZARSKI, R., LANDAU, D. B., COOK, G. J., ROSS, P., MASON, R. & GOH, V. 2015. Assessment of changes in tumor heterogeneity following neoadjuvant chemotherapy in primary esophageal cancer. *Dis Esophagus,* 28**,** 172-9.

YIP, C., LANDAU, D., KOZARSKI, R., GANESHAN, B., THOMAS, R., MICHAELIDOU, A. & GOH, V. 2014. Primary esophageal cancer: heterogeneity as potential prognostic biomarker in patients treated with definitive chemotherapy and radiation therapy. *Radiology,* 270**,** 141-8.

YOSHIHARA, K., SHAHMORADGOLI, M., MARTINEZ, E., VEGESNA, R., KIM, H., TORRES-GARCIA, W., TREVINO, V., SHEN, H., LAIRD, P. W., LEVINE, D. A., CARTER, S. L., GETZ, G., STEMKE-HALE, K., MILLS, G. B. & VERHAAK, R. G. 2013. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun,* 4**,** 2612.

YOUNG, H., BAUM, R., CREMERIUS, U., HERHOLZ, K., HOEKSTRA, O., LAMMERTSMA, A. A., PRUIM, J. & PRICE, P. 1999. Measurement of clinical and subclinical tumour response using [18F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. European Organization for Research and Treatment of Cancer (EORTC) PET Study Group. *Eur J Cancer,* 35**,** 1773-82.

YU, H., CALDWELL, C., MAH, K. & MOZEG, D. 2009a. Coregistered FDG PET/CT-based textural characterization of head and neck cancer for radiation treatment planning. *IEEE Trans Med Imaging,* 28**,** 374-83.

YU, H., CALDWELL, C., MAH, K., POON, I., BALOGH, J., MACKENZIE, R., KHAOUAM, N. & TIRONA, R. 2009b. Automated radiation targeting in head-and-neck cancer using region-based texture analysis of PET and CT images. *Int J Radiat Oncol Biol Phys,* 75**,** 618-25.

YUSHKEVICH, P. A., PIVEN, J., HAZLETT, H. C., SMITH, R. G., HO, S., GEE, J. C. & GERIG, G. 2006. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage,* 31**,** 1116-28.

ZASADNY, K. R. & WAHL, R. L. 1993. Standardized uptake values of normal tissues at PET with 2-[fluorine-18]-fluoro-2-deoxy-D-glucose: variations with body weight and a method for correction. *Radiology,* 189**,** 847-50.

ZHANG, G. M., SUN, H., SHI, B., JIN, Z. Y. & XUE, H. D. 2017. Quantitative CT texture analysis for evaluating histologic grade of urothelial carcinoma. *Abdom Radiol (NY),* 42**,** 561-568.

ZHANG, H., GRAHAM, C. M., ELCI, O., GRISWOLD, M. E., ZHANG, X., KHAN, M. A., PITMAN, K., CAUDELL, J. J., HAMILTON, R. D., GANESHAN, B. & SMITH, A. D. 2013. Locally advanced squamous cell carcinoma of the head and neck: CT texture and histogram analysis allow independent prediction of overall survival in patients treated with induction chemotherapy. *Radiology,* 269**,** 801-9.

ZHANG, S. Z., JING, Y., ZHANG, M. Y., ZHANG, Z. F., MA, P. F., PENG, H. X., SHI, K. X., GAO, W. Q. & ZHUANG, G. L. 2015. Stroma-associated master regulators of molecular subtypes predict patient prognosis in ovarian cancer. *Scientific Reports,* 5.

Presentations

Arshad M, Thornton A, Scarsbrook A, McDermott G, Cook G, Chua S, O'Connor R, Barwick T, Rockall A, Aboagye EO. Identification of Texture Features from FDG PET images that are associated with Survival Outcomes in Non-small Cell Lung Cancer. CRUK-EPSRC National meeting. 12 September 2016.

Arshad M, Thornton A, Rodgers, N, Scarsbrook A, McDermott G, Cook G, Chua S, O'Connor R, Barwick T, Rockall A, Aboagye EO. Identification of Texture Features from FDG PET images that are associated with Survival Outcomes in Non-small Cell Lung Cancer. National Cancer Research Institute Conference. 6-9 November 2016.

Publications

Please find PDFs attached

Radiopharmaceuticals as probes to characterize tumour tissue.
Alam IS, Arshad MA, Nguyen QD, Aboagye EO.
Eur J Nucl Med Mol Imaging. 2015 Apr;42(4):537-61.

Discovery of pre-therapy 2-deoxy-2-18F-fluoro-D-glucose positron emission tomography-based radiomics classifiers of survival outcome in non-small-cell lung cancer patients.
Arshad MA, Thornton A, Lu H, Tam H, Wallitt K, Rodgers N, Scarsbrook A, McDermott G, Cook GJ, Landau D, Chua S, O'Connor R, Dickson J, Power DA, Barwick TD, Rockall A, Aboagye EO.
Eur J Nucl Med Mol Imaging. 2019 Feb;46(2):455-466. Epub 2018 Sep 1.

A mathematical-descriptor of tumor-mesoscopic-structure from computed-tomography images annotates prognostic- and molecular-phenotypes of epithelial ovarian cancer.
Lu H, Arshad M, Thornton A, Avesani G, Cunnea P, Curry E, Kanavati F, Liang J, Nixon K, Williams ST, Hassan MA, Bowtell DDL, Gabra H, Fotopoulou C, Rockall A, Aboagye EO.
Nat Commun. 2019 Feb 15;10(1):764.

Appendices

Appendix A1

Literature review-search strategy

The search was last run on 17/11/2016 (MEDLINE and Web of Science) using the following search terms and combination:

(variability OR heteroge* OR descriptor OR spatial OR feature)

AND

(characteri* OR quantificat*)

AND

(PET OR "Positron Emission Tomography" OR CT OR "Computerised Tomography") AND (cancer OR lesion OR oncolog* OR neoplasm* OR tumour*)

Appendix A2

Literature review-identification, screening and selection of publications for literature review

Records identified through database searching

N= 5670 (web of Science)

N= 2644 (MEDLINE)

Records selected (by title)

N= 175

Records after duplicates removed

N = 133

Records selected (by abstract)

Image Descriptors

N= 20

Full-text article assessed for eligibility

N= 51

Records eligible for literature review

N= 71

Appendix B

This is a detailed table of all the texture features with a description and the mathematical formulae

Name	Description		Mathematical Formula
Size and Shape based features			
SNS_vol	Volume	The volume (V) of the tumour is obtained by multiplying the total number of pixels in the mask by voxel size	
SNS_area	Area	The surface area is calculated by triangulation (i.e. dividing the surface into connected triangles)	$$\sum_{i=1}^{N} \frac{1}{2} \lvert a_i b_i \times a_i c_i \rvert$$
SNS_s2v	Surface to volume		$$\frac{A}{V}$$

	ratio		
SNS_sph	Sphericity		$\dfrac{\pi^{\frac{1}{3}}(6V)^{\frac{2}{3}}}{A}$
SNS_sph_dis	Spherical dispropor tion		$\dfrac{A}{4\pi R^2}$
SNS_com_1	Compactn ess 1		$\dfrac{v}{\sqrt{\pi A^{\frac{2}{3}}}}$
SNS_com_2	Compactn ess 2		$36\pi\dfrac{V^2}{A^3}$
SNS_max3d	Maximum 3d diameter	The maximum three-dimensional tumour diameter is measured as the largest pairwise Euclidean distance, between voxels on the surface of the tumor volume.	
FOS_CV	Coefficien t of Variation		$\dfrac{1}{N}\sum\limits_{v=1}^{N}1(v)\sqrt{\dfrac{1}{N}\sum\limits_{v=1}^{N}(1(v)-I_{mean})^2}$

FOS_Imean	Mean		$\dfrac{1}{N}\displaystyle\sum_{i}^{N} X(i)$
FOS_Imedian	Median	The median intensity value of X	
FOS_Imode	Mode		
FOS_Istd	Standard Deviation		$\left(\dfrac{1}{N-1}\displaystyle\sum_{i=1}^{N}(X(i)-\bar{x})^2\right)^{1/2}$ where $\bar{x}$ is the mean of x
FOS_Imin	Minimum	The minimum intensity value of X	
FOS_Imax	Maximum	The maximum intensity value of X	
FOS_Range	Range	The range of intensity values of X	
FOS_Skew	Skewness	The degree of histogram asymmetry around the mean	$\dfrac{\frac{1}{N}\sum_{i=1}^{N}(X(i)-\bar{x})^3}{\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X(i)-\bar{x})^2}\right)^3}$ where $\bar{x}$ is the mean of x
FOS_Kurt	Kurtosis	Measure of the histogram sharpness.	$\dfrac{\frac{1}{N}\sum_{i=1}^{N}(X(i)-\bar{x})^4}{\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X(i)-\bar{x})^2}\right)^2}$

			where $\bar{x}$ is the mean of x
FOS_ImeanAbsDev	Mean Absolute Deviation	The mean of the absolute deviations of all voxel intensities around the mean intensity value	
FOS_RMS	Root mean square		$$\sqrt{\frac{\sum_i^N X(i)^2}{N}}$$
AUC-CSH	Area under the Curve		
FOS_Entr	Entropy	Measure of information content	$$-\sum_{i=1}^{Ng} p(i)log_2 p(i)$$
FOS_Ener	Energy	Measure of repeated pixel pairs	$$\sum_i^{N_l} X(i)^2$$
GLSZM_SmallZone	Small Zone Emphasis	Emphasizes small zones	$$\sum_{i=1}^{Ng}\sum_{j=1}^{Nz} \left(\frac{Szpcent}{j}\right)^2 p(i,j) \Big/ \sum_{i=1}^{Ng}\sum_{j=1}^{Nz} p(i,j)$$
GLSZM_LargeZone	Large Zone Emphasis	Emphasizes large zones	$$\sum_{i=1}^{Ng}\sum_{j=1}^{Nz} (Szpcent \times j)^2 p(i,j) \Big/ \sum_{i=1}^{Ng}\sum_{j=1}^{Nz} p(i,j)$$

GLSZM_GlNonUnif	Grey-level Non Uniformity	Measure of grey-level variability	$$\sum_{i=1}^{Ng}\left(\sum_{j=1}^{Nz}p(i,j)\right)^2\Bigg/\sum_{i=1}^{Ng}\sum_{j=1}^{Nz}p(i,j)$$
GLSZM_SzNonUnif	Size zone Non uniformity	Measure of Size-Zone variability	$$\sum_{j=1}^{Ng}\left(\sum_{i=1}^{Nz}p(i,j)\right)^2\Bigg/\sum_{i=1}^{Ng}\sum_{j=1}^{Nz}p(i,j)$$
GLSZM_ZonePcent	Zone Percentage	Ratio between the number of zones and the total number of possible zones (number of voxels)	$$\sum_{i=1}^{Ng}\sum_{j=1}^{Nz}p(i,j)\Big/N$$
GLSZM_ZoneLoGl	Zone Low grey-level Emphasis	Emphasizes zones of low grey-level	$$\sum_{i=1}^{Ng}\sum_{j=1}^{Nz}\frac{p(i,j)}{i^2}\Bigg/\sum_{i=1}^{Ng}\sum_{j=1}^{Nz}p(i,j)$$
GLSZM_ZoneHiGl	Zone Low grey-level Emphasis	Emphasizes zones of high grey-level	$$\sum_{i=1}^{Ng}\sum_{j=1}^{Nz}i^2p(i,j)\Bigg/\sum_{i=1}^{Ng}\sum_{j=1}^{Nz}p(i,j)$$
GLSZM_SzoneLoGl	Small Zone Low grey level Zone Emphasis	Emphasizes small zones of low grey-level	$$\sum_{i=1}^{Ng}\sum_{j=1}^{Nz}\left(\frac{Szpcent}{j}\right)^2\frac{p(i,j)}{i^2}\Bigg/\sum_{i=1}^{Ng}\sum_{j=1}^{Nz}p(i,j)$$

GLSZM_SzoneHiGl	Small Zone High grey level Zone Emphasis)	Emphasizes small zones of high grey-level	$$\sum_{i=1}^{Ng}\sum_{j=1}^{Nz}\left(\frac{Szpcent}{j}\right)^2 i^2 p(i,j) \Big/ \sum_{i=1}^{Ng}\sum_{j=1}^{Nz} p(i,j)$$
GLSZM_LzoneLogl	Large Zone Low grey level Emphasis)	Emphasizes large zones of low grey-level	$$\sum_{i=1}^{Ng}\sum_{j=1}^{Nz}\frac{(Szpcent \times j)^2}{i^2} p(i,j) \Big/ \sum_{i=1}^{Ng}\sum_{j=1}^{Nz} p(i,j)$$
GLSZM_LzoneHiGl	Large Zone High grey level Emphasis)	Emphasizes large zones of high grey-level	$$\sum_{i=1}^{Ng}\sum_{j=1}^{Nz}(Szpcent \times j)^2 i^2 p(i,j) \Big/ \sum_{i=1}^{Ng}\sum_{j=1}^{Nz} p(i,j)$$
GLSZM_GlVarianc	Grey Level Variance)	Weighted variance of grey-level	$$\left(\frac{1}{Ng \times Nz}\sum_{i=1}^{Ng}\sum_{j=1}^{Nz}\left(i \times p(i,j) - \mu_{gl}\right)^2\right)^{1/2}$$ with $\mu_{gl} = \frac{1}{Ng \times Nz}\sum_{i=1}^{Ng}\sum_{j=1}^{Nz} i \times p(i,j)$
GLSZM_SzVarianc	Size-Zone Variance)	Weighted variance of size-zone	$$\left(\frac{1}{Ng \times Nz}\sum_{i=1}^{Ng}\sum_{j=1}^{Nz}\left((Szpcent \times j) \times p(i,j) - \mu_{NZ}\right)^2\right)^{1/2}$$ with $\mu_{gl} = \frac{1}{Ng \times Nz}\sum_{i=1}^{Ng}\sum_{j=1}^{Nz}(Szpcent \times j) \times p(i,j)$
NGTDM_Coarse	Coarseness	Measure of texture uniformity	$$\left[\varepsilon + \sum_{i=1}^{N_g} p_i M(i)\right]^{-1}$$
NGTDM_Contra	Contrast	Measure of contrast taking into account the global dynamic	$$\left[\frac{1}{N_t(N_t-1)}\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p_i p_j (i-j)^2\right]\left[\frac{1}{n^2}\sum_{i=1}^{N_g} M(i)\right]$$

		range of grey-levels and local variations	
NGTDM_Busyne	Busyness	Measure the rate of change between neighbourhood intensities weighed by the difference in intensities.	$$\left[\sum_{i=0}^{N_g} p_i s(i)\right] \Big/ \left[\sum_{i=0}^{N_g}\sum_{j=0}^{N_g} ip_i - jp_j\right]$$ $$p_i \neq 0, p_j \neq 0$$
NGTDM_Complex	Complex	Complexity increases with high spatial intensity variations, Contrast and the presence of small clusters of voxels of different intensities	$$\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\left\{\frac{|i-j|}{N_i+N_j}\right\}\{p_i M(i) + p_j M(j)\}$$
NGTDM_Streng	Strength	Measure of 'distinguishability' between	$$\frac{\left[\sum_{i=1}^{N_g}(p_i+p_j)(i-j)^2\right]}{\left[\varepsilon + \sum_{i=1}^{N_g} M(i)\right]} p_i \neq 0, p_j \neq 0$$

		clusters of voxels. This depends on cluster sizes and grey-level differences between patches	
GLRLM_SRE	Short Run Emphasis	Measures the occurrence of short runs	$$\dfrac{\sum_{i=1}^{Ng}\sum_{j=1}^{Nr}\left[\dfrac{p(i,j|\theta)}{j^2}\right]}{\sum_{i=1}^{Ng}\sum_{j=1}^{Nr}p(i,j|\theta)}$$
GLRLM_LRE	Long Run Emphasis	Measures the occurrence of long runs	$$\dfrac{\sum_{i=1}^{Ng}\sum_{j=1}^{Nr}j^2 p(i,j|\theta)}{\sum_{i=1}^{Ng}\sum_{j=1}^{Nr}p(i,j|\theta)}$$
GLRLM_GLN	Grey Level Non-Uniformity	Measures the similarity of pixel intensities	$$\dfrac{\sum_{i=1}^{Ng}\left[\sum_{j=1}^{Nr}p(i,j|\theta)\right]^2}{\sum_{i=1}^{Ng}\sum_{j=1}^{Nr}p(i,j|\theta)}$$
GLRLM_RLN	Run Length Non-Uniformity	Measures the similarity of lengths of pixels	$$\dfrac{\sum_{i=1}^{Nr}\left[\sum_{j=1}^{Ng}p(i,j|\theta)\right]^2}{\sum_{i=1}^{Ng}\sum_{j=1}^{Nr}p(i,j|\theta)}$$
GLRLM_RP	Run Percentag	Measures the percentage	$$\sum_{i=1}^{Ng}\sum_{j=1}^{Nr}\dfrac{p(i,j|\theta)}{N_p}$$

	e	of runs of a given length. The value is 1 if every run of length 1.	
GLRLM_LGLRE	Low Grey Level Run Emphasis	Measures the occurrence of long runs weighted towards low intensity	$$\frac{\sum_{i=1}^{Ng}\sum_{j=1}^{Nr}\left[\frac{p(i,j\mid\theta)}{i^2}\right]}{\sum_{i=1}^{Ng}\sum_{j=1}^{Nr}p(i,j\mid\theta)}$$
GLRLM_HGLRE	High Grey Level Run Emphasis	Measures the distribution of pixels weighted towards high intensity	$$\frac{\sum_{i=1}^{Ng}\sum_{j=1}^{Nr}\left[\frac{p(i,j\mid\theta)}{i^2j^2}\right]}{\sum_{i=1}^{Ng}\sum_{j=1}^{Nr}p(i,j\mid\theta)}$$
GLRLM_SRLGLE	Short Run Low Grey Level Emphasis	Measures the occurrence of short runs weighted towards low intensity	$$\frac{\sum_{i=1}^{Ng}\sum_{j=1}^{Nr}\left[\frac{p(i,j\mid\theta)}{i^2j^2}\right]}{\sum_{i=1}^{Ng}\sum_{j=1}^{Nr}p(i,j\mid\theta)}$$
GLRLM_SRHGLE	Short Run High Grey Level	Measures the occurrence	$$\frac{\sum_{i=1}^{Ng}\sum_{j=1}^{Nr}\left[\frac{p(i,j\mid\theta)i^2}{j^2}\right]}{\sum_{i=1}^{Ng}\sum_{j=1}^{Nr}p(i,j\mid\theta)}$$

	Emphasis	of short runs weighted towards high intensity	
GLRLM_LRLGLE	Long Run Low Grey Level Emphasis	Measures the occurrence of long runs weighted towards low intensity	$$\frac{\sum_{i=1}^{Ng}\sum_{j=1}^{Nr}\left[\frac{p(i,j|\theta)j^2}{i^2}\right]}{\sum_{i=1}^{Ng}\sum_{j=1}^{Nr}p(i,j|\theta)}$$
GLRLM_LRHGLE	Long Run High Grey Level Emphasis	Measures the occurrence of long runs weighted towards high intensity	$$\frac{\sum_{i=1}^{Ng}\sum_{j=1}^{Nr}p(i,j|\theta)i^2j^2}{\sum_{i=1}^{Ng}\sum_{j=1}^{Nr}p(i,j|\theta)}$$
FD_mean	Mean		
FD_sd	Standard deviation		
FD_var	Variance		
FD_lacunarity	Lacunarity		
FD_max	Maximum		
FD_min	Minimum		

GLCM_Varian	Variance	Measure the spread in pixel intensities	$$\sum_{i=1}^{Ng}\sum_{j=1}^{Ng}(i-\mu)^2 \boldsymbol{P}(i,j)$$
GLCM_Correl	Correlation	Measure of linear dependency between intensities of pairs of voxels (0: uncorrelated, 1: linearly correlated)	$$\frac{\sum_{i=1}^{Ng}\sum_{j=1}^{Ng}(ij)p(i,j)-\mu_x(i)\mu_y(j)}{\sigma_x(i)\sigma_y(j)}$$
GLCM_InfCo1		Measure of non-linear dependency between intensities of pairs of voxels "informational coefficient of correlation" (0: independent variables, 1: dependent variables)	$$\frac{HXY-HXY1}{\max\{HX,HY\}}$$ Where HX and HY are entropies of p_x and p_y and $$H = -\sum_{i=1}^{Ng}\sum_{j=1}^{Ng}\boldsymbol{P}(i,j)\log_2[\boldsymbol{P}(i,j)]\text{ be the}$$ entropy of $\boldsymbol{P}(i,j)$, $$HXY = -\sum_{i=1}^{Ng}\sum_{j=1}^{Ng}\boldsymbol{P}(i,j)\log_2[\boldsymbol{P}(i,j)],$$ $$HXY1 = -\sum_{i=1}^{Ng}\sum_{j=1}^{Ng}p_x(i)p_y(j)\log_2[p_x(i)p_y(j)]$$

GLCM_InfCo2			$\sqrt{1 - e^{-2(HXY2-HXY)}}$ Where $HXY = - \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} P(i,j) \log_2[P(i,j)]$ $HXY2 = - \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} P_x(i)P_y(j) \log_2[P_x(i)P_y(j)]$
GLCM_ClShad	Cluster Shade	Measure the skewness of the GLCM. Increases with asymmetry.	$\sum_{i=1}^{Ng} \sum_{j=1}^{Ng} [i + j - \mu_x(i) - \mu_y(j)]^3 P(i,j)$
GLCM_ClProm	Cluster Prominence		$\sum_{i=1}^{Ng} \sum_{j=1}^{Ng} [i + j - \mu_x(i) - \mu_y(j)]^4 P(i,j)$
GLCM_Angsmo	Angular Second Moment	Measure of texture uniformity (1/Ng2 0: heterogenous, 1: homogeneous)	$\sum_{i=1}^{Ng} \sum_{j=1}^{Ng} [P(i,j)]^2$
GLCM_MxProb	Maximum Probability	Reflects the highest number of occurrences of a pixel pair.	$max\{P(i,j)\}$

