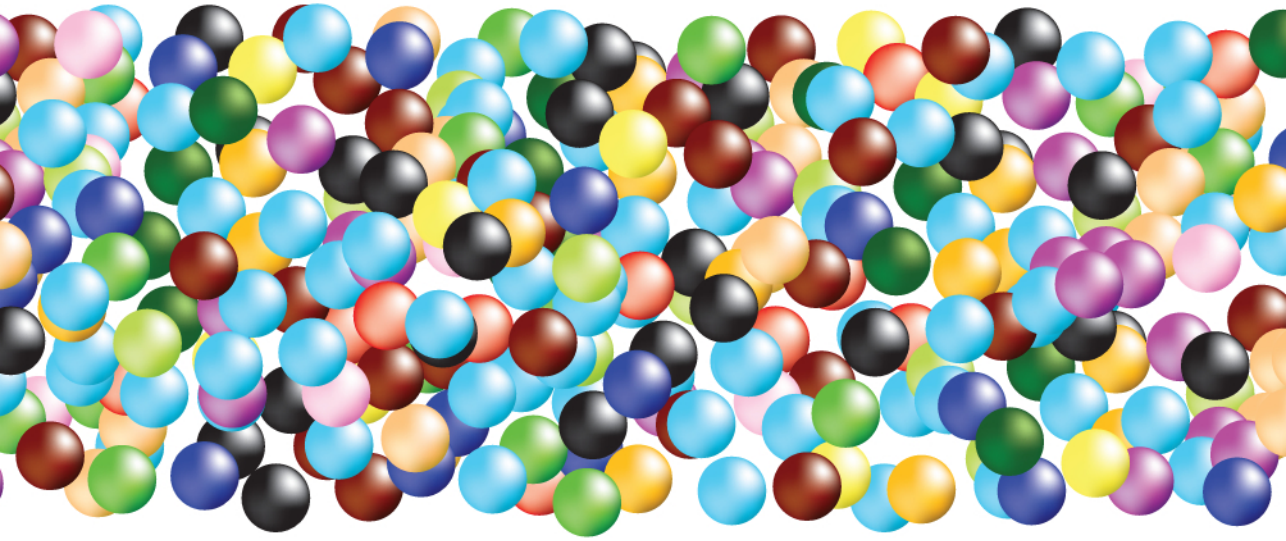
QUANTIFICATION OF TUMOUR HETEROGENEITY IN MRI



LEJLA ALIĆ

QUANTIFICATION OF TUMOUR HETEROGENEITY IN MRI

LEJLA ALIĆ

QUANTIFICATION OF TUMOUR HETEROGENEITY IN MRI

KWANTIFICATIE VAN HETEROGENITEIT IN TUMOREN MET MRI

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam op gezag van de Rector Magnificus

Prof.dr. H.G. Schmidt

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
dinsdag 18 juni 2013 om 13:30 uur

door
Lejla Alić
geboren te Zvornik, Bosnië & Herzegovina

ERASMUS UNIVERSITEIT ROTTERDAM

Promotiecommissie

Promotor: Prof.dr. W.J. Niessen

Overige leden: Prof.dr. Sir M. Brady
Prof.dr.ir. M. de Jong
Prof.dr.ir. B.P.F. Lelieveldt

Copromotor: Dr. J.F. Veenland

Quantification of Tumour Heterogeneity in MRI
Lejla Alić
PhD thesis, Erasmus University Rotterdam, the Netherlands.

ISBN: 978-90-8891-630-4
Cover design by: A.W. Everaers and L. Alić
Printed by: *www.proefschriftmaken.nl*
Published by: *Uitgeverij BoxPress*

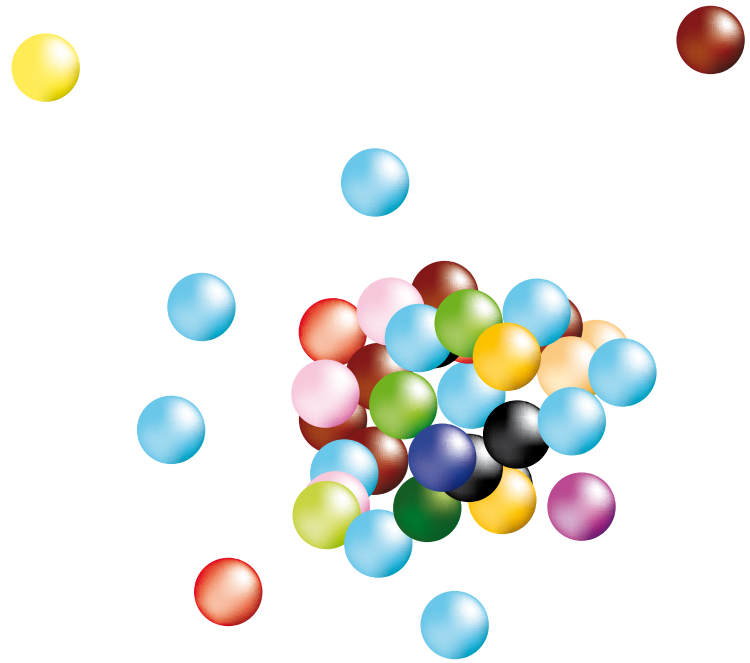
Contents

The research in this thesis was conducted at the Departments of Radiology and Medical Informatics of the Erasmus MC, University Medical Center, Rotterdam, The Netherlands. The research was partly supported by the Netherlands Organization for Scientific Research (*NWO*), under grant number 017.002.019, and by the Dutch Cancer founds (*KWF*) under grant number 2008-4037.

The work was carried out in the *ASCI graduate school*.
ASCI dissertation series number: 276

The research for this thesis was performed within the framework of the *Erasmus Postgraduate School Molecular Medicine.*

Financial support for the publication of this thesis was kindly provided by the Department of Radiology, the Erasmus University, and the ASCI graduate school.



Chapter 1

General introduction

Today's scientists have substituted mathematics for experiments, and they wander off through equation after equation, and eventually build a structure which has no relation to reality.

Nikola Tesla

Cancer epidemiology

Cancer is the leading cause of death that touches us all, either directly or indirectly. It is estimated that the number of newly diagnosed cases in the Netherlands will increase to 123,000 by the year 2020 [1]. General Dutch statistics are similar to those in the UK, i.e. over the last ten years, the age-standardised incidence rate[1] has stabilised at around 355 females and 415 males per 100,000. Figure 1 shows the cancer incidence per gender [2, 3]. In the UK, the rise in lifetime risk of cancer[2] is more than one in three [4] and depends on many factors, including age, lifestyle and genetic makeup.

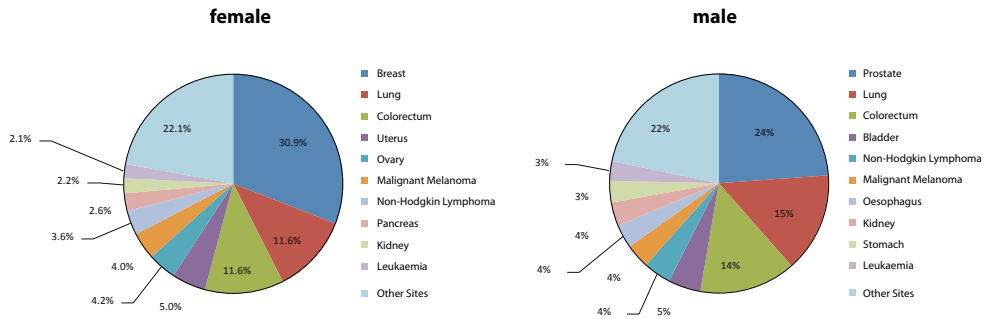


Figure 1. Cancer incidence for the most common cancer types in the EU [2] (left: female population, right: male population).

The average age at the time of diagnosis is 67 years and about 75% of all cancers are diagnosed at an age above 55 [5]. Moreover, with a steadily ageing population in the western world, the absolute numbers of cancer deaths will continue to increase steadily [1]. Forecasted worldwide demographic changes imply that, by the year 2030, the number of people with cancer will probably increase to more than 20 million per year [6].

Current treatment options are chemotherapy, radiation therapy, surgery, hyperthermia, gene therapy, immunotherapy, hormone therapy, and anti-angiogenic therapy. The probable success of these treatment options is highly dependent on the cancer

1 Age-standardisation adjusts rates to take into account how many old or young people are in the population under investigation. When rates are age-standardised the differences in the rates over time, or between geographical areas, do not simply reflect variations in the age structure of the populations. This is important when looking at cancer rates because cancer is a disease that predominantly affects the elderly. If cancer rates are not age-standardised, a higher rate in one country is likely to reflect a greater proportion of older people.

2 The lifetime risk (cancer) is the estimated risk that a newborn will develop cancer at some point during its life. It is based on current incidence and mortality rates and is therefore calculated under the assumption that the current rates (at all ages) will remain constant during the life of the newborn.

type, as shown by Figure 2. Compared to the incidence rate, the mortality is low for breast and prostate cancer (Figure 2: first row), whereas for lung cancer the mortality rate is very high compared to the incidence rate (Figure 2: second row).

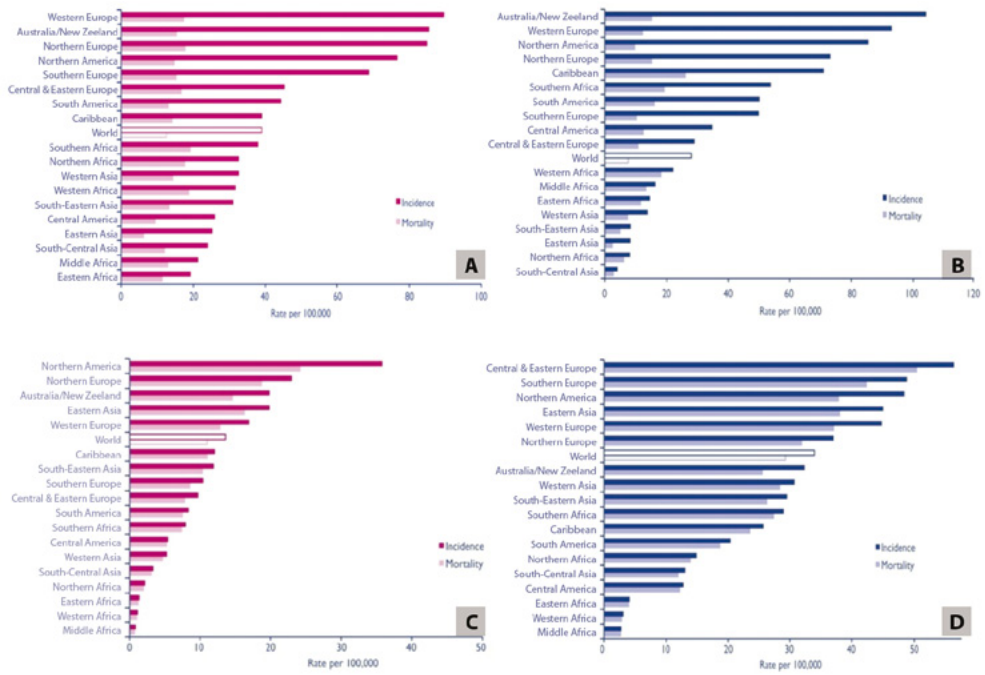


Figure 2. The age-standardised incidence and mortality rates in 2008 per region of the world [2]. This shows that the two most common cancer types for females are breast (A) and lung cancer (C), and for males are prostate (B) and lung cancer (D).

Apart from developing new treatment options, the overall mortality rate could potentially be reduced by primary prevention strategies[3], the implementation of vaccination programmes (for liver and cervical cancer), and early detection programmes (for colorectal, breast, and cervical cancer) [6]. Additional reductions in mortality might be accomplished by increasing access to curative treatment for specific cancer types and by personalizing treatment aiming at specific cancer characteristics.

3 Lifestyle factors: e.g. cessation of smoking, reduction of alcohol consumption, reduction of obesity, increasing physical activity.

Cancer biology

Even though historians disagree about the precise dating of the first description of cancer, there is no doubt that it goes back (at least) to the ancient Greeks [7]. Cancer is defined as an abnormal growth of cells caused by multiple changes in gene expression leading to a deregulated balance of cell proliferation and cell death and, ultimately, evolving into uncontrollable growth and spread of abnormal cells to distant sites [5]. This growth starts by mutations (changes in DNA) that specifically affect genes, initiating unlimited cell growth. Biologically a tumour is a complex system in which distinct populations of cancer cells can interact in a competitive manner [8, 9, 10]. Based on molecular studies, subtypes of the same cancer with large intra-tumour heterogeneity in terms of both biology and response to treatment have been identified [11, 12]. Various types of tumour progression models have been proposed to explain intra-tumour heterogeneity [13, 14, 15], as shown in Figure 3.

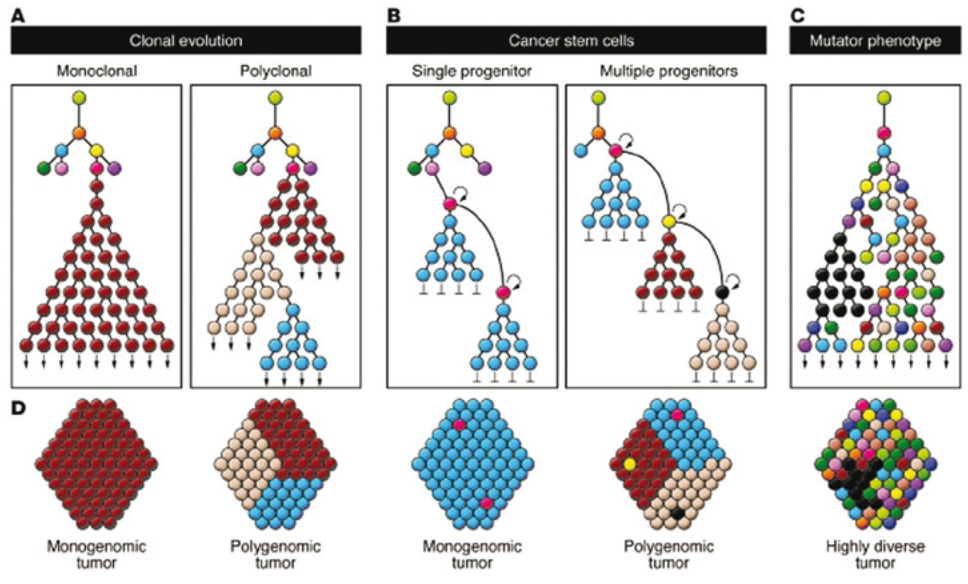


Figure 3. Hypothetical tumour progression models that can explainintra-tumour heterogeneity (A–C): the clonal evolution (A), the cancer stem cell (B), and the mutator phenotype (C) models. The different models result in distinct spatial distributions of cell subpopulations (D).Reprinted with permission from J Clin Invest [15].

The impact of intra-tumour heterogeneity on tumour therapy should not be underestimated [16]. A more heterogeneous tumour is more likely to fail chemotherapy [17]. Multiple cellular subpopulations with different genetic and phenotypic characteristics imply that a specific lesion does not have a single target but multiple oncogenic

targets that must be overcome to achieve optimized therapeutic benefit [18]. Different parts of the tumour can differ in sensitivity to an applied treatment approach. The more aggressive tumour populations, e.g. that proliferate faster, have a higher neo-angiogenesis level or are less sensitive to treatment, will suppress the less aggressive populations and, in this way, evade therapy. The therapy can probably even cause a tumour to become more aggressive by the addition of new mutations and clonal evolution [19]. Therefore, heterogeneity of the tumour and the changes due to treatment should be closely monitored. This is currently facilitated by the rapid development of technologies allowing for *in vivo* and non-invasive tumour examination.

Cancer imaging

Cancer imaging is essential in biomedical research, e.g. for drug discovery/development, and for clinical practice including diagnosis, therapy, assessment of treatment response, and prediction of treatment outcome. Deployment of imaging for drug discovery/development has been discussed in detail elsewhere [20]. This thesis focuses on cancer imaging for diagnosis, treatment monitoring and outcome prediction. Techniques used in cancer imaging include radiography, ultrasound (US) and Doppler imaging, magnetic resonance imaging (MRI), computed tomography (CT), single photon emission tomography (SPECT), positron emission tomography (PET), electron paramagnetic resonance imaging, electromagnetic (EM) imaging, and their variations and combinations [20]. These imaging modalities reflect aspects of the tissue's internal anatomy or a functional aspect of the tissue. Imaging makes it possible to discern a tumour from its environment. In addition, dedicated imaging techniques allow to discriminate between different types of tumours and different stages of the same tumour type.

MRI has a number of distinct advantages for clinical oncology. It has multi-sequence capabilities producing superior contrast among soft tissues, provides full 3D imaging, and does not require ionizing radiation. One of the most important MR techniques in analysing tumour characteristics is dynamic contrast-enhanced MRI (DCE-MRI). DCE-MRI is the acquisition of serial MR images before, during, and after the administration of an intravenous contrast agent. Figure 4A shows the resulting DCE-MRI image series with the time-intensity curve for a particular voxel (Figure 4B). The resulting time-intensity curves can be modelled using pharmacokinetic [21, 22] or heuristic models [23-28], producing parametric maps. Figure 4C presents a pharmacokinetic parametric map (left) and a heuristic parametric map (right).

The suitability of DCE-MRI in combination with different quantification methods to monitor anticancer therapy is undergoing extensive research [29-34]. Correlation with histopathology showed the ability of DCE-MRI parameters to monitor treatment response by identifying areas of residual viable tumour tissue [35]. Parametric maps can be monitored over time, e.g. during the course of therapeutic interventions, to evaluate different anti-angiogenic and antivascular cancer treatments or treatment strategies. The heterogeneity present in the parametric maps, extracted from DCE-MRI, can be quantified using a variety of texture analysis methods [36].

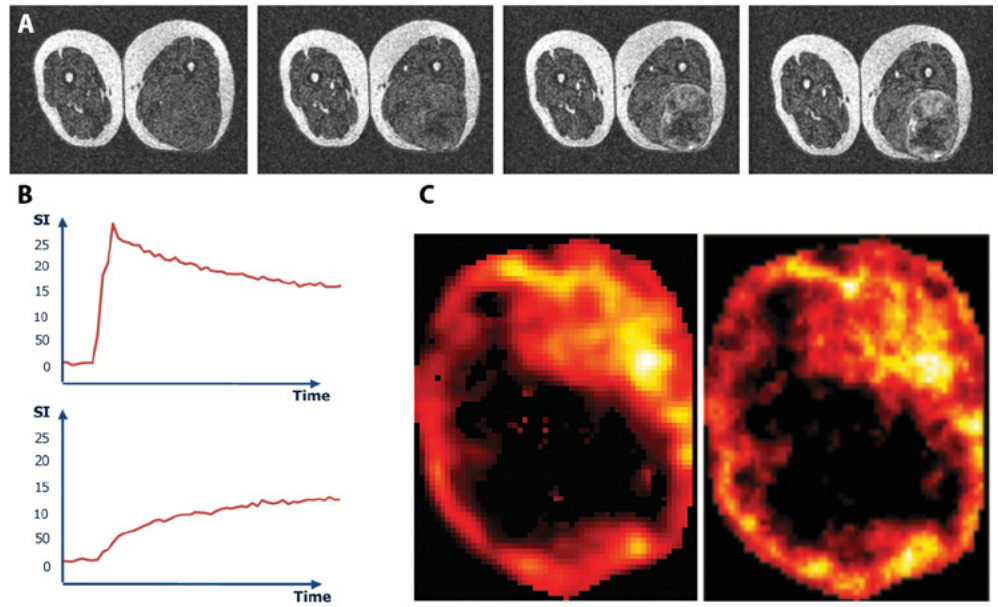


Figure 4. Example of enhancement in a sarcoma (A) and the corresponding time-intensity curves (B) for an enhancing sarcoma region (left) and a muscle region (right). The corresponding parametric map (C) shows the pharmacokinetic (left) and heuristic based (right) parameters.

Texture

Texture is defined as a characteristic intensity variation, which in natural images, for example, often originates from the properties of the object surface. With no formal definition of what a characteristic intensity variation is, this concept can be approached more intuitively. Perfectly-periodic intensity variations are referred to as *periodic pattern.* Similarly, completely random patterns constitute a *noise pattern*. A pattern which shares both properties (randomness and regularity), is what most people would consider a texture. An additional feature of a texture is its *busyness*, i.e. the degree of mix between randomness and regularity. To a certain extent, texture typically express a busy microstructure, but uniform macrostructure

[37]. Additionally, a texture may also vary according to direction, orientation and coarseness. Using these concepts, several authors have provided a systematic description of textual measures [38-40] and have generally divided texture extraction methods into structural and statistical. Structural approaches analyse textures with regular macro-structure, and will not be discussed further in this thesis. Statistical approaches are better suited to characterize micro textures. The first order statistics of the grey-level distribution are often labelled as texture features. However, where an image is represented by one histogram only, the inverse is not true: images with different textures can be characterised by the same histogram (Figure 5). A subsequent step in the analysis of tumour appearance is through texture analysis. Texture descriptors originating from statistical approaches include model-based features (fractals, autoregressive models, fractional differencing models, and Markov random fields), and non-model-based features (co-occurrence matrices, grey-level sum and difference histograms, Laws' masks, frequency domain methods, and Gabor filters).

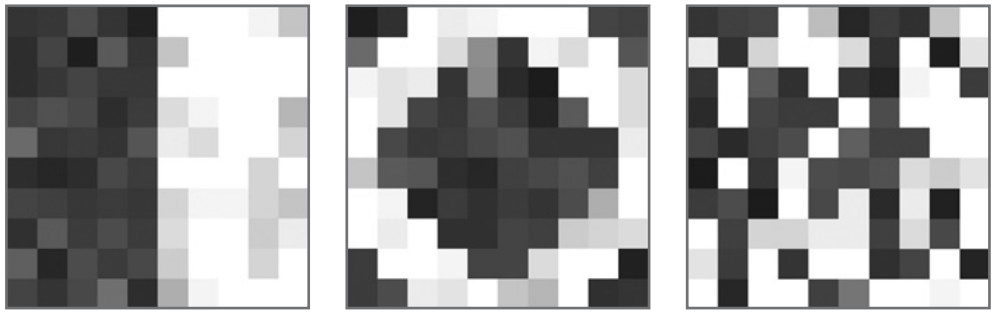


Figure 5. Images with different heterogeneities and similar histograms.

Almost all these texture features can be computed based on radiological images of tumours. The question is: how good are these features in grasping the textural differences between different tumour types and different grades? Is it possible to monitor treatment induced texture changes with these features? Are these texture features related to treatment outcome?

Outline of the thesis

The aim of this thesis is to develop and evaluate tumour heterogeneity quantification techniques and to investigate their importance for tumour treatment monitoring and outcome prediction. In particular, this thesis focuses on the following questions:

- Do MR imaging data reveal the underlying tumour heterogeneity?
- Which analysis methods are used to quantify tumour heterogeneity for diagnostic and/or treatment purposes, and what is the reported performance of these methods?
- Is tumour heterogeneity in DCE-MRI, as quantified with texture analysis methods, sensitive to changes due to therapy, and can patient outcome be predicted?

Chapter 2 presents a method to obtain an accurate 3D relation between high resolution *in vivo* MRI and the corresponding 3D histology of an experimental tumour model [41]. The aim of this study is to relate *in vivo* MR image features to the underlying pathophysiology as reflected in histological sections. The key elements of the methodology are: 1) standardized acquisition and processing, 2) use of an intermediate *ex vivo* MRI, 3) use of a reference cutting plane, 4) dense histological sampling, 5) use of elastic registration, and 6) use of complete 3D datasets. The methodology consists of two separate registration steps, both exploiting a three-step strategy of gradually increasing degrees of freedom (rigid, affine, and elastic transformation). These two registration steps involve *in vivo* MRI to *ex vivo* MRI registration, and *ex vivo* MRI to histology registration. The established 3D correspondence between tumour histology and *in vivo* MRI will allow the extraction of MRI characteristics for histologically confirmed regions.

Chapter 3 provides a systematic review of the literature on radiological image-based quantification of tumour heterogeneity for grading, differentiation, response monitoring and outcome prediction. A systematic search in Medline, Embase, and Cochrane Central was performed. Based on the selected literature, the following questions were explored: Which analysis methods are used for the quantification of heterogeneity or texture in diagnostic tumour imaging, tumour treatment monitoring and outcome prediction? What are the reported performances of the different analysis methods? Is there a relation between reported performance and image modality or analysis method? Can the performance results be generalized? What is the potential clinical impact of the methods? Has the performance also been evaluated in comparison to or in combination with established biomarkers?

In **Chapter 4** two heterogeneity biomarkers are evaluated for their potential of monitoring tumour changes due to treatment and predicting patient outcome. DCE-MRI images of 18 sarcoma patients undergoing isolated limb perfusion (ILP) with

TNF-α and melphalanare acquired at baseline and follow-up [42]. According to the histopathology, the tumours of the patients are classified into responding and non-responding tumours. The pharmacokinetic (K^{trans}) and heuristic model-based parametric maps (slope, max enhancement, AUC) are computed from the DCE-MRI data. The heterogeneity biomarkers are computed for all parametric maps. For each map and each heterogeneity biomarker, the ability to monitor the changes due to treatment and the predict outcome is evaluated.

Chapter 5 presents a study investigating regional heterogeneity changes in DCE-MRI due to treatment with ILP in experimental soft-tissue sarcoma [43]. The focus is on short-term treatment effects, i.e. within hours after treatment. DCE-MRI of drug-treated and sham-treated rats is performed at baseline and 1h after ILP intervention. Data are acquired using a macromolecular contrast medium, albumin-$(Gd\text{-}DTPA)_{45}$. To accurately identify the regional changes, the DCE-MRI at baseline and at follow-up are co-registered. To assess the regional heterogeneity the tumours are divided into 16 tumour sectors, and for each sector cumulative map-volume histograms are computed. The effect of treatment on regions and the variance between the regions is studied for the ILP-treated and sham-treated animals.