GLCM_Entrop	Entropy	Measure of information content (0 : predominant combination of pairs of voxels' intensities, log2(Ng) equal contribution of all pairs)	$$\sum_{i=1}^{Ng}\sum_{j=1}^{Ng} P(i,j)\log_2[P(i,j)]$$
GLCM_Contra	Contrast	Local intensity variations (0: no contrast, (Ng-1)2: maximum contrast)	$$\sum_{i=1}^{Ng}\sum_{j=1}^{Ng}|i-j|^2\, P(i,j)$$
GLCM_Dissim	Dissimilarity	Contrast weighted linearly with the difference of grey-level (0: similarity (diagonal GLCM), Ng-1	$$\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}|i-j|P(i,j)$$

		dissimilarity)	
GLCM_Homoge	Homogeneity	Opposite of contrast (0: dissimilar, 1: similar (diagonal GLCM))	$$\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{P(i,j)}{1+\lvert i-j\rvert^2}$$
GLCM_sumAvg	Sum Average	Average of the pixel intensities	$$\sum_{i=2}^{2N_g} i\boldsymbol{P_{x+y}}(i)$$
GLCM_sumVar	Sum Variance		$$\sum_{i=2}^{2N_g} (i-SE)^2\boldsymbol{P_{x+y}}(i)$$
GLCM_sumEnt	Sum Entropy		$$\sum_{i=2}^{2N_g} \boldsymbol{P_{x+y}}(i)\log_2\left[\boldsymbol{P_{x=y}}(i)\right]$$
GLCM_difVar	Difference in Variance		$$variance\ of\ P_{x-y}$$
GLCM_difEnt	Difference entropy		$$\sum_{i=0}^{N_g-1} P_{x-y}(i)\log_2\left[P_{x-y}(i,j)\right]$$
GLCM_AutoCorrel	Autocorrelation		$$\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} ijP(i,j)$$
GLCM_ClTend	Cluster Tendency	Measure variations in pixel intensities	$$\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\left[i+j-\mu_x(i)-\mu_y(j)\right]^2\boldsymbol{P}(i,j)$$

GLCM_Homoge1	Homogeneity 1	Measure the uniformity of the image	$$\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{P(i,j)}{1+|i-j|}$$
GLCM_IDMN	Inverse Difference Moment Normalised		$$\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{P(i,j)}{1+\left(\frac{|i-j|^2}{N^2}\right)}$$
GLCM_IDN	Inverse Difference Normalised		$$\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{P(i,j)}{1+\left(\frac{|i-j|}{N}\right)}$$
GLCM_invVar	Inverse Variance	Also called Inverse Difference Moment, similar to homogeneity	$$\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{P(i,j)}{|i-j|^2}\ ,i\ \neq j$$

Key

GLCM:

$P(i,j)$ be the co-occurrence matrix for an arbitrary δ and α,

N_g be the number of discrete intensity levels in the image,

μ be the mean of $P(i,j)$,

$p_x(i)= -\sum_{j=1}^{Ng}P(i,j)$ be the marginal row probabilities,

$p_y(i)= -\sum_{i=1}^{Ng}P(i,j)$ be the marginal column probabilities,

μ_x be the mean of p_x,

μ_y be the mean of p_x,

σ_x be the standard deviation of p_x,

σ_y be the standard deviation of p_y,

$p_{x+y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i,j), i + j = k, k = 2,3, ..., 2N_g,$

$p_{x-y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i,j), |i + j| = k, k = 0,1, ..., N_g\text{-}1.$

For run length:

$p(i,j|\theta)$ be the (i,j)th entry in the given run-length matrix p for a direction θ,

N_g the number of discrete intensity values in the image,

N_r the number of different run lengths,

N_p the number of voxels in the image.

NGTDM

$M(i)$ is the ith entry of the NGTDM which is the sum of the intensity differences between each voxel of intensity i belonging to VOI V and its 26 direct neighbours in 3 dimensions,

ε is a small number

N_g is the number of quantisation levels used for normalising the VOI intensities,

N_t is the number of different grey-levels present in the image,

p_i is the probability of occurrence of grey-level I,

N_i is the number of voxels of intensity in the calculation,

N is the total number of voxels used in the calculation.

GLSZM:

Where $P(i)$ is the (i,j)th entry of the GLSZM which records the number of clusters of voxels of intensity i and number of voxels j within the VOI.

Ng is the number of quantisation level used for normalising the VOI intensities

X_z is the number of zones in the image

N_z is the number of zones in the images

N is the number of voxels in the VOI.

Appendix C

PCA for the a) Manufacturer type, b)model, c) slice thickness, d) number of rows and e) columns for reach gray level.

PCA of imaging traits by Manufacturer type and 8gl

PCA of imaging traits by ManufacturerModelName type and 8gl

PCA of imaging traits by Columns type and 8gl

PCA of imaging traits by SliceThickness type and 8gl

PCA of imaging traits by Rows type and 8gl

8GL
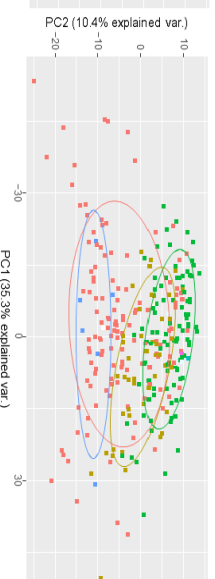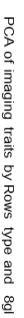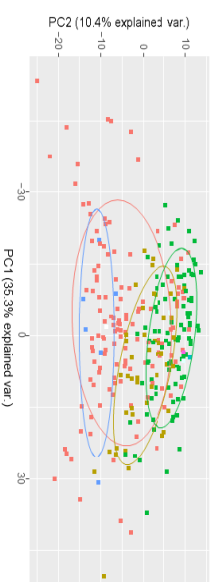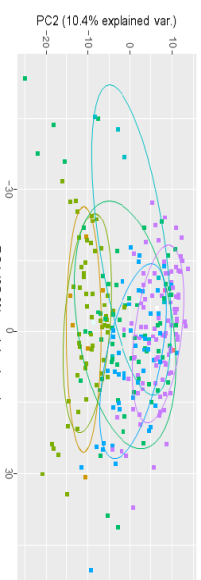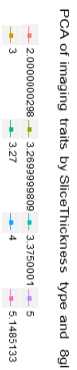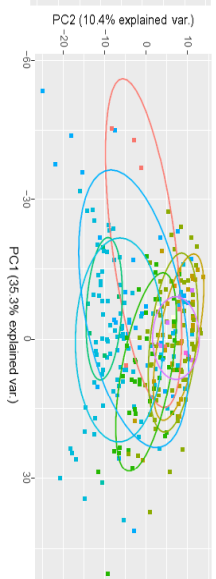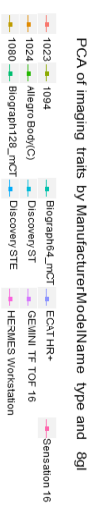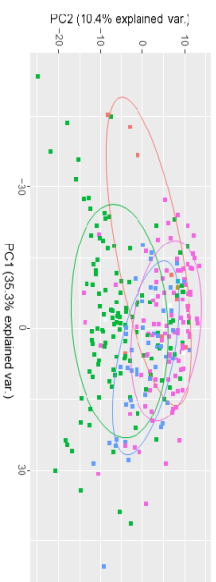
16GL

PCA of imaging traits by Manufacturer type and 16gl
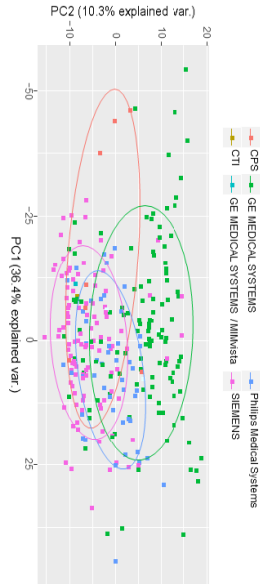
PCA of imaging traits by ManufacturerModelName type and 16gl
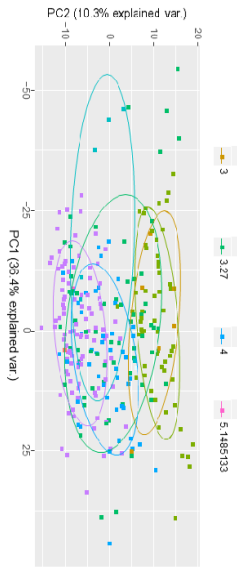
PCA of imaging traits by Rows type and 16gl

PCA of imaging traits by SliceThickness type and 16gl

PCA of imaging traits by Columns type and 16gl
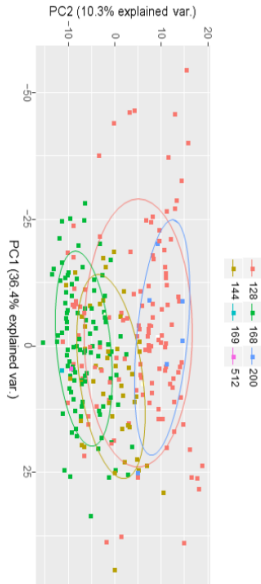
32GL

PCA of imaging traits by Manufacturer type and 32gl

PCA of imaging traits by ManufacturerModelName type and 32gl

PCA of imaging traits by Columns type and 32gl

PCA of imaging traits by Rows type and 32gl

PCA of imaging traits by SliceThickness type and 32gl

128GL

PCA of imaging traits by Manufacturer type and 128gl

PCA of imaging traits by Rows type and 128gl

PCA of imaging traits by SliceThickness type and 128gl

PCA of imaging traits by ManufacturerModelName type and 128gl

PCA of imaging traits by Columns type and 128gl

256GL

PCA of imaging traits by Manufacturer type and 256gl
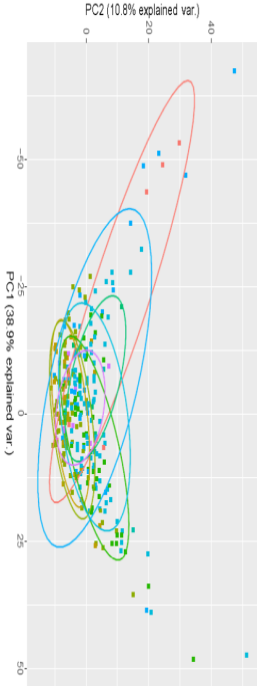
PCA of imaging traits by SliceThickness type and 256gl

PCA of imaging traits by ManufacturerModelName type and 256gl

PCA of imaging traits by Rows type and 256gl

PCA of imaging traits by Columns type and 256gl

Appendix D

The full breakdown of the results when the 5 papers' findings were tested in our dataset.

Paper	Subsection	Grey levels	No of features	Features	weights	Sensitivity	Specificity	PPV	NPV	Accuracy	Sig with test and model (extra) set (group and stage model)	Stage and group validation (with stage extra): stage sig	Stage and group model: : group sig hazard	Stage and group model: group hazard ratio	Stage and group model: hazard ratio
Cook	All	4 gl		Contrast	Median	0.445	0.442	0.49	0.455	0.443	0.0042	0.00664	1.38	0.906	0.978
					Kmroc	0.0373	0.458	0.438	0.433	0.416					
		8g			Median	0.444	0.441	0.49	0.454	0.442	0.0042	0.0037	1.39	0.895	1.02
					Kmroc	0.307	0.451	0.393	0.405	0.381					

16 gl		Median	0.444	0.442	0.49	0.454	0.443	0.0042 2	0.004	1.39	0.945	1.012
		Kmroc	0.244	0.549	0.298	0.432	0.4					
32 gl		Median	0.435	0.432	0.49	0.444	0.433	0.0037 3	0.0083	1.35	0.615	0.91
		Kmroc	0.235	0.54	0.294	0.425	0.391					
64 gl		Median	0.445	0.442	0.49	0.455	0.443	0.0036 3	0.0067	1.35	0.58	0.908
		Kmroc	0.256	0.534	0.314	0.429	0.398					
12 8gl		Median	0.444	0.442	0.49	0.454	0.443	0.0037 7	0.00766	1.35	0.631	0.918
		Kmroc	0.209	0.54	0.27	0.417	0.379					
25 6gl		Median	0.436	0.433	0.49	0.445	0.434	0.0035 6	0.00841	1.35	0.558	0.902
		Kmroc	0.256	0.498	0.329	0.412	0.379					
4gl	Coa	Median	0.427	0.424	0.49	0.436	0.425	0.0023	0.00999	1.327	0.286	0.832

rsen ess							8				
	Kmroc	0.275	0.49	0.348	0.415	0.385					
8gl	Median	0.428	0.424	0.49	0.437	0.426	0.00229	0.00719	1.335	0.269	0.83
	Kmroc	0.47	0.296	0.602	0.369	0.381					
16 gl	Median	0.419	0.415	0.491	0.428	0.417	0.00215	0.00748	1.33	0.245	0.823
	Kmroc	0.244	0.54	0.301	0.428	0.395					
32 gl	Median	0.41	0.405	0.491	0.418	0.408	0.00178	0.00984	1.32	0.19	0.8
	Kmroc	0.454	0.316	0.578	0.377	0.384					
64 gl	Median	0.406	0.402	0.491	0.415	0.404	0.0012	0.0111	1.314	0.114	0.766
	Kmroc	0.446	0.326	0.567	0.381	0.384					
12	Median	0.4	0.396	0.491	0.409	0.398	0.0014	0.008	1.326	0.147	0.785

Group		Measure										
8gl		Kmroc	0.442	0.285	0.597	0.348	0.362	6			0.168	0.801
25 6gl		Median	0.405	0.4	0.491	0.413	0.402					
		Kmroc	0.475	0.319	0.587	0.389	0.395	0.0016 4	0.00324	1.35	0.303	1.19
4gl	Bus yne ss	Median	0.576	0.58	0.487	0.589	0.578					
		Kmroc	0.499	0.717	0.399	0.6	0.61	0.0024 8	0.01	1.38	0.607	1.09
8gl		Median	0.567	0.58	0.483	0.584	0.574					
		Kmroc	0.516	0.707	0.411	0.605	0.614	0.0037 1	0.00505	1.36	0.607	1.09
16 g		Median	0.559	0.561	0.487	0.571	0.56	0.0037 1	0.0039	1.36	0.608	1.09
		Kmroc	0.525	0.68	0.424	0.6	0.604					
32		Median	0.567	0.571	0.487	0.58	0.569	0.0029	0.006	1.34	0.39	1.15

Appendix D

	Contrast								1				
gl		Kmroc	0.542	0.671	0.436	0.606	0.608						
		Median	0.576	0.58	0.487	0.589	0.578	0.0017 8	0.00813	1.33	0.19	1.246
64 gl		Kmroc	0.482	0.734	0.385	0.597	0.611						
		Median	0.566	0.569	0.487	0.579	0.568	0.0025 8	0.0031	1.36	0.321	1.17
12 8gl		Kmroc	0.611	0.57	0.505	0.605	0.59						
		Median	0.509	0.51	0.488	0.521	0.509						
25 6gl		Kmroc	0.269	0.656	0.281	0.484	0.467						
	Con tras t	Median	0.464	0.482	0.486	0.478	0.473	0.105	0.045	1.39	0.6	1.15
SC C 4gl		Kmroc	0.235	0.573	0.287	0.433	0.406						
8gl		Median	0.482	0.482	0.495	0.487	0.482	0.11	0.0467	1.38	0.675	1.11

Appendix D

16 gl	Kmroc	0.271	0.476	0.358	0.4	0.375					
	Median	0.463	0.482	0.486	0.478	0.473	0.114	0.053	1.37	0.74	1.09
32 gl	Kmroc	0.196	0.552	0.28	0.412	0.375					
	Median	0.464	0.463	0.496	0.468	0.463	0.117	0.0884	1.32	0.813	0.941
64 gl	Kmroc	0.233	0.513	0.308	0.405	0.374					
	Median	0.425	0.443	0.485	0.44	0.434	0.0823	0.141	1.265	0.386	0.801
12 8gl	Kmroc	0.136	0.605	0.181	0.416	0.373					
	Median	0.424	0.423	0.496	0.428	0.424	0.105	0.122	1.29	0.602	0.872
25 6gl	Median	0.425	0.423	0.496	0.429	0.424	0.1	0.128	1.28	0.545	0.855
	Kmroc	0.252	0.514	0.325	0.412	0.384					

Coarseness											
	Median	0.406	0.405	0.496	0.41	0.405					
4gl	Kmroc	0.252	0.475	0.342	0.393	0.364	0.095	0.0967	1.29	0.492	0.845
8gl	Median	0.406	0.405	0.496	0.41	0.405	0.106	0.099	1.3	0.62	0.882
	Kmroc	0.425	0.29	0.59	0.34	0.357					
16 gl	Median	0.387	0.385	0.497	0.39	0.386	0.0682	0.132	1.26	0.289	0.768
	Kmroc	0.386	0.327	0.537	0.352	0.356					
32 gl	Median	0.388	0.385	0.497	0.391	0.386	0.0929	0.109	1.28	0.473	0.836
	Kmroc	0.391	0.274	0.583	0.315	0.332					
64 gl	Median	0.354	0.351	0.497	0.356	0.352	0.0637	0.141	1.26	0.262	0.755
	Kmroc	0.353	0.312	0.526	0.329	0.332					
12	Median	0.355	0.352	0.497	0.357	0.353	0.0576	0.152	1.25	0.227	0.734

8gl		Kmroc	0.354	0.294	0.542	0.316	0.323	
25 6gl		Median	0.372	0.369	0.497	0.374	0.37	
		Kmroc	0.352	0.369	0.484	0.367	0.36	
Ad en o 4gl	Con tras t	Median	0.396	0.398	0.473	0.422	0.397	
		Kmroc	0.393	0.342	0.508	0.385	0.367	
8gl		Median	0.37	0.368	0.475	0.394	0.369	
		Kmroc	0.318	0.415	0.408	0.403	0.369	
16 gl		Median	0.396	0.398	0.473	0.422	0.397	
		Kmroc	0.237	0.482	0.307	0.412	0.366	
32 gl		Median	0.368	0.366	0.475	0.391	0.367	

		Kmroc/Median									
	Kmroc	0.236	0.454	0.319	0.398	0.351					
64 gl	Median	0.393	0.395	0.473	0.419	0.394					
	Kmroc	0.188	0.401	0.297	0.354	0.3					
12 8gl	Median	0.366	0.364	0.475	0.389	0.365					
	Kmroc	0.19	0.429	0.285	0.371	0.316					
25 6gl	Median	0.368	0.367	0.475	0.392	0.368	0.0668	0.0666	1.48	0.76	0.91
	Kmroc	0.245	0.386	0.363	0.362	0.319					
Coa rsen ess 4gl	Median	0.403	0.405	0.472	0.43	0.404	0.0538	0.07	1.45	0.47	0.807
4gl	Kmroc	0.24	0.432	0.333	0.387	0.341					
8gl	Median	0.394	0.396	0.473	0.42	0.395	0.0511	0.07	1.44	0.43	0.79
	Kmroc	0.24	0.432	0.333	0.387	0.341					

16 gl	Median	0.394	0.396	0.473	0.42	0.395	0.049	0.07	1.44	0.399	0.78
	Kmroc	0.29	0.384	0.405	0.375	0.34					
32 gl	Median	0.368	0.366	0.475	0.391	0.367	0.03	0.105	1.39	0.196	0.683
	Kmroc	0.29	0.384	0.405	0.375	0.34					
64 gl	Median	0.368	0.366	0.475	0.391	0.367	0.03	0.105	1.39	0.196	0.683
	Kmroc	0.24	0.407	0.347	0.373	0.328					
12 8gl	Median	0.412	0.415	0.472	0.439	0.413	0.0406	0.076	1.43	299	0.741
	Kmroc	0.263	0.458	0.341	0.408	0.365					
25 6gl	Median	0.463	0.472	0.469	0.494	0.468					
	Kmroc	0.209	0.58	0.245	0.449	0.404					
NS 4gl	Median	0.451	0.476		0.53	0.465					
Con							0.0644	0.0494	2.08	0.567	1.42

CLC	trast										
8gl	Kmroc	0.192	0.456	0.244	0.423	0.341					
	Median	0.451	0.476	0.421	0.53	0.465	0.634	0.0451	2.08	0.549	1.426
16 gl	Kmroc	0.192	0.456	0.244	0.423	0.341					
	Median	0.455	0.481	0.421	0.534	0.469	0.0741	0.0799	1.87	0.859	1.109
32 gl	Kmroc	0.192	0.456	0.244	0.423	0.341					
	Median	0.455	0.481	0.421	0.534	0.469	0.0741	0.799	1.87	0.859	1.109
64 gl	Kmroc	0.122	0.603	0.134	0.472	0.394					
	Median	0.455	0.481	0.421	0.534	0.469	0.0741	0.0799	1.87	0.859	1.11
12 8gl	Kmroc	0.061	0.603	0.072	0.455	0.368					
	Median	0.455	0.481	0.421	0.534	0.469	0.0741	0.0799	1.87	0.859	1.11

	Kmroc	0	0.762	0	0.498	0.431					
25 6gl	Median	0.455	0.481	0.421	0.534	0.469					
	Kmroc	0	0.762	0	0.498	0.431	0.0741	0.0799	1.87	0.859	1.11
4gl Coa rsen ess	Median	0.455	0.481	0.421	0.534	0.469					
	Kmroc	0.511	0.316	0.554	0.457	0.401					
	Median	0.455	0.481	0.421	0.534	0.469					
8gl	Kmroc	0.452	0.398	0.466	0.486	0.422					
	Median	0.455	0.481	0.421	0.534	0.469					
16 gl	Median	0.455	0.481	0.421	0.534	0.469					
	Kmroc	0.753	0.0756	0.884	0.285	0.37					
32 gl	Median	0.455	0.481	0.421	0.534	0.469					
	Kmroc	0.753	0.0756	0.884	0.285	0.37					