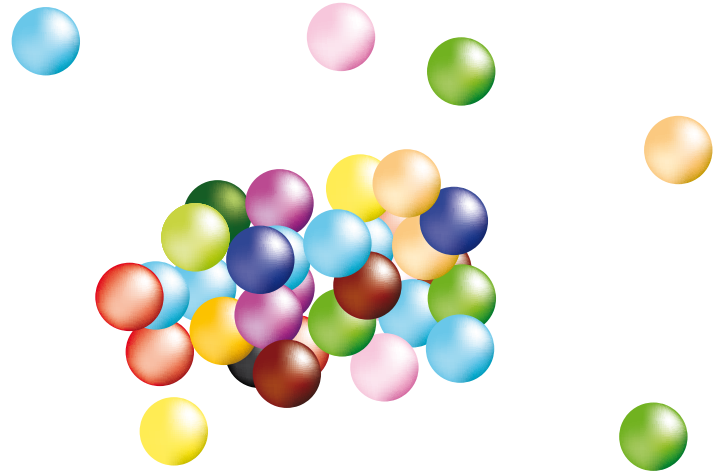
Chapter 6 summarizes the main results and contributions of the thesis, discusses implications for experimental and clinical applications, and offers some recommendations for future research.

Appendix A presents a short review of the automatic registration approach as a process of transforming different datasets into one coordinate system to achieve biological, anatomical or functional correspondence by using image intensities and gradients. The registration is used to correct for different deformations of *ex vivo* tumours with respect to the original *in vivo* shape (Chapter 2), for registration between different MRI sequences, and for registration between baseline and follow-up images (chapter 5).

References

1. KWF, *Kanker in Nederland tot 2020: Trends en prognoses*. 2012.
2. 2012; Available from: http://info.cancerresearchuk.org/cancerstats/incidence/commoncancers/.
3. Statistics, O.f.N., *Cancer statistics registrations: Registrations of cancer diagnosed in 2008*. 2011, National Statistics: London, UK.
4. Sasieni, P.D., et al., *What is the lifetime risk of developing cancer?: the effect of adjusting for multiple primaries.* Br J Cancer, 2011. 105(3): p. 460-5.
5. Ruddon, R.W., *Cancer biology*. 4th ed. 2007, Oxford ; New York: Oxford University Press. xiv, 530 p.
6. Bray, F., et al., *Global cancer transitions according to the Human Development Index (2008-2030): a population-based study.* Lancet Oncol, 2012. 13(8): p. 790-801.
7. Harding, F., *Breast cancer : cause, prevention, cure*. 2006, Aylesbury: Tekline Pub. 436 p.
8. Shibata, Y., et al., *Extrachromosomal microDNAs and chromosomal microdeletions in normal tissues.* Science, 2012. 336(6077): p. 82-6.
9. Shibata, D., *Cancer. Heterogeneity and tumor history.* Science, 2012. 336(6079): p. 304-5.
10. Fisher, R., L. Pusztai, and C. Swanton, *Cancer heterogeneity: implications for targeted therapeutics.* Br J Cancer, 2013.
11. Perou, C.M., et al., *Molecular portraits of human breast tumours.* Nature, 2000. 406(6797): p. 747-52.
12. Sorlie, T., et al., *Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.* Proc Natl Acad Sci U S A, 2001. 98(19): p. 10869-74.
13. Marusyk, A. and K. Polyak, *Tumor heterogeneity: causes and consequences.* Biochim Biophys Acta, 2010. 1805(1): p. 105-17.
14. Navin, N.E. and J. Hicks, *Tracing the tumor lineage.* Mol Oncol, 2010. 4(3): p. 267-83.
15. Russnes, H.G., et al., *Insight into the heterogeneity of breast cancer through next-generation sequencing.* J Clin Invest, 2011. 121(10): p. 3810-8.
16. Rehemtulla, A., *Overcoming intratumor heterogeneity of polygenic cancer drug resistance with improved biomarker integration.* Neoplasia, 2012. 14(12): p. 1278-89.
17. Gerlinger, M., et al., *Intratumor heterogeneity and branched evolution revealed by multiregion sequencing.* N Engl J Med, 2012. 366(10): p. 883-92.
18. Yap, T.A., et al., *Intratumor heterogeneity: seeing the wood for the trees.* Sci Transl Med, 2012. 4(127): p. 127ps10.
19. Ding, L., et al., *Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing.* Nature, 2012. 481(7382): p. 506-10.
20. Hayat, M.A., *Cancer imaging : instrumentation and applications*. 2008, Amsterdam ; Boston: Elsevier, Academic Press. lvii, 733 p.
21. Tofts, P.S., *Modeling tracer kinetics in dynamic Gd-DTPA MR Imaging.* Journal of Magnetic Resonance Imaging, 1997. 7: p. 91-101.
22. Tofts, P.S., et al., *Estimating kinetic parameters from dynamic contrast-enhanced T1-weighted MRI of a diffusable tracer: standardized quantities and symbols.* JMRI-Journal of Magnetic Resonance Imaging, 1999. 10: p. 223-232.
23. Brown, J., *Magnetic resonance imaging screening in women at genetic risk of breast cancer: imaging and analysis protocol for the UK multicentre study. UK MRI Breast Screening Study Advisory Group.* Magn Reson Imaging 2000. 18(7): p. 765-11
24. Alic, L., et al. *Quantification of Heterogeneity in Dynamic Contrast Enhancement MRI Data for Tumor Treatment Assessment*. 2006. IEEE International Symposium on Biomedical Imaging: From Nano to Macro. Arlington, Virginia.
25. Martel, A.L., *A fast method of generating pharmacokinetic maps from dynamic contrast-enhanced images of the breast.* Med Image Comput Comput Assist Interv, 2006. 9(Pt 2): p. 101-8.

26. Lavini, C., et al., *Pixel-by-pixel analysis of DCE MRI curve patterns and an illustration of its application to the imaging of the musculoskeletal system.* Magn Reson Imaging, 2007. 25(5): p. 604-12.
27. Eyal, E., et al., *Principal component analysis of breast DCE-MRI adjusted with a model-based method.* J Magn Reson Imaging, 2009. 30(5): p. 989-98.
28. Koh, T.S., et al., *Independent component analysis of dynamic contrast-enhanced magnetic resonance images of breast carcinoma: a feasibility study.* J Magn Reson Imaging, 2008. 28(1): p. 271-7.
29. Padhani, A.R., *MRI for assessing antivascular cancer treatments.* British Journal of Radiology, 2003. 76: p. S60-S80.
30. Johansen, R., et al., *Predicting survival and early clinical response to primary chemotherapy for patients with locally advanced breast cancer using DCE-MRI.* J Magn Reson Imaging, 2009. 29(6): p. 1300-7.
31. Kim, J.H., et al., *Dynamic contrast-enhanced 3-T MR imaging in cervical cancer before and after concurrent chemoradiotherapy.* Eur Radiol, 2012. 22(11): p. 2533-9.
32. Chikui, T., et al., *Pharmacokinetic analysis based on dynamic contrast-enhanced MRI for evaluating tumor response to preoperative therapy for oral cancer.* J Magn Reson Imaging, 2012. 36(3): p. 589-97.
33. Machiels, J.P., et al., *Phase II study of sunitinib in recurrent or metastatic squamous cell carcinoma of the head and neck: GORTEC 2006-01.* J Clin Oncol, 2010. 28(1): p. 21-8.
34. Mayr, N.A., et al., *Longitudinal changes in tumor perfusion pattern during the radiation therapy course and its clinical impact in cervical cancer.* Int J Radiat Oncol Biol Phys, 2010. 77(2): p. 502-8.
35. Ellingsen, C., et al., *Dynamic contrast-enhanced magnetic resonance imaging of human cervical carcinoma xenografts: pharmacokinetic analysis and correlation to tumor histomorphology.* Radiother Oncol., 2010. 97(2): p. 217-24.
36. Bushberg, J.T., *The essential physics of medical imaging*. 3rd ed. 2012, Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins. xii, 1030 p.
37. Mirmehdi, M., X. Xie, and J.S. Suri, *Handbook of texture analysis*. 2008, London Singapore ; Hackensack, NJ: Imperial College Press; Distributed by World Scientific. x, 413 p.
38. Haralick, R.M., *Statistical and structural approaches to texture.* Proceedings of the IEEE 1979. 67(5): p. 786-804.
39. Reed, T.R. and J.M.H. du Buf, *A review of recent texture segmentation and feature extraction techniques.* CVGIP: Image Understanding, 1993. 57(3): p. 259-372.
40. Chen, C.H., L.F. Pau, and P.S.P. Wang, *Handbook of pattern recognition & computer vision*. 2nd ed. 1999, River Edge, NJ: World Scientific. xxiii, 1019 p.
41. Alic, L., et al., *Facilitating tumor functional assessment by spatially relating 3D tumor histology and in vivo MRI: image registration approach.* PLoS One, 2011. 6(8): e22835.
42. Alic, L., et al., *Heterogeneity in DCE-MRI parametric maps: a biomarker for treatment response?* Phys Med Biol, 2011. 56(6): p. 1601-16.
43. Alic, L., et al., *Regional heterogeneity changes in DCE-MRI as response to isolated limb perfusion in experimental soft-tissue sarcomas.* Contrast Media and Molecular Imaging, 2013. 8(4): p. 340-349



Chapter 2

Facilitating tumor functional assessment by spatially relating 3D tumor histology and *in vivo* MRI: image registration approach

This chapter is based upon:
L Alić, JC Haeck, K Bol, S Klein, ST Van Tiel, PA Wielopolski, M Bijster, M Bernsen, M de Jong, WJ Niessen, JF Veenland. Facilitating tumor functional assessment by spatially relating 3D tumor histology and *in vivo* MRI: Image registration approach. *PLoS ONE, 2011. 6(8): e22835.*

Abstract

Magnetic resonance imaging (MRI), together with histology, is widely used to diagnose and to monitor treatment in oncology. Spatial correspondence between these modalities provides information about the ability of MRI to characterize cancerous tissue. However, registration is complicated by deformations during pathological processing, and differences in scale and information content.

This study proposes a methodology for establishing an accurate 3D relation between histological sections and high resolution *in vivo* MRI tumour data. The key features of the methodology are: 1) standardized acquisition and processing, 2) use of an intermediate *ex vivo* MRI, 3) use of a reference cutting plane, 4) dense histological sampling, 5) elastic registration, and 6) use of complete 3D data sets. Five rat pancreatic tumours imaged by T2*-w MRI were used to evaluate the proposed methodology. The registration accuracy was assessed by root mean squared (RMS) distances between manually annotated landmark points in both modalities. After elastic registration the average RMS distance decreased from 1.4 to 0.7 mm. The intermediate *ex vivo* MRI and the reference cutting plane shared by all three 3D images (*in vivo* MRI, *ex vivo* MRI, and 3D histology data) were found to be crucial for the accurate co-registration between the 3D histological data set and *in vivo* MRI. The MR intensity in necrotic regions, as manually annotated in 3D histology, was significantly different from other histologically confirmed regions (i.e., viable and hemorrhagic). However, the viable and the hemorrhagic regions showed a large overlap in T2*-w MRI signal intensity.

The established 3D correspondence between tumour histology and *in vivo* MRI enables extraction of MRI characteristics for histologically confirmed regions. The proposed methodology allows the creation of a tumour database of spatially registered multi-spectral MR images and multi-stained 3D histology.

Introduction

Recognizing the impact of the tumour microenvironment ononcogenic processes [1] led to the awareness that successful cancer management involves not only the tumour cells, but also needs to target the tumour microenvironment itself. Therefore, understanding and quantifying of the complex molecular and cellular interactions in cancer tissue is of paramount importance. Hence, the imaging of local tumour properties is becoming increasingly important to diagnose, monitor and predict tumour treatment [2, 3]. Magnetic resonance imaging (MRI) has considerable potential in non-invasive tumour characterization, as a multitude of scanning techniques can be employed. However, the exact relation between the signal intensities in MRI and the underlying pathophysiology is not always understood. Thorough understanding of the MRI oncogenic signatures involves an accurate spatial correlation of MRI and histology, offering a means to verify MRI findings. On the other hand, to create histological images the tumour tissue undergoes excision, fixation by formalin followed by dehydration, paraffin embedding, sectioning, and rehydratation during staining. An important side effect of this process is the significant tissue deformation which inevitably changes the tumour appearance. This severely complicates the registration of *in vivo* MRI to histological sections. Besides the loss of the tumour 3D integrity, the registration is also complicated by the inherent differences in image characteristics between colour histological images and gray scale MRI images.

Although the field of multi-modality registration has evolved considerably, the literature specifically dealing with registration of MRI to histology is limited, especially for *in vivo* MRI acquisitions. The first attempts to register histology and MRI were part of an effort to establish brain atlases, starting with affine registration [4] and advancing to piece-wise affine models [5]. Although affine registration achieved good initial results in these applications, they are inadequate to deal with non-linear distortions that occur during tissue excision and histological processing. Elastic registration for linking MRI with histology using surface matching has also been considered [6, 7]. Unfortunately, the reported results are limited to global matching of MRI volumes. Other studies [8] included point-based registration using manually placed landmarks. Besides being time consuming, these studies are also prone to intra-observer variability due to involvement of human interaction.

In oncological applications, co-localization of histology and MRI is often based on simple visual evaluation of local tissue features [9] and is therefore subjective and limited to a small number of histological sections. To facilitate rigid alignment several fiducial marker systems have been introduced [10-12]. These markers are physical implants that are clearly visible in all imaging modalities. Even though they might be useful for animal imaging, the use of fiducial markers in clinical applications is rarely possible. Therefore, as an alternative, distinctive image features (within or at

the surface) of the object under registration can be used to facilitate image alignment. For example, *in vivo* MRI of whole rat brain [13] and human prostate [14, 15] was related to their histological sections by point-based registration using manually placed [13, 15] or automatically established [14] landmark points. Although these internal landmarks have successfully assisted the registration of a complete organ, this compromises the registration accuracy within the tumour as it registers the organ instead of the tumour. Even though these methods solve part of the registration problem by using block-face images, they fail to account for 3D deformation as they use a limited number of histological sections.

To overcome the limitations of these methods, we propose the registration of complete 3D histology with *in vivo* MR images of the tumour tissue, i.e. excluding surrounding tissue. The aim of this work is to develop a methodology for establishing an accurate 3D relation between high resolution *in vivo* MRI and corresponding 3D histology of tumour tissue. The key features of the methodology are: a standardized imaging and histology method, acquisition of an intermediate *ex vivo* MRI, use of a reference cutting plane, a dense histological sampling, elastic (B-spline) registration, and use of the complete 3D data set.

Material and methods

Figure 1 is a schematic overview of the proposed methodology, which consists of a number of image acquisition steps (top-to-bottom) and image registration (bottom-up) steps. To facilitate the registration of *in vivo*, *ex vivo* and histology images, we kept track of the tumour orientation by colour coding the different tumour surfaces and by creating a reference cutting plane. This reference plane was created, after fixation, by slicing of a thin section of the whole tumour volume along the longest tumour axis and perpendicular to the subcutaneous side of the tumour. Although the reference plane is not physically present in *in vivo* MRI, the knowledge of its orientation is crucial to perform image resampling prior to registering *in vivo* MRI with *ex vivo* MRI [16, 17]. Figure 2 shows the tumour at onset of dissection, and the location of the reference plane in the volume rendered tumour in MRI.

Animal and tumour model

For this study, approval from the Ethical Committee of the Erasmus MC was obtained (Erasmus MC OZP 112-08-06). All investigations were carried in accordance with the requirements of the institution concerned, and also conform to the general requirements in the Netherlands regarding animal studies. Five male Lewis rats (Harlan-CPB, Austerlitz, The Netherlands), with a mean body weight of 300 g, were inoculated subcutaneously in the right hind limb with 106 pancreatic (CA20948) tumour cells

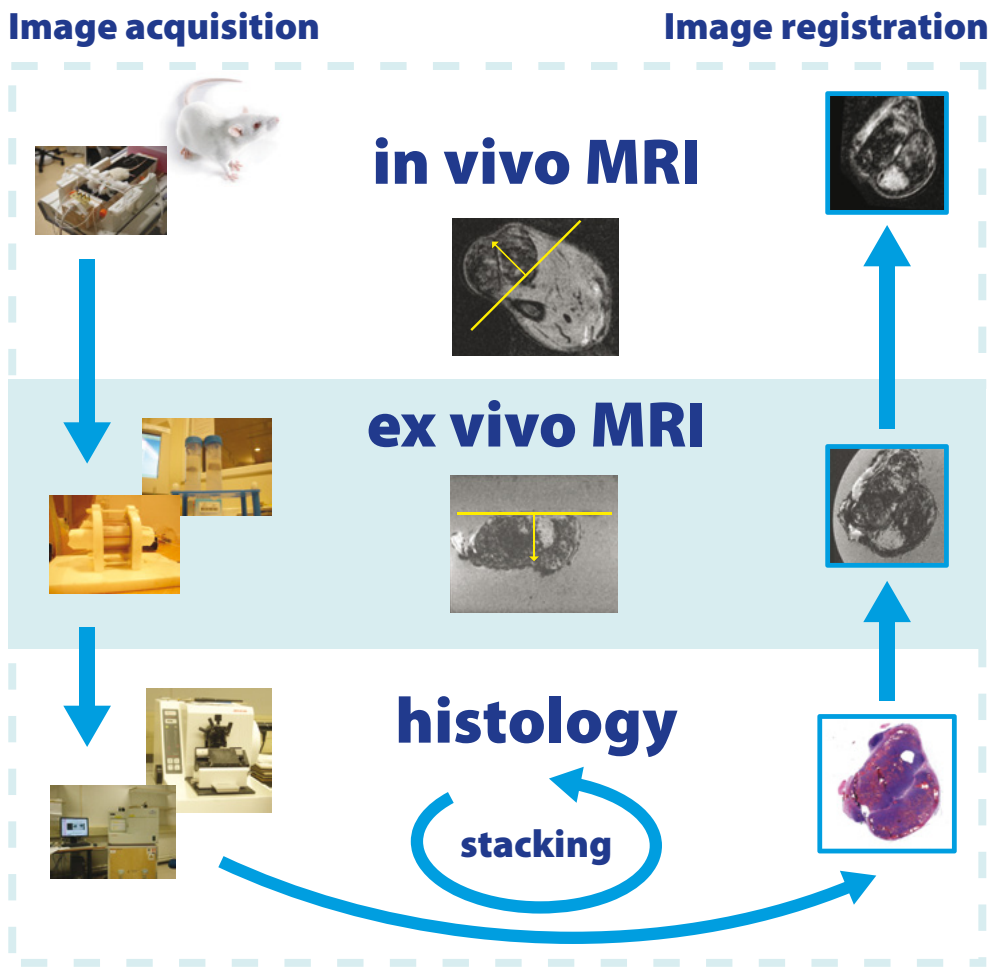


Figure 1. Overview of the processing steps (left-hand side) and the image registration and stacking procedures (right-hand side).

suspended in Hanks' balanced salt solution. The inoculated pancreatic tumours grow just beneath the skin as an encapsulated mass on top of the muscle tissue, with a preferred growth direction parallel to the skin (see Figure 2A). The tumour boundaries are well defined and the tumour is easy to separate from surrounding tissue. The animals were inspected daily for tumour growth and general appearance. The tumours were imaged using MRI when they reached approximately 10 mm in diameter. Before MRI, the animals were anesthetized by intra-peritoneal injection of medetomidine (Sedator, Eurovet Animal Health B.V., Bladel, The Netherlands) and sufentanil (Sufenta forte, Janssen-Cilag B.V., Tilburg, The Netherlands). During the imaging, the animals were kept at a temperature of 38–39°C by warm water mattresses. After

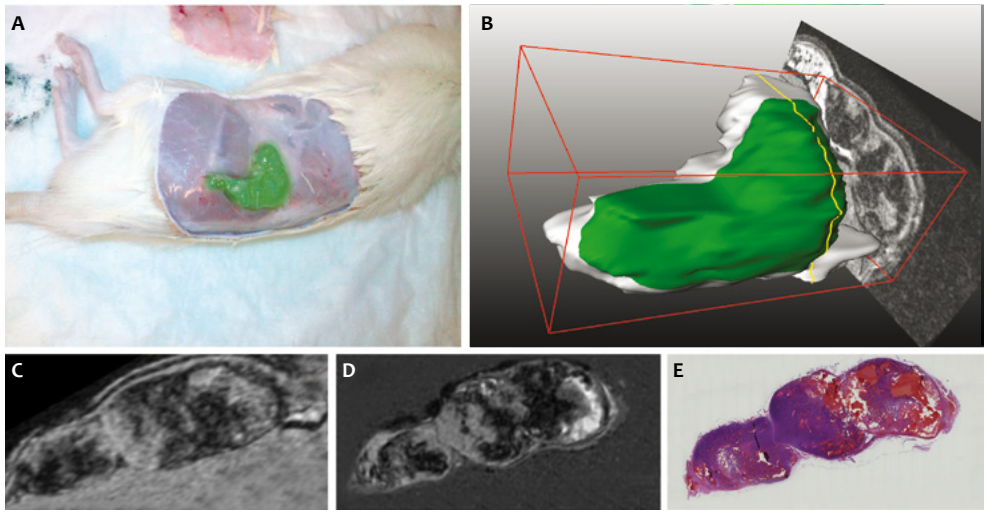


Figure 2. Illustration of subcutaneous tumour position. Tumour at onset of dissection (A) and as a 3D *in vivo* MRI tumour volume rendering (B), the subcutaneous side of the tumour is marked in green. A yellow line represents the cutting plane orientation along the longest tumour axis and perpendicular to the subcutaneous tumour side. The second row images show the corresponding slices of *in vivo* MRI (C), *ex vivo* MRI (D) and as histological section (E).

in vivo MRI, animals were euthanized, and the complete undamaged tumours were dissected. During the dissection, the tumour surfaces were dyed to track the *in vivo* tumour orientation by marking the subcutaneous, the head, the tail and dorsal side of the tumour. Figure 2A shows the subcutaneous tumour position at onset of dissection. Immediately after dissection, tumours were placed in 200 ml 10% buffered formalin (Boom, The Netherlands). A crucial step to facilitate alignment between *in vivo* MRI, and 3D histology stack is the knowledge of tumour orientation in all imaging modalities concerned [16, 17]. We created a reference plane by slicing of a thin section of the whole tumour volume along the longest tumour axis and perpendicular to the subcutaneous side of the tumour. The reference plane is illustrated in Figure 2B as a yellow line. The tumours were washed first to avoid possible T2*-artefacts due to remaining formalin concentrated on the tumour surface. Washing the tumours was achieved by sinking them into saline solution and drying the remaining moisture by paper towels. Subsequently, tumours were suspended in 1% agar dissolved in phosphate buffered saline (PBS, AbDSerotec, MorphoSys, Munich, Germany) to facilitate *ex vivo* MRI acquisition by restricting tissue motion and air-tissue MRI artefacts.

Magnetic resonance imaging

For the *in vivo* MRI acquisition parameters were: TR/TE =23.2/8.9 ms, flip angle of 10°, field-of-view (FOV) of 50x50 mm^2, image acquisition matrix of 320x256 with a

slice thickness of 0.4 mm (acquired voxel resolution = 0.156x0.195x0.4 mm^3) and a resampled matrix of 512x512 using zero-filling for a reconstructed voxel size of 0.098x0.098x0.2 mm^3. For the *ex vivo* MRI acquisition parameters were: TR/TE=42.2/20.9 ms, flip angle of 15°, field-of-view (FOV) of 50x50 mm^2,image acquisition matrix of 320x256 with a slice thickness of 0.4 mm (acquired voxel resolution = 0.094x0.118x0.4 mm^3) and a resampled matrix of 512x512 using zero-filling for are constructed voxel size of 0.059x0.059x0.2 mm^3. For both *in vivo* and *ex vivo* MRI read bandwidth was 48.8 Hz/voxel, no flow compensation or saturation pulse, two averages, frequency encoding = left-right, and the phase encoding direction = up-down. The total acquisition time was less than 20 minutes for both *in vivo* and *ex vivo* acquisitions. No acceleration was used for imaging.

Histological processing

Following the *ex vivo* MR imaging, tumours were processed in a Histokinette, and subsequently embedded in paraffin. The histological data consisted of 4-mm thick sections (cut from the reference plane onwards, see Figure 1 and Figure 2) mounted on glass slides, and stained with hematoxylin and eosin (H&E). Depending on the tumour size, up to 40 sections (4-mm thickness each) were mounted at intervals of 80 mm. The procedure also enables to acquire histological sections with different stains. The slides were digitized using the NanoZoomer Digital Pathology (C9600, Hamamatsu, Japan) at 20x magnification, which resulted in a pixel size of 3.64 mm.

Registration

We first provide an outline of the different parts in the automatic registration procedure which were performed using Elastix [18]. The details of the image registration are included in Appendix A [19-27] with the basic components of the registration framework are illustrated. Between the different image acquisition steps (Figure 1) a tumour undergoes deformations with respect to its original *in vivo* shape. As these deformations differ in nature and scale, the registration procedure consists of three distinct parts. All registrations use contrast in image intensities to perform the registration automatically.

1. Reconstruction of tumour 3D histology by rigid registration of digitized adjacent H&E sections and adjustment of the slice thickness, referred to as stacking.
2. Volumetric alignment of 3D histology stack and 3D *ex vivo* MRI using a three-step strategy (rigid, affine, and elastic registration),referred to as stack2ex.
3. Volumetric alignment of 3D *ex vivo* MRI to 3D *in vivo* MRI using a three-step strategy (rigid, affine, and elastic registration), referred to as ex2in.

All separate registrations were performed using Elastix [18]. To achieve the desired volumetric alignment of 3D histology to *in vivo* MRI, the separate transformations (the results from stack2ex and ex2in registrations) were concatenated automatically. The final concatenated geometric transformation, referred to as stack2in,was applied to the 3D colour histology stack which aligns it to the *in vivo* MRI.

Stacking. As the first step in the automatic registration process, we automatically reconstructed 3D histological volume by rigid registration of adjacent H&E stained images. To optimally exploit the digital image information, considering the necrotic and viable tissue, the information content of separate image channels was evaluated. We used the red image channel in the registration as it provides the best separation between signal intensities of necrotic and viable volumes of interest (VOIs) and presumably the best image contrast (Figure 3 presents histograms of these VOIs). The series of 2D histological slices (red channel) were reconstructed iteratively into a 3D volumetric image. The resulting transformations were applied to the other two (green and blue) image channels, resulting in a 3D colour histology stack. Subsequently, the slice thickness was set to 80 mm, i.e. the physical distance between subsequent sections.

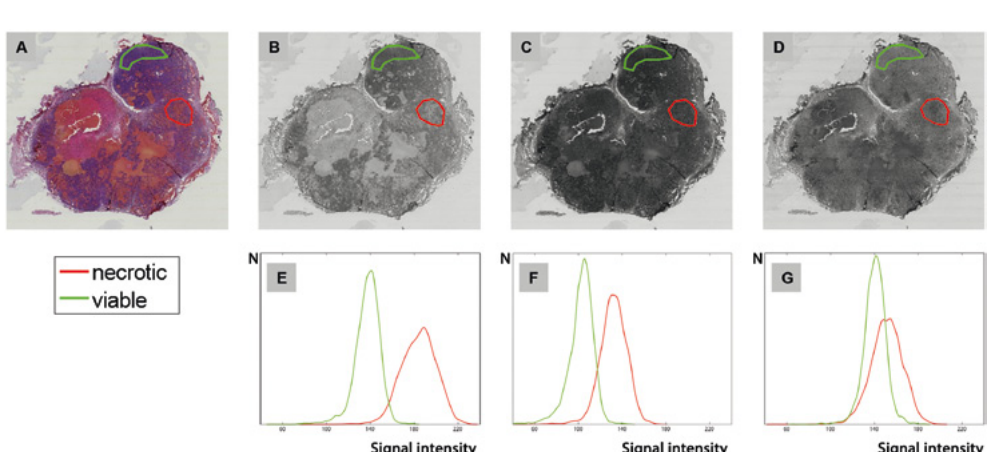


Figure 3. The distribution of separated image channels from a H&E section. We used the red image channel to perform the registration as it provides the best separation between signal intensities of necrotic and viable volumes of interest (VOIs) and presumably the best image contrast. A H&E stained histological section (A) and three separate colour channels, red-green-blue, (B–D) with corresponding histogram distributions of the vital (green) and necrotic (red) tumour regions (E–G).

stack2ex. The second automatic registration step, aligning the 3D histology with *ex vivo* MRI, is greatly facilitated by the definition of the reference cutting plane (see Figure 2), i.e. both images (*ex vivo* MRI and histology stack) start at the same position (reference plane). This provides the initialization for a three-step registration strategy of gradually increasing degrees of freedom, starting as rigid registration, followed by affine registration, and finalized by elastic refinement.

ex2in. Prior to the third automatic registration step, the knowledge of the reference plane within *in vivo* 3D-T2*w MRI (see Figure 1) was used to realign and resample the *in vivo* data according to the e*x vivo* MRI orientation. This ensures similar orientation and rough alignment of *in vivo* MRI and *ex vivo* MRI. First, rigid registration was performed, followed by affine transformation allowing isotropic scaling to account for volume changes, and finalized by elastic registration.

Evaluations

Evaluation of registration accuracy. The resulting alignment of *in vivo* 3D-T2*w MRI with 3D histology stack was qualitatively evaluated by two observers using visual inspection with a moving quadrant view, and quantitatively evaluated using anatomical landmarks (e.g., characteristic features in the tumour and on the contour). For the quantitative evaluation, ten clearly identifiable anatomical landmarks were initially defined on the colour 3D histology stack. Subsequently, two observers independently annotated the corresponding anatomical landmarks in the *in vivo* 3D-T2*w MRI. To evaluate registration accuracy, the root mean squared (RMS) distance between the corresponding points in the *in vivo* MRI and 3D histology was calculated before registration, and after the two registration steps (i.e., rigid and elastic). Furthermore, the inter observer variability was estimated by computing the RMS distance between the corresponding points of the two observers on the MRI.

Evaluation of reference plane. The reference plane greatly facilitates the registration procedure. The difference in reference plane position between two 3D images after registration, measures the initial reference plane error. To quantify the error in reference plane positioning, the out-of-plane angulation is estimated as the rotation component of the rigid registration for both steps (stack2ex and ex2in).

Tumour volume change. Tumour global volume change between *in vivo* MRI, *ex vivo* MRI and histology was established by computing the determinant of the corresponding affine transformation for both registration steps (stack2ex and ex2in). The tumour local volume change for the different histological regions was also estimated. For this purpose, three volumes of interest (VOIs) representing viable, necrotic and hemorrhagic regions were delineated in the colour 3D histology stack. This provides three masks which were warped using the transformation, provided by the corresponding registration step, to match the *in vivo* MRI.

For each region (viable, necrotic and hemorrhagic) and both registration steps (stack2ex and ex2in), the change in volume was estimated before and after registration: see Eq. 1.

$$\Delta_V = 2\,\frac{|V_R - V_o|}{|V_R - V_o|} \qquad \text{(Eq. 1)}$$

where V_R, and V_O represent the VOIs before and after registration, respectively.

Facilitating MRI characteristics identification. To identify image characteristics of *in vivo* 3D-T2*w MRI, histograms of histologically confirmed VOIs were used to estimate the probability density function (pdf). For each VOI's histogram, the pdf interquartile range was then used for automatic segmentation of the *in vivo* 3D-T2*w MRI.

Results

Evaluation of registration accuracy

Figure 4 shows the results of the separate registration steps (stack2ex and ex2in) and the concatenation of those registrations (stack2in) for the five tumours. The checker board view (Figure 4; fourth column) of the registered *in vivo* 3D-T2*w MRI and the 3D histology shows that good alignment has been achieved. For all five tumours, the final registration (stack2in) was evaluated as excellent for 25% and good for 53% of the registered slices. For 13% of the slices the registration was evaluated as fair, and for the remaining 9% as poor. The registration of the slices towards the tumour borders was in general less accurate than the registration of central slices.

Table 1 presents the RMS distance error for 10 landmark positions averaged over all five tumours after final registration (stack2in). By utilization of the reference plane, the initial average accuracy was already 1.4 mm. After registration, the average accuracy increased from 1.4 mm to 0.7 mm. When compared with the *in vivo* pixel size, the average accuracy increased from 15 to 7 pixels. The final accuracy of 0.7 mm corresponds on average with 30–50 cells. To assess the uncertainty of the manual annotations, we computed the inter-observer variation, which was in the order of 0.7 to 0.9 mm.

Table 1. Average root mean squared distances (mm) for the different registration steps, averaged over all 5 subjects.

	Observer 1	Observer 2	Average	Inter-observer
Initial	1.2 ±0.6	1.6 ±0.7	1.4 ± 0.6	0.9 ± 0.6
Rigid	1.1 ±0.5	1.0 ±0.4	1.0 ± 0.4	0.7 ± 0.3
Elastic	0.8 ±0.3	0.6 ±0.2	0.7 ± 0.3	0.7 ± 0.4

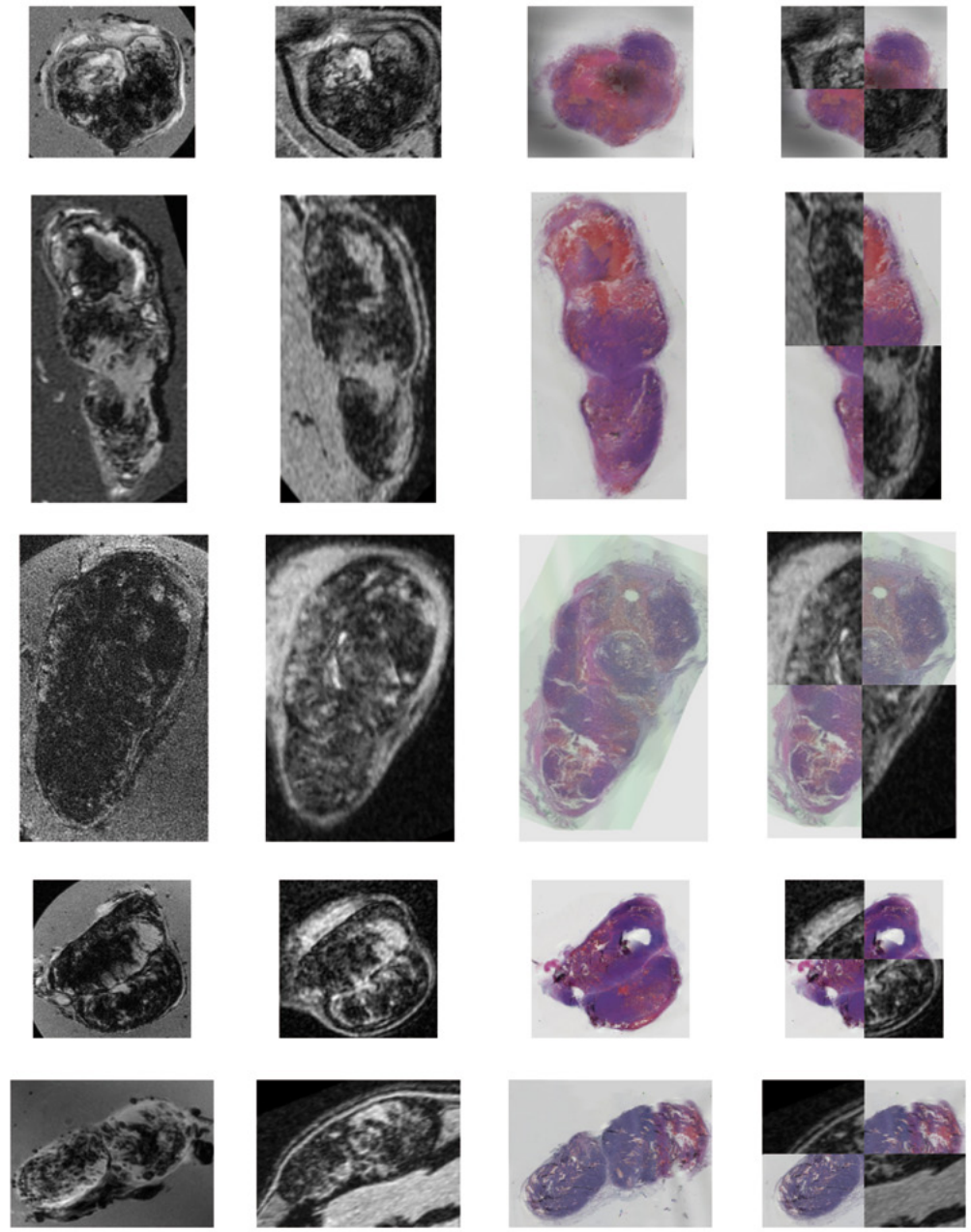


Figure 4. Final registration results for five tumours. Registered *ex vivo* T2*-w MRI (first column), *in vivo* T2*-w MRI (second column), registered colour 3D histology (third column), and checkerboard view of *in vivo* and registered histology (fourth column).

Evaluation of reference plane

Error in the reference plane positioning, measuring the remaining 3D mismatch, was established for both registration steps separately. Table 2 summarizes the angulation as averaged over all five subjects. The absolute angulation for st2ex registration was 1.4±1.30%, ranging from 22.52 to 3.08, and for ex2in registration was 2.3±1.34% with a range of 22.79 to 4.00. This shows that directional mismatch between the resampled *in vivo* MRI and *ex vivo* MRI, and between *ex vivo* MRI and histological sections were minimal.

Table 2. Summarized registration results averaged over all five subjects.

Registration step	% Global Volume change	Angulation [°]	% Δ_v Necrotic	% Δ_v Viable	% Δ_v Hemorrhagic
Stack2ex	-1.9 ± 0.07	1.4 ± 1.30	12.2 ± 6.4	11.9 ± 6.4	11.0 ± 8.7
Ex2in	13.2 ± 0.05	2.3 ± 1.34	16.9 ± 6.8	16.0 ± 4.4	11.2 ± 15.0

Tumour volume change

On average, the global tumour volume expanded 1.9% after sectioning. The same specimens shrank on average 13.2% after chemical fixation. Table 2 summarizes global and local volume change (per VOI) averaged over all five subjects. All histologically different regions (i.e., viable, necrotic, and hemorrhagic) expanded similarly after sectioning. On the other hand, we observed a significant difference in deformation between different histologically confirmed regions. That is, the shrinkage after chemical fixation is different for the hemorrhagic region compared with the necrotic and viable regions.

Facilitating MRI characteristics identification

The 3D correspondence of tumour histology and *in vivo* MRI enables extraction of MRI characteristics for histologically defined regions. This is illustrated using the histogram-based pdf of the registered histology (Figure 5C) which clearly separated the different tissue types in the H&E stained images. The corresponding pdf of *in vivo* 3D-T2*w MRI (Figure 5A) demonstrates that viable and hemorrhagic regions cannot be separated using solely *in vivo* 3D-T2*w MRI signal intensities. Nevertheless, necrotic regions can be effectively separated from the other two histologically confirmed regions. Figure 6 evaluates the pdf of *in vivo* 3D-T2*-w MRI for all subjects demonstrating similar gray value ranges for each VOI. When considering the pdf for all tumours, the necrotic regions were significantly different from other histologically confirmed regions (i.e., viable and hemorrhagic). However, the viable and the hemorrhagic regions showed a large overlap in T2*-w MRI signal intensity.

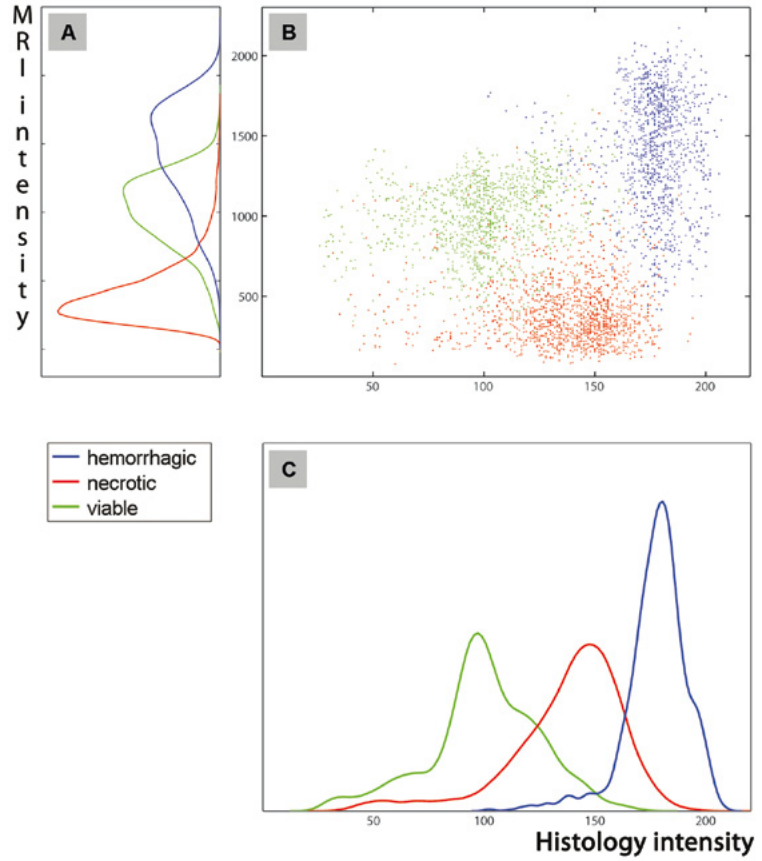


Figure 5. The illustration of signal intensity correspondence between *in vivo* T2*-W MRI and registered 3D histology for three VOIs (e.g., necrotic-red, viable-green, and hemorrhagic-blue). The 3D correspondence of tumour histology and *in vivo* MRI enables extraction of MRI characteristics for histologically defined regions. This is illustrated in scatter plot (B) and using the histogram-based probability density function of the registered histology (C) which clearly separated the different tissue types in the H&E stained images. The corresponding probability density function of *in vivo* 3D-T2*w MRI (A) demonstrates that viable and hemorrhagic regions cannot be separated using solely *in vivo* 3D-T2*w MRI signal intensities. Nevertheless, necrotic regions can be effectively separated from the other two histologically confirmed regions.

To confirm the findings shown in Figure 5 we used the histogram-based pdf to define the interquartile intensity ranges for histologically confirmed regions. These ranges were used for automatic segmentation of *in vivo* 3D-T2*w MRI signal intensities. Figure 7 illustrates the necrotic segmentation superimposed on 3DT2*-w MRI. The viable and hemorrhagic tissues cannot be separated based on the T2*-w MRI intensity.

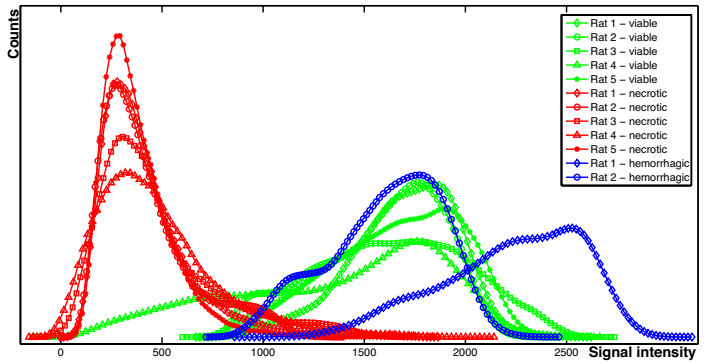


Figure 6. The group-wise probability density functions distributions of *in vivo* 3D-T2*-w MRI. Three VOIs (e.g., necrotic-red, viable-green, and hemorrhagic-blue) were annotated in histological sections and used for segmentation of automatically aligned *in vivo* MRI. The excessive hemorrhagic regions are visible in two out of five subjects. It demonstrates the similar gray value ranges for each VOI. When considering the probability density function for all tumours, the necrotic regions were significantly different from other histologically confirmed regions (i.e., viable and hemorrhagic). However, the viable and the hemorrhagic regions showed a large overlap in T2*-w MRI signal intensity.

Discussion

This proposed methodology, i.e. aligning histological tissues sections to *in vivo* MRI, consists of a number of image acquisition and image registration steps that have been evaluated. The methodology is assembled around two separate registration steps, both exploiting a three-step strategy of gradually increasing degrees of freedom (rigid, affine, and elastic transformation), which allow for a coarse-to-fine scheme. To enable the registrations, we kept track of the tumour orientation by colour coding the different tumour surfaces and by creating a reference plane. Qualitative and quantitative evaluation of the registration and protocol accuracy was performed.

During the registration evaluation, the alignment of tumour surface and internal structures was qualitatively evaluated as accurate. Quantitatively, we achieved an average accuracy of 0.7 mm after the registration. The results involving two observers to estimate the RMS error showed similar trends in increasing accuracy with increasing degrees of freedom. The inter-observer variation of the manual annotation was approximately 0.7 mm. This is an indication of the limitation of the measurement method; smaller distances could not reliably be measured. The RMS distance after elastic registration is of the same order. Evaluation of the protocol accuracy shows that a 3D-registration method complemented by standardized acquisition is essential to accurately align histology to *in vivo* MRI. Excision and fixation of the tumour resulted in an average shrinkage of 13%. However, the sectioning of the tumour enlarged the tissue by 1.9%.

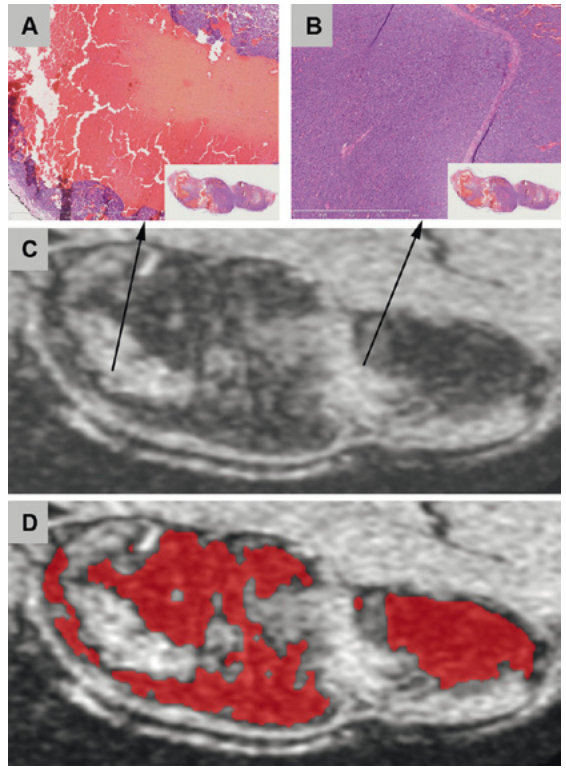


Figure 7. Details from a H&E stained section and its corresponding MRI slice. Histological section (A–B) shows the difference in histological appearance, whereas the MRI appearance in 3D T2*-w MRI (C) is similar. The necrotic segmentation, superimposed on 3D T2*-w, is shown in red (D).

Park et al. registered prostates imaged using *in vivo* MRI, *ex vivo* MRI after prostatectomy, block-face photographs, and histological sections [15]. They used block-face photographs to reconstruct the original histology. Registration was performed by point-based registration using manually placed landmarks. They moved towards 3D registration using three consecutive slices during histology-to-MRI registration. Although studies have registered whole-prostate histology to *in vivo* MRI, to our knowledge the present study is the first attempt to register pancreatic tumours. Our methodology intentionally excludes the use of block-face images as this would complicate image acquisition and registration when acquiring large number of histological sections. Compared to the method proposed by Park et al. [15] our method uses denser histological sampling, no user interaction is required during the registration procedure, and the whole image content of the tumour volume is utilized for registration.

This study presents the successful development and careful evaluation of a combined methodology for alignment of tumour histological sections to *in vivo* MRI. At the same time, it demonstrates the importance of integrated methodology between imaging and registration. The established 3D correspondence between tumour histology and *in vivo* MRI enables extraction of MRI characteristics for histologically

confirmed regions. We showed that, based on T2*-w MRI signal intensity, automatic identification of necrotic tissue is feasible. However, based on T2*-w MRI, the separation of hemorrhagic and viable tissue was not possible. The hypo-intense areas in T2*-w MRI seem to correspond to necrotic tissue, see Figure 7. However, this conclusion should be taken cautiously as deoxyhemoglobin and hemosiderin can also cause low intensity on T2*-w MRI [28]. As those may be undistinguishable in the T2*-w MRI, tumour necrosis may have been overestimated by MRI analysis.

This work is a first step in MRI tumour characterization. When the basic correspondence between *in vivo* MRI and 3D H&E histology can be established, the extension to multi-spectral MR images and multi-stained histological sections is a logical next step. Different histological stains highlight different aspects of the tumour, in Figure 8 the spatial correspondence between the *in vivo* MRI, the *ex vivo* MRI and multi-stained histological sections is shown. This work can be used to create a database consisting of multi-spectral MRI images and multi-stained 3D reconstructed histology that may be an essential and valuable source for understanding MR images, and highly beneficial in the process of identifying MRI tumour characteristics.

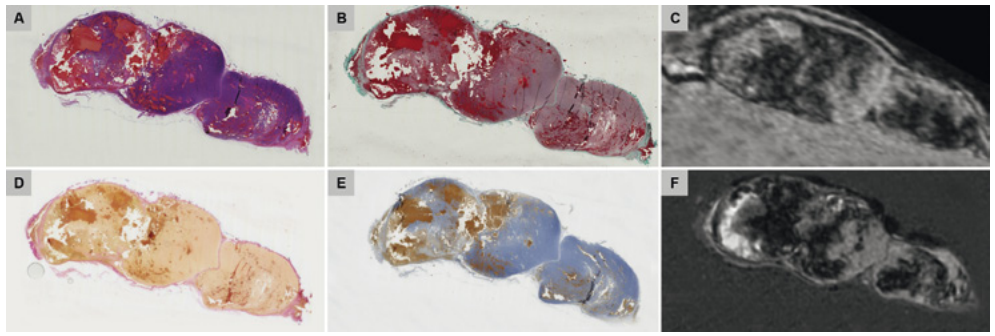


Figure 8. Multi-stained histology dataset. H&E (A), Goldner (B), van Gieson (C) and Peroxidase (D) stained consecutive sections with the corresponding *in vivo* MRI (E) and intermediate *ex vivo* MRI (F). The slice thickness of all histology sections is 4 mm, and the distance between the consecutive histology sections (A–D) is 8 mm.

Some modifications are envisioned which need exploring, as they will increase the robustness and accuracy of the technique without significantly increasing processing time. In the protocol used, the hyper-intense regions cannot be specified based on solelyT2*-w MRI as shown in Figure 7 and Figure 8. The use of multimodality MRI images is expected to enable a more detailed differentiation between tissue types by combining the different contrast mechanisms present in the MRI sequences. For example, contrast enhanced (CE) MRA or DWI-MRI may create a contrast between vital and hemorrhagic regions. Multi-modality MRI images will therefore refine the registration and offer a more detailed biological profile of the tumour.