Group		Stat										
64 gl		Median	0.455	0.481	0.421	0.534	0.469					
16 onU SzN		Kmroc	0.452	0.398	0.466	0.486	0.422	0.469				
12 8gl		Median	0.455	0.481	0.421	0.534						
		Kmroc	0.407	0.815	0.277	0.641	0.637					
25 6gl		Median	0.518	0.563	0.414	0.604	0.544					
		Kmroc	0.366	1	0.219	0.672	0.724					
Cheng 2015 16 gl nif		Median	0.594	0.598	0.498	0.596	0.596	0.000163	0.011	1.28	0.0084	1.48
		Kmroc	0.495	0.775	0.389	0.606	0.635					
32 onU nif SzN gl		Median	0.592	0.596	0.498	0.594	0.594	0.00104	0.0122	1.28	0.07	1.31

Study	Bin	Feature	Stat										
			Kmroc	0.456	0.772	0.371	0.587	0.614					
	64 gl	SzN onU nif	Median	0.607	0.611	0.498	0.609	0.609	0.0001 15	0.025	1.25	0.005 8	1.15
			Kmroc	0.518	0.74	0.412	0.606	0.629	0.0016 9	0.0098	1.29	0.128	1.26
	16 gl	GINONUnif	Median	0.588	0.592	0.498	0.59	0.59					
			Kmroc	0.411	0.837	0.329	0.587	0.624					
	32 gl	GLSZM	Median	0.593	0.596	0.498	0.594	0.594	0.0004 21	0.0241	1.256	0.024 5	1.41
			Kmroc	0.461	0.799	0.366	0.597	0.63	0.0031				
Cheng 2013	4gl	GLCM Angsmo	Median	0.573	0.576	0.498	0.575	0.575	0.0031 7	0.02	1.283	0.303	1.183
			Kmroc	0.523	0.643	0.448	0.574	0.583	0.0031 7	0.02	1.283	0.303	1.183
Hatt 2015	64	GLCM	Median	0.57	0.574	0.498	0.572	0.572	0.0017	0.011	1.29	0.13	1.26

Group	Feature	Metric										
gl	Entrop	Kmroc	0.543	0.692	0.439	0.602	0.617	2				
64 gl	GLCM Entrop	Median	0.44	0.444	0.498	0.443	0.442	0.0031 8	0.0105	1.299	0.303	0.852
		Kmroc	0.226	0.579	0.281	0.428	0.402					
64 gl Gl	Lzo neHi	Median	0.518	0.522	0.498	0.52	0.52					
		Kmroc	0.415	0.71	0.369	0.548	0.563	0.0026 5	0.01	1.296	0.233	0.833
64 gl	GLSZM ZonePcent	Median	0.438	0.441	0.498	0.44	0.44					
		Kmroc	0.221	0.574	0.278	0.425	0.398					
4gl 2	SNS com	Median	0.56	0.658	0.459	0.599	0.609	0.0002 13	0.018	1.27	0.011	1.46
Aerts 2015												

FOS Ener	Kmroc	0.641	0.742	0.407	0.722	0.697					
GLRLM GLN											
GLRLM GLN HLH											
8gl	Median	0.45	0.761	0.372	0.581	0.606	8.31x10-5	0.033	1.24	0.0035	1.57
	Kmroc	0.513	0.871	0.319	0.692	0.712					
16 gl	Median	0.437	0.768	0.362	0.577	0.602	0.000179	0.0285	1.25	0.0082	1.51
	Kmroc	0.436	0.871	0.285	0.66	0.678					
32 gl	Median	0.425	0.786	0.351	0.578	0.606	0.000148	0.0342	1.24	0.0066	1.54
	Kmroc	0.41	0.871	0.272	0.65	0.667	0.0001				
64	Median	0.425	0.786	0.351	0.578	0.606	0.0002	0.0228	1.26	0.011	1.49

gl	Kmroc	0.41	0.871	0.272	0.62	0.667	32			0.018 2	1.44
12 8gl	Median	0.437	0.768	0.362	0.577	0.602	0.0003 57	0.0147	1.27		
	Kmroc	0.487	0.839	0.316	0.673	0.683					
25 6gl	Median	0.444	0.775	0.364	0.583	0.61	0.0001 83	0.0164	1.27	0.008 27	
	Kmroc	0.487	0.839	0.316	0.673	0.683					1.5

Appendix E1: Ovarian CRF

Pt code name:	Radiologist Initials:

CT findings

Mediastinal nodes	No/ Equivocal/Yes/Not Available
Pre cardiac nodes	No /Equivocal/Yes
Solid Pleural disease	No/Yes
Pleural effusion	No/Yes
Diaphragms	No/Diffuse/ Focal Nodule (s) /Massive
Liver capsule right	No/Diffuse/ Focal Nodule (s) /Massive
Liver capsule left	No/Diffuse/ Focal Nodule (s) /Massive
Liver parenchyma	No/Equivocal/Yes
Liver hilum	No/Moderate/ Large
Falciform ligament	No/Yes
Morrison's pouch	No/Focal/Moderate/Large
Spleen (subcapsule)	No/Diffuse/ Focal Nodule (s) /Massive
Spleen (parenchyma)	No/Equivocal/Yes
Gastrosplenic ligament	No/Diffuse/ Focal Nodule (s) /Massive
Pancreas	No/Equivocal/Yes

Lesser sac   No/Equivocal/Yes

Stomach wall   No/ Abutment/Diffuse Serosal/

Invasion/Massive

Supracolic omentum   No/Diffuse/ Focal Nodule (s) /Massive

Right colon/caecum   No/ Abutment/Diffuse Serosal/

Invasion/Massive

Appendix   Not Seen/No/Mass

Rt paracolic peritoneum   No/Equivocal/Focal/Diffuse/Massive

Trans colon   No/ Abutment/Diffuse Serosal/

Invasion/Massive

Left colon   No/ Abutment/Diffuse Serosal/

Invasion/Massive

Lt paracolic peritoneum   No/Equivocal/Focal/Diffuse/Massive

Sigmoid   No/ Abutment/Diffuse Serosal/

Invasion/Massive

Pelvic peritoneum   No/Equivocal/Focal/Diffuse/Massive

Rectum   No/ Abutment/Diffuse Serosal/

Invasion/Massive

Small bowel   No/ Abutment/Diffuse Serosal

Invasion/Massive

Small bowel mesentery   No/Equivocal/Nodes/Focal Nodule/Mass

Invading Root /Diffuse

Greater Omentum	No/Equivocal/ Focal Nodule (S)/Cake
Anterior abdominal/umbilical	No/Equivocal/Focal/Large
Paraaortic nodes	Not Enlarged/Borderline/Enlarged
Pelvic nodes	Not Enlarged/Borderline/Enlarged
Groin nodes	Not Enlarged/Borderline/Enlarged
Uterus	Not Involved/Equivocal/Involved
Fallopian tubes	Not Involved/Equivocal/Involved

Right ovary   1. No lesion/not seen/benign lesion/ highly likely benign/not sure/highly likely malignant

2. Normal/ not breaching capsule/breaching capsule

3. Outline: Smooth/Irregular

4. Texture: cystic/solid/mixed predominantly cystic/mixed predominantly solid

5. Enhancement pattern: homogenous/heterogeneous

6. Presence: thick septations/papillary projections/calcifications

Left ovary   1. No lesion/not seen/benign lesion/ highly likely benign/not sure/highly likely malignant

2. Normal/ not breaching capsule/breaching capsule

3. Outline: Smooth/Irregular

4. Texture: cystic/solid/mixed predominantly cystic/mixed

predominantly solid

5. Enhancement pattern: homogenous/heterogeneous

6. Presence: thick septations/papillary projections/calcifications

Ascites   No or physiological/previously drained/less than 1 litre/more

than 1 litre

Appendix E2: Peritoneal Cancer Index

Taken from (Gilly et al., 2006)

Peritoneal Cancer Index
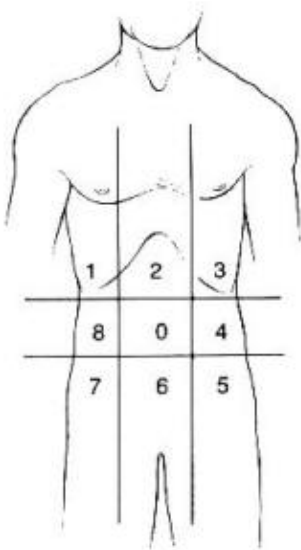
Regions	Lesion Size
0 Central	____
1 Right Upper	____
2 Epigastrium	____
3 Left Upper	____
4 Left Flank	____
5 Left Lower	____
6 Pelvis	____
7 Right Lower	____
8 Right Flank	____
9 Upper Jejunum	____
10 Lower Jejunum	____
11 Upper Ileum	____
12 Lower Ileum	____

Lesion Size Score

LS 0 No tumor seen
LS 1 Tumor up to 0.5 cm
LS 2 Tumor up to 5.0 cm
LS 3 Tumor > 5.0 cm
        or confluence

PCI
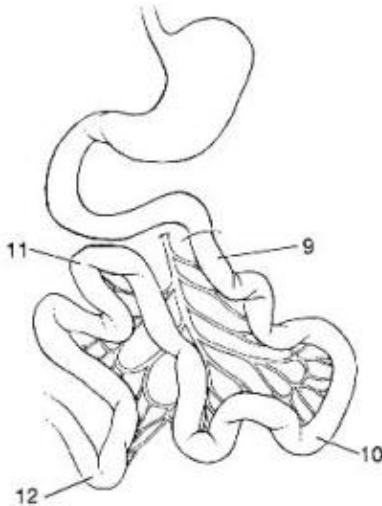
Radiopharmaceuticals as probes to characterize tumour tissue.

Alam IS, Arshad MA, Nguyen QD, Aboagye EO.

ORIGINAL ARTICLE

CrossMark

Discovery of pre-therapy 2-deoxy-2-^{18}F-fluoro-D-glucose positron emission tomography-based radiomics classifiers of survival outcome in non-small-cell lung cancer patients

Mubarik A. Arshad [1,2,3] · Andrew Thornton [1] · Haonan Lu [1] · Henry Tam [2,3] · Kathryn Wallitt [2,3] · Nicola Rodgers [1] · Andrew Scarsbrook [4,5] · Garry McDermott [4] · Gary J. Cook [6] · David Landau [6] · Sue Chua [7] · Richard O'Connor [8] · Jeanette Dickson [9] · Danielle A. Power [2,3] · Tara D. Barwick [1,2,3] · Andrea Rockall [1,2,3] · Eric O. Aboagye [1]

Abstract

Purpose The aim of this multi-center study was to discover and validate radiomics classifiers as image-derived biomarkers for risk stratification of non-small-cell lung cancer (NSCLC).

Patients and methods Pre-therapy PET scans from a total of 358 Stage I–III NSCLC patients scheduled for radiotherapy/chemoradiotherapy acquired between October 2008 and December 2013 were included in this seven-institution study. A semi-automatic threshold method was used to segment the primary tumors. Radiomics predictive classifiers were derived from a training set of 133 scans using TexLAB v2. Least absolute shrinkage and selection operator (LASSO) regression analysis was used for data dimension reduction and radiomics feature vector (FV) discovery. Multivariable analysis was performed to establish the relationship between FV, stage and overall survival (OS). Performance of the optimal FV was tested in an independent validation set of 204 patients, and a further independent set of 21 (TESTI) patients.

Results Of 358 patients, 249 died within the follow-up period [median 22 (range 0–85) months]. From each primary tumor, 665 three-dimensional radiomics features from each of seven gray levels were extracted. The most predictive feature vector discovered (FVX) was independent of known prognostic factors, such as stage and tumor volume, and of interest to multi-center studies, invariant to the type of PET/CT manufacturer. Using the median cut-off, FVX predicted a 14-month survival difference in the validation cohort ($N = 204$, $p = 0.00465$; HR = 1.61, 95% CI 1.16–2.24). In the TESTI cohort, a smaller cohort that presented with unusually poor survival of stage I cancers, FVX correctly indicated a lack of survival difference ($N = 21$, $p = 0.501$). In contrast to the radiomics classifier, clinically routine PET variables including SUV_{max}, SUV_{mean} and SUV_{peak} lacked any prognostic information.

✉ Eric O. Aboagye
eric.aboagye@imperial.ac.uk

1 Imperial College London Cancer Imaging Centre, Department of Surgery & Cancer, Hammersmith Hospital, Du Cane Road, London W12 0NN, UK

2 Imperial College Healthcare NHS Trust, Departments of Clinical Oncology, Radiology and Nuclear Medicine, Hammersmith Hospital, Du Cane Road, London W12 0HS, UK

3 Charing Cross Hospital, Fulham Palace Road, London W6 8RF, UK

4 Department of Nuclear Medicine, Level 1, Bexley Wing, St James's University Hospital, Beckett Street, Leeds LS9 7TF, UK

5 Leeds Institute of Cancer and Pathology, School of Medicine, University of Leeds, Leeds, UK

6 Department of Cancer Imaging, School of Biomedical Engineering and Imaging Sciences, King's College London, St. Thomas' Hospital, Westminster Bridge Rd, London SE1 7EH, UK

7 Department of Nuclear Medicine, The Royal Marsden Hospital, Downs Rd, Sutton, London SM2 5PT, UK

8 Department of Nuclear Medicine, Queen's Medical Centre, Nottingham University Hospital, Derby Rd, Nottingham NG7 2UH, UK

9 Department of Clinical Oncology, Mount Vernon Hospital, Rickmansworth Road, Northwood HA6 2RN, UK

Springer

Conclusion PET-based radiomics classifiers derived from routine pre-treatment imaging possess intrinsic prognostic information for risk stratification of NSCLC patients to radiotherapy/chemo-radiotherapy.

Keywords Radiomics · NSCLC · Survival · PET · Risk stratification

Introduction

Lung malignancy is a leading cause of cancer-related death, with a predicted 5-year survival rate of 8–13% [1]. Worldwide, approximately 1.8 million new cases were diagnosed in 2012. Distinct from histology, stage, and performance status, the ability to provide prognosis on the basis of tumor biology is often lacking in current clinical practice. More recently, DNA sequencing from several tumor regions has been undertaken to highlight spatio-temporal mutational heterogeneity [2, 3]. Currently, imaging identifies the sites of disease and response to treatment by assessing the change in size but other than TNM staging provides limited prognostic information. In addition, outcomes of patients within each TNM staging group can vary widely highlighting the need for more accurate prognostic markers. Potential interventional methods to assess genetic heterogeneity will probably employ multi-core invasive biopsy, which limits its safe use for routine prognosis determination. The micro- and macro-structure of tumors, however, also harbor heterogeneous phenotypes due to factors such as hypoxia, necrosis, directional/non-directional tumor cell growth, vascular density, and immune infiltration. It is hypothesized that the asymmetric local, regional, and global density and architectural distortions of tumor phenotypes could have prognostic value, and this has resulted in a new 'omics' — radiomics [4–6] — whereby quantitative features describing tumor phenotypes are extracted in high-throughput from routine radiologic images and further processed by machine learning methods for prognostication; such high-dimensional output of tumor phenotypic heterogeneity is thought to have important prognostic value, with drug resistance and potential for development of metastatic spread implied.

2-deoxy-2-^{18}Fluorine-fluoro-D-glucose positron emission tomography-computed tomography (FDG-PET/CT) is routinely used for staging lung cancer prior to consideration of radical treatment such as surgery or chemo-radiotherapy including the use of stereotactic body radiotherapy. Indeed, radiomics classifiers based on the CT component have been investigated for predicting lung cancer histology and shown to have moderate prediction accuracy [7]. Beyond the use of FDG-PET/CT for staging, we investigated in the present study whether pre-therapy radiomics features derived from routine FDG-PET/CT examinations of non-small-cell lung cancer (NSCLC) patients who were subsequently treated with radiotherapy/chemo-radiotherapy across multiple hospitals might harbor useful prognostic information.

Patients and methods

Patients and procedures

The inclusion criteria were consecutive patients with non-small-cell lung cancer (NSCLC), or entire available cohort for The Cancer Imaging Archive (TCIA) (http://cancerimagingarchive.net/, last accessed June 2015), with a target lesions ≥ 5 ml who had a pre-therapy FDG-PET/CT scan available and underwent radical radiotherapy with or without chemotherapy between October 2008 and December 2013. The minimum lesion volume of interest (VOI) of 5 ml was selected, in accordance with work carried out by Soussan et al. [8]. Exclusion criteria were patients undergoing surgery or palliative treatment. Institutional ethical approval for retrospective analysis was obtained, and informed consent was waived.

The following hospitals took part in the trial (Fig. 1): Imperial College Healthcare NHS Trust, London, St James's University Hospital, Leeds, Guy's and St. Thomas' Hospitals, London, The Royal Marsden Hospital, Sutton, Nottingham University Hospital, Nottingham, and Mount Vernon Hospital, Northwood; a dataset was also obtained from TCIA. This work was carried out sequentially with training and validation followed by TESTI. Data from the four hospitals and The Cancer Imaging Archive (Imperial, Kings, Leeds, and Royal Marsden, and TCIA) patients were collated and randomly split into two (by computer) as training set ($n = 133$) and validation set ($n = 134$). A power calculation based on the training set (HR = 1.78, median survival: 2.92 years, censoring rate: 0.012, median follow-up: 2.17 years) suggested a sample size of 203 was needed to obtain the alpha of 0.05 and beta of 0.25. Therefore, all 70 cases from another centre were added to the validation set to make a total of 204. This validation set was only used for testing the findings from the training set. We used the maximum number of patients in the TCIA database that were available at the time. The original number of patients screened and basis for exclusion are indicated in Supplementary Table 1.

Pre-therapy clinico-pathologic data were obtained from medical records (Table 1). Overall survival was defined as number of months from commencement of treatment to date of death. Patients who were alive were censored at last follow-up to 31st July 2016. The hospital records were used to determine who was still alive at the time of cut-off. This was a multi-institutional analysis and so patients were examined on

Fig. 1 Overview of centers and PET images. **a** Overview of the centers contributing to the study and how the data were randomly divided into training, validation or independent test set. TCIA, The Cancer Imaging Archive. **b** Typical images from the PET/CT scans of two patients including PET, CT, and fusion images. Patient 1 (age 74, squamous cell carcinoma, stage IIA, tumour volume 22.6, overall survival 8 months) with the lower stage and smaller volume primary lesion had a worse survival outcome than patient 2 (age 77, squamous cell carcinoma, stage IIIA, tumour volume 26.5, overall survival 33 months) with the higher tumour stage
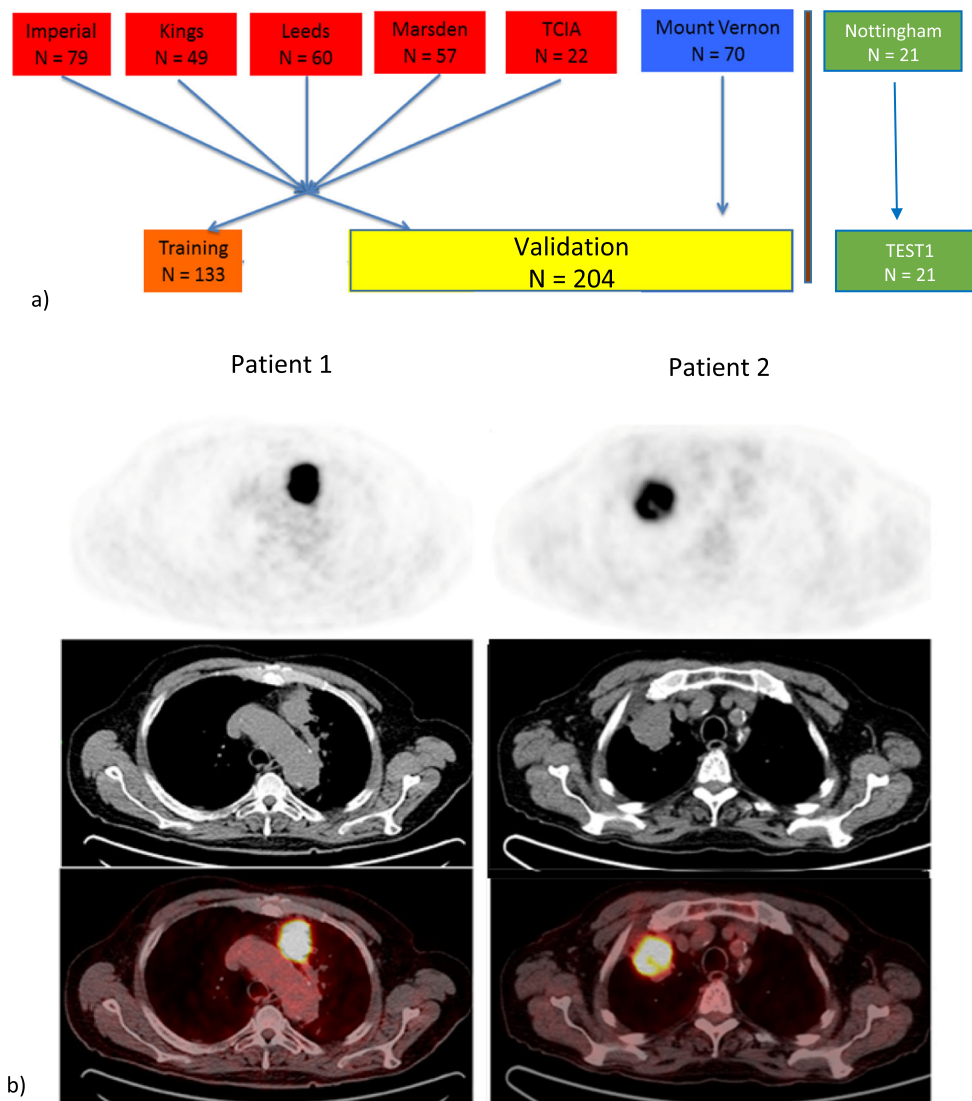
different PET/CT scanners including Phillips Allegro Body, Phillips Gemini TF TOF 16 (Phillips Medical Systems, Amsterdam, Netherlands), Siemens Biograph 64 mCT, Siemens Biograph 128 mCT (Siemens Healthcare, Erlangen, Germany), GE Healthcare Discovery ST, GE Discovery STE (GE Healthcare, Waukesha, Wisconsin, USA), CTI ECAT HR+ (CTI PET Systems Inc., Knoxville, Tennessee, USA), and CPS/Siemens Sensation 16. For PET, slice thickness ranged between 2 and 5.15 mm; the matrix size ranged between 128^2 and 512^2. After injection of 350–500 MBq ^{18}F-FDG [9], emission data were acquired (five or six bed positions, 2–4 min per bed position) after a 60–90 min uptake period. In all cases, PET/CT scans were performed from upper thighs to the base of the skull following $\geq$ 4–6-h fast, and had a measured blood glucose level < 11.0 mmol/l at the time of injection. CT was acquired without oral or intravenous contrast agent. The PET data were reconstructed using OSEM

iterative reconstruction and were attenuation-corrected using the CT data.