References

1. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. 100(1): p. 57-70.
2. Evans, P.M., *Anatomical imaging for radiotherapy.* Phys Med Biol, 2008. 53(12): p. R151-91.
3. Schlemmer, H.P., *Imaging for new radiotherapy techniques.* Cancer Imaging, 2010. 10: p. S73.
4. Ourselin, S., et al., *Fusion of Histological Sections and MR Images: Towards the Construction of an Atlas of the Human Basal Ganglia*, in *MICCAI*, W.J.a.V. Niessen, M., Editor. 2001, Springer-Verlag Berlin Heidelberg. p. 743-751.
5. Pitiot, A., et al., *Piecewise affine registration of biological images for volume reconstruction.* Med Image Anal, 2006. 10(3): p. 465-483.
6. Jacobs, M.A., et al., *Registration and warping of magnetic resonance images to histological sections.* Med Phys, 1999. 26(8): p. 1568-1578.
7. Schormann, T. and K. Zilles, *Three-dimensional linear and nonlinear transformations: an integration of light microscopical and MRI data.* Hum Brain Mapp, 1998. 6(5-6): p. 339-47.
8. Li, G., S. Nikolova, and R. Bartha, *Registration of in magnetic resonance T1-weighted brain images to triphenyltetrazolium chloride stained sections in small animals.* J Neurosci Methods, 2006. 156(1-2): p. 368-375.
9. Wang, H., et al., *Treatment of rodent liver tumor with combretastatin a4 phosphate: noninvasive therapeutic evaluation using multiparametric magnetic resonance imaging in correlation with microangiography and histology.* Invest Radiol, 2009. 44(1): p. 44-53.
10. Humm, J.L., et al., *A stereotactic method for the three-dimensional registration of multi-modality biologic images in animals: NMR, PET, histology, and autoradiography.* Med Phys, 2003. 30(9): p. 2303-2314.
11. Lazebnik, R.S., et al., *Volume registration using needle paths and point landmarks for evaluation of interventional MRI treatments.* IEEE Trans Med Imaging, 2003. 22(5): p. 653-660.
12. Breen, M.S., et al., *Three-dimensional method for comparing in vivo interventional MR images of thermally ablated tissue with tissue response.* J Magn Reson Imaging, 2003. 18(1): p. 90-102.
13. Meyer, C.R., et al., *A methodology for registration of a histological slide and in vivo MRI volume based on optimizing mutual information.* Mol Imaging, 2006. 5(1): p. 16-23.
14. Zhan, Y., et al., *Registering histologic and MR images of prostate for image-based cancer detection.* Acad Radiol, 2007. 14(11): p. 1367-1381.
15. Park, H., et al., *Registration methodology for histological sections and in vivo imaging of human prostate.* Acad Radiol, 2008. 15(8): p. 1027-1039.
16. Alic, L., et al., *Multi-modal image registration: matching MRI with histology*, in *SPIE Medical Imaging*. 2010: San Diego
17. Bol, K., et al., *Developing a tool for the validation of quantitative DCE-MRI.*, in *SPIE Medical Imaging*. 2011.
18. Klein, S., et al., *elastix: a toolbox for intensity-based medical image registration.* IEEE Trans Med Imaging, 2010. 29(1): p. 196-205.
19. Modersitzki, J., *Numerical Methods for Image Registration*. 2003: Oxford University Press.
20. Klein, S., M. Staring, and J.P. Pluim, *Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines.* IEEE Trans Image Process, 2007. 16(12): p. 2879-2890.
21. Ibanez, L., et al., *The ITK Software Guide*. 2005: Kitware, Inc.
22. Rueckert, D., et al., *Nonrigid registration using free-form deformations: application to breast MR images.* IEEE Trans Med Imaging, 1999. 18(8): p. 20.
23. Maes, F., et al., *Multimodality image registration by maximization of mutual information.* IEEE Trans Med Imaging, 1997. 16(2): p. 187-198.
24. Thevenaz, P. and M. Unser, *Optimization of mutual information for multiresolution image registration.* IEEE Trans Image Process, 2000. 9(12): p. 2083-2099.
25. Klein, S., et al., *Adaptive stochastic gradient descent optimisation for image registration.* International Journal of Computer Vision,, 2009. 3(81): p. 227-239.
26. Pluim, J.P.W., J.B.A. Maintz, and M.A. Viergever, *Interpolation artefacts in mutual information-based image registration.* Computer Vision and Image Understanding, 2000. 77(2): p. 211-232.
27. Lester, H.A., S.R, *A survey of hierarchical non-linear medical image registration.* Pattern Recognition, 1999. 32(1): p. 21.
28. Imaizumi, T., et al., *Detection of hemosiderin deposition by T2*-weighted MRI after subarachnoid hemorrhage.* Stroke, 2003. 34(7): p. 1693-8.

Chapter 3

Quantification of heterogeneity as a biomarker in tumour imaging: a systematic review

This chapter is based upon:
L Alić, WJ Niessen, JF Veenland. Quantification of heterogeneity as a biomarker in tumour imaging: a systematic review. *Submitted.*

Abstract

Tumours frequently demonstrate heterogeneity in structure, function and response to treatment wich may be visualised en quantified by imaging techniques. This systematic literature review aims to answer the following questions: Which analysis methods are used for the quantification of heterogeneity or texture in diagnostic tumour imaging, outcome prediction and tumour treatment monitoring? What are the reported performances of the different analysis methods? Is there a relation between reported performance and analysis method? Can the performance results be generalized? What is the potential clinical impact of the methods? Has the performance also been evaluated in comparison to or in combination with established biomarkers?

The databases Ovid, Embase and Cochrane Central were searched up to 24 January 2013. Heterogeneity analysis methods were divided into four categories: non-spatial methods (NSM), spatial grey level methods (SGLM), fractal analysis (FA) methods, and filters and transforms (F&T). The reported results are divided into: publications reporting classification experiments, and those reporting significance testing.

From the 8,956 potentially relevant publications, 192 reporting on 170 studies were included. Generally, about 60% of the studies use NSM, 49% use SGLM, 11% use FA, and 28% use F&T. Differential diagnosis, grading or outcome prediction was the goal in 86% of the studies. In 72% of these studies NSM or SGLM was performed, and 36% of the studies were based on MRI. For the response monitoring NSM was the most frequently used method, i.e. in 73%. Classification results were reported in 68% of the studies, statistical outcomes in 30%, and no outcome in 2%. Practically no papers evaluated the additional value of the heterogeneity biomarker on top of the available clinical markers.

No relation was found between the discriminative power and the quantification methods, or between the discriminative power and the imaging modality. The reported AUC ranged from 0.5 to 1 with a median of 0.89. A negative correlation was found between the AUC and the number of features estimated per tumour, which is probably caused by overfitting in small datasets. In only 53.4% of the classification studies was the use of cross-validation reported. Many studies report on the potential of tumor heterogeneity for grading, differentiation, outcome prediction and treatment response monitoring. However, none report on the use of an external validation set to test their findings. Retrospective analyses were conducted in 60% of the studies, but without a clear description of the inclusion criteria. Only 12% of the studies had a prospective study design.

To enable the translation of imaging biomarkers from the research stage to clinical practice, research should focus more on prospective studies, use external datasets for validation, and evaluate the added value of the proposed heterogeneity biomarker on top of the clinical markers.

Introduction

Tumours are not always homogeneous. Regional variations in cell death, metabolic activity, proliferation, and vascular structure can be observed. Heterogeneity is associated with malignancy, disease progression and therapeutic response [1]. For example, parameters in hot spots as quantified with DCE-MRI are reported to be more relevant for monitoring tumour response than parameters averaged over the whole tumour [2-4]. These findings are also supported by the discovery that distinct populations of cancer cells interact in a competitive manner [5]. For example, more aggressive cancer populations (fast proliferating populations with a higher neo-angiogenesis) are less sensitive to treatment and will, therefore, suppress the less fit populations. There is also clinical evidence that recurrent tumours are more malignant than the primary tumour: the more aggressive populations have survived. In this respect, visualization and quantification of tumour heterogeneity is a useful tool in grading, differentiation, monitoring and predicting tumour treatment response.

Several methods have been developed and used to quantify tumour heterogeneity from imaging data. Many studies use histogram-based features such as percentile values, standard deviation and enhancing fraction. However, these features do not take into account the spatial distribution of the intensity values. Texture methods do take spatial information into account, by quantifying the spatial variations in the images. An important advantage of texture-based methods is the independence of the absolute values in the images. Therefore, texture analysis can provide additional and independent information compared to absolute histogram-based measures.

The present systematic review investigates the performance of different heterogeneity imaging biomarkers extracted from diagnostic tumour images for tumour grading, differentiation, outcome prediction or response monitoring. The following research questions were formulated:

- Which analysis methods are used for quantifying heterogeneity or texture in diagnostic tumour imaging, outcome prediction and tumour treatment monitoring?
- What are the reported performances of the different analysis methods? Is there a relation between performance and analysis method?
- What is the potential clinical impact of the methods? Can the performance results be generalized? Is the performance evaluated in addition to established imaging biomarkers?

Methods

Data sources and search method

This review was performed in accordance with the Prisma guidelines [6]. The investigators wrote a protocol and registered it with the International Prospective Register of Systematic Reviews (identification number: CRD42013003634) in January 2013 [7]. A systematic literature search was conducted in the databases of Medline, Embase and Cochrane Central; the search was performed on January 24th 2013.

The following search entries were used in Medline as full text or MeSH Terms when available cancer[1]:

- neoplasms or cancer or tumour or tumor
- heterogeneity or texture
- MRI or MRS or CT or PET or SPECT or ultrasonography

Study selection

Two authors (LA and JFV) independently reviewed the titles and abstracts, and screened the full-text articles. Subsequently, the publications were passed to the data extraction phase. Any discrepancies about study inclusion during the title and abstract review were resolved by passing the publication in question to full-text screening. Any discrepancy at the following stages was resolved by discussion. The bibliographies of seminal review papers [1, 8, 9] were reviewed to identify additional relevant articles.

Inclusion/exclusion criteria

Publications related to diagnostic imaging were only included when they reported quantification of tumour heterogeneity or tumour texture with the goal of tumour grading, differentiation between benign/malign, outcome prediction, and tumour response monitoring. No restrictions were made on the basis of the location, type, stage or the grade of malignancy. Prior to review a decision was formulated to exclude any study with too few participants, i.e. for patient studies (n<10) and for animal studies (n<5). Therefore, all case studies and those that contained no information on the number of subjects were excluded. Additionally, all the following studies were excluded:

- publications based on non-tumour images;
- publications not based on quantitative assessment of heterogeneity or texture in images

1 The full details of the searches can be obtained from the first author.

- publications without one of the following goals: tumour differentiation, tumour grading, outcome prediction or treatment monitoring;
- publications not based on *in vivo* studies (histology, phantom, *ex vivo*, synthetic data);
- publications describing non original research (editorial, letters to the editor, review, meta-analysis, opinion publications).

Data extraction

A data extraction form was designed. All publications were independently reviewed and data extraction was cross-checked. Disagreements between the reviewers were resolved by consensus. The following data were extracted from the full paper: year of publication, human or animal study, type of study (retrospective or prospective), number of subjects, number of tumours, location of tumour, imaging modality, tracer/contrast agent, goal of heterogeneity/texture analysis, and the type of heterogeneity/texture quantification method used. For studies reporting on the same analysis method based on the same dataset, only the latest publication was included. For publications reporting classification experiments the following features were extracted: number of candidate features for classification, number of selected classification features, the results of the best classification experiment, i.e., accuracy, sensitivity, specificity, area under the receiver operator curve (AUC), type of cross-validation used, and use of an external validation set. For the publications reporting significance testing, the number of candidate features, the number of significant features and the number of significant features after Holm-Bonferroni correction [10] were extracted.

Data synthesis and analysis

The imaging modality was divided into four categories, i.e., MRI, CT, PET/SPECT, and US. No further subdivision on the type of imaging protocol or use of contrast agent was made. Image analysis methods to estimate tumour heterogeneity were divided into four categories: non-spatial methods, local spatial distribution methods, fractal analysis, and a category consisting of filters and transforms.

Non-spatial methods (NSM). These methods characterize tumour heterogeneity by non-spatial descriptors, such as descriptors of the grey level frequency distributions: standard deviation, skewness, range, peak height, peak position, and percentile values.

Spatial grey level methods (SGLM). Methods included here extract the local spatial image intensity distribution. This category includes grey-tone spatial-dependence matrix (GTSDM) [11], neighbourhood-grey-tone-difference matrix (NGTDM) [12], run-length matrix (RLM) [13], and local binary patterns (LBP) [14]. The GTSDM, originally proposed by Haralick et al. [11] is often referred to as co-occurrence or the second-order histogram. When divided by the total number of neighbouring pixels in the image, this matrix becomes the estimate of the joint probability of two pixels at a distance along a given direction having a particular grey value. The NGTDM, originally proposed by Amadasm and King [12], is based on spatial changes in grey values by inspecting the difference between grey levels of a specific pixel and the average grey level of their surrounding neighbours. The RLM, originally proposed by Galloway [13], is subsidiary to the observation that a coarse texture would have relatively longer grey level runs compared to a fine texture. This matrix provides the information about runs of pixels with the same grey-level values in a given direction. LBP, originally proposed by Ojala et al. [14] and later modified to a rotation and scale invariant approach [15], represents local texture. In its simplest form, it labels the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary number.

Fractal analysis (FA). The third category consists of fractal analysis methods that overcome the scale problem by providing a statistical measure reflecting pattern changes as a function of scale. The two basic parameters in fractal analysis are fractal dimension (FD) and lacunarity [16]. An often used method to estimate FD is box counting [16]. This procedure systematically overlays an image with a series of grids with increasing/decreasing sizes. For each step, this procedure captures the predefined relevant features [17]. Another frequently used technique in fractal analysis is the blanket method [16], which is often used in its extended form, as described by Peleg et al. [18]. This method estimates the surface area by measuring the volume between an upper and lower blanket.

Filters and transforms (F&T). The fourth category consists of a collection of image processing algorithms that extract texture features. Examples are methods that use techniques defined in the spatial domain such as filters (Gabor filters or Law's filters) or transformations to other domains (Fourier transform, Wavelet transform, S-transform, discrete cosine transform). Since the various methods have only been used in a limited number of publications included in the present analysis, these methods were grouped together.

Publications reporting classification experiments. Publications were considered classification studies if they reported a classification result such as accuracy, sensitivity, specificity or AUC value. Only the publications where the results of the classification experiments were solely based on texture parameters were further analysed. These studies often utilise a high number of candidate features to describe a tumour. To keep the system manageable, dimensionality reduction techniques were commonly applied to select a subset of features that are relevant for the classification problem. The ratio between the number of tumours classified and the dimensionality of the feature space (e.g., the number of selected features) should be chosen sensibly [19]. Therefore, we evaluated the number of candidate features, the number of selected features, and the ratio between the number of tumours included in the study and the number of selected classification features.

Publications reporting on significance testing. A commonly used approach to test the validity of the selected features is significance testing. For heterogeneity analysis, many publications compute a large number of features. As multiple comparisons generally require a stronger level of evidence to be considered as significant, the Holm-Bonferroni correction can be applied [10]. This correction allows for the significance levels for single and multiple comparisons to be directly compared. In the included publications we evaluated whether a Holm-Bonferroni correction was applied and, if this was not the case, computed the number of significant features after correction using the available data.

Results

Figure 1 shows the results of the systematic search, with the number of articles retrieved, exclusion criteria applied, and the final number of included publications. In summary, 8,956 potentially relevant articles were identified through our search strategy. Based on the basis of the review of their titles and abstracts, 488 (5.5%) publications were considered for inclusion. All these publications were retrieved for full paper screening. In this stage, an additional 296 publications were excluded, and the remaining 192 original research publications [2, 20-208] entered the data extraction phase. In this phase, an additional 22 papers were excluded as they report the results of a similar analysis method on the same dataset as in another publication. For these publications, the most recent one was included in the analysis. In total, data from 170 studies were extracted.

General characteristics

Table 1 lists the characteristics of the included publications (after removing double publications). A publication may include more than one imaging modality, analysis method, or goal. One study (0.6%) reported on two imaging modalities, 51 studies (30%) reported on two or more analysis methods, and two studies (1.2%) reported both goals (diagnose/grading and treatment response monitoring). In total, 30 studies used dynamic contrast-enhanced imaging (17.7%), wich were conducted using MRI (n=27) or CT (n=3).

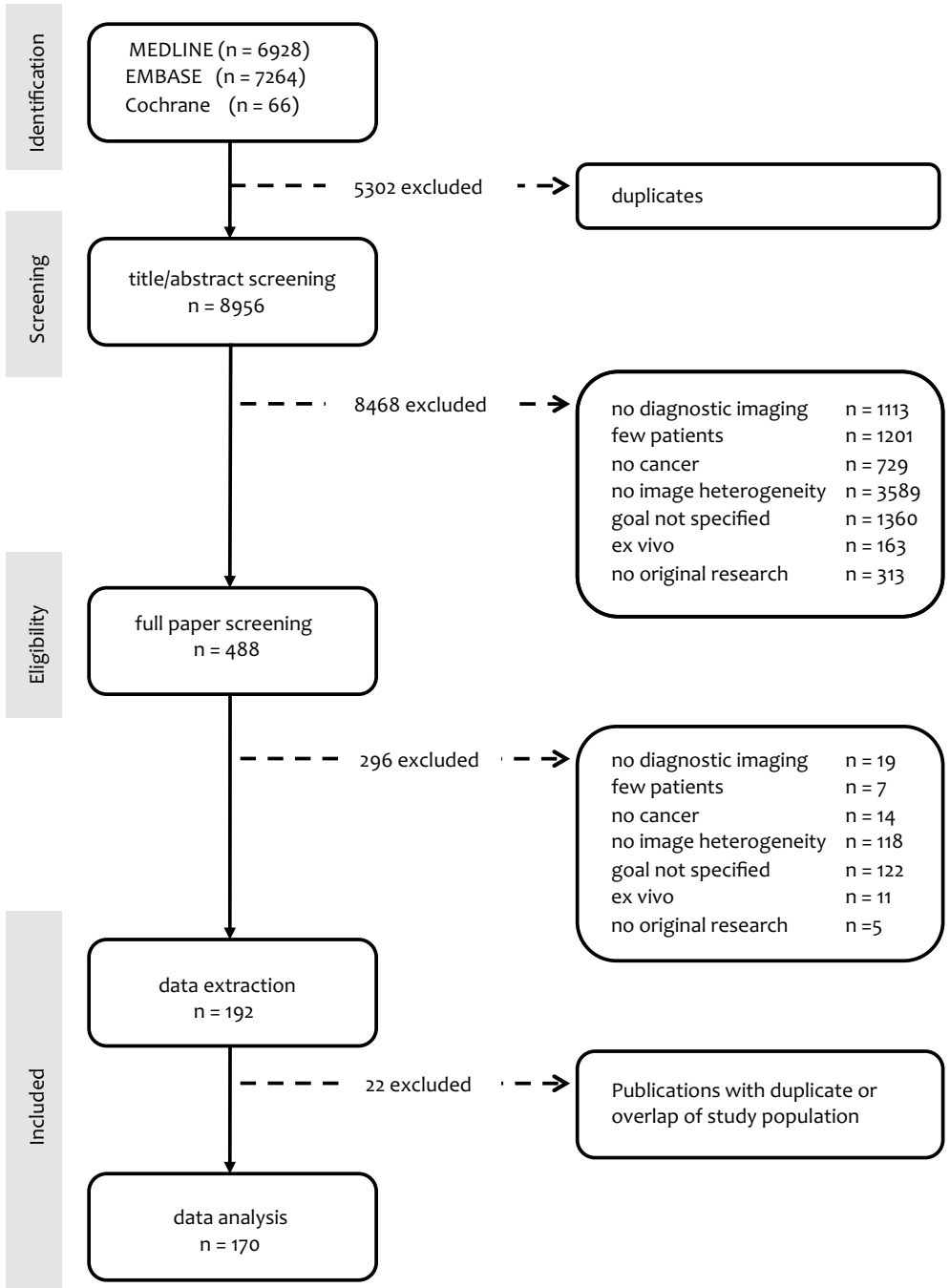


Figure 1. Prisma flow diagram [6] of articles (included and excluded) in the systematic review.

The number of imaging studies quantifying tumour heterogeneity has clearly increased since 2009 to over 40 publications biannually (Figure 2A). Figure 2B-D show (biannually) the number of publications per modality, per quantification method and per goal. Prior to 2007, heterogeneity was primarily studied based on ultrasound data. After 2007, the number of studies using MRI to quantify tumour heterogeneity increased (Figure 2B). The non-spatial method and the spatial grey-level methods are generally the most frequently used methods to analyse tumour heterogeneity. Although the number of publications using these methods has increased since 2007, their contribution to heterogeneity literature is relatively stable (Figure 2C). Over the years, the number of studies reporting tumour response monitoring ranges from 6% to 25% (Figure 2D).

Table 1. Characteristics of the included publications (n=170).

Characteristic		n	%
Imaging method	MRI	61	36%
	CT	32	19%
	PET/SPECT	13	8%
	US	65	38%
Analysis method	NSM	102	60%
	SGLM	82	48%
	FA	19	12%
	F&T	47	28%
Study goal	Diagnosis/grading/outcome prediction	146	86%
	Response monitoring	24	14%
Study type	Retrospective	102	59%
	Retrospective (with inclusion criteria)	48	28%
	Prospective	20	12%
Type of subjects	Human	160	94%
	Animal	10	6%
Type of experiment	Classification	115	68%
	Significance testing	51	30%
	Neither	4	2%

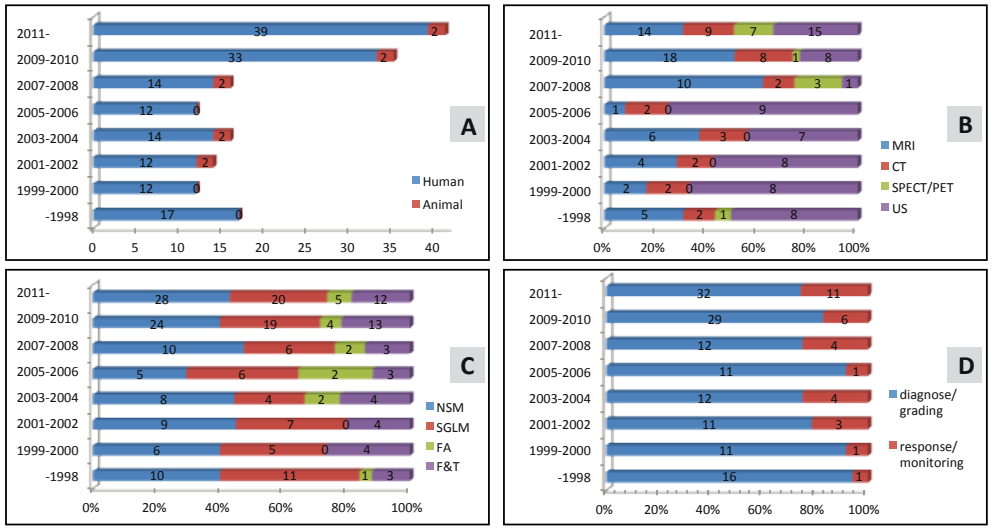


Figure 2. Number of publications reporting on tumour heterogeneity analysis for all publications bi-annually. Total number of publications (A), publications per imaging modality (B), publications per analysis method (C), and publications per goal (D).

Breast tumours have been studied in 37% of the publications. Figure 3 shows the distribution of studies per tumour location. Figure 3A shows the use of imaging modalities for quantification of tumour heterogeneity. MRI is used primarily for brain and breast tumours, CT for lung and GIST tumours, SPECT/PET for sarcoma and gynaecological tumours, and US for breast tumours and head and neck (H&N) tumours. Heterogeneity analysis of brain tumours was performed almost exclusively on MRI, while for breast tumours both MRI and US were used. Figure 3B shows the analysis methods used per tumour location. For almost all locations all methods have been used.

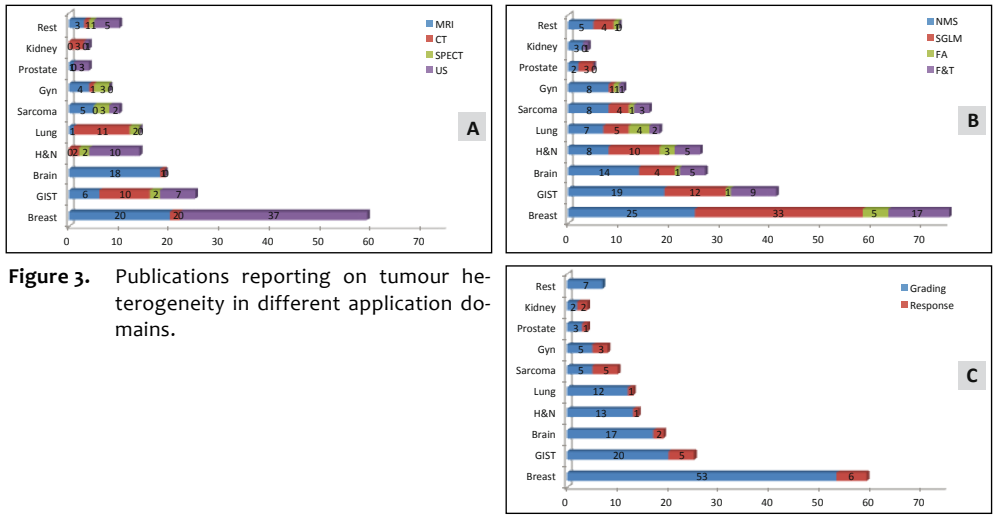
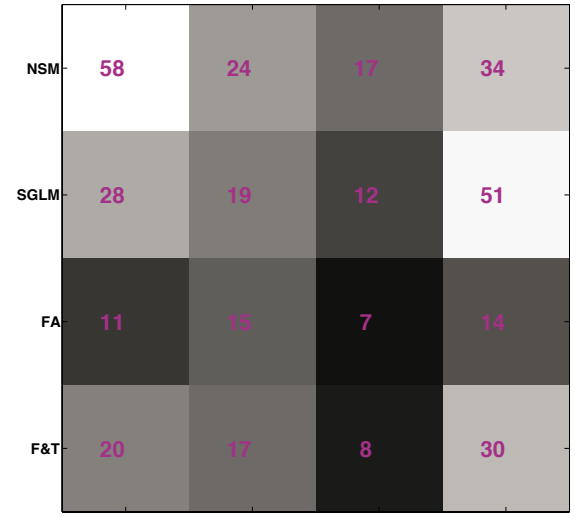


Figure 3. Publications reporting on tumour heterogeneity in different application domains.