PET analysis

Central analyses of all PET/CT data were conducted at Imperial College London by a semi-automated adaptive threshold method. The primary tumor was initially delineated using an initial threshold of 40% of the SUV_{max} on semi-automated software (Hermes Gold3; Hermes Medical Solutions Ltd., London, UK) and VOIs drawn. The PET volume was correlated with the primary tumor on CT, and underestimation was determined by checking if the PET tumour VOI encompassed the whole tumour on the CT component of the PET. If the VOI did not cover the tumour visually, a lower threshold was used [10]. Manual adjustment was employed when the VOI incorporated adjacent normal

Table 1 Characteristics of the training, validation and test datasets

	Training set	Validation set	Test set I
Number	133	204	21
Mean age (range) years	69 (35–89)	71 (42–91)	71 (53–101)
Male (%)	82 (61.7)	126 (61.7)	10 (47.6)
Stage I (%)	24 (18)	33 (16.2)	4 (19)
Stage II (%)	34 (25.6)	37 (18.1)	4 (19)
Stage III (%)	75 (56.4)	134 (65.7)	13 (61.9)
Histology: SCC (%)	69 (51.9)	95 (46.7)	14 (66.7)
Histology: adeno(%)	41 (30.8)	77 (37.7)	5 (23.8)
Histology: NSCLC *NOS* (%)	18 (13.5)	25 (12.3)	2 (9.5)
Histology: other (%)	5 (3.8)	7 (3.4)	0
SUVmean (range)	8.25 (1.78–17.4)	8.44 (2.11–23.7)	7.75 (4.44–16.8)
SUVmax (range)	16.5 (4.9–42.8)	15.9 (3.26–49.5)	13.6 (6.66–39.2)
SUVpeak (range)	14.2 (3.8–35.4)	14.2 (2.9–43.1)	12.5 (6.26–34)
MTV (range) mls	40.4 (5.13–467)	33.7 (5.27–525)	30.8 (7.03–230)
TLG (range)	344 (16.2–5.45×10^3)	315.2 (19.4–5.7×10^3)	266 (40.5–2.59×10^3)
Median overall survival (months)	25 (0–83)	21.0 (0–85)	20 (2–37)
Number of deaths (%)	88 (66.2)	145 (71.1)	17 (81%)
Length of follow-up (median + IQR in months)	26 (12–39)	22.0 (11–36)	21 (8–31)

SCC squamous cell carcinoma, *Adeno* adenocarcinoma, *NSCLC* non-small-cell lung cancer (not otherwise specified, i.e., not classified into squamous or adenocarcinoma), *MTV* metabolic tumour volume, *TLG* total lesion glycolysis, *IQR* interquartile range. Stage AJCC/UICC 7

structures such as adjacent myocardial activity [11]. All segmentations were made by the same operator (observer 1, a radionuclide specialist radiologist with 4 years' experience of tumor delineation).

The SUV$_{mean}$, SUV$_{max}$, SUV$_{peak}$, metabolic tumor volume (MTV), and total lesion glycolysis (SUV$_{mean}$ × MTV)(TLG) of the primary tumor were recorded. Using Youdens's J to find the optimal cut-off from the ROC for median survival, Kaplan–Meier curves were generated. The VOIs were extracted and imported into the radiomics software. To assess intra- and inter-observer variability of the segmentation method, 18 patients were selected at random by SPSS, and segmentation of the tumor was performed (at 128 Gy level) by two additional experienced operators (observers 2 and 3, with 6 and 10 years' experience of tumor delineation respectively) blinded to the original results and clinical data. Lymph nodes were excluded from statistical analyses.

The interclass coefficient was used to assess intra- (by observer 1) and inter-observer (by observers 1, 2, and 3) differences in texture. The differences between the observers were performed by a 2-way ANOVA repeated measures model using Bonferroni correction.

Radiomics analysis

Radiomics analysis (Supplementary Fig. 1) was performed at seven different quantisation/gray levels — 4, 8, 16, 32, 64,128

and 256 Gy — on TexLAB v2, which was developed and implemented in-house within Matlab R2015b (MathWorks Inc., Natick, MA, USA). From each primary tumor, 665 radiomic features (listed in Supplementary Table 3) were extracted from segmented VOIs using local, regional, global, fractal, and wavelet techniques. These included intensity features, shape features, and texture features [gray level co-occurrence matrix (GLCM), gray level run length matrix (GLRLM) and neighbourhood gray difference matrix (NGTDM)] with or without wavelet transformation, as previously reported [5, 6]. Radiomics features were determined from 133 PET scans (training set) using TexLAB *v2*.

Feature selection and radiomics signature discovery

As with other high-throughput analyses, it is important to reduce the total number of features for prediction purposes in order to eliminate Type 1 errors and instead learn the true basis of a decision. We initially identified highly correlated features for elimination using heatmaps, as highly correlated features suggested that some feature reduction could be undertaken without information being lost. Heatmaps were created using R software (http://www.r-project.org/; Version 3.03 Vienna, Austria). It is known that there is correlation of several texture features with volume [12]. Using Spearman's rank correlation, features that had a high correlation with volume (≥ 0.7) were normalised by dividing the feature value by

volume to obtain volume-invariant texture features (notably, the two features included in the final analysis did not correlate with volume, and thus, did not require normalisation to volume).

From the 665 sets of features at each gray-level, we used least absolute shrinkage and selection operator (LASSO) regression analysis for data dimension reduction, radiomics feature vector (composite feature) discovery, generating Kaplan–Meier curves and computing the Cox regression analysis. LASSO is a form of penalised regression used to reduce the problem of multi-collinearity. Briefly, the non-contributory variables were assigned zero-weighting, and numerous iterations were performed to link the non-zero contributory variables to the chosen clinical outcome (in this example, overall survival) [13]. Analyses were conducted with R software; the packages in R used for our analysis are indicated in Supplementary Table 4. Two-sided statistical significance levels were used, and $p \leq 0.05$ was considered statistically significant. SPSS for Statistics Version 22 (IBM, Armonk, NY, USA) was used for interclass correlation and 2-way ANOVA.

The most predictive feature vectors (FVX) were computed by linear combination of selected statistical features of the matrices weighted by their respective coefficients and by comparison with overall survival (OS). Survival curves were plotted using Kaplan–Meier (KM) methods, stage-specific or Youden's J cut-off on the receiver operator curve for the median survival in the case of FVX. Kaplan–Meier curves were plotted for overall survival using the 'survfit' function from the 'survival' package in R using the median cut-off for the MTV, TLG, and FVX. The statistical significance of the difference in the survival curves was calculated using the logrank test implemented in the 'survdiff' function. The survival curves were evaluated using a log-rank test (Cox Regression). Multivariable analysis of the FVX, stage, MTV, and TLG were compared with each other using a stepwise backward procedure to determine significantly independent survival indicators. P values of ≤ 0.05 were considered statistically significant, and 95% confidence intervals were calculated. A continuous Cox regression and the C-index, was computed for each prognosticator in the univariate analysis, and for the multivariable analysis with and without FVX. All four variables (FVX, stage, MTV, and TLG) were used as continuous variables in the analysis.

Independent validation and testing

Performance of the FVX and stage were tested by comparison to OS in an independent validation set of 204 patients, and a further independent set of 21 (TESTI; the final institutional dataset to be accepted into the study) patients. Similar survival comparisons were made with routine PET variables including SUV_{mean}, SUV_{max}, SUV_{peak}, MTV, and TLG.

Results

## Patient characteristics and PET analysis

Patient characteristics are displayed in Table 1. There were no significant differences in the proportion of males to females except in TESTI, which was a very small dataset. The majority of patients were, as expected for such a cohort, stage III. Typical PET images are shown in Fig. 1b. Primary tumor SUV_{max} ranged from 3.3 to 49.5 (Table 1). All patients were treated with radiotherapy with or without chemotherapy, and median survival values were not significantly different between the training, validation, and independent test cohorts, except in the small cohort in TESTI. No cases treated primarily by surgery were analysed, as the inclusion criterion was the cohort having radiotherapy with or without chemotherapy rather than surgery. Admittedly, some patients with stage I or II disease would have been unfit for surgery, while others would have elected for radiotherapy with or without chemotherapy in preference to surgery due to factors including patient choice; we do not have accurate data for the reasons for this choice. Of 358 patients, 249 died within the follow-up period [median 22 (range 0–85) months]. The comprehensive data for scanners are provided in Supplementary Table 2.

Segmentation is an important source of variability in radiomics analysis [14]. The most prevalent threshold cut-off values were 40% (47.8% of cases) and 30% (27.7% of cases); together these accounted for 75.5% of all the thresholds; 24.5% required a lower threshold value in order to encompass the whole tumour as defined by the CT component of the PET. Furthermore, 9.5% of cases required additional manual adjustment, after setting the initial threshold, to achieve optimal segmentation. Intra- and inter-observer variability of radiomics features from 18 randomly selected patients are displayed in Supplementary Tables 5 and 6. There was near-perfect [15, 16] intra- and inter-observer variability in the PET-derived radiomics features. The intra-observer variability of the radiomics features alone and when combined with PET features was 0.9 and 0.92 respectively. Corresponding inter-observer variability values were 0.86 and 0.88 respectively.

Radiomics feature vector selection

Radiomics predictive classifiers (665) were derived from TexLAB *v2*. Generation of a heatmap (Fig. 2) from all the patient data — both training and validation sets — visually indicated multi-collinearity (when many features are related), and suggested that some features could be reduced without information being lost. From the 665 radiomics features returned by the software in the training set of 133 patient scans, one FV was selected as the optimal predictor (FVX) — a weighted linear combination of the statistical features of size-variance of the gray-level size zone matrix at 64 Gy levels (GLSZM_SzVarianc_64gl;
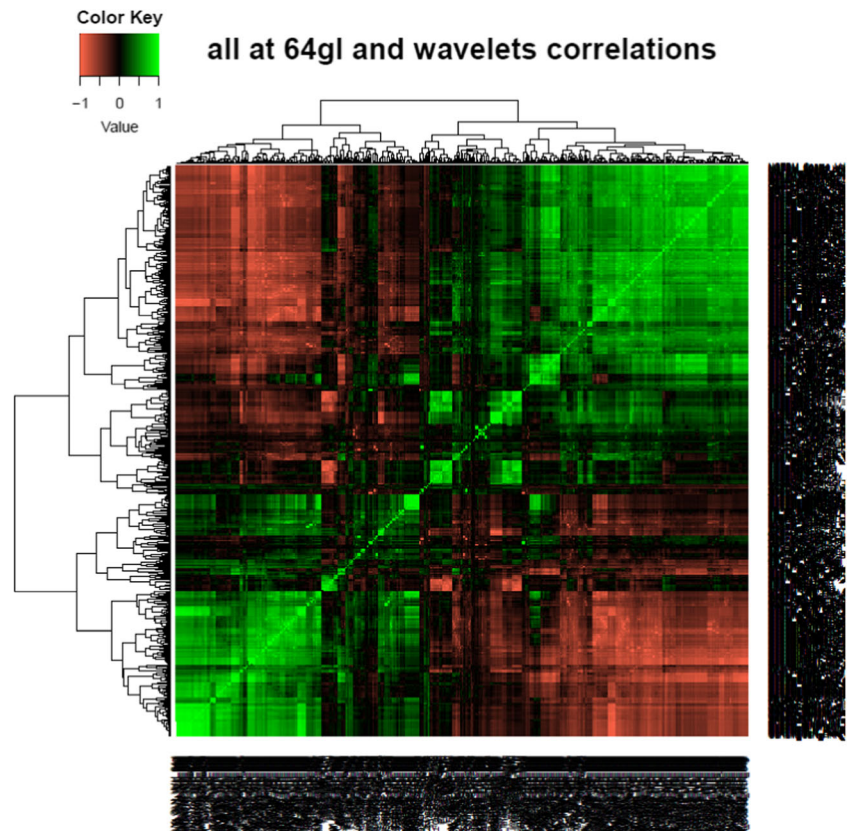
weighted by 0.128) and complexity of the neighbourhood gray-
tone difference matrix (NGTDM) at 64 Gy levels
(NGTDM_Complex_64gl; weighted by −0.018) — using
LASSO conducted independently at each gray level (example
shown in Supplementary Fig. 2). Previous studies have indicated
that primary tumor volume is an important predictor of survival
in lung cancer [17]. A multivariable analysis was performed
which included the new FVX, volume (SNS_{vol} variable) and
stage; volume was not significant, and was not further consid-
ered. Both the stage and FVX were significant and, most impor-
tantly, there was no correlation between the two ($p = 0.22$).

We tested the influence of PET scanner equipment proper-
ties on the FVX. Principal component analysis of FVX (at
64 Gy levels) was used to assess the congruence of data across
different manufacturer types, manufacturer models, slice
thickness, number of rows, or number of columns (Fig. 3;
Supplementary Figs. 3–5 and Supplementary Table 7). All
elements of the data were tightly clustered around each other
(minimal variance), suggesting that FVX was invariant to the
type of PET/CT manufacturer or slice thickness; thus, no cor-
rection was made for sets of data from different institutions.
Other FVs were dependent on scanner type (*data not shown*).

Performance of radiomics feature vector

We tested the performance of FVX in an independent valida-
tion cohort comprised of 204 patient scans by comparison to
OS. Kaplan–Meier (KM) plots for stage and FVX are shown
in Fig. 4. FVX was significantly associated with OS in the
validation set when dichotomised at median ($p = 0.00465$),
optimal cut-off by Kmroc ($p = 0.00116$) or as a continuous
variable ($p = 0.00429$), with hazard ratios (HRs) of 1.61
(1.16–2.24), 1.74 (1.25–2.44), and 5.30 (1.69–16.6) respec-
tively. In the TESTI cohort that presented with an unusually
poor survival of the four stage I cancers (Supplementary
Figs. 6 and 7), FVX correctly indicated a lack of survival
difference ($p = 0.501$). FVX values for image data presented
in Fig. 1b, for instance, were − 29.9 and − 03.1 for patients 1
and 2 respectively, thus correctly predicting survival relative
to stage. In contrast to the radiomics classifier, clinically rou-
tine PET variables including SUV_{max}, SUV_{mean}, and SUV_{peak}
lacked any predictive information (Supplementary Fig. 8).
The MTV and TLG were significant on the KM plots; surpris-
ingly, MTV was also significant on the TESTI KM plot (Fig.
5). The MTV and TLG were highly correlated with each other
(Supplementary Table 8), but neither the TLG nor MTV when
tested separately with the FVX and stage were significant on
the multivariable Cox regression (Supplementary Table 9).

FVX, stage, MTV, and TLG were the only potential prog-
nosticators that showed significance in the univariate analysis.
The Cox regression analysis associating FVX, stage, MTV,
and TLG with overall survival in three datasets are
summarised in Supplementary Table 9, with both univariate
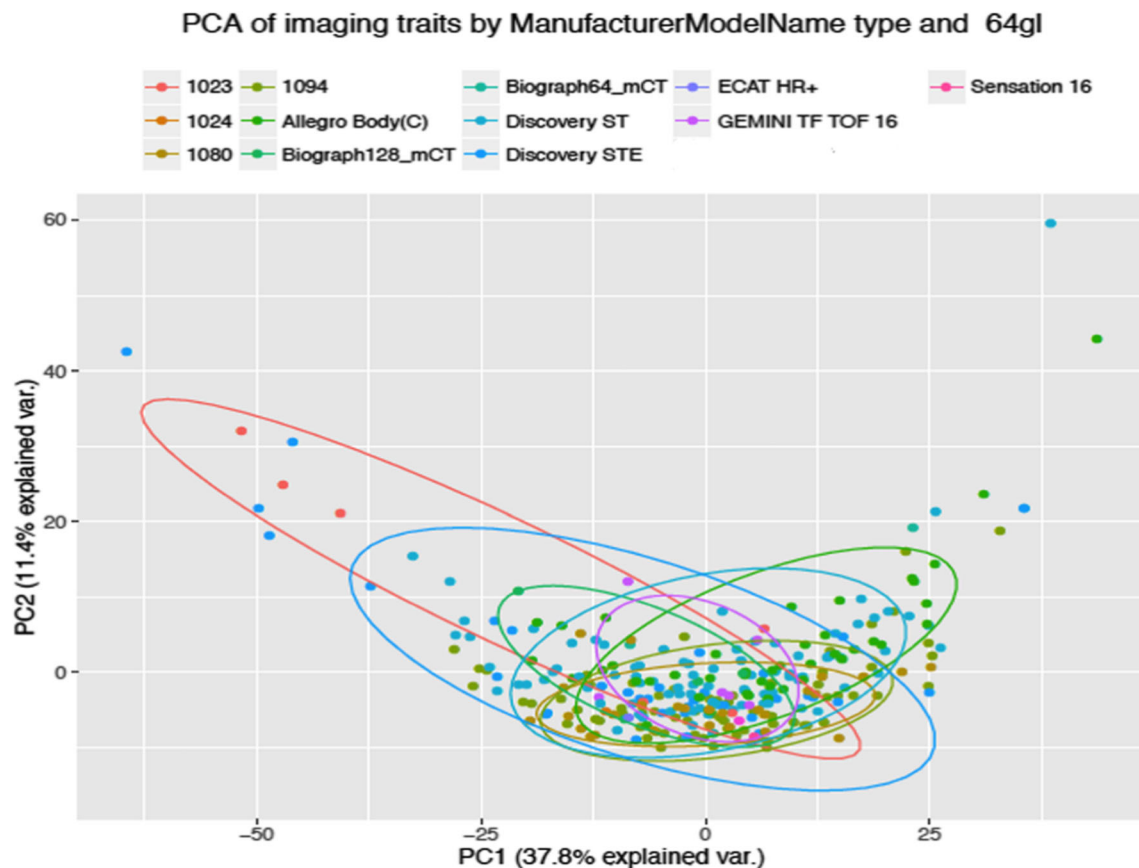analysis and multivariable analysis combining all the four

Fig. 3 Principal component analysis (explained variance) of PET radiomics features (at 64 Gy level) to assess congruence of data from different manufacturer models: CPS 1023, CPS 1024, Siemens 1080, 1094, Phillips Allegro Body (C), Siemens Biograph 64 mCT, Siemens Biograph 128 mCT, GE Discovery ST, GE Discovery STE, CTI ECAT HR+, Phillips Gemini TF TOF 16, and CPS/Siemens Sensation 16 respectively

variables. FVX was prognostic, independent of stage, MTV, and TLG in both training and validation sets. MTV and TLG were not significant once combined with FVX in the multivariable model, suggesting that FVX is a significantly better prognosticator than MTV and TLG.

Discussion

This multi-institution retrospective study showed that a radiomics feature vector, FVX, derived from analysis of FDG-PET data of primary tumors ≥ 5 ml in patients with NSCLC is invariant to PET scanner properties and predicts OS. Accurate prognostic information is crucial in stratifying newly diagnosed NSCLC patients to different treatments or best supportive care. Currently, TNM staging is the primary method to stratify treatment approach in NSCLC; however, it offers imprecise prognostic information, leading to both under-treatment and over-treatment in some patients. Other established prognostic factors for lung cancer include performance status (Karnofsky or ECOG (Eastern Cooperative Oncology Group) classification), weight loss (e.g., > 5%)

and systemic inflammation (C-reactive protein or modified Glasgow Prognostic Score) [18–20]. While factors such as EGFR (epidermal growth factor receptor) mutation predict response to targeted therapy [21–23], tumor-specific prognostic factors are lacking. In the current work, we assessed the role of radiomics features as prognostic factors in NSCLC. A machine-learning-enabled weighted linear combination of the statistical features of GLSZM_SzVarianc_64gl and NGTDM_Complex_64gl — FVX — was found to possess prognostic information and importantly was invariant to scanner properties investigated (Supplementary Table 7). The features do not have immediate physiological relevance. GLSZM is a regional 'homogeneity' texture feature that calculates lengths of uniform pixels (picture elements) in a 2D image, or in our case, directionally-independent groups of uniform voxels (volume elements) in each of the 26 available directions in 3D space; GLSZM_SzVarianc_64gl (size variance of the GLSZM at 64 Gy level) examines the variance in the number blocks by size (independent of the gray-level) and is negatively correlated with survival, possibly identifying hypoxic or necrotic regions with poor prognosis [24–26].
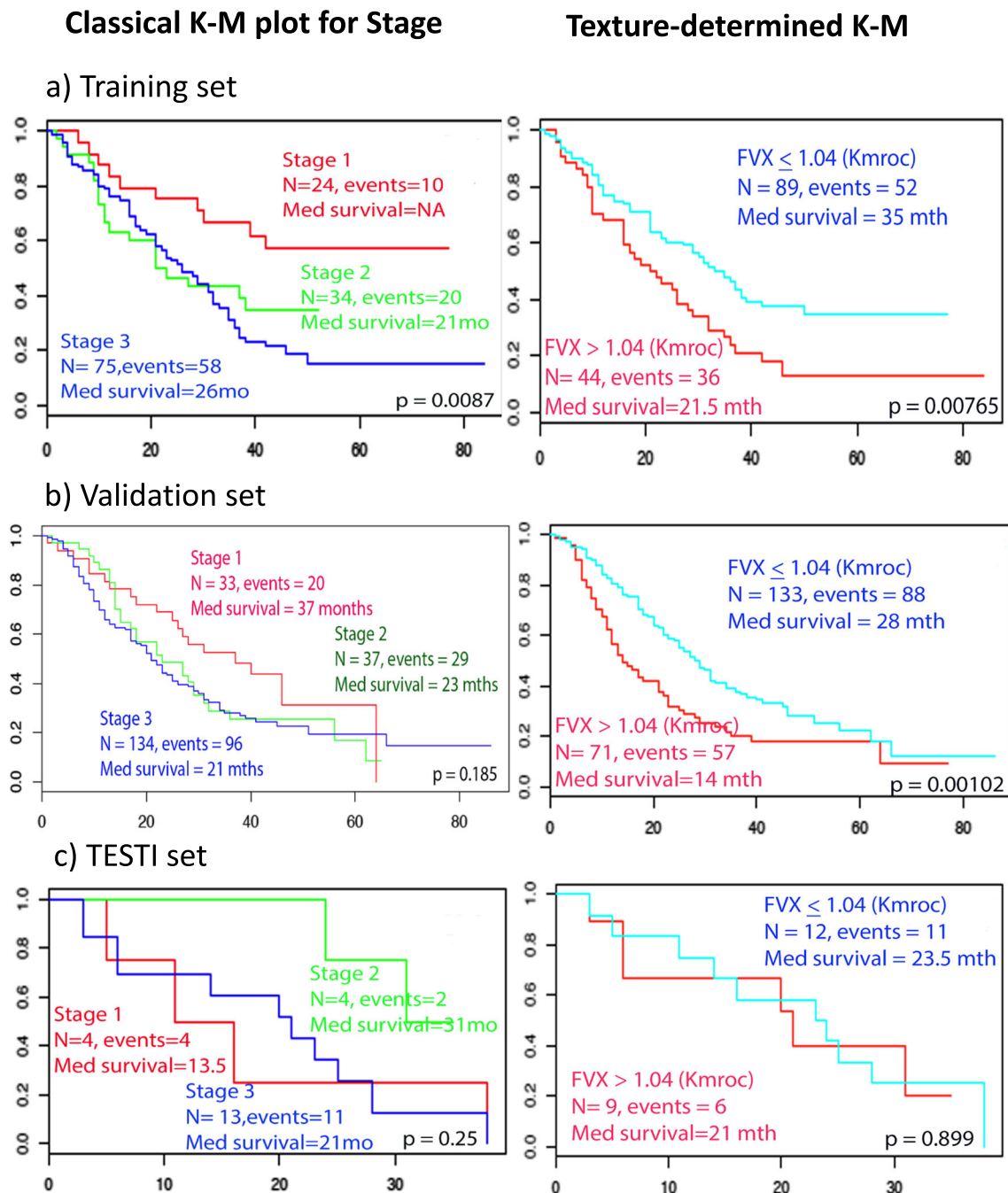
Classical K-M plot for Stage   **Texture-determined K-M**

a) Training set

b) Validation set

c) TESTI set

Fig. 4 Survival analysis based on composite radiomics feature dichotomized using ROC. Kaplan–Meier plots of **a** training, **b** independent validation, and **b** TESTI. Note that the validation dataset has longer follow-up period. *K–M* = Kaplan–Meier, *N* = number of subjects, *mths, mo, mth* = months, *Med* = median

NGTDM represents contrast, and is determined by examining changes in intensity between a target voxel and the surrounding neighbors to enable calculation of apparent difference between neighboring regions of voxel intensities. Contrast is related to the information content of an image and is a mathematical measure of heterogeneity; non-responding tumors with poor prognosis tend to have higher contrast [27]. NGTDM_Complex_64gl (complexity of the NGTDM for 64gl), which refers to the average visual complexity within the volume, is positively correlated with survival, although with less of a magnitude than SzVarianc, and perhaps acting as a balance on SzVarianc.

This is one of the first reports of a whole tumor image-derived lung cancer prognostic factor. In our analysis, there was a higher hazard of death (1.74; $p = 0.00116$) when the median FVX was used as input. The implicit assumption here is that a set of mathematically-derived tumor phenotypes correlate with survival. It should be noted, however, that death as

MTV based K-M plot

TLG K-M plot

a) Training set

b) Validation set
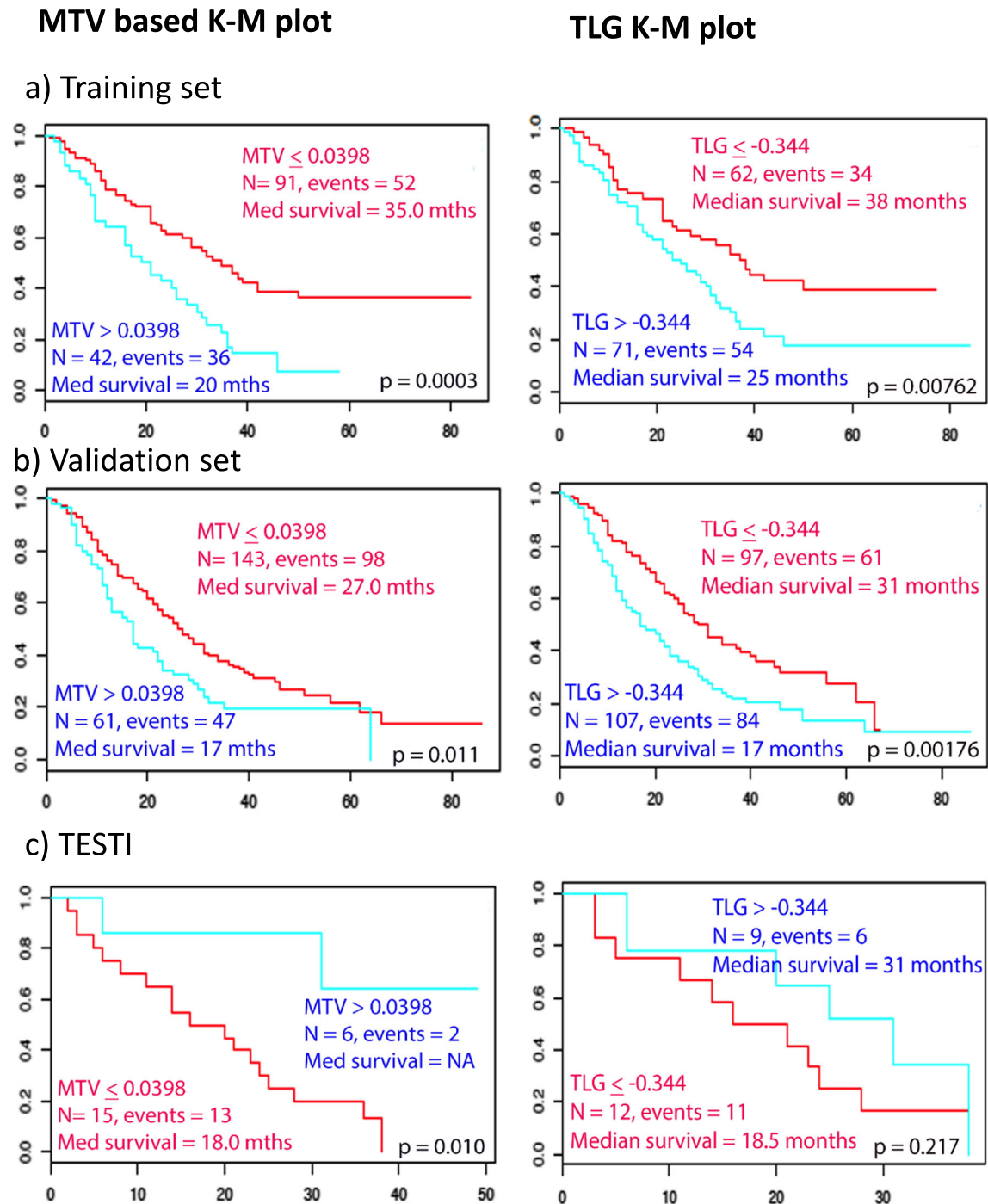
c) TESTI

Fig. 5 Survival analysis based on the SUV variables, MTV and TLG, dichotomized using ROC. Kaplan–Meier plots of **a** training dataset, **b** independent validation set, and **c** independent TESTI. Note that the validation dataset has a longer follow-up period. *MTV* metabolic tumour volume, *TLG* total lesion glycolysis

an endpoint could have occurred by other means — indirect consequence of the tumor or non-tumor related — or may have been subjected to different variations of physician choice of chemotherapy/chemo-radiotherapy, making this analysis the more interesting. Furthermore, routine PET variables, while useful for staging, did not possess prognostic information. Baseline primary tumor SUV_{max} has been reported by some groups, but not all, to predict outcome in NSCLC patients. For resectable NSCLC patients, a meta-analysis of 13 studies showed that primary tumor SUV_{max} has significant prognostic value on patient survival [28]. A more recent meta-analysis, assessing the prognostic value of primary tumor SUV_{max} prior to radiotherapy in NSCLC, reported that higher tumor SUV_{max} was correlated with shorter

OS, particularly in stage I NSCLC receiving stereotactic body radiotherapy (SBRT) [29].

Volumetric parameters, such as MTV and TLG, which consider the whole tumor volume, have been reported to be prognostic in NSCLC. Secondary analysis of the large multicentre prospective American trial of 196 inoperable Stage IIb/III NSCLC has reported MTV, TLG to be strongly prognostic for OS, while SUV_{max} was not [30]. The TLG and MTV, which were shown in our study to highly correlate with each other, were not significant on multivariable analysis.

As this was a large multi-centre study acquired on different scanners, voxel sizes differed (Supplementary Table 2); we did not standardise the voxel sizes. However, slice thickness and matrix size did not significantly affect the FVX. The robustness of FVX, invariant to instrument factors including slice thickness, permits this variable to be applied in multi-institutional studies. Previous work on scanner types [31–33] have yielded mixed results in terms of texture stability across model and manufacturer type, although limited models have been used. In addition, much thought has been given in the methodology to reduce Type 1 errors and false discovery which have entered the published literature [34]. Compared to CT technology that has seen substantive reduction in slice thickness, PET FWHM (full width at half-maximum; a measure of resolution) has not seen such substantive change over the past decade, and this could have led to the scanner invariance of our study. We set a threshold of 5 ml, in keeping with earlier work of Soussan [8]. It is likely that inclusion of smaller tumors would have led to higher variability given the poor resolution of PET (compared to CT or MRI). Hence, the inferences from this study are limited to the group of patients presenting with medium-large lesions. The classification of patient subgroups in the Kaplan–Meier analysis was based on the FVX value calculated from combination of weighted radiomic features. FVX was a continuous variable, and we have demonstrated that FVX was linearly correlated with overall survival in the training and validation sets using a continuous Cox regression analysis. It should be noted that other analyses were performed: Cox regression based on dichotomised median FVX, Kmeans clustering and optimal cut-off (Youden's J) from the ROC curve, and showed consistent results (data not shown).

TESTI was an unusual dataset in terms of size and heterogeneity with a high number of stage 1 tumors and high mortality. The fact that these patients had radiotherapy ± chemotherapy instead of surgery indicates that there was probably associated poor performance status. Unfortunately, information on performance status was not available. However it was felt important to attempt to test the radiomics signature against this unusual dataset, as ideally the 'real-life' prediction of the radiomics feature vector should work irrespective of sample size.

A recent single-institution study of a PET/CT radiomics signature for prediction of disease-free survival (DFS) in NSCLC undergoing surgery with curative intent reported that image derived parameters outperformed TNM staging in predicting DFS. However, although promising, this was a single-centre study, utilising what appears to be unenhanced CT scans without external validation and in a different cohort of patients to our study [35].

Limitations of the present study should be highlighted. 1) This is a retrospective, albeit multi-institutional, study and future prospective studies in similarly large cohorts will be needed to verify this novel endpoint. 2) We did not consider other prognostic factors as these were not consistently available from all institutions. Addition of other prognostic factors in future studies will enable more rigorous assessment of events likely to have caused death. 3) We did not consider EGFR mutations/expression or other genetic outcomes, as these were not available for all patients. 4) Approximately 14% of the initial 535 patients screened were excluded for having tumours < 5 ml (Supplementary Table 1). The choice of 5 ml reflects a statistical limitation of applying radiomics to PET data (less data-points within the VOI compared to CT). Thus, with regard to generalisation of our study to patients having chemo-radiotherapy, we would caution the exploitation of our findings to smaller tumors. Irrespective of these limitations, we highlight a massive opportunity for physicians and patients, whereby mathematically-derived features from scans that newly diagnosed NSCLC patients would normally have as part of routine care can be 're-purposed' to predict prognosis. Only software implementation and computing power are required for incorporation into patient management pathways; thus, we envisage easy acceptance of this potentially cost-effective methodology for use with existing prognostic methods. As it is tumor-specific, patients stratified to poor prognostic FVX groups could be candidates for earlier follow-up or a lower threshold in change of therapy [36].

In summary, we have discovered a scanner-invariant radiomics feature vector that performs well in independent validation and test datasets. This multi-institutional study provides new opportunities for prospective assessment of radiomics features for prognosis in patients with NSCLC.

Compliance with ethical standards

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the principles of the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. Institutional ethical approval for retrospective analysis of human data was obtained and informed consent was waived, typical for retrospective analysis of anonymised imaging data. This article does not describe any studies performed on animals.

Conflict of interest The authors declare they have no conflict of interest according to the subject and matter of the present article.

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin. 2011;61(2):69–90.
2. de Bruin EC, McGranahan N, Mitter R, Salm M, Wedge DC, Yates L, et al. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. Science. 2014;346(6206):251–6.
3. Zhang J, Fujimoto J, Zhang J, Wedge DC, Song X, Zhang J, et al. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. Science. 2014;346(6206):256–9.
4. Segal E, Sirlin CB, Ooi C, Adler AS, Gollub J, Chen X, et al. Decoding global gene expression programs in liver cancer by non-invasive imaging. Nat Biotechnol. 2007;25(6):675–80.
5. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Cavalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;5:4006.
6. Willaime JM, Turkheimer FE, Kenny LM, Aboagye EO. Quantification of intra-tumour cell proliferation heterogeneity using imaging descriptors of 18F fluorothymidine-positron emission tomography. Phys Med Biol. 2013;58(2):187–203.
7. 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;6:71
8. Soussan M, Orlhac F, Boubaya M, Zelek L, Ziol M, Eder V, et al. Relationship between tumor heterogeneity measured on FDG-PET/CT and pathological prognostic factors in invasive breast cancer. PLoS One. 2014;9(4):e94017.
9. Boellaard R, Delgado-Bolton R, Oyen WJG, Giammarile F, Tatsch K, Eschner W, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. Eur J Nucl Med Mol Imaging. 2015;42:328–54.
10. Biehl KJ, Kong FM, Dehdashti F, Jin JY, Mutic S, El Naqa I, et al. 18F-FDG PET definition of gross tumor volume for radiotherapy of non-small cell lung cancer: is a single standardized uptake value threshold approach appropriate? J Nucl Med. 2006;47(11):1808–12.
11. Foster B, Bagci U, Mansoor A, Xu Z, Mollura DJ. A review on segmentation of positron emission tomography images. Comput Biol Med. 2014;50:76–96.
12. Brooks FJ, Grigsby PW. FDG uptake heterogeneity in FIGO IIb cervical carcinoma does not predict pelvic lymph node involvement. Radiat Oncol. 2013;8:294.
13. Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med. 1997;16(4):385–95.
14. Leijenaar RT, Nalbantov G, Carvalho S, van Elmpt WJ, Troost EG, Boellaard R, et al. The effect of SUV discretization in quantitative FDG-PET radiomics: the need for standardized methodology in tumor texture analysis. Sci Rep. 2015;5:11075.
15. Büyükdereli G, Güler M, Şeydaoğlu G. Interobserver and Intraobserver variability among measurements of FDG PET/CT parameters in pulmonary tumors. Balkan Med J. 2016;33(3):308–15.
16. Armitage P, Berry G, Matthews JNS. Statistical methods in medical research. 4th ed: Oxford, Blackwell Science; 2001.
17. Ball DL, Fisher RJ, Burmeister BH, Poulsen MG, Graham PH, Penniment MG, et al. 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(3):305–11.
18. Lee ES, Son DS, Kim SH, Lee J, Jo J, Han J, et al. Prediction of recurrence-free survival in postoperative non-small cell lung cancer patients by using an integrated model of clinical information and gene expression. Clinical Cancer Res. 2008;14(22):7397–404.
19. Detterbeck FC, Postmus PE, Tanoue LT. The stage classification of lung cancer: diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest. 2013;143(5 Suppl):e191S–210S.
20. Alberg AJ, Brock MV, Ford JG, Samet JM, Spivack SD. Epidemiology of lung cancer: diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest. 2013;143(5 Suppl):e1S–29S.
21. Majem M, Remon J. Tumor heterogeneity: evolution through space and time in EGFR mutant non small cell lung cancer patients. Transl Lung Cancer Res. 2013;2(3):226–37.
22. Nana-Sinkam SP, Powell CA. Molecular biology of lung cancer: diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest. 2013;143(5 Suppl):e30S–9S.
23. Chen Z, Fillmore CM, Hammerman PS, Kim CF, Wong KK. Non-small-cell lung cancers: a heterogeneous set of diseases. Nat Rev Cancer. 2014;14(8):535–46.
24. Overgaard J. Hypoxic radiosensitization: adored and ignored. J Clin Oncol. 2007;25(26):4066–74.
25. Wilson WR, Hay MP. Targeting hypoxia in cancer therapy. Nat Rev Cancer. 2011;11(6):393–410.
26. Thibault G, Fertil B, Navarro C, Pereira S, Cau P, Levy N, et al. Texture indexes and gray level size zone matrix application to cell nuclei classification. PRIP. 2009; 2009:140–5.
27. Sun CJ, Wee WG. Neighboring gray level dependence matrix for texture classification. Comput Vision Graph. 1983;23(3):341–52.
28. Berghmans T, Dusart M, Paesmans M, Hossein-Foucher C, Buvat I, Castaigne C, et al. Primary tumor standardized uptake value (SUVmax) measured on fluorodeoxyglucose positron emission tomography (FDG-PET) is of prognostic value for survival in non-small cell lung cancer (NSCLC): a systematic review and meta-analysis (MA) by the European Lung Cancer Working Party for the IASLC Lung Cancer Staging Project. J Thorac Oncol. 2008;3(1):6–12.
29. Na F, Wang J, Li C, Deng L, Xue J, Lu Y. Primary tumor standardized uptake value measured on F18-Fluorodeoxyglucose positron emission tomography is of prediction value for survival and local control in non-small-cell lung cancer receiving radiotherapy: meta-analysis. J Thorac Oncol. 2014;9(6):834–42.

30. Salavati A, Duan F, Snyder BS, Wei B, Houshmand S, Khiewvan B, et al. Optimal FDG PET/CT volumetric parameters for risk stratification in patients with locally advanced non-small cell lung cancer: results from the ACRIN 6668/RTOG 0235 trial. Eur J Nucl Med Mol Imaging. 2017; 44(12):1969–83

31. Reuzé S, Orlhac F, Chargari C, Nioche C, Limkin E, Riet F, et al. Prediction of cervical cancer recurrence using textural features extracted from (18)F-FDG PET images acquired with different scanners. Oncotarget. 2017;8(26):43169–79.

32. Galavis PE, Hollensen C, Jallow N, Paliwal B, Jeraj R. Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters. Acta Oncol. 2010;49(7): 1012–6.

33. 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(7): e102107.

34. Chalkidou A, O'Doherty MJ, Marsden PK. False discovery rates in PET and CT studies with texture features: a systematic review. PLoS One. 2015;10(5):e0124165.

35. Kirienko M, Cozzi L, Antunovic L, Lozza L, Fogliata A, Voulaz E, et al. Prediction of disease-free survival by the PET/CT radiomic signature in non-small cell lung cancer patients undergoing surgery. Eur J Nucl Med Mol Imaging. 2018;45(2):207–17.

36. Okimoto RA, Bivona TG. Recent advances in personalized lung cancer medicine. Pers Med. 2014;11(3):309–21.

ARTICLE

A mathematical-descriptor of tumor-mesoscopic-structure from computed-tomography images annotates prognostic- and molecular-phenotypes of epithelial ovarian cancer

Haonan Lu [1,2], Mubarik Arshad[2], Andrew Thornton[1], Giacomo Avesani [2], Paula Cunnea [1], Ed Curry[1], Fahdi Kanavati[2], Jack Liang[2], Katherine Nixon[1], Sophie T. Williams[1], Mona Ali Hassan[1], David D.L. Bowtell[3,4], Hani Gabra[1,5], Christina Fotopoulou[1], Andrea Rockall[2,6,7] & Eric O. Aboagye [2]

The five-year survival rate of epithelial ovarian cancer (EOC) is approximately 35–40% despite maximal treatment efforts, highlighting a need for stratification biomarkers for personalized treatment. Here we extract 657 quantitative mathematical descriptors from the preoperative CT images of 364 EOC patients at their initial presentation. Using machine learning, we derive a non-invasive summary-statistic of the primary ovarian tumor based on 4 descriptors, which we name "Radiomic Prognostic Vector" (RPV). RPV reliably identifies the 5% of patients with median overall survival less than 2 years, significantly improves established prognostic methods, and is validated in two independent, multi-center cohorts. Furthermore, genetic, transcriptomic and proteomic analysis from two independent datasets elucidate that stromal phenotype and DNA damage response pathways are activated in RPV-stratified tumors. RPV and its associated analysis platform could be exploited to guide personalized therapy of EOC and is potentially transferrable to other cancer types.