For prostate, H&N and breast analysis the spatial grey level methods were most frequently used. Heterogeneity analyses for tumour-response monitoring were primarily performed in GIST, sarcoma, and breast (Figure 3C). Figure 4 shows the distribution (in a matrix form) over different imaging modalities and analysis method for all the included publications.

Figure 4. Numbers of publications for a specific imaging modality and analysis method.



For publications with the goal of tumour diagnosis and grading, the relation between imaging modality and analysis methods is shown in Figure 5A. In general, about 76% of all studies on tumour diagnosis and grading used either MRI or US. With 36%, SGLM is the most frequently used analysis method to grade and diagnose tumours. Figure 5B shows the relation between imaging modality and analysis method for the publications (n=24) with the goal of tumour response monitoring. MRI was used in 82% of these studies and SPECT/PET in 9%. In 3% of the studies, US based heterogeneity quantification was used for tumour-response monitoring. With 73%, NSM is the most frequently used analysis method for tumour-response monitoring.

Only 12% of all studies utilized a prospective study design. Figure 6A shows the relation between imaging modality and analysis method used for tumour diagnosis, grading and outcome prediction in prospective studies (n=6). All modalities appeared to be primarily analysed by NSM. Figure 6B shows the relation between imaging modality and analysis method, for publications reporting a prospective study design for tumour-response monitoring (n=14). Again, the majority of all data were analysed by NSM. US and CT are rarely used for heterogeneity quantification in prospective tumour-response monitoring studies.

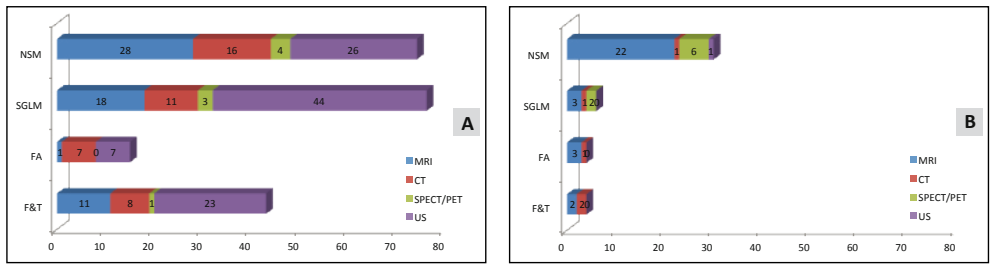


Figure 5. All included publications reporting diagnosis/grading/outcome prediction (A) and tumour-response monitoring (B).

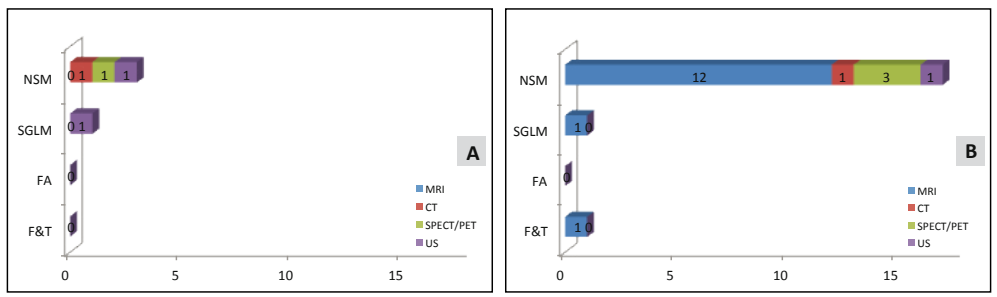


Figure 6. Imaging modalities and analysis method for tumour diagnosis/grading/outcome prediction (A) and tumour-response monitoring (B) in studies with a prospective study design.

Publications reporting classification experiments

Of all included studies, 68% reported classification experiments and 30% reported significance testing. The remaining 2% either reported no quantitative results or the experiments were incompletely described. Twenty studies only reported the results of classification experiments in which texture features were combined with non-texture features. For these papers, because it was not possible to extract the performance of the texture features or the added value of the texture features, the results were not included in further analysis. Additional eight papers with the exact number of generated or selected features unclear, were excluded from further analysis. Of the papers reporting on classification, 47% used US and 38% used MRI. In 41% of the classification papers, features originating from different texture analysis methods were combined. Figure 7 shows the relation between the number of feature candidates and selected features used in classification experiments for different imaging modalities and different analysis methods. For the papers indicated on the dotted line, no feature selection procedure was performed. The number of candidate features ranged from 1 to 1560 (median=20) while the number of selected features ranged from 1 to 208 (median=3). About half of the papers describing a classification experiment reported the use of cross-validation or training-

test sets as a technique to limit the effect of overfitting on the available data. Figure 7B shows that the papers using a combination of texture analysis methods generally use more features than papers that use one method. Moreover, in the multi-methods papers dimensionality reduction techniques are not as rigorously applied as in the SGLM papers.

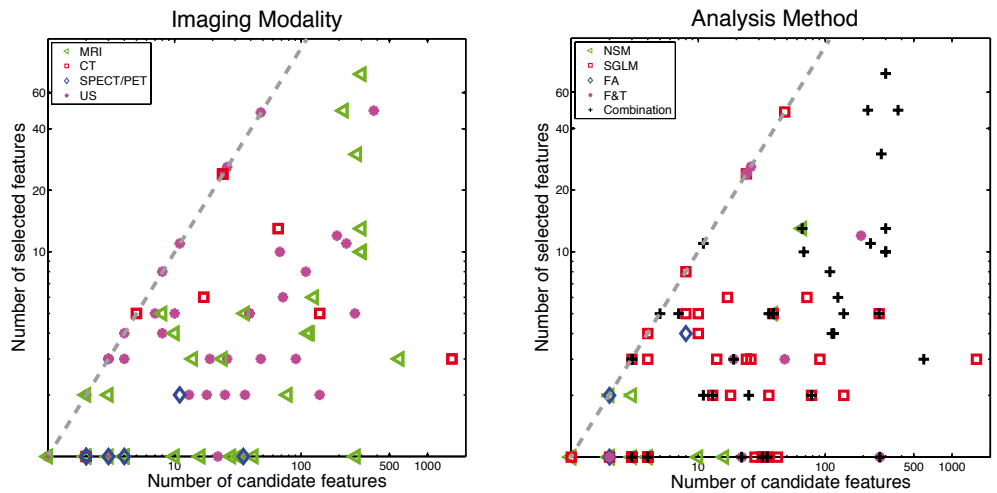


Figure 7. Number of features used in classification experiments for different imaging modalities (left) and for different analysis methods (right).

In the classification experiments, one or more of the following performance measures were reported: sensitivity, specificity, accuracy, or AUC. Figure 8 shows the performance measures per imaging modality, and Figure 9 the performance measures per analysis method. In both figures, the reported performance is depicted as a function of the tumour-feature ratio. It is clear that the combination of features from different methods leads to a high number of candidate features. In general, the tumour-feature ratio ranged from 0.26 to 502 (median=21), with (on average) 28% of the publications showing a tumour-feature ratio below 10. Studies using the combined method and those using F&T had the most papers with a tumour-feature ratio ≤ 10, i.e. 42% and 30%, respectively. Regarding imaging modality, CT has the highest percentage (50%) of classification experiments with a tumour-feature ratio < 10.

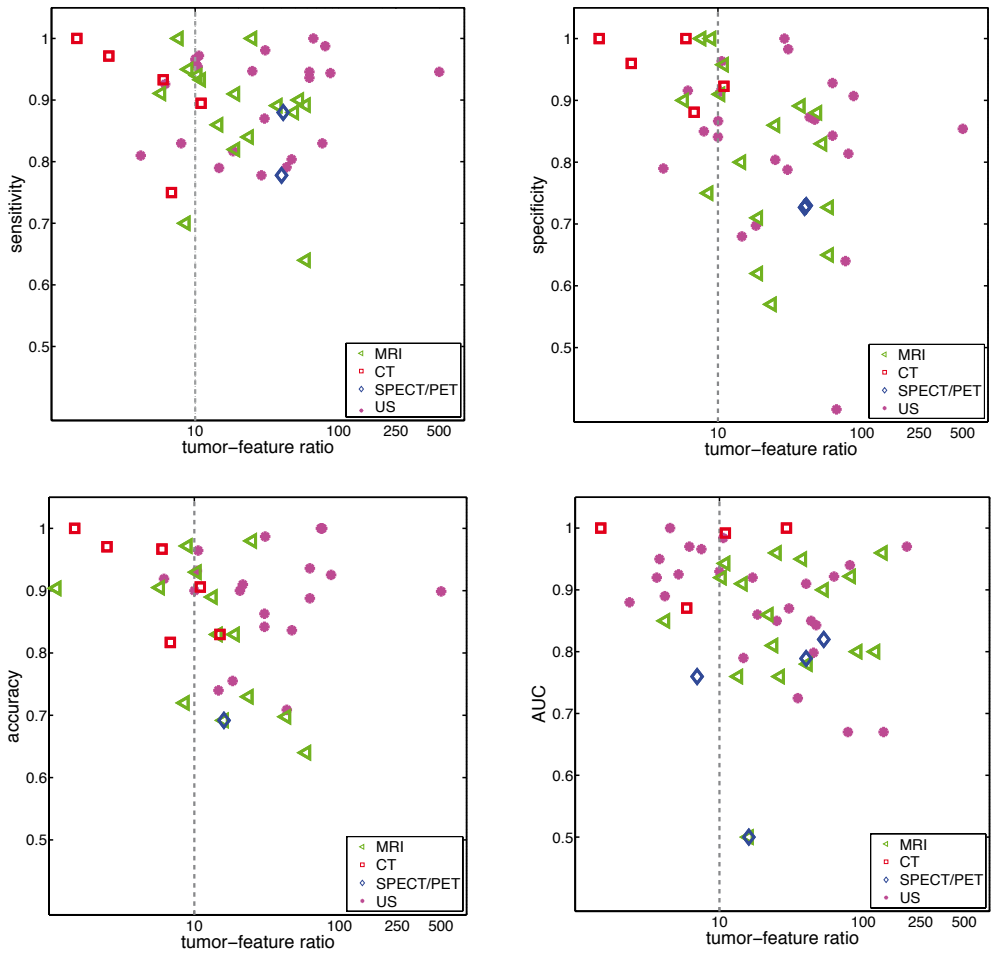


Figure 8. Performance measures (sensitivity, specificity, accuracy and AUC) as a function of tumour-feature ratio in the classification experiments. The scatter plot shows each imaging modality separately. Dotted line represents the ratio of 10 tumours per selected feature.

Only two papers reported an AUC of 0.5; all other papers reported higher values. This is most likely caused by publication bias: only the positive performance of heterogeneity features reach the journals. No relation was observed between performance measures and the imaging modality or the type of analysis method used. However, there is a negative correlation between the logarithm of the number of tumour-feature ratio and the AUC (r=-0.29, p<0.05). This correlation can be the result of overfitting when less tumours per feature are available.

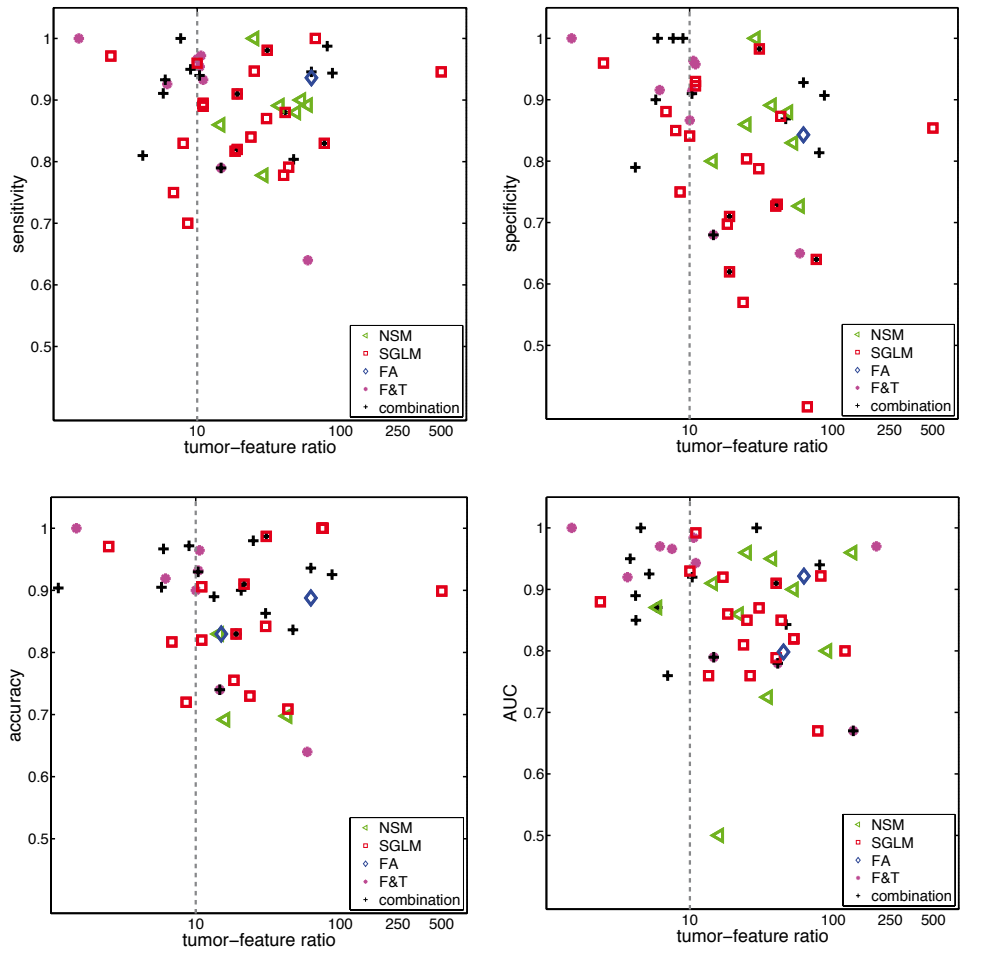


Figure 9. Performance measures (sensitivity, specificity, accuracy and AUC) as a function of tumour-feature ratio in the classification experiments. The scatter plot shows each analysis method separately. Dotted line represents the ratio of 10 tumours per selected feature.

Publications reporting significance testing

Thirty percent (N=43) reported significance testing with the number of features ranging from 1 to 34 (median= 4). Of all papers reporting significance testing included in this study, 34% were based on MRI, 30% on CT, 16% on SPECT/PET, and 20% on US. Similarly, in 58% of the cases data were analysed using NSM, 9% using SGLM, 5% using FA, 7% using F&T, and 21% using a combination of methods. The number of significant features, as reported by the authors, ranged from 0 to 12. Since multiple comparisons generally require a stronger level of evidence to be considered as significant, the Holm-Bonferroni correction [10] was applied by the original research authors, or

by the authors of this review paper. This correction allows for the significance levels for single and multiple comparisons to be directly compared. For 7 papers the correction could not be performed due to missing information. After Holm-Bonferroni correction, the number of significant features ranged from 0 to 6. Figure 10 shows the number of significant features before and after Holm-Bonferroni correction per imaging modality (left) and per analysis method (right). In 36% of the papers the number of significant features decreased significantly after correction. No relation with imaging modality or analysis method was observed.

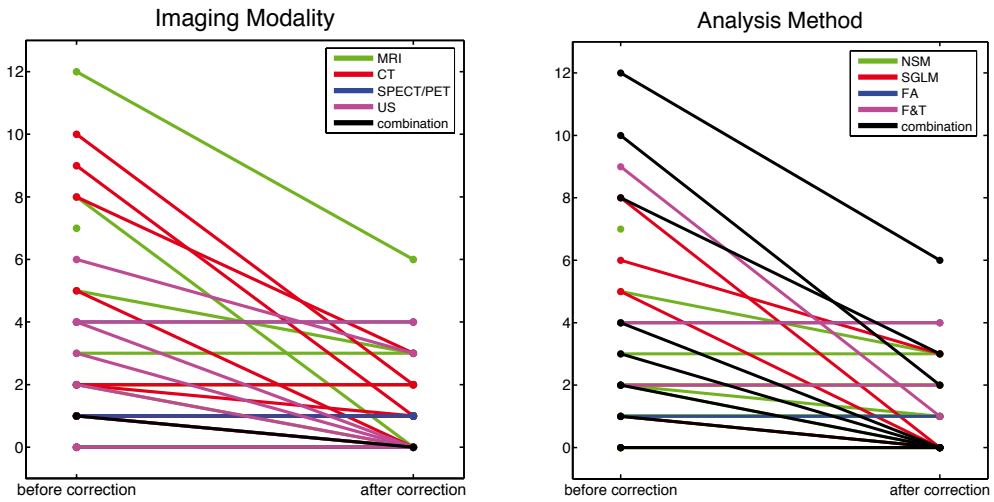


Figure 10. Number of significant features before and after Holm-Bonferroni correction in publications reporting on significance testing per image modality (left) and per analysis method (right).

Discussion

In this systematic review we investigated the use and performance of heterogeneity or texture quantification methods in radiological images for tumour grading, differentiation, outcome prediction and treatment response monitoring. Of the 8,956 unique studies identified in the database search, 170 studies reported on heterogeneity as an imaging biomarker in tumour imaging. Since 2009 there has been a growing number of publications reporting on tumour heterogeneity quantification. Up to 2006 most heterogeneity papers were based on US, whereas after 2007 the number of papers using MRI has increased. The NSM and the SGLM are the methods most frequently used during the studied period. Most papers focus on heterogeneity quantification for differentiation, grading or outcome prediction. However, the number of papers with the goal of response monitoring has increased recently. In

tumour heterogeneity quantification, US is the most frequently used imaging modality for tumour differentiation and grading, and MRI is the most frequently used modality for treatment response monitoring. For monitoring of treatment response, NSM is the most frequently used method. For tumour grading and differentiation, all methods are evenly distributed over all modalities.

The performance of the heterogeneity features was mostly (68%) evaluated by classification experiments reporting performance measures such as accuracy, sensitivity, specificity and AUC. Papers reporting only on the results of combined texture features with other features, were excluded from the analysis. Some authors selectively report on sensitivity without mentioning the specificity. The AUC is the preferred measure to report performance, as it is more comprehensive compared to a measure based on a single threshold such as accuracy. Only 59% of the publications reporting classification results reported the use of cross-validation as a technique to limit the effect of overfitting on the available data. We found no relation between the performance measures and the modality nor with the analysis method used. However, a negative correlation was found between the number of tumours per selected feature and the AUC. If more tumours were available per selected feature the accuracy was lower. This correlation might be the result of overfitting of the data when less tumours per feature are available.

Publications reporting significance testing often did not perform a correction of the significance levels for multiple comparisons. For 7 papers, due to missing information, retrospective Holm-Bonferroni correction could not be performed by the authors. For 36% of the papers the number of significant features decreased after Holm-Bonferroni correction.

The number of prospective studies is small, i.e. only 12% of all studies. These studies are primarily based on MRI and report NSM features. Although the use of retrospectively collected data is necessary to develop, test and evaluate heterogeneity as a biomarker for tumour grading, differentiation, outcome prediction and treatment response monitoring, the real test is to evaluate the performance of the developed features in a prospective study design. In a retrospective study design the criteria for the inclusion of cases are often not (or not clearly) described, so that the performance of the heterogeneity feature may be overestimated. By using a prospective study design with clear inclusion criteria the real performance of heterogeneity features can be more reliably assessed.

Moreover, in most studies the performance of the heterogeneity feature is evaluated without taking into account currently accepted clinical features, such as mean signal intensity, tumour size, grade or border regularity. Other studies report only the combined classification performance of heterogeneity and clinical features.

A large number of publications even use the mean signal intensity as a feature to estimate tumour heterogeneity, even though this is clearly not a heterogeneity measure (i.e., mean signal intensity does not measure intra-tumour heterogeneity). Based on these types of studies, it is not possible to evaluate the added value of heterogeneity over and above the currently accepted clinical features. While the researcher is interested in the performance of the feature in itself, the clinician is interested in the additional value of the feature on top of the already available clinical features. Since the quantification of heterogeneity is usually more complex and computationally more costly than computing the mean intensity, the benefit of the added effort for characterization of heterogeneity should be justified.

To enable the translation of imaging biomarkers from the research stage to clinical practice, research should focus on studies that investigate the additional value of the proposed heterogeneity biomarker on top of the already established clinical markers. These studies should have a well-considered design by taking into account the requirements from pattern recognition: i.e. a balanced number of subjects and features, cross-validation, independent test data sets, and a prospective study design. Satisfying these requirements implies that the benefit of heterogeneity features can be assessed more reliably.

References

1. Yang, X. and M.V. Knopp, *Quantifying tumor vascular heterogeneity with dynamic contrast-enhanced magnetic resonance imaging: a review.* J Biomed Biotechnol, 2011. 2011: p. 732848.
2. Hayes, C., A.R. Padhani, and M.O. Leach, *Assessing changes in tumour vascular function using dynamic contrast-enhanced magnetic resonance imaging.* NMR Biomed, 2002. 15(2): p. 154-63.
3. van Rijswijk, C.S., et al., *Dynamic contrast-enhanced MR imaging in monitoring response to isolated limb perfusion in high-grade soft tissue sarcoma: initial results.* Eur Radiol, 2003. 13(8): p. 1849-58.
4. Pickles, M.D., et al., *Prognostic value of pre-treatment DCE-MRI parameters in predicting disease free and overall survival for breast cancer patients undergoing neoadjuvant chemotherapy.* Eur J Radiol, 2009. 71(3): p. 498-505.
5. Ng, C.K., H.N. Pemberton, and J.S. Reis-Filho, *Breast cancer intratumor genetic heterogeneity: causes and implications.* Expert Rev Anticancer Ther, 2012. 12(8): p. 1021-32.
6. Moher, D., et al., *Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement.* PLoS Med, 2009. 6(7): p. e1000097.
7. Alic, L., J.F. Veenland, and W.J. Niessen, *Quantification of heterogeneity as a biomarker in tumour imaging: a systematic review*. 2013, PROSPERO. http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42013003634.
8. Asselin, M.C., et al., *Quantifying heterogeneity in human tumours using MRI and PET.* Eur J Cancer, 2012. 48(4): p. 447-55.
9. Davnall, F., et al., *Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice?* Insights Imaging, 2012. 3(6): p. 573-89.
10. Holm, S., *A simple sequentially rejective multiple test procedure.* Scandinavian Journal of Statistics, 1979. 6(2): p. 65-70.
11. Haralick, R.M., K. Shanmugam, and J. Dinstein, *Textural features for image classification.* IEEE Trans Syst Man Cybern, 1973. 6: p. 12.
12. Amadasun, M. and R. King, *Textural features corresponding to textural properties.* IEEE Transactions on Systems, Man and Cybernetics, 1989. 19(5): p. 1264-1273.
13. Galloway, M.M., *Texture analysis using gray level run lengths.* Computer Graphics and Image Processing, 1975. 4(2): p. 172-179.
14. Ojala, T., M. Pietikäinen, and D. Harwood, *A comparative study of texture measures with classification based on feature distributions.* Pattern Recognition, 1996. 29(1): p. 51-59.
15. Ojala, T., M. Pietikäinen, and T. Mäenpää, *Multiresolution gray-scale and rotation invariant texture classification with local binary patterns.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002. 24(7): p. 971-987.
16. Mandelbrot, B.B., *The fractal geometry of nature*. Updated and augm. ed. 1983, New York: W.H. Freeman. 468 p.
17. Smith, T.G., Jr., G.D. Lange, and W.B. Marks, *Fractal methods and results in cellular morphology--dimensions, lacunarity and multifractals.* J Neurosci Methods, 1996. 69(2): p. 123-36.
18. Peleg, S., et al., *Multiple resolution texture analysis and classification.* IEEE Trans Pattern Anal Mach Intell, 1984. 6(4): p. 518-23.
19. Young, T.Y. and T.W. Calvert, *Classification, estimation, and pattern recognition*. 1974, New York,: American Elsevier Pub. Co. xiv, 366 p.
20. Chang, R.F., et al., *Support vector machines for diagnosis of breast tumors on US images.* Acad Radiol, 2003. 10(2): p. 189-97.
21. Chen, C.Y., et al., *Computer-aided diagnosis of soft-tissue tumors using sonographic morphologic and texture features.* Acad Radiol, 2009. 16(12): p. 1531-8.