[1] Ovarian Cancer Action Research Centre, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, London W12 0HS, UK. [2] Cancer Imaging Centre, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, London W12 0HS, UK. [3] Peter MacCallum Cancer Centre, Melbourne 3010 VIC, Australia. [4] Sir Peter MacCallum Department of Oncology, The University of Melbourne, Melbourne 3010 VIC, Australia. [5] Early Clinical Development, iMED Biotech Unit, AstraZeneca, Cambridge SG8 6HB, UK. [6] Department of Radiology, Imperial College Healthcare NHS Trust, London W12 0HS, UK. [7] Department of Radiology, The Royal Marsden NHS Foundation Trust, London SW3 6JJ, UK. These authors contributed equally: Haonan Lu, Mubarik Arshad. Correspondence and requests for materials should be addressed to E.O.A. (email: eric.aboagye@imperial.ac.uk)

"Radiomics" quantifies mesoscopic tumor phenotype from anatomic or functional images by defining tumor spatial complexity—including first and higher order statistics, fractal and shape features—generating disease features not appreciated by the naked eye[1–3]. The development of a radiomics approach for disease phenotyping, using routine pre-surgical computed tomography (CT), as an extension of current imaging semantics is therefore promising[4–6].

Epithelial ovarian cancer (EOC) is the sixth most common cancer among women in the UK and has the highest mortality of all gynecological cancers, accounting for 4% of all cancer deaths in women[7]. High-grade serous ovarian cancer (HGSOC) represents the most dominant (70% of EOC patients) and most lethal histological subtype[8]. Although it is well known that HGSOC patients have a heterogeneous response to treatment and prognosis, extensive cytoreductive surgery combined with platinum-based chemotherapy are currently the standard treatments for most patients without consideration of individual prognostic and predictive biomarkers. Recently, a number of studies including the Cancer Genome Atlas (TCGA) project have obtained a comprehensive genomic profile of HGSOC, resulting in several molecular prognostic biomarker discoveries[9]. For instance, *CCNE1* amplification is commonly associated with platinum-resistant and refractory disease[10,11]; HGSOCs were classified into prognostically distinct molecular subtypes according to gene expression profiling[12–14]. More recently, large sets of microRNAs have been exploited to determine the risk profile of EOC[15]. It remains challenging, however, to translate these molecularly determined characteristics into clinically relevant biomarkers due to intratumor heterogeneity, additional high assay cost, and time delays. Therefore, a noninvasive, real-time, and cost-effective prognostic marker approach is warranted to reliably guide personalized treatment of EOC patients.

In the current study, a novel radiomics-determined mathematical descriptor of EOC tumor risk phenotype with a reliable, convincing predictive value is discovered and validated, and further insights into the biological basis of the descriptor is provided through investigation of correlated transcriptomics, proteomics and copy-number alterations (CNAs).

Results

Characteristics of data and patients. We developed TexLab 2.0, a software program that summarized 657 features relating to the shape and size, intensity, texture and wavelet decompositions of 364 preoperative contrast-enhanced CT scans[16] (Table 1 and Supplementary Figure 1). All the radiomic features are summarized in Supplementary Data 1. A comprehensive molecular profile including gene expression, copy-number, and protein expression was analyzed for a subset of patients (Table 1). The study workflow is summarized in Supplementary Figure 2.

We evaluated 294 primary EOC patients with fresh frozen tissue treated within the Hammersmith Hospital, Imperial College Healthcare NHS Trust, London, UK between 2004 and 2015 as well as 70 EOC patients from the TCGA project (Supplementary Table 1, Supplementary Figure 2).

Overview of radiomic profile in epithelial ovarian cancer. We wished to investigate the data structure within the radiomic profiles derived from primary tumors of EOC patients in relation to clinical and genetic features. For samples with both radiomics and CNA data, we performed a spectral clustering analysis based on the Pearson correlation coefficients between each samples' radiomic profile (Fig. 1a). There was a clear division of samples into three major groups with each group characterized by high feature similarity but largely distinct from those in other groups. Notably, one of these groups (Group 1) was found to be significantly enriched for HGSOC (Fig. 1b). EOC, particularly the HGSOC subtype, frequently features CNAs[17]. We found that Group 1 was enriched for tumors with high CNAs (Fig. 1c). This group had a worse outcome as measured by progression-free survival (PFS) (Supplementary Figure 3).

To further understand the radiomic characteristics of the HGSOC subtype, we performed unsupervised hierarchical clustering analysis using the radiomic profiles in the HH cohort. We found two distinct clusters within this population based purely on the radiomic profile (Fig. 1d). Cluster 2 was significantly associated with the presence of ascites ($p = 0.00729$, chi-squared test) and poor PFS ($p = 0.022$, log-rank test; Fig. 1e), marginally associated with higher tumor stage ($p = 0.0686$, Fisher's exact test), but not associated with postoperative residual disease or molecular subtype (Fig. 1d). Of interest, 96% of bilateral tumors from patient were assigned to the same cluster, revealing a close radiomic similarity (Supplementary Note 1).

In aggregate, unsupervised analysis highlighted an intrinsic association between radiomic profile, genetic background, and clinical characteristics, warranting further characterization.

Radiomic prognostic vector predicts survival. We used three datasets to assess the prognostic potential of the radiomic profile for HGSOC patients: HGSOC cases from the HH cohort were split into the HH discovery ($n = 136$) and the HH validation datasets ($n = 77$), and examined in parallel with the TCGA validation dataset ($n = 70$) (Supplementary Figure 2 and Supplementary Table 1). We firstly performed Cox regression with overall survival (OS) examining each radiomic feature in turn, using data from primary tumors in the HH discovery dataset (Supplementary Figure 2). Forty-two radiomic features were found to be significantly associated with OS (false discovery rate < 0.05; Fig. 2a; Supplementary Data 2). The 42 radiomic features were further reduced to 4 weighted features using least absolute shrinkage and selection operator (LASSO[18]) (Fig. 2b, c and Supplementary Table 2). The weighted sum of these four radiomic features gave a RPV score for each tumor.

With an unsupervised k-means clustering approach, we split all the patients from the three cohorts based on their RPV into three subgroups (low risk, medium risk, and high risk; Supplementary Table 3). The patient groups stratified by RPV had distinct OS differences in the discovery dataset ($N = 136$, $p < 0.0001$, log-rank test; Fig. 2d). Using the same RPV decision boundaries, OS differences were confirmed in two independent validation datasets, the TCGA validation dataset ($N = 70$, $p = 0.000105$, log-rank test; Fig. 2e) and the HH validation dataset ($N = 77$, $p = 0.0274$, log-rank test; Fig. 2f).

In a multivariable Cox regression model with age, stage, postoperative residual disease, neo-adjuvant chemotherapy, and
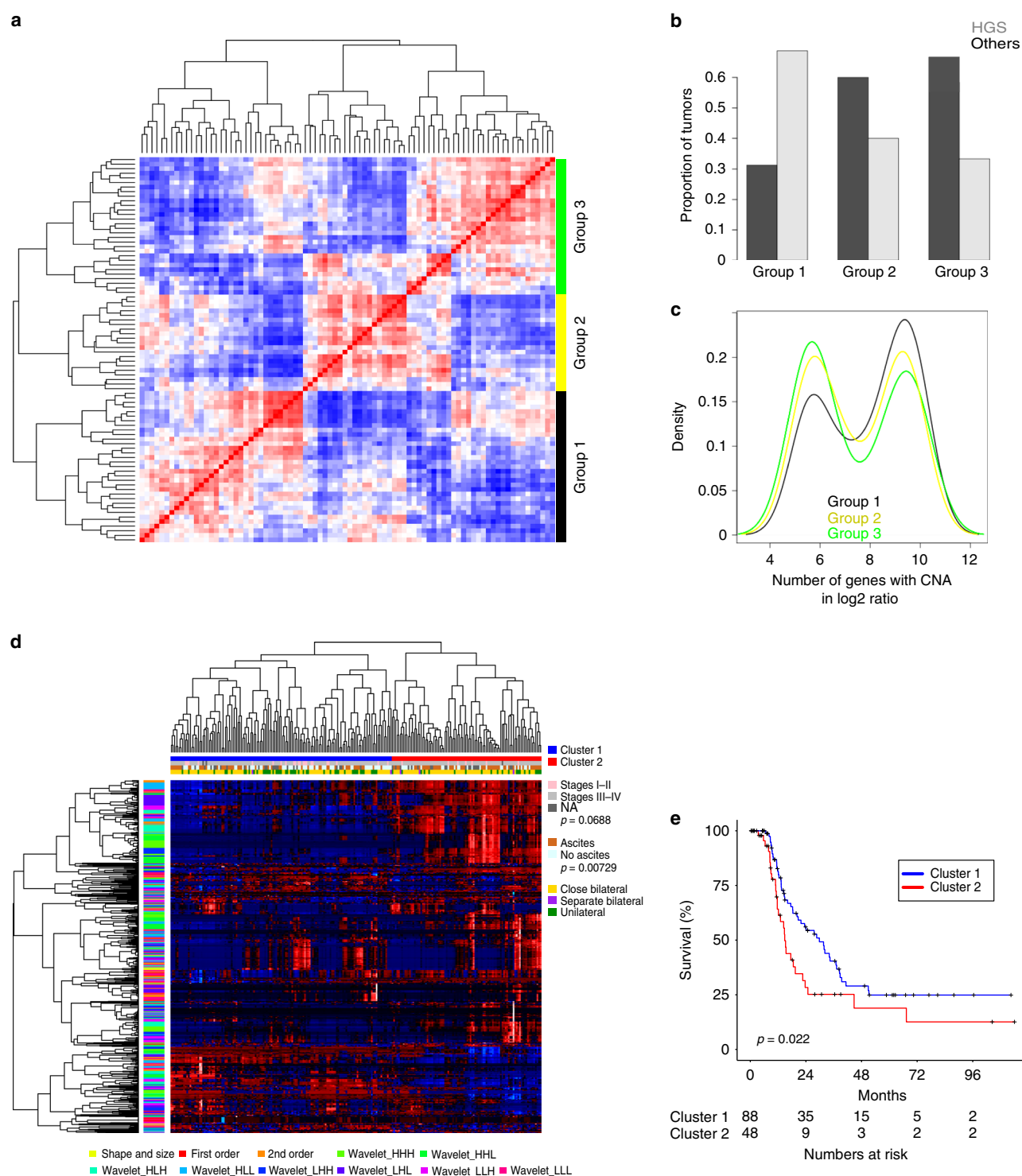
Table 1 Summary of data produced

Data type	Cohort	Platforms	Features	Cases
Radiomic profile	HH	TexLab 2.0	657	294
	TCGA			70
DNA copy number	HH	Affymetrix SNP6	Whole genome	84
	TCGA			70
Protein expression	HH	RPPA	299	198
	TCGA		199	48
mRNA expression	HH	Illumina MiSeq	68	173
	TCGA	Affymetrix U133	Whole genome	70

HH Hammersmith Hospital, *TCGA* The Cancer Genome Atlas

Fig. 1 Unsupervised clustering analysis of radiomic data in EOC. **a** Heatmap illustrating clustered matrix of sample-wise similarities (HH cohort) based on whole tumor radiomic profiles of primary ovarian tumors from all EOC histology. Black bar, group 1; yellow bar, group 2; green bar, group 3. **b** Distribution of high-grade serous ovarian carcinomas over patient groupings defined by similarities of radiomic profile ($n = 84$, $p = 0.02$, Fisher's exact test). **c** Differences in the numbers of genes affected by copy-number aberration in tumors with spectral radiomic clusters. Black line, group 1; yellow line, group 2; green line, group 3. **d** Unsupervised hierarchical clustering of radiomic profile from primary HGSOC identified two distinct subgroups (blue and red as shown on the top row above heatmap). The associations between radiomic subgroups with the presence of ascites, lateral and tumor stage are indicated on the right. A summary of radiomic features are given on the y-axis. Blue bar, cluster 1; red bar, cluster 2. **e** Kaplan–Meier analysis of the radiomic subgroups with progression-free survival ($n = 136$). Blue line, cluster 1; red line, cluster 2. p value from log-rank test is included. HH Hammersmith Hospital, EOC epithelial ovarian cancer, HGSOC high-grade serous ovarian cancer
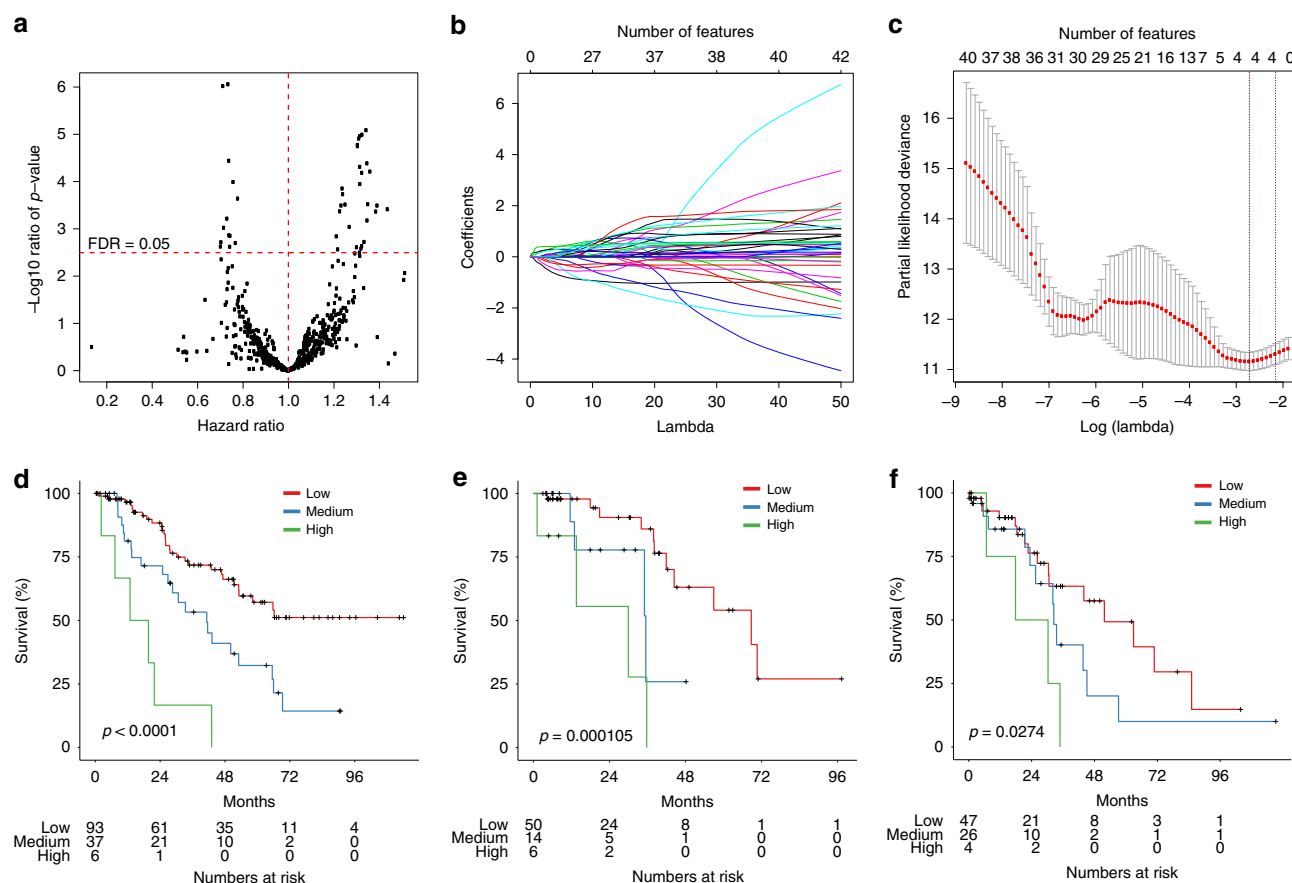
Fig. 2 Prognostic model based on radiomic profile in HGSOC. **a** Summary of univariate Cox regression between each radiomic feature and overall survival in the discovery set. Each black point represents the *p* value (*y*-axis) and hazard ratio (HR; *x*-axis) of a radiomic feature. Red horizontal dashed line indicates the false discovery rate (FDR) of 0.05; red vertical dashed line indicates HR of 1. **b** Least absolute shrinkage and selection operator (LASSO) regression analysis was performed to select radiomic features for prognostic model-building for HGSOC patients. Feature coefficients were plotted against shrinkage parameter (Lambda). **c** Partial likelihood deviance from Cox regression (*y*-axis) was generated under different shrinkage parameters (*x*-axis). Number of features selected corresponding to each lambda are given above the plot. Kaplan−Meier analyses were performed between radiomic prognostic vector (RPV) and overall survival in **d** HH discovery cohort (*n* = 136), **e** TCGA validation cohort (*n* = 70) and **f** HH validation cohort (*n* = 77). Red line, RPV low; green line, RPV medium; blue line, RPV high. *p* values are given by log-rank test. HGSOC high-grade serous ovarian cancer, TCGA the Cancer Genome Atlas

the potential structured noise in the datasets (scan thickness), RPV remained significantly and continuously associated with OS in the discovery dataset (hazard ratio (HR): 3.83, 95% confidence interval (CI) (2.27–6.46), $p = 5.11 \times 10^{-7}$; RPV range: −0.322 to 3.16), as well as the TCGA validation dataset (HR: 4.87, 95% CI (1.67–14.2), $p = 0.00380$) and the HH validation dataset (HR: 7.36, 95% CI (1.29–41.9), $p = 0.0245$; Table 2). The addition of RPV improved the clinically available prognostic methods (stage, age, and postoperative residual disease) in all three datasets as measured by the concordance index (C-index)[19] (HH discovery: from 0.658 to 0.739; TCGA validation: from 0.549 to 0.690; HH validation: from 0.659 to 0.679). Age, stage, and postoperative residual disease were significantly associated with OS in either uni- or multivariable analysis in the combined HH cohort while RPV remained the strongest prognostic factor, suggesting RPV is prognostic in a representative HGSOC cohort. RPV was also found associated with OS independent of performance status in a subset of patients (Supplementary Table 4). We excluded performance status from the multivariable analysis to avoid misinterpretation in the presence of insufficient data, given that we only had the performance status of 62 out of the total 213 patients in the HH cohort, and less than 20 of them had a performance status > 1. For that reason, any statistical

conclusions relating to performance status will not be valid due to the very small sample size. Notably, RPV possessed a better prognostic power when compared to the existing prognostic markers including CA125 and the transcriptome-based molecular subtype and potentially synergizes with existing CT-based morphological approaches (Supplementary Tables 5–7; Supplementary Note 1). Apart from prognosis, high RPV was found significantly associated with primary chemotherapy resistance, shorter PFS, and poor surgical outcome (Fig. 3e and Supplementary Figure 7g, Supplementary Note 1), suggesting RPV as a potential predictive marker in HGSOC.

Taking advantage of the gene expression profiles collected in parallel with radiomic profiles, we constructed a surrogate marker of RPV based on a weighted list of mRNA expressions in the TCGA validation dataset where both CT scans and gene expression profiles were available (eRPV; Supplementary Note 1). eRPV strongly correlated with RPV ($r = 0.720$) in the TCGA validation dataset and significantly interacted with RPV in the Cox regression model (Supplementary Figure 10c). It showed a similar prognostic potential as RPV in two additional cohorts (TCGA dataset without publicly available CT scans: $n = 448$, HR = 2.19, 95% CI (1.23–4.25), $p = 0.0208$; Tothill dataset: $n = 228$, HR = 7.94, 95% CI (2.02–31.3), $p = 0.00303$; adjusted

	Variables	Univariate		Multivariable	
	HR (95% CI)	p value	HR (95% CI)	p value	
HH discovery ($n = 136$)	RPV	4.08 (2.48–6.71)	3.37e-08	3.86 (2.30–6.46)	3.04×10^{-7}
	Stage	2.03 (1.37–3.00)	0.000426	1.88 (1.24–2.86)	0.00305
	Residual disease	1.75 (1.03–2.99)	0.0393	1.40 (0.803–2.44)	0.235
	Age[a]	1.25 (0.741–2.11)	0.404	1.47 (0.865–2.51)	0.154
HH validation ($n = 77$)	RPV	2.05 (1.01–4.18)	0.0485	5.08 (1.03–25.2)	0.0465
	Stage	1.32 (0.775–2.24)	0.309	1.32 (0.664–2.64)	0.425
	Residual disease	1.78 (0.777–4.08)	0.173	1.28 (0.514–3.21)	0.593
	Age[a]	2.10 (0.940–4.68)	0.0704	3.44 (1.19–9.94)	0.0228
HH cohort combined[b] ($n = 213$)	RPV	2.94 (2.02–4.26)	1.54×10^{-8}	3.32 (2.16–5.10)	4.91×10^{-8}
	Stage	1.82 (1.33–2.48)	0.00017	1.75 (1.24–2.50)	0.0017
	Residual disease	1.72 (1.11–2.69)	0.0163	1.36 (0.855–2.15)	0.196
	Age[a]	1.46 (0.951–2.24)	0.0835	1.74 (1.10–2.76)	0.0183
TCGA validation ($n = 70$)	RPV	4.94 (2.06–11.8)	0.00034	6.21 (2.06–18.7)	0.00117
	Stage	1.75 (0.913–3.34)	0.0921	1.03 (0.309–3.44)	0.960
	Residual disease	1.34 (0.480–3.74)	0.576	1.45 (0.414–5.05)	0.564
	Age[a]	1.08 (0.435–2.66)	0.874	0.500 (0.154–1.63)	0.249

Table 2 Summary of Cox regression analysis of RPV in three datasets. RPV was used as a continuous variable in the Cox regression analysis

HR hazard ratio, *CI* confidence interval, *RPV* radiomic prognostic vector, *HH* Hammersmith Hospital, *TCGA* the Cancer Genome Atlas
[a]Age has been dichotomized at 60 years
[b]Combining HH discovery and HH validation datasets

for stage, grade, residual disease, age and neo-adjuvant chemotherapy). We thus considered eRPV as a surrogate of RPV and subsequently used eRPV in a subset of the TCGA dataset without publicly available CT scans, as an extension of RPV (Noted as "eRPV" in Fig. 3d, e, Supplementary Figure 7a, 7c, 7e and h-j).

Overall, we observed RPV to be associated with OS, independent of known clinical prognostic factors, suggesting that it may reflect distinct aspects of clinically relevant variation across HGSOC.

Biological interpretation of the radiomic prognostic vector. To understand tumor biological characteristics linked to RPV, we evaluated enrichments of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways from Spearman correlation coefficients of gene expression with RPV (Fig. 3a, b; false discovery rate (FDR) < 0.05); the full lists of pathways are given in Supplementary Data 3 and 4.

We found that ECM−receptor interaction and focal adhesion were the two pathways most significantly enriched for associations with high RPV. These two pathways contained ECM components ($TIMP3$ ($r = 0.530$), $COL11A1$ ($r = 0.460$)) and focal adhesion receptors ($ITGA5$ ($r = 0.368$), $ITGB5$ ($r = 0.387$)), and from previous studies both pathways were enriched in stroma[20,21]. Accordingly, genes with expression correlated to high RPV were significantly enriched for genes expressed in the stromal component (Fig. 3c, chi-squared test $p < 0.0001$). Additionally, RPV was positively correlated with a stroma marker, fibronectin, at the protein level in both the TCGA and the HH cohorts (Fig. 3d and Supplementary Figure 7a). Furthermore, high RPV was associated with high proportion of tumor-associated stromal cells, evidenced from both histological data (Fig. 3e) and stroma score estimated from transcriptomic data[22] (Supplementary Figure 4b). A lower tumor cell content is inversely related to high stromal content. Consistent with previous results, we noted that higher RPV was associated with lower tumor cellularity (Fig. 3e) in the TCGA cohort and the same trend was observed in the HH cohort (Supplementary Figure 7d). These associations between molecular and histological characteristics with RPV were also observed with eRPV in a subset of the TCGA dataset without publicly available CT scans (Supplementary Figure 7a, 7c and 7e).

Besides stroma-related pathways, a number of proliferation and DNA damage response (DDR) pathways, including DNA replication, cell cycle, mismatch repair, base excision repair, nucleotide excision repair and homologous recombination, were among the top pathways activated in the RPV-low tumors (Fig. 3b). To verify the validity of the pathway analysis, we analyzed reverse phase protein array (RPPA) data from both HH and TCGA cohorts and found the expression of proliferation and DDR pathway marker proteins including Stathmin 1, FoxM1 and Rad51 to be higher at the protein level in tumors with low RPV in the two independent datasets (Fig. 3d and Supplementary Figure 7a), which was consistent with our transcriptomic and pathway analysis. Existence of highly proliferative cancer cells with impaired DDR mechanism (e.g. $TP53$ mutation) could elicit accumulation of DNA damage[23]. Accordingly, higher tumor mutational burden and CNA burden were observed in RPV-low tumors (Supplementary Figure 7k-7l). Collectively, these molecular features suggest that RPV-low patients may benefit from DDR inhibitors (PARPi) and immunotherapy (anti-PD1/PD-L1)[24]. Potential alternative therapeutic targets based on the molecular characteristics associated with RPV are listed in Supplementary Table 8.

Molecular subtype, $BRCA1/2$ mutations and $CCNE1$ amplification are well-established molecular characteristics contributing to primary chemotherapy response and prognosis. However, they were not found correlated with RPV, highlighting the independent disease mechanisms associated with RPV (Fig. 3e; Supplementary Note 1).

Overall, stromal phenotype on one hand, and proliferation and DDR pathways on the other, were respectively activated in RPV-high and RPV-low tumors, all of which are potential actionable therapeutic targets in HGSOC.

The reliability and reproducibility of the radiomic profile. To determine the reliability and reproducibility of the radiomic profile, we assessed potential sources of error during radiomic data preparation. Firstly, we assessed the batch effect of CT scanner types on radiomic profile and RPV. Principal component analysis of the radiomic profile for all tumors showed no association at all with the five vendors or two types of matrix and was only moderately
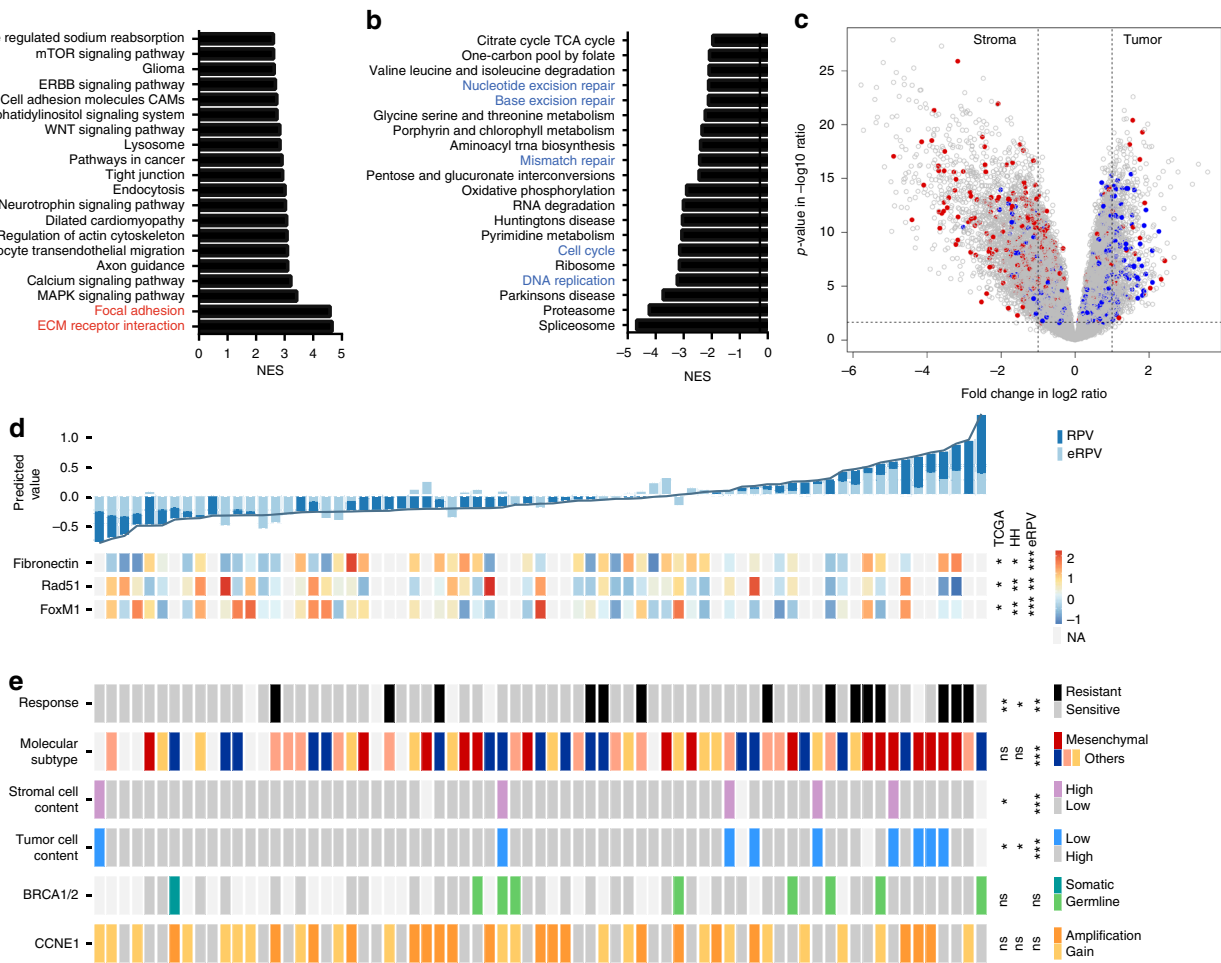
Fig. 3 Molecular characteristics associated with RPV in HGSOC. Gene set enrichment analysis identified **a** RPV-positively correlated biological pathways and **b** RPV-negatively correlated biological pathways from KEGG pathway database (FDR < 0.05). NES normalized enrichment score. **c** Volcano plot showing the differential expressed genes between stroma and tumor epithelial component from a public dataset, GSE40595. The genes that positively correlated with RPV are highlighted in red (r > 0.3, Spearman correlation); genes that are negatively correlated with RPV are highlighted in blue (r < −0.3, Spearman correlation). **d** Heatmap showing correlation of protein expression (Fibronectin, Rad51 and FoxM1) with RPV for 47 cases in the TCGA validation dataset. The significance between these protein features with RPV was indicated for 119 cases in the HH cohort and with eRPV from 353 additional TCGA cases. Top panel, RPV ranked from low to high (left to right) and their corresponding eRPV (light blue). Lower panel, protein expression level of Fibronectin, Rad51 and FoxM1. p values are given by one-sided Spearman's correlation test as validation of the transcriptomic analyses. **e** Clinical, histological and genetic characteristics associated with RPV in the TCGA and HH cohorts. Each rectangle block represents one patient in the TCGA validation dataset. The significance of association between these characteristics with RPV in the TCGA validation dataset, HH cohort and eRPV in additional TCGA dataset is indicated on the right side. The significance is indicated on the right from Kruskal−Wallis test (molecular subtype) or two-tailed Wilcoxon rank-sum test (others). The association between RPV and stromal component is shown in (**a**), (**c**), (**d**) and (**e**); The association between RPV and proliferation or DNA damage response is highlighted in (**b**) and (**d**). ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, ns $p > 0.1$. RPV radiomic prognostic vector, HGSOC high-grade serous ovarian cancer, TCGA the Cancer Genome Atlas, KEGG Kyoto Encyclopedia of Genes and Genomes

associated with the scan thickness (Fig. 4a and Supplementary Figure 12a-12c; Range: 1–10 mm). Thus, RPV is deemed to be unaffected by the types of CT scanner investigated.

To assess the reliability of radiomic data generated, we investigated the feature-wise correlation in HH and TCGA cohorts (Fig. 4b). A consistent feature-wise correlation across independent studies is an indicator of high reliability. The feature-wise correlation in HH and TCGA cohorts were strongly correlated ($r = 0.817$, $p < 0.0001$, Pearson correlation), signifying a relatively consistent structure within the radiomic profile compared with molecular profiles from RPPA (Supplementary Figure 12c).

In the present study, the primary tumors from CT scans were initially segmented by radiologists, then analyzed by the TexLab 2.0 software. The segmentation process could potentially cause

interobserver errors due to the manual nature of the procedure; therefore, we investigated the effect of eight deformations (from −4 to +4 voxels) of the original segmentations on RPV (Fig. 4c). The difference between the deformed and original RPV are shown for each deformation from 106 scans in Fig. 4d. We found that erosion of the segmentation generally amplified the original RPV and dilation had an opposite effect, which resulted in an inverse correlation between the difference in RPV and increase of voxels. Importantly, the variation in RPV was unremarkable within the range of 1-voxel erosion (mean difference: 0.105, sd: 0.167) and 3-voxel dilation (mean difference: −0.0418, sd: 0.125). The interobserver variation in RPV, determined from segmentation made by two independent radiologists for 21 scans, fitted well within this range (Supplementary Figure 12e).
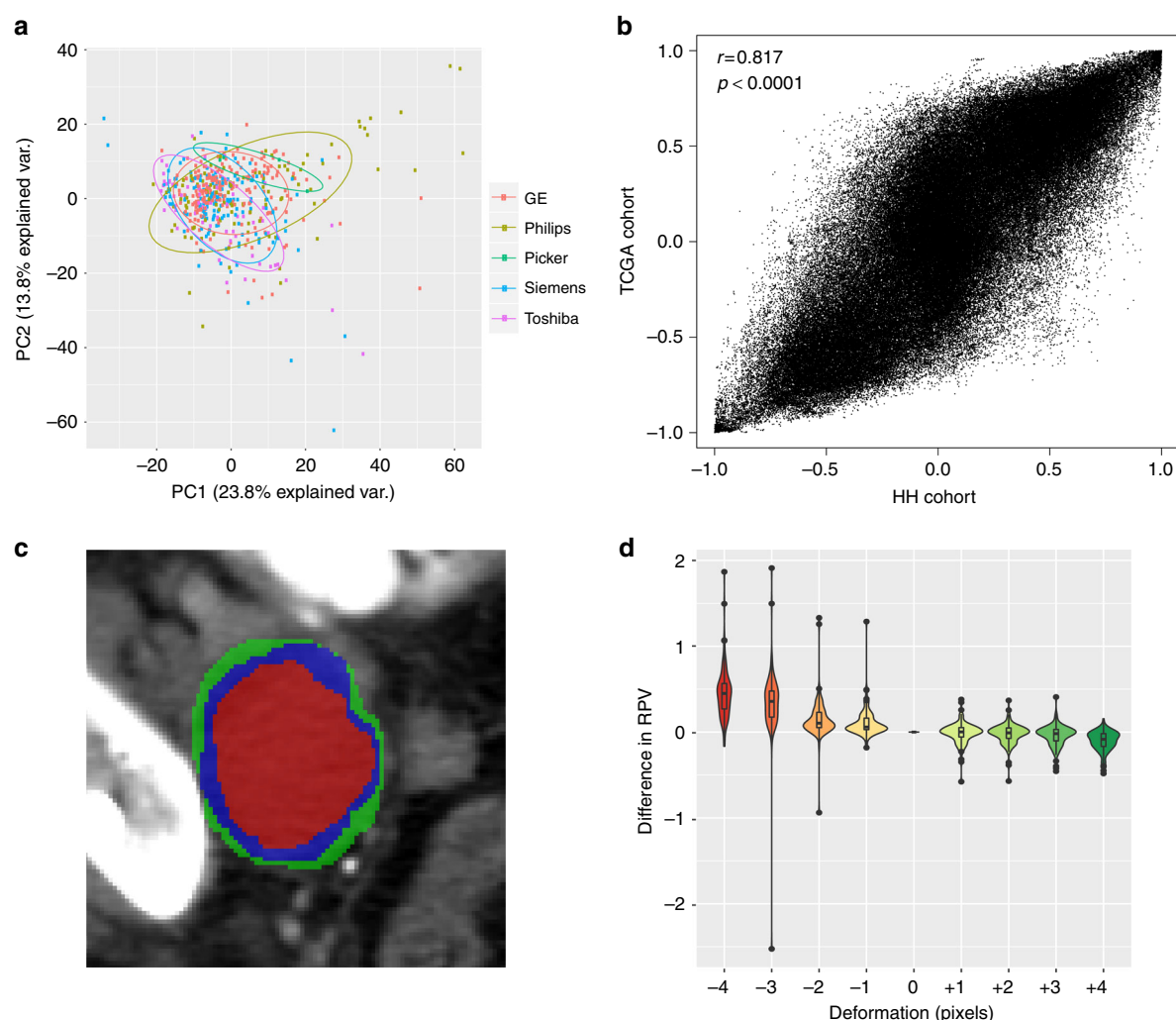
Fig. 4 The reliability and reproducibility of radiomic profile. **a** Principal component analysis (PCA) plot of radiomic profile by scanner manufacturers. **b** Radiomic feature-wise correlations of two independent cohorts. x-axis indicates the correlation between any two radiomic features in the HH cohort, y-axis indicates the corresponding feature-wise correlation in the TCGA cohort. The Pearson correlation coefficient and p value is indicated. **c** An illustration of erosion (red) and dilation (green) on the original segmented CT scan (blue). **d** Impact of segmentation deformations on RPV. x-axis indicates the range of deformations from erosion by 4 voxels (−4) to dilation by 4 voxels (+4) and 0 is the original segmentation. y-axis indicates the difference between deformed RPV and original RPV. HH Hammersmith Hospital, TCGA the Cancer Genome Atlas, CT computed tomography, RPV radiomic prognostic vector

Discussion

In the present study, we obtained and analyzed a comprehensive radiomic profile containing 657 features for 364 EOC cases in total—the largest study of its kind for EOC—and we discovered a novel radiomics-based prognostic signature, RPV, that not only has strong prognostic power (HR > 3), but is also noninvasive and readily accessible, compared to the existing molecular profiles and clinical factors deemed prognostically relevant[12,14]. In contrast to previous studies that lacked interpretation of the prognostic signature, we comprehensively profiled biological and clinical features associated with RPV that will help guide future clinical decision processes in a reliable and reproducible fashion.

Several previous studies have attempted to develop predictive and prognostic tools based on molecular profiles from tumor biopsies such as gene expression, DNA methylation, CNA, and more recently microRNA and circulating tumor DNA[12,14,15,25–27]. These molecular prognostic models are challenging to translate into routine clinical use due to the invasiveness of a biopsy, insufficient prognostic power due to the vast intratumor heterogeneity, high assay costs, and most importantly, the significant time constraints that are associated with the molecular assay

procedures. The prognostic model we propose is simple, built solely on the information extracted from a patient's routine pre-operative CT scan at the presentation of the disease and hence readily accessible without additional costs or time delays, knowing that majority of the HGSOC patients will have CT scans prior to the treatment (compared to PET, MRI or ultrasound). The entire primary ovarian mass is segmented, signifying that any prognostic or biological information extracted is more representative of the disease compared to a single site biopsy. Moreover, RPV is stable across the CT scanner types and the segmentation process, thus limiting the number of potential restrictions for clinical exploitation in the future. We have constructed a software pipeline which is able to compute the RPV of 80 EOC datasets within 5 min on a standard computer. Beyond RPV, the dataset could be mined in a supervised manner for new gene- or protein-radiomics interactions.

We employed manually engineered features as the main component of radiomic analyses; this approach is backed by the current literature in the field[1]. While some studies[28] have started to investigate the application of deep learning for radiomics via the extraction of thousands of deep features from convolutional

networks, the small sample size coupled with the curse of dimensionality in machine learning pose a hurdle for proper evaluation of deep learning features for radiomic analysis. The availability of thousands of annotated or segmented 3D medical datasets would allow for a more robust evaluation and open the possibility for applying transfer learning on 3D medical images, as is currently done with 2D images[29–31].

RPV consists of four radiomic features: (a) FD_max_25HUgl (coefficient: −0.0876), (b) GLRLM_SRLGLE_LLL_25HUgl (coefficient: 0.0869), (c) NGTDM_Contra_HLL_25HUgl (coefficient: 0.165), and (d) FOS_Imedian_LHH (coefficient: 0.250). All the features appear to have approximately even weighting and relate to tumor macro-architecture at the 25 Hounsfield Unit gray level (and discrete wavelet filters). In biological terms, the individual components of RPV combine to define the tumor-mesoscopic structure: (a) maximal fractal dimension of the tumor and its microenvironment, which was negatively correlated with survival, together with the following positively correlated features; (b) proportions of runs that have short lengths in the low pass filtered image; a function which gives coarse low-density textures, e.g. intermixed fibrotic stroma and tumor cells; (c) the average visual contrast across the tumor weighted by sharpening in the x-axis, and blurring in the y and z axes reflecting local heterogeneity, and (d) the median of the distribution of voxel intensities across the entire tumor weighted by blurring in the x-axis and sharpening in the y and z axes, reflecting global heterogeneity, respectively. A visual representation of the four radiomic features is shown in Supplementary Figure 13.

In addition to building a prognostic model, we further demonstrated that the radiomics-derived signature is closely linked to a stromal phenotype and DNA-damage response through genetic, transcriptomic, proteomic and histological analysis. This finding is consistent with the poor prognostic value of cancer stromal phenotype identified in many cancers including ovarian[32–36], pancreatic[37], prostate[38], colorectal[35,39], gastric[32], lung[35] and breast cancer[40]. Tumor stroma consists of immune cells, endothelial cells, fibroblasts and extracellular-matrix (ECM)[41] all of which could directly contribute to outcome via distinct mechanisms in EOC[42–45]. We demonstrate, based on the strong association between RPV and response to primary chemotherapy or surgery, that patients with high RPV have a significantly high risk of failing quality surgery or systemic strategies and suggest that they possibly need to be directed towards alternative therapeutic approaches including stroma modifying therapies (e.g. ClinicalTrials.gov ID: NCT03363867).

Interestingly, in our HGSOC cohort we did not observe a strong association between RPV and any single cancer driver events including ovarian cancer "molecular subtypes", specific gene mutations or CNA, suggesting that the RPV phenotype and related poor prognosis may be shaped by noncanonical genetic alterations or pathways.

There are some limitations of the present study: firstly, the study had a retrospective design albeit with two independent validation datasets. A future prospective study or analysis of retrospective randomized clinical trial data is required to validate RPV in a more general HGSOC population. Secondly, as the stromal component contains a mixture of cells of different origins and ECM composition, the exact elements in the stroma measured by RPV remain unclear. A study to associate RPV with each component in stroma including fibroblast activation, immune cell infiltration and ECM density is necessary to better understand the basis of the prognostic power of RPV. In addition, EOC patients often present with bilateral disease and one tumor was chosen to represent the patient in this study. Further investigation into heterogeneity in RPV for bilateral tumors may further help optimize the prognostic model.

In summary, we have discovered and validated a novel mathematical descriptor of tumor phenotype and prognosis that convincingly fulfills an unmet need in the management of patients with EOC, and have demonstrated a disruptive technology that opens the way for multiple classifications of patients and rapid patient entry into clinical trials at the point of care.

Methods

Patient cohort and biospecimen collection. This is an observational study of patient data (including data related to fresh frozen tissue, imaging and clinical annotations) from the Hammersmith Hospital (HH), Imperial College Healthcare NHS Trust and from the TCGA study. All procedures involving human participants were done in accordance with the ethical standards of the institutional and/or national research committee and with the principles of the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. Ethical approval for retrospective analysis of human data was obtained under the Hammersmith and Queen Charlotte's & Chelsea Research Ethics Committee approval 05/QO406/178 and informed consent was waived, typical for retrospective analysis of anonymized imaging data.

EOC patients included in the Hammersmith cohort were treated at the Hammersmith Hospital (HH), Imperial College London NHS Trust between June 2004 and November 2015. The patients were identified based on the availability of fresh frozen tumor tissue samples and preoperative CT images.

Patient demographics, surgical and tumor related data were collected retrospectively from medical records and the multidisciplinary team (MDT) notes by the clinical members of the team are summarized in Supplementary Table 1. PFS and OS were defined as the time from the date of surgery until the date of first relapse or death, respectively. Staging was defined according to FIGO-criteria for ovarian epithelial carcinoma. Optimal debulking was defined by postoperative residual disease < 10 mm since this criterion was applied to majority of the retrospective patients. Primary chemotherapy resistance was defined as stable disease, a partial response or progressive disease during the first-line chemotherapy.