22. Gensure, R.H., et al., *Evaluation of hepatic tumor response to yttrium-90 radioembolization therapy using texture signatures generated from contrast-enhanced CT images.* Acad Radiol, 2012. 19(10): p. 1201-7.
23. Giger, M.L., et al., *Computerized analysis of lesions in US images of the breast.* Acad Radiol, 1999. 6(11): p. 665-74.
24. Holli, K., et al., *Characterization of breast cancer types by texture analysis of magnetic resonance images.* Acad Radiol, 2010. 17(2): p. 135-41.
25. Huang, Y.L., J.H. Chen, and W.C. Shen, *Diagnosis of hepatic tumors with texture analysis in nonenhanced computed tomography images.* Acad Radiol, 2006. 13(6): p. 713-20.
26. Kim, K.G., et al., *Computerized scheme for assessing ultrasonographic features of breast masses.* Acad Radiol, 2005. 12(1): p. 58-66.
27. Kuo, W.J., et al., *Computer-aided diagnosis of breast tumors with different US systems.* Acad Radiol, 2002. 9(7): p. 793-9.
28. Liu, F., et al., *Optimization of time-to-peak analysis for differentiating malignant and benign breast lesions with dynamic contrast-enhanced MRI.* Acad Radiol, 2011. 18(6): p. 694-704.
29. McLaren, C.E., et al., *Prediction of malignant breast lesions from MRI features: a comparison of artificial neural network and logistic regression techniques.* Acad Radiol, 2009. 16(7): p. 842-51.
30. Nie, K., et al., *Quantitative analysis of lesion morphology and texture features for diagnostic prediction in breast MRI.* Acad Radiol, 2008. 15(12): p. 1513-25.
31. Shah, S.K., et al., *Computer-aided diagnosis of the solitary pulmonary nodule.* Acad Radiol, 2005. 12(5): p. 570-5.
32. Wu, W.J. and W.K. Moon, *Ultrasound breast tumor image computer-aided diagnosis with texture and morphological features.* Acad Radiol, 2008. 15(7): p. 873-80.
33. Thijssen, J.M., et al., *Echographic differentiation of intraocular melanomas by computer analysis.* Acta Ophthalmol Suppl, 1992(204): p. 26-34.
34. Abramyuk, A., et al., *Preliminary assessment of dynamic contrast-enhanced CT implementation in pretreatment FDG-PET/CT for outcome prediction in head and neck tumors.* Acta Radiol, 2010. 51(7): p. 793-9.
35. Jung, S.C., J.Y. Cho, and S.H. Kim, *Subtype differentiation of small renal cell carcinomas on three-phase MDCT: usefulness of the measurement of degree and heterogeneity of enhancement.* Acta Radiol, 2012. 53(1): p. 112-8.
36. Kim, D.Y., et al., *Pulmonary nodule detection using chest CT images.* Acta Radiol, 2003. 44(3): p. 252-7.
37. Kjaer, L., et al., *Texture analysis in quantitative MR imaging. Tissue characterisation of normal brain and intracranial tumours at 1.5 T.* Acta Radiol, 1995. 36(2): p. 127-35.
38. Rehn, S., et al., *Quantification of inhomogeneities in malignancy grading of non-Hodgkin lymphoma with MR imaging.* Acta Radiol, 1993. 34(1): p. 3-9.
39. Barajas, R.F., Jr., et al., *Diffusion-weighted MR imaging derived apparent diffusion coefficient is predictive of clinical outcome in primary central nervous system lymphoma.* AJNR Am J Neuroradiol, 2010. 31(1): p. 60-6.
40. Ma, J.H., et al., *Differentiation among glioblastoma multiforme, solitary metastatic tumor, and lymphoma using whole-tumor histogram analysis of the normalized cerebral blood volume in enhancing and perienhancing lesions.* AJNR Am J Neuroradiol, 2010. 31(9): p. 1699-706.
41. Adler, R.S., et al., *Evaluation of soft-tissue masses using segmented color Doppler velocity images: preliminary observations.* AJR Am J Roentgenol, 1999. 172(3): p. 781-8.
42. Herts, B.R., et al., *Enhancement characteristics of papillary renal neoplasms revealed on triphasic helical CT of the kidneys.* AJR Am J Roentgenol, 2002. 178(2): p. 367-72.
43. Hirano, M., et al., *Diffusion-weighted imaging of breast masses: comparison of diagnostic performance using various apparent diffusion coefficient parameters.* AJR Am J Roentgenol, 2012. 198(3): p. 717-22.
44. Huang, B., et al., *Nasopharyngeal carcinoma: investigation of intratumoral heterogeneity with FDG PET/CT.* AJR Am J Roentgenol, 2012. 199(1): p. 169-74.

45. Plant, R.L., *Image analysis of benign and malignant neck masses.* Ann Otol Rhinol Laryngol, 1998. 107(8): p. 689-96.
46. Eliat, P.A., et al., *Is magnetic resonance imaging texture analysis a useful tool for cell therapy in vivo monitoring?* Anticancer Res, 2001. 21(6A): p. 3857-60.
47. Chang, R.F., et al., *Computer-aided diagnosis for surgical office-based breast ultrasound.* Arch Surg, 2000. 135(6): p. 696-9.
48. Chen, D.R., et al., *Computer-aided diagnosis for 3-dimensional breast ultrasonography.* Arch Surg, 2003. 138(3): p. 296-302.
49. Iakovidis, D.K., E.G. Keramidas, and D. Maroulis, *Fusion of fuzzy statistical distributions for classification of thyroid ultrasound patterns.* Artif Intell Med, 2010. 50(1): p. 33-41.
50. Mougiakakou, S.G., et al., *Differential diagnosis of CT focal liver lesions using texture features, feature selection and ensemble driven classifiers.* Artif Intell Med, 2007. 41(1): p. 25-37.
51. Lee, C.C. and C.Y. Shih, *Learning patterns of liver masses using improved RBF networks.* Biomedical Engineering - Applications, Basis and Communications, 2010. 22(2): p. 137-147.
52. O'Sullivan, F., S. Roy, and J. Eary, *A statistical measure of tissue heterogeneity with application to 3D PET sarcoma data.* Biostatistics, 2003. 4(3): p. 433-48.
53. Farace, P., et al., *Inhibition of tyrosine kinase receptors by SU6668 promotes abnormal stromal development at the periphery of carcinomas.* Br J Cancer, 2009. 100(10): p. 1575-80.
54. O'Connor, J.P., et al., *DCE-MRI biomarkers of tumour heterogeneity predict CRC liver metastasis shrinkage following bevacizumab and FOLFOX-6.* Br J Cancer, 2011. 105(1): p. 139-45.
55. Karahaliou, A., et al., *Assessing heterogeneity of lesion enhancement kinetics in dynamic contrast-enhanced MRI for breast cancer diagnosis.* Br J Radiol, 2010. 83(988): p. 296-309.
56. Tse, D.M., et al., *A computer-aided algorithm to quantitatively predict lymph node status on MRI in rectal cancer.* Br J Radiol, 2012. 85(1017): p. 1272-8.
57. Ganeshan, B., et al., *Texture analysis of non-small cell lung cancer on unenhanced computed tomography: initial evidence for a relationship with tumour glucose metabolism and stage.* Cancer Imaging, 2010. 10: p. 137-43.
58. Nguyen, P., et al., *Optical differentiation between malignant and benign lymphadenopathy by grey scale texture analysis of endobronchial ultrasound convex probe images.* Chest, 2012. 141(3): p. 709-15.
59. Brown, R., et al., *The use of magnetic resonance imaging to noninvasively detect genetic signatures in oligodendroglioma.* Clin Cancer Res, 2008. 14(8): p. 2357-62.
60. Kidd, E.A. and P.W. Grigsby, *Intratumoral metabolic heterogeneity of cervical cancer.* Clin Cancer Res, 2008. 14(16): p. 5236-41.
61. Chen, D.R., et al., *Classification of breast ultrasound images using fractal feature.* Clin Imaging, 2005. 29(4): p. 235-45.
62. Chen, S.J., et al., *Characterizing the major sonographic textural difference between metastatic and common benign lymph nodes using support vector machine with histopathologic correlation.* Clin Imaging, 2012. 36(4): p. 353-9 e2.
63. Ganeshan, B., et al., *Tumour heterogeneity in oesophageal cancer assessed by CT texture analysis: preliminary evidence of an association with tumour metabolism, stage, and survival.* Clin Radiol, 2012. 67(2): p. 157-64.
64. Retico, A., et al., *Pleural nodule identification in low-dose and thin-slice lung computed tomography.* Comput Biol Med, 2009. 39(12): p. 1137-44.
65. Mitrea, D., et al., *Abdominal tumor characterization and recognition using superior-order cooccurrence matrices, based on ultrasound images.* Comput Math Methods Med, 2012. 2012: p. 348135.
66. Chen, D.R., Y.L. Huang, and S.H. Lin, *Computer-aided diagnosis with textural features for breast lesions in sonograms.* Comput Med Imaging Graph, 2011. 35(3): p. 220-6.
67. McNitt-Gray, M.F., et al., *The effects of co-occurrence matrix based texture parameters on the classification of solitary pulmonary nodules imaged on computed tomography.* Comput Med Imaging Graph, 1999. 23(6): p. 339-48.

68. Mittal, D., et al., *Neural network based focal liver lesion diagnosis using ultrasound images.* Comput Med Imaging Graph, 2011. 35(4): p. 315-23.
69. Piliouras, N., et al., *Development of the cubic least squares mapping linear-kernel support vector machine classifier for improving the characterization of breast lesions on ultrasound.* Comput Med Imaging Graph, 2004. 28(5): p. 247-55.
70. Sivaramakrishna, R., et al., *Texture analysis of lesions in breast ultrasound images.* Comput Med Imaging Graph, 2002. 26(5): p. 303-7.
71. Strzelecki, M., et al., *Classification and segmentation of intracardiac masses in cardiac tumor echocardiograms.* Comput Med Imaging Graph, 2006. 30(2): p. 95-107.
72. Wu, W.J., S.W. Lin, and W.K. Moon, *Combining support vector machine with genetic algorithm to classify ultrasound breast tumor images.* Comput Med Imaging Graph, 2012. 36(8): p. 627-33.
73. Acharya, U.R., et al., *ThyroScreen system: high resolution ultrasound thyroid image characterization into benign and malignant classes using novel combination of texture and discrete wavelet transform.* Comput Methods Programs Biomed, 2012. 107(2): p. 233-41.
74. Glotsos, D., et al., *A multi-classifier system for the characterization of normal, infectious, and cancerous prostate tissues employing transrectal ultrasound images.* Comput Methods Programs Biomed, 2010. 97(1): p. 53-61.
75. Rodrigues, P.S., et al., *A new methodology based on q-entropy for breast lesion classification in 3-D ultrasound images.* Conf Proc IEEE Eng Med Biol Soc, 2006. 1: p. 1048-51.
76. Valavanis, I.K., et al., *Evaluation of texture features in hepatic tissue characterization from non-enhanced CT images.* Conf Proc IEEE Eng Med Biol Soc, 2007. 2007: p. 3741-4.
77. Raja, J.V., et al., *Texture analysis of CT images in the characterization of oral cancers involving buccal mucosa.* Dentomaxillofac Radiol, 2012. 41(6): p. 475-80.
78. Yoshiura, K., et al., *Ultrasonographic texture characterization of salivary and neck masses using two-dimensional gray-scale clustering.* Dentomaxillofac Radiol, 1997. 26(6): p. 332-6.
79. Cao, K., et al., *Apparent diffusion coefficient of diffusion weighted MRI in endometrial carcinoma-Relationship with local invasiveness.* Eur J Radiol, 2012. 81(8): p. 1926-30.
80. Hirning, T., et al., *Quantification and classification of echographic findings in the thyroid gland by computerized B-mode texture analysis.* Eur J Radiol, 1989. 9(4): p. 244-7.
81. Ng, F., et al., *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, 2013. 82(2): p. 342-8.
82. Wang, H., et al., *Multilevel binomial logistic prediction model for malignant pulmonary nodules based on texture features of CT image.* Eur J Radiol, 2010. 74(1): p. 124-9.
83. Engelbrecht, M.R., et al., *Follow-up of Wilms' tumour during pre-operative chemotherapy by qualitative and quantitative sonography.* Eur J Ultrasound, 1998. 8(3): p. 157-65.
84. Cui, C., et al., *Quantitative analysis and prediction of regional lymph node status in rectal cancer based on computed tomography imaging.* Eur Radiol, 2011. 21(11): p. 2318-25.
85. Ganeshan, B., et al., *Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: a potential marker of survival.* Eur Radiol, 2012. 22(4): p. 796-802.
86. Newell, D., et al., *Selection of diagnostic features on breast MRI to differentiate between malignant and benign lesions using computer-aided diagnosis: differences in lesions presenting as mass and non-mass-like enhancement.* Eur Radiol, 2010. 20(4): p. 771-81.
87. Loren, D.E., et al., *Computer-assisted analysis of lymph nodes detected by EUS in patients with esophageal carcinoma.* Gastrointest Endosc, 2002. 56(5): p. 742-6.
88. Al-Kadi, O.S. and D. Watson, *Texture analysis of aggressive and nonaggressive lung tumor CE CT images.* IEEE Trans Biomed Eng, 2008. 55(7): p. 1822-30.
89. Gletsos, M., et al., *A computer-aided diagnostic system to characterize CT focal liver lesions: design and optimization of a neural network classifier.* IEEE Trans Inf Technol Biomed, 2003. 7(3): p. 153-62.
90. O'Sullivan, F., et al., *A statistical modeling approach to the analysis of spatial patterns of FDG-PET uptake in human sarcoma.* IEEE Trans Med Imaging, 2011. 30(12): p. 2059-71.

91. Donohue, K.D., et al., *Analysis and classification of tissue with scatterer structure templates.* IEEE Trans Ultrason Ferroelectr Freq Control, 1999. 46(2): p. 300-10.
92. Sachdeva, J., et al., *A dual neural network ensemble approach for multiclass brain tumor classification.* Int j numer method biomed eng, 2012. 28(11): p. 1107-20.
93. Cao, Y., et al., *Survival prediction in high-grade gliomas by MRI perfusion before and during early stage of RT [corrected].* Int J Radiat Oncol Biol Phys, 2006. 64(3): p. 876-85.
94. de Lussanet, Q.G., et al., *Dynamic contrast-enhanced magnetic resonance imaging of radiation therapy-induced microcirculation changes in rectal cancer.* Int J Radiat Oncol Biol Phys, 2005. 63(5): p. 1309-15.
95. Howe, F.A., et al., *Vessel size index magnetic resonance imaging to monitor the effect of antivascular treatment in a rodent tumor model.* Int J Radiat Oncol Biol Phys, 2008. 71(5): p. 1470-6.
96. Jansen, J.F., et al., *Tumor metabolism and perfusion in head and neck squamous cell carcinoma: pretreatment multimodality imaging with 1H magnetic resonance spectroscopy, dynamic contrast-enhanced MRI, and [18F]FDG-PET.* Int J Radiat Oncol Biol Phys, 2012. 82(1): p. 299-307.
97. Klein, H.M., et al., *Pattern recognition system for focal liver lesions using "crisp" and "fuzzy" classifiers.* Invest Radiol, 1996. 31(1): p. 6-10.
98. Issa, B., D.L. Buckley, and L.W. Turnbull, *Heterogeneity analysis of Gd-DTPA uptake: improvement in breast lesion differentiation.* J Comput Assist Tomogr, 1999. 23(4): p. 615-21.
99. Kido, S., et al., *Fractal analysis of small peripheral pulmonary nodules in thin-section CT: evaluation of the lung-nodule interfaces.* J Comput Assist Tomogr, 2002. 26(4): p. 573-8.
100. Kido, S., et al., *Fractal analysis of internal and peripheral textures of small peripheral bronchogenic carcinomas in thin-section computed tomography: comparison of bronchioloalveolar cell carcinomas with nonbronchioloalveolar cell carcinomas.* J Comput Assist Tomogr, 2003. 27(1): p. 56-61.
101. Mussurakis, S., P. Gibbs, and A. Horsman, *Peripheral enhancement and spatial contrast uptake heterogeneity of primary breast tumours: quantitative assessment with dynamic MRI.* J Comput Assist Tomogr, 1998. 22(1): p. 35-46.
102. Agner, S.C., et al., *Textural kinetics: a novel dynamic contrast-enhanced (DCE)-MRI feature for breast lesion classification.* J Digit Imaging, 2011. 24(3): p. 446-63.
103. Kim, K.G., J.H. Kim, and B.G. Min, *Comparative analysis of texture characteristics of malignant and benign tumors in breast ultrasonograms.* J Digit Imaging, 2001. 14(2 Suppl 1): p. 208-10.
104. Prescott, J.W., et al., *Temporal analysis of tumor heterogeneity and volume for cervical cancer treatment outcome prediction: preliminary evaluation.* J Digit Imaging, 2010. 23(3): p. 342-57.
105. Zhu, Y., et al., *Feature selection and performance evaluation of support vector machine (SVM)-based classifier for differentiating benign and malignant pulmonary nodules by computed tomography.* J Digit Imaging, 2010. 23(1): p. 51-65.
106. Harrison, L.C., et al., *Non-Hodgkin lymphoma response evaluation with MRI texture classification.* J Exp Clin Cancer Res, 2009. 28: p. 87.
107. Juntu, J., et al., *Machine learning study of several classifiers trained with texture analysis features to differentiate benign from malignant soft-tissue tumors in T1-MRI images.* J Magn Reson Imaging, 2010. 31(3): p. 680-9.
108. Li, X., et al., *Analysis of the spatial characteristics of metabolic abnormalities in newly diagnosed glioma patients.* J Magn Reson Imaging, 2002. 16(3): p. 229-37.
109. Mayerhoefer, M.E., et al., *Texture-based classification of focal liver lesions on MRI at 3.0 Tesla: a feasibility study in cysts and hemangiomas.* J Magn Reson Imaging, 2010. 32(2): p. 352-9.
110. Mayr, N.A., et al., *Pixel analysis of MR perfusion imaging in predicting radiation therapy outcome in cervical cancer.* J Magn Reson Imaging, 2000. 12(6): p. 1027-33.
111. Meinel, L.A., et al., *Breast MRI lesion classification: improved performance of human readers with a backpropagation neural network computer-aided diagnosis (CAD) system.* J Magn Reson Imaging, 2007. 25(1): p. 89-95.
112. Peeters, F., et al., *Early (72-hour) detection of radiotherapy-induced changes in an experimental tumor model using diffusion-weighted imaging, diffusion tensor imaging, and Q-space imaging parameters: a comparative study.* J Magn Reson Imaging, 2012. 35(2): p. 409-17.

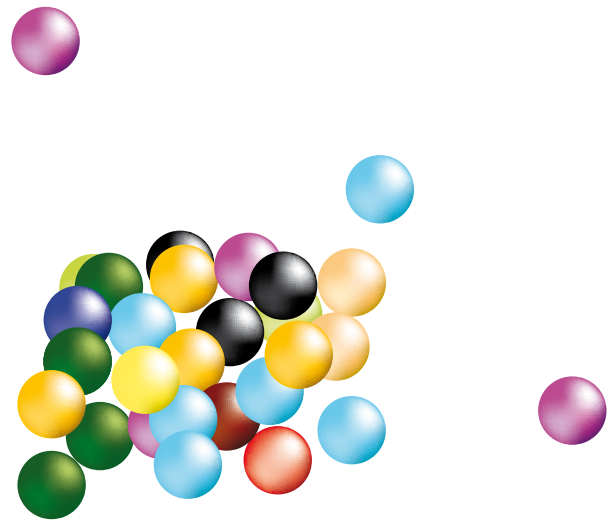
113. Sinha, S., et al., *Multifeature analysis of Gd-enhanced MR images of breast lesions.* J Magn Reson Imaging, 1997. 7(6): p. 1016-26.
114. Padma, A. and R. Sukanesh, *Combined texture feature analysis of segmentation and classification of benign and malignant tumour CT slices.* J Med Eng Technol, 2013. 37(1): p. 1-9.
115. Sasikala, M. and N. Kumaravel, *A wavelet-based optimal texture feature set for classification of brain tumours.* J Med Eng Technol, 2008. 32(3): p. 198-205.
116. Liu, Y., et al., *Computer aided diagnosis system for breast cancer based on color Doppler flow imaging.* J Med Syst, 2012. 36(6): p. 3975-82.
117. Chang, S.M., et al., *Integration of preoperative anatomic and metabolic physiologic imaging of newly diagnosed glioma.* J Neurooncol, 2009. 92(3): p. 401-15.
118. Mouthuy, N., et al., *Multiparametric magnetic resonance imaging to differentiate high-grade gliomas and brain metastases.* J Neuroradiol, 2012. 39(5): p. 301-7.
119. Benz, M.R., et al., *Treatment monitoring by 18F-FDG PET/CT in patients with sarcomas: interobserver variability of quantitative parameters in treatment-induced changes in histopathologically responding and nonresponding tumors.* J Nucl Med, 2008. 49(7): p. 1038-46.
120. Cook, G.J., et al., *Are Pretreatment 18F-FDG PET Tumor Textural Features in Non-Small Cell Lung Cancer Associated with Response and Survival After Chemoradiotherapy?* J Nucl Med, 2013. 54(1): p. 19-26.
121. de Langen, A.J., et al., *Monitoring response to antiangiogenic therapy in non-small cell lung cancer using imaging markers derived from PET and dynamic contrast-enhanced MRI.* J Nucl Med, 2011. 52(1): p. 48-55.
122. Eary, J.F., et al., *Spatial heterogeneity in sarcoma 18F-FDG uptake as a predictor of patient outcome.* J Nucl Med, 2008. 49(12): p. 1973-9.
123. Tixier, F., et al., *Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer.* J Nucl Med, 2011. 52(3): p. 369-78.
124. Chen, S.J., et al., *Quantitatively characterizing the textural features of sonographic images for breast cancer with histopathologic correlation.* J Ultrasound Med, 2005. 24(5): p. 651-61.
125. Ding, J., et al., *Quantitative measurement for thyroid cancer characterization based on elastography.* J Ultrasound Med, 2011. 30(9): p. 1259-66.
126. Nguyen, V.X., et al., *Digital image analysis is a useful adjunct to endoscopic ultrasonographic diagnosis of subepithelial lesions of the gastrointestinal tract.* J Ultrasound Med, 2010. 29(9): p. 1345-51.
127. Chen, S.J., et al., *The Representations of Sonographic Image Texture for Breast Cancer Using Co-occurrence Matrix.* Journal of Medical and Biological Engineering 2005. 25(4): p. 193-199.
128. Chen, D.R., et al., *Computerized quantitative assessment of sonomammographic homogeneity of fibroadenoma and breast carcinoma* Journal of Medical Ultrasound 1999. 7(3): p. 157-162.
129. Sarty, G.E., et al., *Magnetic resonance diffusion imaging of ovarian masses: a first experience with 12 cases.* MAGMA, 2004. 16(4): p. 182-93.
130. Checkley, D., et al., *Dynamic contrast-enhanced MRI of vascular changes induced by the VEGF-signalling inhibitor ZD4190 in human tumour xenografts.* Magn Reson Imaging, 2003. 21(5): p. 475-82.
131. Georgiadis, P., et al., *Enhancing the discrimination accuracy between metastases, gliomas and meningiomas on brain MRI by volumetric textural features and ensemble pattern recognition methods.* Magn Reson Imaging, 2009. 27(1): p. 120-30.
132. Georgiadis, P., et al., *Quantitative combination of volumetric MR imaging and MR spectroscopy data for the discrimination of meningiomas from metastatic brain tumors by means of pattern recognition.* Magn Reson Imaging, 2011. 29(4): p. 525-35.
133. Kurki, T., et al., *MR classification of brain gliomas: value of magnetization transfer and conventional imaging.* Magn Reson Imaging, 1995. 13(4): p. 501-11.
134. Mayerhoefer, M.E., et al., *Are signal intensity and homogeneity useful parameters for distinguishing between benign and malignant soft tissue masses on MR images? Objective evaluation by means of texture analysis.* Magn Reson Imaging, 2008. 26(9): p. 1316-22.

135. Yang, X., et al., *Microcirculatory fraction (MCF(I)) as a potential imaging marker for tumor heterogeneity in breast cancer.* Magn Reson Imaging, 2012. 30(8): p. 1059-67.
136. Berry, L.R., et al., *Quantification of viable tumor microvascular characteristics by multispectral analysis.* Magn Reson Med, 2008. 60(1): p. 64-72.
137. Chen, W., et al., *Volumetric texture analysis of breast lesions on contrast-enhanced magnetic resonance images.* Magn Reson Med, 2007. 58(3): p. 562-71.
138. Gibbs, P. and L.W. Turnbull, *Textural analysis of contrast-enhanced MR images of the breast.* Magn Reson Med, 2003. 50(1): p. 92-8.
139. Rose, C.J., et al., *Quantifying spatial heterogeneity in dynamic contrast-enhanced MRI parameter maps.* Magn Reson Med, 2009. 62(2): p. 488-99.
140. Zacharaki, E.I., et al., *Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme.* Magn Reson Med, 2009. 62(6): p. 1609-18.
141. Levner, I., et al., *Predicting MGMT methylation status of glioblastomas from MRI texture.* Med Image Comput Comput Assist Interv, 2009. 12(Pt 2): p. 522-30.
142. Zheng, Y., et al., *Segmentation and classification of breast tumor using dynamic contrast-enhanced MR images.* Med Image Comput Comput Assist Interv, 2007. 10(Pt 2): p. 393-401.
143. Alvarenga, A.V., et al., *Complexity curve and grey level co-occurrence matrix in the texture evaluation of breast tumor on ultrasound images.* Med Phys, 2007. 34(2): p. 379-87.
144. Bilello, M., et al., *Automatic detection and classification of hypodense hepatic lesions on contrast-enhanced venous-phase CT.* Med Phys, 2004. 31(9): p. 2584-93.
145. Chen, W., et al., *Automatic identification and classification of characteristic kinetic curves of breast lesions on DCE-MRI.* Med Phys, 2006. 33(8): p. 2878-87.
146. Cui, J., et al., *A new automated method for the segmentation and characterization of breast masses on ultrasound images.* Med Phys, 2009. 36(5): p. 1553-65.
147. Goldberg, V., et al., *Improvement in specificity of ultrasonography for diagnosis of breast tumors by means of artificial intelligence.* Med Phys, 1992. 19(6): p. 1475-81.
148. Horsch, K., et al., *Computerized diagnosis of breast lesions on ultrasound.* Med Phys, 2002. 29(2): p. 157-64.
149. Liao, Y.Y., et al., *Classification of scattering media within benign and malignant breast tumors based on ultrasound texture-feature-based and Nakagami-parameter images.* Med Phys, 2011. 38(4): p. 2198-207.
150. McNitt-Gray, M.F., et al., *A pattern classification approach to characterizing solitary pulmonary nodules imaged on high resolution CT: preliminary results.* Med Phys, 1999. 26(6): p. 880-8.
151. Sahiner, B., et al., *Computerized characterization of breast masses on three-dimensional ultrasound volumes.* Med Phys, 2004. 31(4): p. 744-54.
152. Way, T.W., et al., *Computer-aided diagnosis of pulmonary nodules on CT scans: segmentation and classification using 3D active contours.* Med Phys, 2006. 33(7): p. 2323-37.
153. Way, T.W., et al., *Computer-aided diagnosis of pulmonary nodules on CT scans: improvement of classification performance with nodule surface features.* Med Phys, 2009. 36(7): p. 3086-98.
154. Yan, K., et al., *Clinical study of a noninvasive multimodal sono-contrast induced spectroscopy system for breast cancer diagnosis.* Med Phys, 2012. 39(3): p. 1571-8.
155. Zheng, Y., et al., *STEP: spatiotemporal enhancement pattern for MR-based breast tumor diagnosis.* Med Phys, 2009. 36(7): p. 3192-204.
156. Stantz, K.M., et al., *Monitoring the longitudinal intra-tumor physiological impulse response to VEGFR2 blockade in breast tumors using DCE-CT.* Mol Imaging Biol, 2011. 13(6): p. 1183-95.
157. Drabycz, S., et al., *An analysis of image texture, tumor location, and MGMT promoter methylation in glioblastoma using magnetic resonance imaging.* Neuroimage, 2010. 49(2): p. 1398-405.
158. Eliat, P.A., et al., *Can dynamic contrast-enhanced magnetic resonance imaging combined with texture analysis differentiate malignant glioneuronal tumors from other glioblastoma?* Neurol Res Int, 2012. 2012: p. 195176.
159. Peng, S.L., et al., *Analysis of parametric histogram from dynamic contrast-enhanced MRI: application in evaluating brain tumor response to radiotherapy.* NMR Biomed, 2012.