Tumor cellularity was quantified from hematoxylin and eosin-stained sections by an experienced pathologist. Based on the multidimensional scaling analysis we performed on the RPPA data, only samples with more than 30% tumor cellularity were included in the RPPA analysis.

A subset of EOC patients from TCGA study were used as the validation cohort. The preoperative CT images for these cases were downloaded from the cancer imaging archive[46] (http://www.cancerimagingarchive.net/). This was a multicenter cohort with patients originating from Memorial Sloan Kettering (30 cases), Mayo Clinic—Rochester (4 cases), University of Pittsburgh (10 cases), UCSF (16 cases) and Washington University (9 cases). The clinical and histological data were downloaded from UCSC cancer browser (https://genome-cancer.ucsu.edu/).

Clinical and surgical pathways. The management of all patients and the indications for surgery were discussed within a multidisciplinary team as per the UK National Health Service (NHS) guidelines. All operations were performed through a midline laparotomy by a specialized dedicated multidisciplinary team within a maximal effort approach aiming to achieve total macroscopic tumor clearance. Standard surgical procedures included peritoneal cytology, extrafascial hysterectomy, bilateral salpingoophorectomy and infra-gastric omentectomy. When indicated, additional procedures, such as dissection of macroscopically suspicious pelvic and paraaortic lymph nodes, bowel resection, splenectomy, diaphragmatic stripping/resection and/or partial resection of other affected organs (e.g. urinary bladder, liver/liver capsule, pancreas, lesser sack) were performed in order to achieve optimal tumor debulking. No systematic pelvic and paraaortic lymph node dissection was performed routinely in the absence of suspicious bulky lymph nodes (<1 cm).

Ninety-seven percent of patients were treated with a platinum-based chemotherapy mainly in a combination regimen with paclitaxel or as monotherapy in isolated cases.

Clinical follow-up of patients. Patients were regularly evaluated at the end of their treatment for evidence of disease recurrence. Clinical examination and CA-125 assessment (if the preoperative value was elevated) were performed every 3 months for the first 2 years and then 6-monthly. A CT/MRI-scan was ordered if the above examinations revealed any pathology. An isolated CA-125 increase was not regarded as a recurrence.

CT segmentation and radiomic analysis. As patients were referred to the cancer center from a network of cancer units, contrast-enhanced CT scans were acquired at multiple institutions using different manufacturers and different imaging protocols.

For both the HH and the TCGA datasets, the primary tumor masses were segmented separately by experienced radiologists (M.A., G.A.) using ITK snap (Version 3.2, 2015) and then all segmentations were checked in consensus with a radiologist with over 16 years' experience of ovarian cancer imaging (A.R.). We included the entire primary tubo-ovarian mass (cystic and solid components). If both

adnexae were involved, then both were included in the analysis, either as two separate segmentations or as a single segmentation if the mass was confluent. We segmented the entire primary mass including cystic and solid components, but excluded ascites. The segmentations only included tissue that was considered highly likely to be cancer by the expert reader. Areas of doubt on CT were not included in any segmentations. Inter-observer variation was also measured by comparing independent segmentation from two radiologists using the TCGA cohort.

For this study, the primary tumor mass segmentations were used as input for the in-house texture analysis software package (TextLAB 2.0) developed in MATLAB 2015b (Mathworks Inc., Nathick, Massachusetts, USA)[16].

Using methodologies for feature extractions[1,47–53], we defined 657 radiomic image features that describe tumor characteristics. The features can be divided into several groups: 1. Shape and Size features; 2. First-order statistics; 3. Second-order statistics; 4. Wavelet features.

The first group relates to statistics based on the shape of the tumor, e.g. compactness or sphericity. The second group quantified tumor voxel intensity characteristics. Group 3 consists of textural features that quantify different measures of three-dimensional intratumoral heterogeneity. The wavelet features group calculates the features in groups 2 and 3 after performing wavelet decompositions of the original image using high-pass or low-pass filters from the coiflet 1 family of wavelets. All feature algorithms were implemented within MATLAB.

Transcriptomic, proteomic and copy-number analysis. Frozen tumor tissue pieces ($n = 314$) were placed into ceramic bead tubes (Stretton Scientific) for protein extraction by the Functional Proteomics RPPA Core Facility, MD Anderson, USA. Protein concentration was determined following extraction and adjusted to 1.5 µg/µl. Proteins were denatured by 1% SDS plus beta-Mercaptoethanol and serially diluted for subsequent Reverse Phase Protein arrays.

For each frozen tumor in the study, one frozen tumor piece was placed into a tube containing 500 µl RLT buffer from RNeasy kit (QIAGEN) and one Retsch 6 mm steel core bead. Tubes were placed into well adapters of a Tissuelyser II (QIAGEN) and tissues were lysed at 15 Hz for 2 min. Tubes were centrifuged briefly and 320 µl was removed for subsequent RNA extraction using the RNeasy kit (QIAGEN) according to the manufacturer's instructions. RNA concentrations were quantified using the Bioanalyzer system (Agilent).

For DNA extraction, 450 µl of Buffer ATL from the QIAAMP DNA kit (QIAGEN) was added to the centrifuge tube, and DNA was extracted following the manufacturer's instructions and quantified using QuBit (Thermo Fisher Scientific).

RPPA arrays were carried out and analyzed by MD Anderson Cancer Center[54]. Briefly, protein lysates were diluted and loaded onto nitrocellulose-coated slides that had been preconjugated with primary antibodies. Each protein was then visualized via a colorimetric reaction and quantified by Array-Pro Analyzer. The raw expression values were then normalized to protein loading and quantified by means of standard curves. Log2 transformed and median-centered data were used for the downstream analyses.

To perform molecular subtyping, total RNA from each individual case was reverse transcribed into cDNA, followed by amplification with a pool of indexed primers that target a predefined gene list (42 genes)[13]. The primers were selected from the Illumina DesignStudios. The cleaned PCR product underwent QC by Tapestation (Agilent) to confirm the amplicon sizes. Forty-eight samples were multiplexed in one single MiSeq run. SR 50 bp were used to generate approximately 20 million reads per run.

Copy-number estimates for *CCNE1* in 131 tumor samples from the HH cohort were obtained through quantitative qPCR[10]. ΔCt values for tumor samples (*CCNE1* relative to the endogenous control *LINE1*) were normalized to equivalent ΔCt values from reference (normal Fallopian tube cell line DNA) with an assumed *CCNE1* copy number of 2.

Unsupervised clustering and signature discovery. A simple spectral cluster analysis was performed using radiomic data from patients with both radiomic and genomic copy-number data available. First a similarity measure was computed for each patient with radiomic and CNA profiles as the average Spearman correlation coefficient. The profiles of pair-wise similarity were then used to compute the Euclidean distance between each pair of patients. Visual inspection of a hierarchical clustering dendrogram was used to select three clusters of patients, so that patients from a given cluster tended to share correlated radiomic profiles.

The number of genes affected by CNA was calculated for each tumor sample, so that the distribution of the logarithm of these numbers could be compared for tumors from patients belonging to different clusters.

Kaplan−Meier curves were drawn for PFS and OS using the "survfit" function from the "survival" package in R. The statistical significance of the difference in these survival measures across the three patient clusters was calculated using the log-rank test implemented in the "survdiff" function.

Unsupervised hierarchical clustering of radiomic profiles were performed using hclust and cutree function in R 3.3.1. The raw radiomic data were firstly scaled by mean and centered. Pearson correlation-based distance and complete linkage was used to obtain the final clusters indicated in Fig. 1e. The resulting clusters were confirmed by repeating the clustering analysis using Euclidean's distance. The optimization of the radiomic clusters is indicated in Supplementary Figure 14. The heatmap was generated using the "heatmap.plus" package in R 3.3.1.

Least absolute shrinkage and selection operator (LASSO) analysis was performed to build a prognostic model for OS using radiomic data. We first selected for the discovery dataset (HH discovery) a group of HGSOC patients who had primary debulking surgery as well as patients not in the unsupervised subgroup 2 (which had different slice thickness compared to other subgroups, Supplementary Figure 14). All the other HGSOC patients were used as the HH validation dataset and the HGSOC patients in the TCGA cohort were used as the TCGA validation dataset. We selected a large pilot dataset for discovery (HH discovery, $n = 136$) and the number of patients in the two validation datasets fulfilled sample size estimate (73 cases needed after accepting the alpha of 0.05 and beta of 0.25. HR of 2.78, 31.6% cases in the high-risk group, median survival of 5 years in the low-risk group and median follow-up as 5 years). To generate a prognostic model of OS, a univariate Cox regression was performed between individual radiomic features and OS, which was adjusted for stage, slice thickness and residual disease in the HH discovery dataset. Since it was not possible to decide the more prognosis-related tumor for bilateral tumors and we had demonstrated close similarity between the two tumors, we included both bilateral tumors at the model-building stage. The radiomic features with FDR < 5% were selected as input for LASSO regression using glmnet package in R 3.3.1. "Cox" was set as the family in the model. Ten-fold cross-validation was performed using cv.glmnet function to select lambda minimum to give the minimum cross-validated error. The resulting four radiomic features with coefficients were used to calculate a predictive index—RPV—for each patient. The RPV was used to perform subsequent continuous Cox regression and Kaplan−Meier analysis with OS and PFS. For patients with bilateral tumors, the tumor that gave the higher RPV was selected since it resulted in better performance compared with the one with lower RPV. After considering the distribution of RPV and number of patients in each subgroup, K-means clustering was applied to split the patients into three subgroups (low risk: min−0.0950, medium risk: 0.0950–0.658, high risk: 0.658−max). The same criteria were used to obtain subgroups in the validation cohorts. For validation, the radiomic data from the TCGA dataset and the HH validation dataset were initially scaled and centered. The RPV was calculated using the four radiomic features with coefficients derived from the discovery set. For those cases with bilateral tumors which resulted in two RPV values, the higher RPV was selected for the survival analysis. For multiple Cox regression of RPV, the slice thickness was introduced as additional variable. Only cases with complete clinical information (stage, age and postoperative residual disease) and slice thickness were included in the multivariable Cox regression analysis. REMARK guidelines were followed when reporting RPV as a prognostic marker in this study[55].

Similar procedure was applied to generate eRPV with some modifications. Gene expression profile from Affymetrix HT Human Genome U133a (Level 2) and Agilent 244K custom gene expression G4502A_07_3 (Level 3) from the TCGA study were downloaded from UCSC cancer browser (https://genome-cancer.ucsc.edu/) and TCGA data portal (https://cancergenome.nih.gov/). Spearman correlation was applied to obtain a list of genes correlated with RPV (FDR < 0.25 for Affymetrix and FDR < 0.1 for Agilent). The gene list obtained was used to perform feature selection and linear regression with RPV using "glmnet" package. "gaussian" was set as the family in the model and a tenfold cross-validation was applied. The resulting weighted gene lists contributing to eRPV are given in Supplementary Data 5 and 6.

Gene set enrichment analysis (GSEA) was performed for RPV-correlated genes in the TCGA dataset. The Level 3 RNA-sequencing dataset of EOC from the TCGA project was downloaded from UCSC cancer browser (https://genome-cancer.ucsc.edu/) and the gene-level transcription estimates were obtained in reads per kilobase million. Spearman correlation coefficient was determined from RPV and all the genes respectively. The full list of correlation coefficients was used as the pre-ranked list in GSEA 2.1.0 with KEGG database, 1000 of permutations and classic enrichment statistic.

Differential gene expression between tumor and stroma in HGSOC was analyzed using limma package from Bioconductor in R 3.3.1. The Robust Multi-array Average (RMA) normalized microarray dataset (GSE40595) was downloaded from GEO (https://www.ncbi.nlm.nih.gov/geo/). The empirical Bayes moderated t-statistics were computed comparing gene expression from HGSOC stroma and tumor epithelial component. The p value derived was adjusted for multiple testing using Benjamini–Hochberg procedure.

Statistical analysis. Standard statistical analysis was applied to all the figures as appropriate and indicated in the figure legends. All samples were used once. Multiple testing was corrected with FDR method. All the statistical analyses were conducted in R 3.3.1.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Code availability. The R script that was used to reproduce the key findings and generate figures are publically accessible in Mendeley Data with the identifier https://doi.org/10.17632/4c5znk5m8t.1.

Data availability

The radiomics, clinical, RNA-sequencing and proteomics data generated in this study have been deposited into the Mendeley database under the accession code:

References

1. Aerts, H. J. W. L. et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* **5**, 4006 (2014).
2. Aerts, H. J. The potential of radiomic-based phenotyping in precision edicine: a review. *JAMA Oncol.* **2**, 1636–1642 (2016).
3. Segal, E. et al. Decoding global gene expression programs in liver cancer by noninvasive imaging. *Nat. Biotechnol.* **25**, 675–680 (2007).
4. Nougaret, S. et al. High-grade serous ovarian cancer: associations between BRCA mutation status, CT imaging phenotypes, and clinical outcomes. *Radiology* **285**, 472–481 (2017).
5. Vargas, H. A. et al. Association between morphologic CT imaging traits and prognostically relevant gene signatures in women with high-grade serous ovarian cancer: a hypothesis-generating study. *Radiology* **274**, 742–751 (2015).
6. Heinzmann, K., Carter, L. M., Lewis, J. S. & Aboagye, E. O. Multiplexed imaging for diagnosis and therapy. *Nat. Biomed. Eng.* **1**, 697–713 (2017).
7. Agarwal, R. & Kaye, S. B. Ovarian cancer: strategies for overcoming resistance to chemotherapy. *Nat. Rev. Cancer* **3**, 502–516 (2003).
8. Vaughan, S. et al. Rethinking ovarian cancer: recommendations for improving outcomes. *Nat. Rev. Cancer* **11**, 719–725 (2011).
9. Cancer Genome Atlas Research, N. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008).
10. Etemadmoghadam, D. et al. Integrated genome-wide DNA copy number and expression analysis identifies distinct mechanisms of primary chemoresistance in ovarian carcinomas. *Clin. Cancer Res.* **15**, 1417–1427 (2009).
11. Patch, A. M. et al. Whole-genome characterization of chemoresistant ovarian cancer. *Nature* **521**, 489–494 (2015).
12. Verhaak, R. G. et al. Prognostically relevant gene signatures of high-grade serous ovarian carcinoma. *J. Clin. Invest.* **123**, 517–525 (2013).
13. Leong, H. S. et al. Efficient molecular subtype classification of high-grade serous ovarian cancer. *J. Pathol.* **236**, 272–277 (2015).
14. Tothill, R. W. et al. Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clin. Cancer Res.* **14**, 5198–5208 (2008).
15. Bagnoli, M. et al. Development and validation of a microRNA-based signature (MiROvaR) to predict early relapse or progression of epithelial ovarian cancer: a cohort study. *Lancet Oncol.* **17**, 1137–1146 (2016).
16. Arshad, M. A. et al. Discovery of pre-therapy 2-deoxy-2-(18)F-fluoro-D-glucose positron emission tomography-based radiomics classifiers of survival outcome in non-small-cell lung cancer patients. *Eur. J. Nucl. Med. Mol. Imaging* **46**, 455–466 (2019).
17. Cancer Genome Atlas Research, N. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).
18. Tibshirani, R. The lasso method for variable selection in the Cox model. *Stat. Med.* **16**, 385–395 (1997).
19. Harrell, F. E. Jr., Lee, K. L. & Mark, D. B. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat. Med.* **15**, 361–387 (1996).
20. Finak, G. et al. Stromal gene expression predicts clinical outcome in breast cancer. *Nat. Med.* **14**, 518–527 (2008).
21. Malhotra, D. et al. Transcriptional profiling of stroma from inflamed and resting lymph nodes defines immunological hallmarks. *Nat. Immunol.* **13**, 499–510 (2012).
22. Yoshihara, K. et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612 (2013).
23. Gaillard, H., Garcia-Muse, T. & Aguilera, A. Replication stress and cancer. *Nat. Rev. Cancer* **15**, 276–289 (2015).
24. Rizvi, N. A. et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348**, 124–128 (2015).
25. Wei, S. H. et al. Prognostic DNA methylation biomarkers in ovarian cancer. *Clin. Cancer Res.* **12**, 2788–2794 (2006).
26. Macintyre, G. et al. Copy-number signatures and mutational processes in ovarian carcinoma. *bioRxiv* https://doi.org/10.1101/174201 (2017).
27. Parkinson, C. A. et al. Exploratory analysis of TP53 mutations in circulating tumour DNA as biomarkers of treatment response for patients with relapsed high-grade serous ovarian cancer: a retrospective study. *PLoS Med.* **13**, e1002198 (2016).
28. Lao, J. et al. A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. *Sci. Rep.* **7**, 10353 (2017).
29. Esteva, A. et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115–118 (2017).
30. De Fauw, J. et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat. Med.* **24**, 1342–1350 (2018).
31. Titano, J. J. et al. Automated deep-neural-network surveillance of cranial images for acute neurologic events. *Nat. Med.* **24**, 1337–1341 (2018).
32. Busuttil, R. A. et al. A signature predicting poor prognosis in gastric and ovarian cancer represents a coordinated macrophage and stromal response. *Clin. Cancer Res.* **20**, 2761–2772 (2014).
33. Zhang, S. Z. et al. Stroma-associated master regulators of molecular subtypes predict patient prognosis in ovarian cancer. *Sci. Rep.-Uk* **5**, 16066 (2015).
34. Chen, P. et al. Identification of prognostic groups in high-grade serous ovarian cancer treated with platinum-taxane chemotherapy. *Cancer Res.* **75**, 2987–2998 (2015).
35. Chen, J. L. Y. et al. Stromal responses among common carcinomas correlated with clinicopathologic features. *Clin. Cancer Res.* **19**, 5127–5135 (2013).
36. Bonome, T. et al. Expression profiling of serous low malignant potential, low-grade, and high-grade tumors of the ovary. *Cancer Res.* **65**, 10602–10612 (2005).
37. Wang, L. M. et al. The prognostic role of desmoplastic stroma in pancreatic ductal adenocarcinoma. *Oncotarget* **7**, 4183–4194 (2016).
38. Mo, F. et al. Stromal gene expression is predictive for metastatic primary prostate cancer. *Eur. Urol.* **73**, 524–532 (2018).
39. Isella, C. et al. Stromal contribution to the colorectal cancer transcriptome. *Nat. Genet.* **47**, 312–319 (2015).
40. Frings, O. et al. Prognostic significance in breast cancer of a gene signature capturing stromal PDGF signaling. *Am. J. Pathol.* **182**, 2037–2047 (2013).
41. Kalluri, R. & Zeisberg, M. Fibroblasts in cancer. *Nat. Rev. Cancer* **6**, 392–401 (2006).
42. Mhawech-Fauceglia, P. et al. Stromal expression of fibroblast activation protein alpha (FAP) predicts platinum resistance and shorter recurrence in patients with epithelial ovarian cancer. *Cancer Microenviron.* **8**, 23–31 (2015).
43. Davidson, B., Trope, C. G. & Reich, R. The role of the tumor stroma in ovarian cancer. *Front. Oncol.* **4**, 104 (2014).
44. Sherman-Baust, C. A. et al. Remodeling of the extracellular matrix through overexpression of collagen VI contributes to cisplatin resistance in ovarian cancer cells. *Cancer Cell* **3**, 377–386 (2003).
45. Wang, W. et al. Effector T cells abrogate stroma-mediated chemoresistance in ovarian. *Cancer Cell.* **165**, 1092–1105 (2016).
46. Clark, K. et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J. Digit Imaging* **26**, 1045–1057 (2013).
47. Kristensen, V. N. et al. Principles and methods of integrative genomic analyses in cancer. *Nat. Rev. Cancer* **14**, 299–313 (2014).
48. Jena, R. et al. A novel algorithm for the morphometric assessment of radiotherapy treatment planning volumes. *Br. J. Radiol.* **83**, 44–51 (2010).
49. Cook, G. J. R. et al. Are pretreatment F-18-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).
50. Cheng, N. M. et al. Textural features of pretreatment F-18-FDG PET/CT images: prognostic significance in patients with advanced T-stage oropharyngeal squamous cell carcinoma. *J. Nucl. Med.* **54**, 1703–1709 (2013).
51. Amadasun, M. & King, R. Textural features corresponding to textural properties. *IEEE T. Syst. Man Cyb.* **19**, 1264–1274 (1989).
52. Cheng, N. M. et al. Zone-size nonuniformity of F-18-FDG PET regional textural features predicts survival in patients with oropharyngeal cancer. *Eur. J. Nucl. Med. Mol. I* **42**, 419–428 (2015).
53. Hatt, M. et al. F-18-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.* **56**, 38–44 (2015).
54. Tibes, R. et al. Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells. *Mol. Cancer Ther.* **5**, 2512–2521 (2006).
55. McShane, L. M. et al. Reporting recommendations for tumor marker prognostic studies (REMARK). *J. Natl. Cancer Inst.* **97**, 1180–1184 (2005).
56. Bell, D. et al. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).

57. Yeung, T. L. et al. TGF-beta modulates ovarian cancer invasion by upregulating CAF-derived versican in the tumor microenvironment. *Cancer Res.* **73**, 5016–5028 (2013).

Acknowledgements

Author contributions

E.O.A., A.R., H.G. and C.F. conceived and designed the project. A.T. developed TexLab 2.0. M.A., G.A., J.L. and A.R. collected and segmented the radiological data. A.T., K.N., S.T.W., M.A.H., H.L. and C.F. collected the clinical data. H.L., P.C. and J.L. collected the RNA-sequencing, RPPA and qPCR data. H.L., E.C. and F.K. provided bioinformatics and computational analysis. E.O.A., H.L., M.A., A.T., P.C., E.C., D.D.L.B., H.G., C.F. and A.R. contributed to the interpretation of data. H.L. and E.O.A. wrote the manuscript. All authors edited the manuscript.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-08718-9.

Competing interests: H.G. is an employee of AstraZeneca. The other authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Journal peer review information: *Nature Communications* thanks Gordon Jayson and the other anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.