160. Tozer, D.J., et al., *Apparent diffusion coefficient histograms may predict low-grade glioma subtype.* NMR Biomed, 2007. 20(1): p. 49-57.
161. Dong, X., et al., *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, 2013. 34(1): p. 40-6.
162. Su, Y., et al., *Automatic detection and classification of breast tumors in ultrasonic images using texture and morphological features.* Open Med Inform J, 2011. 5(Suppl 1): p. 26-37.
163. El Naqa, I., et al., *Exploring feature-based approaches in PET images for predicting cancer treatment outcomes.* Pattern Recognit, 2009. 42(6): p. 1162-1171.
164. Alic, L., et al., *Heterogeneity in DCE-MRI parametric maps: a biomarker for treatment response?* Phys Med Biol, 2011. 56(6): p. 1601-16.
165. Chang, Y.C., et al., *Angiogenic response of locally advanced breast cancer to neoadjuvant chemotherapy evaluated with parametric histogram from dynamic contrast-enhanced MRI.* Phys Med Biol, 2004. 49(16): p. 3593-602.
166. Yoshida, H., et al., *Wavelet-packet-based texture analysis for differentiation between benign and malignant liver tumours in ultrasound images.* Phys Med Biol, 2003. 48(22): p. 3735-53.
167. Brooks, F.J. and P.W. Grigsby, *Current measures of metabolic heterogeneity within cervical cancer do not predict disease outcome.* Radiat Oncol, 2011. 6: p. 69.
168. Yu, H., et al., *A longitudinal study of radiation-induced changes in tumor vasculature by contrast-enhanced magnetic resonance imaging.* Radiat Res, 2002. 158(2): p. 152-8.
169. Yoshida, H., et al., *Computer-aided diagnosis scheme for detection of polyps at CT colonography.* Radiographics, 2002. 22(4): p. 963-79.
170. Zhang, X., et al., *Application of an artificial neural network to the computer-aided differentiation of focal liver disease in MR imaging.* Radiol Phys Technol, 2009. 2(2): p. 175-82.
171. Chen, D.R., R.F. Chang, and Y.L. Huang, *Computer-aided diagnosis applied to US of solid breast nodules by using neural networks.* Radiology, 1999. 213(2): p. 407-12.
172. Dyke, J.P., et al., *Osteogenic and Ewing sarcomas: estimation of necrotic fraction during induction chemotherapy with dynamic contrast-enhanced MR imaging.* Radiology, 2003. 228(1): p. 271-8.
173. Emblem, K.E., et al., *Glioma grading by using histogram analysis of blood volume heterogeneity from MR-derived cerebral blood volume maps.* Radiology, 2008. 247(3): p. 808-17.
174. Goh, V., et al., *Assessment of response to tyrosine kinase inhibitors in metastatic renal cell cancer: CT texture as a predictive biomarker.* Radiology, 2011. 261(1): p. 165-71.
175. Miles, K.A., et al., *Colorectal cancer: texture analysis of portal phase hepatic CT images as a potential marker of survival.* Radiology, 2009. 250(2): p. 444-52.
176. Negendank, W.G., et al., *Lymphomas: MR imaging contrast characteristics with clinical-pathologic correlations.* Radiology, 1990. 177(1): p. 209-16.
177. Ng, F., et al., *Assessment of primary colorectal cancer heterogeneity by using whole-tumor texture analysis: contrast-enhanced CT texture as a biomarker of 5-year survival.* Radiology, 2013. 266(1): p. 177-84.
178. Vaidya, M., et al., *Combined PET/CT image characteristics for radiotherapy tumor response in lung cancer.* Radiother Oncol, 2012. 102(2): p. 239-45.
179. Chen, D.R., et al., *Texture analysis of breast tumors on sonograms.* Semin Ultrasound CT MR, 2000. 21(4): p. 308-16.
180. Acharya, U.R., et al., *Cost-effective and non-invasive automated benign and malignant thyroid lesion classification in 3D contrast-enhanced ultrasound using combination of wavelets and textures: a class of ThyroScan algorithms.* Technol Cancer Res Treat, 2011. 10(4): p. 371-80.
181. Tuma, J., et al., *[Image analysis in the differential diagnosis of renal parenchyma lesions].* Ultraschall Med, 2011. 32(3): p. 286-92.
182. Alam, S.K., et al., *Ultrasonic multi-feature analysis procedure for computer-aided diagnosis of solid breast lesions.* Ultrason Imaging, 2011. 33(1): p. 17-38.

183. Basset, O., et al., *Texture analysis of ultrasonic images of the prostate by means of co-occurrence matrices.* Ultrason Imaging, 1993. 15(3): p. 218-37.
184. Garra, B.S., et al., *Improving the distinction between benign and malignant breast lesions: the value of sonographic texture analysis.* Ultrason Imaging, 1993. 15(4): p. 267-85.
185. Acharya, U.R., et al., *Non-invasive automated 3D thyroid lesion classification in ultrasound: a class of ThyroScan systems.* Ultrasonics, 2012. 52(4): p. 508-20.
186. Alacam, B., et al., *Breast tissue characterization using FARMA modeling of ultrasonic RF echo.* Ultrasound Med Biol, 2004. 30(10): p. 1397-407.
187. Chang, R.F., et al., *Improvement in breast tumor discrimination by support vector machines and speckle-emphasis texture analysis.* Ultrasound Med Biol, 2003. 29(5): p. 679-86.
188. Chen, D., R.F. Chang, and Y.L. Huang, *Breast cancer diagnosis using self-organizing map for sonography.* Ultrasound Med Biol, 2000. 26(3): p. 405-11.
189. Chen, D.R., et al., *Diagnosis of breast tumors with sonographic texture analysis using wavelet transform and neural networks.* Ultrasound Med Biol, 2002. 28(10): p. 1301-10.
190. Chen, D.R., et al., *Use of the bootstrap technique with small training sets for computer-aided diagnosis in breast ultrasound.* Ultrasound Med Biol, 2002. 28(7): p. 897-902.
191. Chen, S.J., et al., *Classification of the thyroid nodules based on characteristic sonographic textural feature and correlated histopathology using hierarchical support vector machines.* Ultrasound Med Biol, 2010. 36(12): p. 2018-26.
192. Chen, S.J., et al., *Characterization of the major histopathological components of thyroid nodules using sonographic textural features for clinical diagnosis and management.* Ultrasound Med Biol, 2009. 35(2): p. 201-8.
193. Chikui, T., et al., *Sonographic texture characterization of salivary gland tumors by fractal analyses.* Ultrasound Med Biol, 2005. 31(10): p. 1297-304.
194. Donohue, K.D., et al., *Tissue classification with generalized spectrum parameters.* Ultrasound Med Biol, 2001. 27(11): p. 1505-14.
195. Huber, S., et al., *Relevance of sonographic B-mode criteria and computer-aided ultrasonic tissue characterization in differential/diagnosis of solid breast masses.* Ultrasound Med Biol, 2000. 26(8): p. 1243-52.
196. Kuo, W.J., et al., *Retrieval technique for the diagnosis of solid breast tumors on sonogram.* Ultrasound Med Biol, 2002. 28(7): p. 903-9.
197. Lefebvre, F., et al., *Computerized ultrasound B-scan characterization of breast nodules.* Ultrasound Med Biol, 2000. 26(9): p. 1421-8.
198. Maruyama, H., et al., *Heterogeneity of microbubble accumulation: a novel approach to discriminate between well-differentiated hepatocellular carcinomas and regenerative nodules.* Ultrasound Med Biol, 2012. 38(3): p. 383-8.
199. Moon, W.K., et al., *Computer-aided diagnosis for the classification of breast masses in automated whole breast ultrasound images.* Ultrasound Med Biol, 2011. 37(4): p. 539-48.
200. Scheipers, U., et al., *Sonohistology for the computerized differentiation of parotid gland tumors.* Ultrasound Med Biol, 2005. 31(10): p. 1287-96.
201. Siebers, S., et al., *Computer aided diagnosis of parotid gland lesions using ultrasonic multi-feature tissue characterization.* Ultrasound Med Biol, 2010. 36(9): p. 1525-34.
202. Sujana, H., S. Swarnamani, and S. Suresh, *Application of artificial neural networks for the classification of liver lesions by image texture parameters.* Ultrasound Med Biol, 1996. 22(9): p. 1177-81.
203. Thijssen, J.M., et al., *Echographic differentiation of histological types of intraocular melanoma.* Ultrasound Med Biol, 1991. 17(2): p. 127-38.
204. Tsantis, S., et al., *Development of a support vector machine-based image analysis system for assessing the thyroid nodule malignancy risk on ultrasound.* Ultrasound Med Biol, 2005. 31(11): p. 1451-9.
205. Bader, W., et al., *Does texture analysis improve breast ultrasound precision?* Ultrasound Obstet Gynecol, 2000. 15(4): p. 311-6.

206. Huang, Y.L., et al., *Image retrieval with principal component analysis for breast cancer diagnosis on various ultrasonic systems.* Ultrasound Obstet Gynecol, 2005. 26(5): p. 558-66.
207. Kratzik, C., et al., *Texture analysis--a new method of differentiating prostatic carcinoma from prostatic hypertrophy.* Urol Res, 1988. 16(5): p. 395-7.
208. Spector, D.I., A.J. Fischetti, and J.R. Kovak-McClaran, *Computed tomographic characteristics of intrapelvic masses in dogs.* Vet Radiol Ultrasound, 2011. 52(1): p. 71-4.



Chapter 4

Heterogeneity in DCE-MRI parametric maps: a biomarker for treatment response?

This chapter is based upon:
L Alić, M van Vliet, CF van Dijke, AMM Eggermont, JF Veenland, WJ Niessen. Heterogeneity in DCE-MRI parametric maps: a biomarker for treatment response? *Physics in Medicine and Biology*

Abstract

This study aims to quantify the heterogeneity of tumour enhancement in dynamic contrast-enhanced MRI (DCE-MRI) using texture analysis methods.

The suitability of the coherence and the fractal dimension to monitor tumour response was evaluated in 18 patients with limb sarcomas imaged by DCE-MRI at baseline and follow-up. According to the histopathology, tumours were classified into two response categories (responders and non-responders). Pharmacokinetic (K^{trans}) and heuristic model-based parametric maps (slope, max enhancement, AUC) were computed from the DCE-MRI data.

A substantial correlation was found between the pharmacokinetic and heuristic model-based parametric maps: ρ=0.56 for the slope, ρ=0.44 for maximum enhancement, and ρ=0.61 for AUC. From all four parametric maps, the enhancing fraction, and the heterogeneity features (i.e. coherence and fractal dimension) were determined. In terms of monitoring tumour response, using both baseline and follow-up DCE-MRI, the enhancing fraction and the coherence showed significant differences between the response categories (i.e. the highest sensitivity (91%) for K^{trans}, and the highest specificity (83%) for max enhancement). In terms of treatment prediction, using solely the baseline DCE-MRI, the enhancing fraction and coherence discriminated between responders and nonresponders. For prediction, the highest sensitivity (91%) was shared by K^{trans}, slope and max enhancement, and the highest specificity (71%) was achieved by K^{trans}. On average, tumours that responded to treatment showed a high enhancing fraction and high coherence at the baseline.

These results suggest that specific heterogeneity features, computed from both pharmacokinetic and heuristic model-based parametric maps, show potential as a biomarker for monitoring tumour response.

Introduction

Dynamic contrast-enhanced MR imaging (DCE-MRI) is a non-invasive technique to study *in vivo* microvascular characteristics such as blood volume, blood flow and endothelial permeability in tumours. DCE-MRI has emerged as a valuable clinical tool to diagnose and stage tumours [1-3]. Pharmacokinetic and heuristic model-based parameters have been shown to correlate with histopathological tumour grade and measures of angiogenesis. DCE-MRI is also capable of monitoring and quantifying the treatment effects of radiotherapy, chemotherapy and antivascular drugs [4, 5].

It is important to develop predictors of treatment outcome which help to tailor treatment for the individual patient. For breast cancer, the potential of DCE-MRI as an early predictor of response has been investigated [6]. More recently, the use of baseline DCE-MRI for response prediction has been investigated for breast cancer [7, 8], colorectal metastases [9] and mesothelioma [10]. Using pharmacokinetic modelling or a heuristic curve-fitting approach, the voxel-wise contrast uptake curve, as measured by DCE-MRI, can be summarized with a small number of descriptive parameters. However, quantification of the parametric maps remains a challenge. A commonly-used method is the computation of an average of the parametric map over the whole tumour [5, 9]. Due to averaging, variations in parts of the tumour may be cancelled out [11], though. Human solid tumours are biologically heterogeneous [12], and heterogeneous areas of DCE-MRI enhancement are reported to be diagnostically important [13].

A logical first step in quantifying the heterogeneity in parametric maps is to identify regions of interest, or 'hot spot' selection. Parameters in hot spots are reported to be more relevant for quantifying tumour response than parameters averaged over the whole tumour [7, 8, 14, 15]. A disadvantage is that the selection of hot spots is usually performed manually and therefore subject to inter- and intra-operator variability. For treatment monitoring, this approach is not recommended [16]. Another method is to compute a binary version of the map, e.g. by thresholding using a parameter value or a fit-statistic. Using the binary map, the number of 'enhancing' voxels can be computed [17-19]. The non-enhancing voxels are then assumed to represent necrotic tissue. An additional analysis can be performed on the enhancing voxels using histogram analysis. There are indications that the histogram of the parametric map can be of added value to improve diagnostic accuracy [20], to grade tumours [21], to predict overall survival [22], and to assess treatment effect [14, 23-25].

However, this method does not take into account the spatial distribution of the different contrast uptake parameters. To specifically extract information on the spatial coherence of enhancing regions, texture analysis can be performed on parametric maps. Additionally, texture analysis methods have been used for diagnostic tasks.

For example, Rose *et al* [26] calculated fractal dimensions and performed geometrical-based texture analysis to grade gliomas. Karahaliou *et al* [27] computed the grey-tone spatial-dependence matrix (GTSDM) features to discriminate malignant from benign breast lesions. Moreover, texture analysis can also be used for treatment monitoring. Alic *et al* [28] describe how the GTSDM-based method can be used to monitor treatment effect.

An important advantage of texture-based analysis is the independence of the absolute values of the parametric maps. Texture methods quantify the spatial variations in parametric maps, not the absolute values of the maps. Therefore, texture analysis can provide additional and independent information compared to histogram-based measures of parametric maps. Furthermore, in case the absolute values of contrast uptake parameters at baseline and follow-up cannot be reliably compared (due to variations in the acquisition process of the MRI data, or to selection of the arterial input function), the heterogeneity of the parametric maps within the tumour can still be computed and compared. In clinical practice, calibration sequences necessary for the estimation of absolute pharmacokinetic parameters are not always acquired. To facilitate the analysis of clinical data, texture analysis can be considered as a means to estimate parametric map heterogeneity.

This study assesses the potential of heterogeneity in DCE-MRI parametric maps as a biomarker for treatment response. Both pharmacokinetic and heuristic model-based parametric maps are computed from the DCE-MRI data of patients with soft tissue sarcomas treated with isolated limb perfusion (ILP). The heterogeneity of DCE-MRI parametric maps is quantified by texture features based on the GTSDM-based method and fractal dimension.

Heterogeneity

Since texture information is hidden in the coherence of intensity values rather than in the absolute values, texture features grasp this information independently from the absolute values of the parametric maps. According to Haralick *et al* [29], texture features should be invariant under monotonic intensity transformations. The invariance under linear transformations of intensity values is the basic criterion that should be fulfilled at all times. This means that the texture features should be independent of the mean and variance of the intensity distribution. Satisfying this condition enables investigation whether texture analysis can provide additional and independent information compared to histogram-based measures. However, most texture analysis methods do not meet this strict criterion.

This study utilizes two texture analysis methods to quantify the heterogeneity in DCE-MRI parametric maps: the GTSDM-based (also referred to as co-occurrence-based) method, and the fractal-dimension-based method. A widely used method for texture analysis, the GTSDM-based method [29], secures the linear transformation invariance by performing histogram equalization as a preprocessing step. Fractal dimensions are also often used to analyse texture. Figure 1 shows three images with decreasing coherence. For this study, we used the Blanket method [30] that secures the linear transformation invariance by estimating the fractal dimension of the intensity surface of the parametric map. Matlab scripts for both methods (the GTSDM-based and fractal dimension-based) are publicly available at *http://DCEanalysis.bigr.nl*.

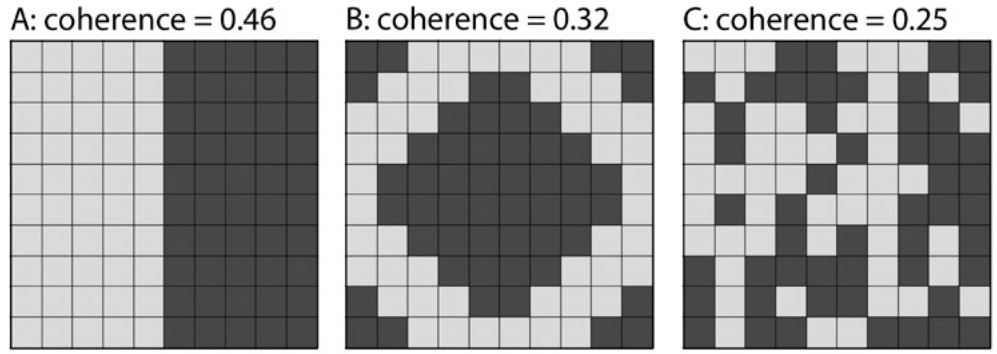


Figure 1. Images with decreasing coherence.

Coherence

In GTSDM-based analysis, histogram equalization is performed first. Thereby, the number of intensity values is reduced to a fixed number *N*. The original intensity values are redistributed in such a way that the incidence probability for all new intensity values is equal (1/*N*). As the second step the GTSDM $M_{d,\theta}$ is computed; the elements of $M_{d,\theta}(i,j)$ contain the number of incidences having intensity values *i* and *j* over a distance *d*, and at an angle θ. The matrix is normalized for the total number of incidences. Over the normalized matrix, several texture features can be computed. Since we are primarily interested in the coherence of the higher image intensities, we divided all the image voxels into two groups based on their median intensity. As none of the original texture features defined by Haralick *et al* [29] quantifies the coherence among the image voxels, we define the image coherence in Eq. 1.

$$\text{Coherence} = \frac{M_{d,\theta}(1,1)}{\sum_{i=0,j=0}^{1} M_{d,\theta}(i,j)} \qquad \text{(Eq. 1)}$$

Figure 1 shows typical images with decreasing coherence. Light grey values in the images demonstrate the high intensity voxels, and dark grey are low intensity voxels. Coherence quantifies the cohesion of the high intensity voxels, with the coherence being high for the spatially close voxels, and low for the more scattered voxels.

Fractal dimension

Although a wide variety of methods are available to estimate the fractal dimension, all these methods are based on the same principle: an image characteristic is measured as a function of a scale parameter. For fractal images, the relation between these two quantities can be described by a straight line in a log–log domain. The slope, computed by linear regression analysis, is then linearly related to the fractal dimension. The various estimation methods differ in the definition of the image characteristic and of the scale parameter. To estimate the fractal dimension of a surface, Peleg *et al* [30] extended the Blanket method described by Mandelbrot [31] to three dimensions.

For this study, we used the Blanket method which estimates the surface area by measuring the volume between the maximum (upper blanket) $u_\varepsilon(i,j)$ and the minimum (lower blanket) $b_\varepsilon(i,j)$ gray levels for a certain resolution [32]. The blankets are not further than a distance ε above and below the surface to be measured [30, 33-35]. With with $|(m,n)-(i,j)| \leq 1$, the upper and lower blankets are defined in Eq. 2.

$$u_\varepsilon(i,j)=MAX\{u_{\varepsilon-1}(m,n)\} \quad \text{(Eq. 2-A)}$$

$$b_\varepsilon(i,j)=MAX\{b_{\varepsilon-1}(m,n)\} \quad \text{(Eq. 2-B)}$$

$$V(\varepsilon)=\sum\{u_\varepsilon(i,j)-b_\varepsilon(i,j)\} \quad \text{(Eq. 2-C)}$$

$$A(\varepsilon)=\tfrac{1}{2}\{V(\varepsilon)-V(\varepsilon-1)\} \quad \text{(Eq. 2-D)}$$

with $V(\varepsilon)$ representing the volume between the blankets, and $A(\varepsilon)$ the surface area. Using the property: $A(\varepsilon) \propto \varepsilon^{2D}$, the fractal dimension can be established as the linear fit of [log $A(\varepsilon)$, log ε].

Material and methods

Patients

In the period 2001 to 2006, patients referred to our cancer centre for ILP of irresectable extremity soft tissue sarcoma, received a DCE-MRI sequence in addition to the standard MRI protocol. Patients were included in this study if they had received a DCE-MRI within 6 weeks before and after the ILP procedure. The study was approved by the local Research Ethics Committee.

MRI data were obtained from 21 patients with sarcomas which were histologically confirmed by biopsies. From the available data, DCE-MRI of two tumours was excluded from further analysis because they were scanned at different tumour locations before and after ILP, and one for using fat suppression during the DCE-MRI protocol. Therefore, our study group consisted of 18 patients (9 men and 9 women) with a median age of 54.3 years. Eleven sarcomas were located in the leg, and seven in the arm. All tumours (median volume of 30 cm^3, with an interquartile range of 17–246 cm^3) were considered primarily irresectable because of size, or fixation to the neurovascular bundle or bone. To render the tumours resectable for limb salvage, patients were treated with ILP. After resection of the tumour remnant, the percentage of necrosis was estimated by a pathologist based on macroscopic evaluation of necrotic tissue. The standard classification was used: if the tumour had 50–100% necrosis, the patient was classified as responding to treatment; if the tumour was <50% necrosis, the patient was classified as non-responding to treatment [36]. The tumour size before treatment did not differ significantly between the responding and the non-responding patients.

Treatment protocol

ILP is based on the technique developed by Creech *et al* [37] and allows for the delivery of high doses of cytostatic drugs to a tumour-bearing extremity that is isolated from the systemic circulation. Hence, without any systemic toxicity, a regional chemotherapeutic concentration can be increased to 20 times the systemically tolerable dose. ILP is used for irresectable extremity soft tissue sarcomas, multiple in-transit melanoma metastases, or a variety of other limb-threatening tumours [38-40].

For the treatment of soft tissue sarcomas, two drugs were used: Melphalan (Burroughs Wellcome, London, UK) and TNF-α (Boehringer, Ingelheim, Germany). Melphalan is a cytostatic drug, mainly focused on tumour cells. TNF-α is an antivascular agent, added because of its selective destruction of tumour vessels and its effect of increasing uptake of Melphalan by the tumour. Due to the combination of Melphalan with TNF-α, ILP is an excellent method to locally treat advanced tumours rendering a non-resectable tumour resectable, without the need for amputation, and reducing the local recurrence rate.

MRI imaging

Imaging was performed on a 1.5 T Vision MR system (Siemens Medical Systems, Erlangen, Germany) using a Body, FlexLarg or SpArray coil, depending on the location and tumour size. Standard non-enhanced T1- and T2-weighted sequences were followed by a 2- to 16-slice DCE-MRI sequence covering at least the tumour centre. The dynamic series were started immediately after bolus injection of Gd-DTPA (Magnevist®, Gadopentetate dimeglumine, Schering, Berlin, Germany, 0.1 mmol/kg) followed by N-saline flush. The DCE-MRI sequence had the following scan parameters: T1-weighted TurboFLASH with TR/TE = 5.8/2.4 ms, a flip angle of 10°, a matrix of 256 × 256 and a slice thickness of 10 mm. A series of 100 time points were acquired with a temporal resolution of 2 s with a total scanning time of 3.28 min.

Data analysis

DCE-MRI data analysis was performed using in-house developed software based on MATLAB (MathWorks, Natick, MA, USA). The tumour and the feeding vessel were manually outlined. First, fuzzy c-means (FCM) clustering [41] was applied to the region of interest of the feeding vessel. The goal of FCM clustering is to locate the three phases in the feeding vessel enhancement: baseline, the wash-in and the wash-out phase [28]. The pre-enhancement timing information was used to construct a mean tumour pre-enhancement image which is subtracted from the subsequent images in the dynamic sequence. The volume transfer constant (K^{trans}) was derived from these normalized dynamic images using the compartmental model [42] assuming that the signal enhancement curve is proportional to the concentration–time curve [43]. The heuristic model-based parameters were derived from the normalized dynamic images using FCM clustering to partition (for each voxel) the concentration–time curve into the three phases [28]. Subsequently, based on these partitions, we compute (voxel-wise) the heuristic model-based parameters: i.e., the slope, maximum enhancement, and area under the curve (AUC). All parameters were visualized in a parametric map. The correlation between the K^{trans} and the three heuristic model-based parameters was computed for the enhancing voxels and the following features were computed over all parametric maps.

Enhancing faction. Some of the voxels in the DCE-MRI parametric maps showed no signal enhancement and/or low goodness of fit values ($r^2 < 0.5$). The enhancing fraction of a parametric map was defined as follows:

- for K^{trans}, the fraction of the tumour volume with high goodness of fit ($r^2 > 0.5$),
- for heuristic model-based parameters, the fraction of the tumour volume with non-zero voxels in the parameter maps.

Coherence. Earlier studies [35] showed that short distances have a higher discriminative power than larger distances and that, when the texture has no distinct direction, the influence of the angle θ is small. Thus, the matrix was averaged over the four angles (0°, 90°, 270°, 180°) and determined on a distance 1. For each of the four parametric maps, K^{trans}, slope, maximum enhancement and AUC, the enhancing fraction of the tumour was segmented based on the median value of the enhancing voxels into two groups of voxels: strongly enhancing and enhancing. The non-enhancing voxels were not taken into account, and therefore the coherence is independent of the enhancing fraction. The coherence was then determined for all acquisitions (at baseline and follow-up) and for all maps (pharmacokinetic and heuristic model-based).

Fractal dimension. Since the image size is limited, the fractal dimension is estimated over a limited range of scales. We used the first four scales (ε) to estimate the fractal dimension from all the parametric maps.

Statistical analysis

Statistical analysis was performed using SPSS (SPSS for Windows, Version 17.0, SPSS Inc., Chicago, IL, USA). Spearman's rank correlation coefficient was used to measure correlation of the parametric maps derived from the different models (pharmacokinetic and heuristic). Two response categories were defined: responding and non-responding tumours. The Wilcoxon signed rank test was used to assess the differences between the two response categories characterized by three heterogeneity texture features (enhancement fraction, coherence and fractal dimension). The heterogeneity features were calculated using four parametric maps, e.g. K^{trans}, slope, maximum enhancement and AUC based on the DCE-MRI acquired before and after treatment. The Mann–Whitney *U*-test was used to assess the differences between the two response categories characterized by the same heterogeneity texture features based on the DCE-MRI acquired before treatment.

To determine the accuracy, sensitivity and specificity of all texture features, a classification experiment was performed, using support vector machine [44] in a leave-one out approach. A leave-one out approach can provide a reliable estimation of the accuracy, sensitivity and specificity of the classifier when the data set is too small to be divided into a learning and test set.

Results

Correlation between the different Parameters

Since we are interested in the enhancing parts of the parametric maps, the statistical dependence between the K^{trans} and the three heuristic model-based parameters was measured for the enhancing voxels using Spearman's rank correlation coefficient.

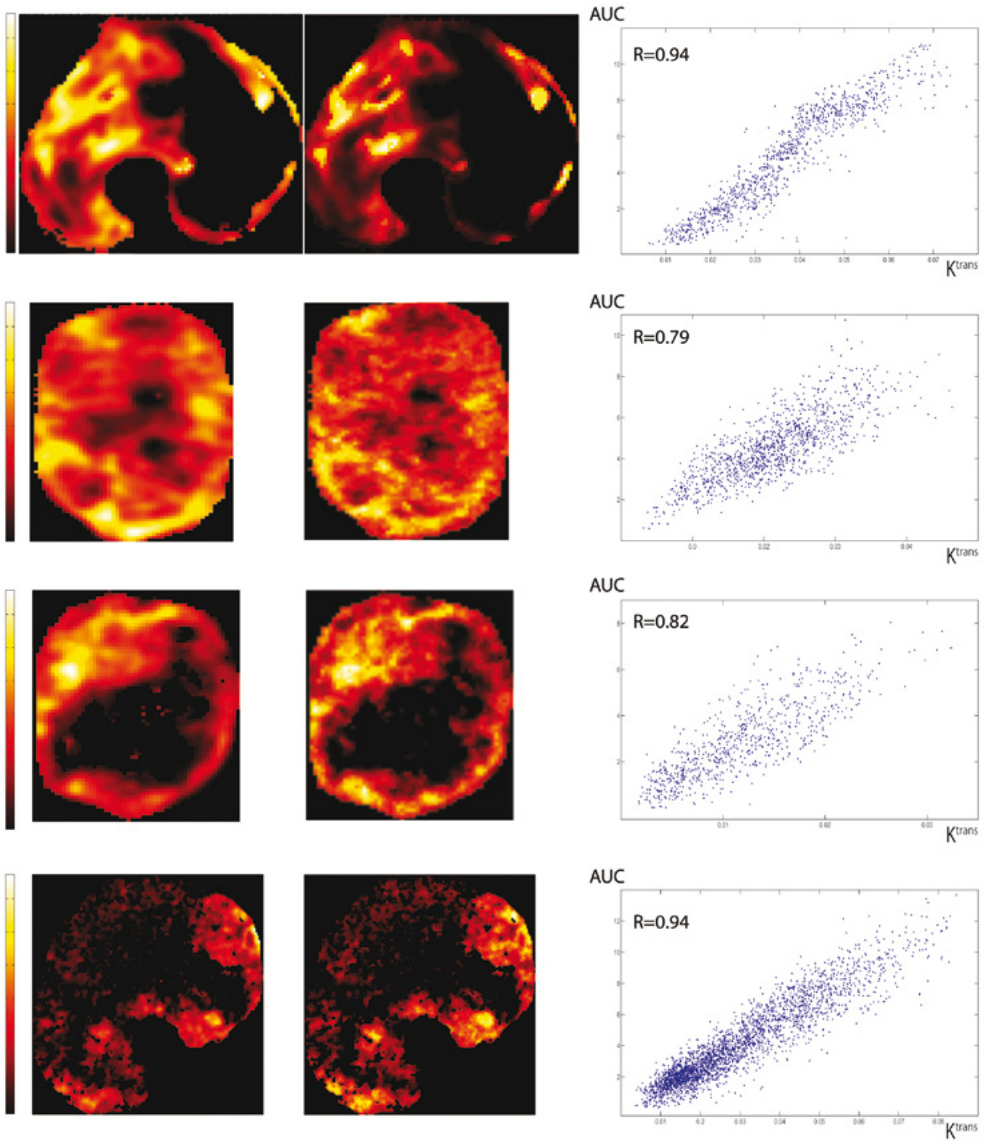


Figure 2. The K^{trans} parametric map (column 1) and AUC parametric map (column 2) with the scatter plots of AUC-K^{trans} (column 3).

This assesses the monotonic relation between two variables. Spearman's rank correlation coefficient, averaged over all tumours, was ρ=0.56 for the slope, ρ=0.44 for maximum enhancement, and ρ=0.61 for AUC. Figure 2 shows a significant spatial correspondence between the AUC map and the K^{trans} map for four selected tumours. As illustrated, not only the non-enhancing part in both maps is comparable but also the heterogeneity in the enhancing part of the tumour shows excellent correspondence.

The smoothness of the parametric maps differs between the K^{trans} and the heuristic model parameters. This is caused by the difference in the estimation procedure: the fitting procedure used to estimate the K^{trans} parameter cancels out smaller variations in the signal enhancement. In contrast, the curve partitioning as used for the heuristic model, is more sensitive for these variations resulting in a more granular image.

Monitoring of tumour treatment response

Figure 3 shows that a sarcoma sometimes contains a large core that hardly takes any contrast. This core can either be present exclusively after treatment (Figure 3A), or before and after treatment (Figure 3B). The parametric map (slope) and corresponding strongly enhancing voxels for pre- and post-treatment acquisitions demonstrate two response categories: responding (Figure 3A) and non-responding (Figure 3B) tumours. The differences in parametric map heterogeneity between DCE-MRI acquisitions (at baseline and follow-up) were assessed using the Wilcoxon signed rank test

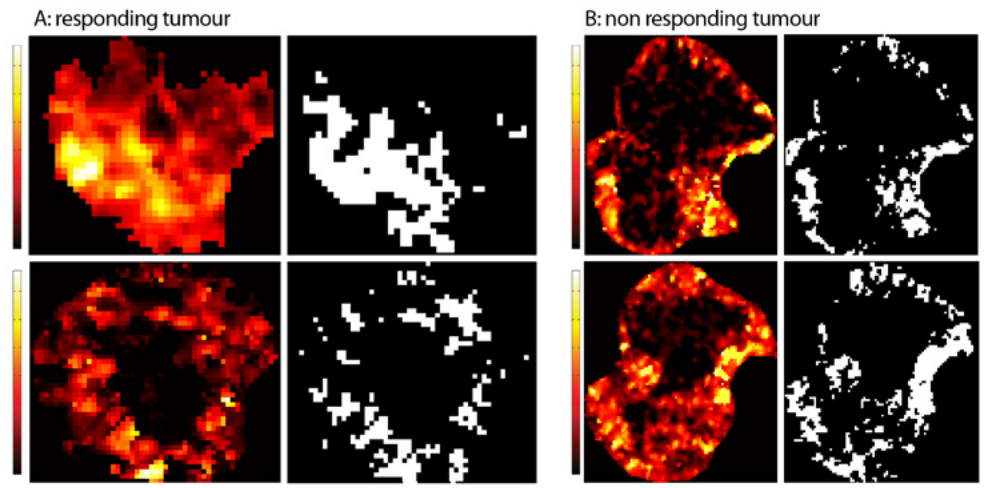


Figure 3. DCE-MRI parametric map (slope) and corresponding strongly enhancing voxels (the median value of enhancing part is used as threshold) for a responding (A) and a non-responding tumour (B): top row is before treatment, bottom row is after treatment.

for two separate groups, i.e. responders and non-responders. The two heterogeneity texture features (coherence and fractal dimension) were estimated upon four parametric maps, e.g. K^{trans}, slope, maximum enhancement and AUC.

The enhancing fraction showed a significant difference between the baseline and follow-up scans for the group that responded to therapy, for both the pharmacokinetic ($Z = -2.981$, $p < 0.003$) and the heuristic model-based approach ($Z = -2.599$, $p < 0.009$). For the non-responding group, no significant differences were found. The first row of Figure 4 shows the change in the enhancing fraction between baseline and follow-up scans for the responding and the non-responding group using both the pharmacokinetic model (Figure 4A) and the heuristic (Figure 4B) model-based approach. For both parametric maps, the responding group (on the left-hand side) shows significant lowering of the enhancing fraction after treatment. For the non-responding group, no clear trend can be distinguished.

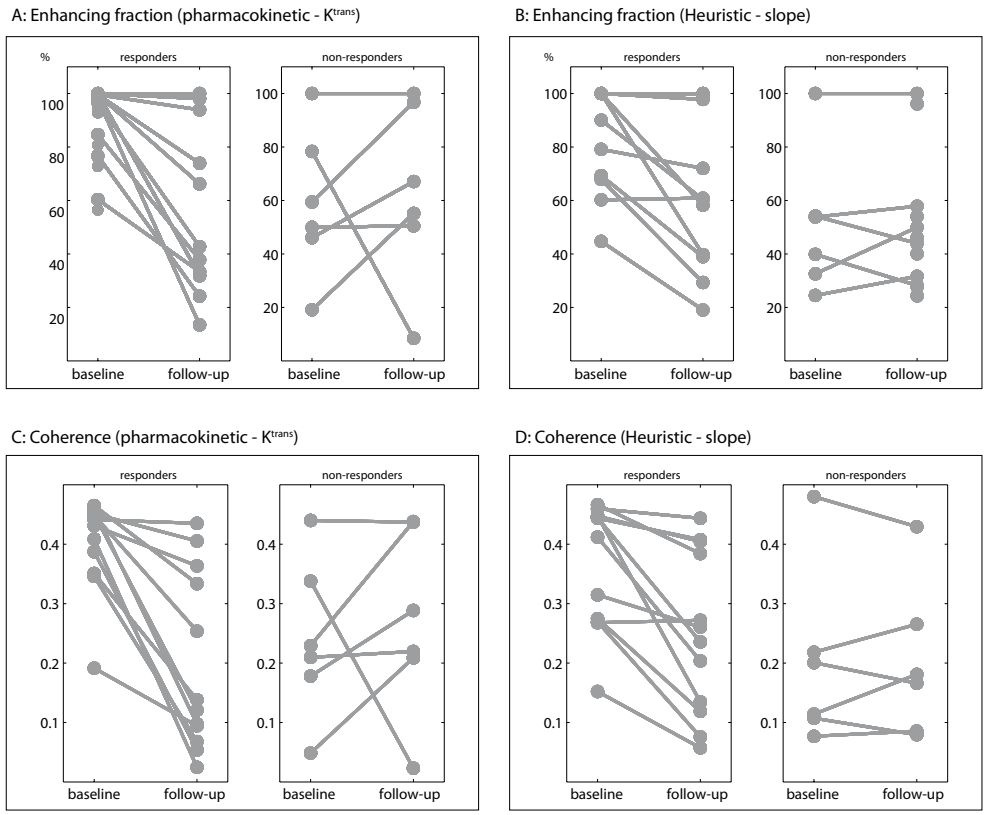


Figure 4. The enhancing fraction (row 1) and the coherence (row 2) for the pharmacokinetic (first column) and the heuristic (second column) model-based approach at baseline and follow-up. The responding group of patients is shown in all panels (A–D) on the left hand side, the non-responding group on the right-hand side.

In the second row of Figure 4, the change in coherence between pre- and post-treatment scans is shown for the responding and the non-responding group for the K^{trans} (Figure 4C) and slope (Figure 4D) maps. For both parametric maps, the responding group (on the left-hand side), reveals significant lowering of the enhancing fraction after treatment. For the non-responding group, no clear trend can be discerned

The results for coherence and fractal dimension, as a feature measuring parametric map heterogeneity, is summarized in Table 1 for all four parametric maps (K^{trans}, slope, maximum enhancement, and AUC). The coherence of all four parametric maps was significantly different between baseline and follow-up for the responding group. For the non-responding group, no significant differences were found. The fractal dimension showed no significant differences between baseline and follow-up parametric maps.

Table 1. Difference in coherence and fractal dimension between baseline and follow-up parametric maps for both response categories as tested with the Wilcoxon signed rank test. Z-values in bold indicate a significant difference.

	Coherence Z (p-value)		**Fractal dimension Z (p-value)**	
	Responders	Non-responders	Responders	Non-responders
K^{trans}	**-3.059 (0.002)**	-0.734 (0.463)	-0.078 (0.937)	-0.524 (0.600)
Slope	**-2.981 (0.003)**	-0.105 (0.917)	-1.020 (0.308)	-0.314 (0.753)
Max Enhancement	**-2.903 (0.004)**	-0.105 (0.917)	-0.706 (0.480)	-1.153 (0.249)
AUC	**-2.824 (0.005)**	-0.105 (0.917)	-0.078 (0.937)	-0.524 (0.600)

The Mann–Whitney *U*-test was used to assess differences in parametric map heterogeneity between the response categories (responders, non-responders) as measured at baseline. The heterogeneity of the parametric maps (K^{trans}, slope, maximum enhancement, and AUC) was assessed using the enhancing fraction, coherence and fractal dimension.

In the baseline scan, the enhancing fraction showed a significant difference between the two response categories using both the pharmacokinetic ($Z = -2.212$, $p < 0.027$) and the heuristic model-based approach ($Z = -2.412$, $p < 0.016$). On average, the tumours that respond to treatment have a high enhancing fraction before treatment, regardless of the exact nature of the parametric map. In the follow-up scan, no significant differences were found between responders and non-responders.

For baseline scans, significant differences were found in coherence between both response categories using all the DCE-MRI parametric maps, i.e. the pharmacokinetic

and the three heuristic model parameters (Tabel 2). At baseline, the tumours that responded to treatment showed a high coherence, irrespective of the exact nature of the parametric map. There was no significant difference between the groups in the fractal dimension, based on either baseline or follow-up data.

Tabel 2. Results of the testing for differences in coherence and fractal dimension between both response categories in baseline and follow-up parametric maps. Z-values in bold indicate a significant difference.

	Coherence		**Fractal dimension**	
	Baseline	Follow-up	Baseline	Follow-up
	Z (p-value)	Z (p-value)	Z (p-value)	Z (p-value)
K^{trans}	**-2.529 (0.011)**	-0.937 (0.340)	-1.030 (0.303)	-0.656 (0.512)
Slope	**-2.060 (0.039)**	-0.562 (0.574)	-0.656 (0.512)	-1.405 (0.160)
Max enhancement	**-2.060 (0.039)**	-0.843 (0.388)	-1.124 (0.261)	-1.124 (0.261)
AUC	**-2.154 (0.031)**	-0.562 (0.574)	-0.749 (0.454)	-1.592 (0.111)

Figure 5 illustrates the results for the pharmacokinetic and the heuristic model-based approach at baseline and follow-up and for both response categories (responding and non-responding). The first row (panels A and B) illustrates the results for the enhancing fraction, and the remaining panels (C-F) illustrate the results for the coherence. Whereas at the follow-up no significant differences can be found between the two response groups, both the pharmacokinetic and the heuristic model-based approach result in significant differences at the baseline. At the baseline significant differences were found for all parametric maps.

Table 3 shows the classification results for the heterogeneity features that are statistically significant in separating responders and non-responders. For the treatment monitoring, the coherence derived from the K^{trans} parametric map shows the highest sensitivity, whereas the coherence of the enhancement has the highest specificity. The accuracy of the coherence of both maps is similar, 83%. Founded on solely baseline DCE-MRI, the accuracy and sensitivity of the coherence derived from the parametric maps range from 67% to 83%. However, the specificity for the heuristic model-based approach was 43%.

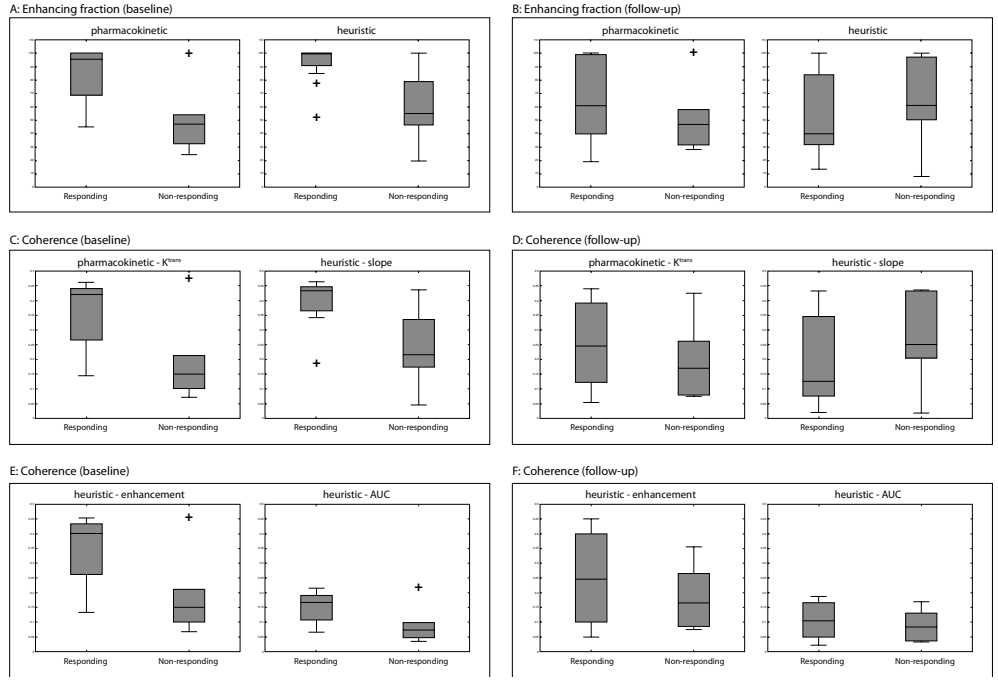


Figure 5. Differences in enhancing fraction (A, B) and coherence (C, D, E and F) for the pharmacokinetic and heuristic model-based approach for both response categories at baseline and follow-up.

Table 3. Results of the classification for the coherence feature

	Δ (Follow-up – baseline)			**Baseline**		
	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)
K^{trans}	83	91	71	83	91	71
Slope	56	64	43	72	91	43
Max enhancement	83	83	83	72	91	43
AUC	56	71	43	67	82	43

Discussion

Currently, volume measurements on standard non-enhanced MR images are used to assess treatment effects of ILP. However, tumour volume changes are not a reliable predictor of treatment effect in soft tissue sarcomas [15]. Since the functional changes of the tumour precede tumour volume changes, DCE-MRI can be a valuable tool

to provide insight into the changes in blood flow and permeability following treatment. Pharmacokinetic and model-free parameters to quantify DCE-MRI data have been extensively described. However, DCE-MRI data analysis is relatively complex [45]. For example, the acquisition strategy, the contrast agent used, selection of the arterial input function, and image analysis method, all have considerable influence on the absolute values of the estimated parameters, hampering the comparability of parameters derived from DCE-MRI. Consensus meetings have emphasized the need for standardized acquisition and analysis methods [16]. However, information is not only present in the absolute values of the DCE-MRI derived parametric maps, but also in the heterogeneity of the map. Human solid tumours can be biologically heterogeneous [12], and this is reflected in the parametric map. This heterogeneity can be both diagnostically and therapeutically important. Using texture analysis methods, the heterogeneity of a parametric map can be assessed independently from the absolute values of the parametric map. The additional analysis provides a more complete quantitative description of the contrast uptake in the tumour. Especially in situations where DCE-MRI analysis cannot yield absolute measurements, owing to the lack of calibration sequences, texture analysis can add important information.

In this study, we investigated the potential of heterogeneity in DCE-MRI parametric maps, quantified by texture analysis, as a biomarker for treatment response. The study used both a pharmacokinetic and a heuristic model-based approach for assessment of heterogeneity in contrast uptake. Whereas in the literature the pharmacokinetic approach is widely used, in clinical routine the heuristic model parameters are commonly available for visual evaluation of DCE-MRI images. The heuristic model-based parameters are reported to have poor reproducibility. However, these reproducibility studies were based on the absolute values of the parameters and not on the spatial distribution of the parameters. This study investigated whether heterogeneity of contrast uptake parameters from routinely acquired DCE-MRI scans, irrespective of the modelling approach, can be used to monitor treatment effect.

The first experiment investigated the correlation between the visual appearance of both pharmacokinetic and heuristic parametric maps: we spatially correlated the parametric values. The correlations ranged from 0.44 for maximum enhancement to 0.61 for AUC (Figure 1 shows the similarity in visual appearance). It is clear that due to the fitting procedure the pharmacokinetic parameter is less sensitive to noise, but the overall visual heterogeneity is very similar.

The monitoring study demonstrated that, regardless of the origin of the estimated parameter (pharmacokinetic or heuristic model-based), the enhancing fraction and coherence showed a significant difference between pre- and post-treatment acquisitions for the group that responded to therapy. Both the enhancing fraction and the coherence were on average lowered due to treatment in the responding group.

For the non-responders, no clear trend was observed. However, this group has an outlier showing clear similarity to the responding group. After careful evaluation of the DCE-MRI data, the computed pixel maps and the estimated features, we are fully convinced that, even though this patient was classified as non-responding, the MRI appearance was consistent with that of the responding group. Since the coherence is computed only over the enhancing voxels, the enhancing fraction and coherence are by definition independent. As the non-enhancing voxels are assumed to represent necrotic tissue, it seems that (in general) due to treatment the necrotic fraction of the tumour becomes larger and the viable tissue becomes less coherent in the responding group. In terms of classification performance, the coherence feature yielded a comparable accuracy, sensitivity and specificity for both mapping approaches, i.e. accuracy of 83%, sensitivity 83% to 92%, and specificity 71% to 83%. These values are comparable with absolute DCE-parametric values as reported in the literature [46, 47].

When evaluating the enhancing fraction and coherence on the pre-treatment scan, it was found that both features differentiate between outcome categories. It seems that patients with a large viable tumour fraction and a high coherence respond well to therapy, whereas patients with a sarcoma that already has large necrotic areas and a low coherence shows less response to treatment. In terms of classification performance, our method yielded a comparable accuracy, sensitivity and specificity for both mapping approaches, i.e., accuracy of 83%, sensitivity 91%, and specificity 43–71%. Post-treatment, no significant differences in the enhancing fraction or coherence between the responding and the non-responding group could be found: the sarcomas of the responding group have become similar to those of the non-responding group.

Even though the discriminating features (enhancing fraction and coherence) were computed independently, they demonstrate a similar tendency, i.e. for the responding group those features were on average high in pre-treatment acquisitions and lowered by treatment. That might indicate that both features are related to the same underlying biological process. In the response to treatment, enhancing fraction and heterogeneity may even be a key factor: well-perfused and coherent areas seem to respond differently to treatment than areas with a low perfusion and low coherence. The fractal dimension failed both tests on parametric map heterogeneity differentiation. Whereas Rose *et al* [26] found that the fractal dimension differed significantly between low and high-grade gliomas, in the case of sarcomas the fractal dimension did not change significantly due to treatment and did not differentiate between responders and non-responders. In this study, this might be due to the fact that the texture features (coherence and fractal dimension) are computed only over the enhancing part of the tumour. Whereas the coherence is not hampered by this, the fractal dimension might be sensitive for this ‘black hole’ in the image.

Some researchers choose to perform texture analysis on post-contrast enhanced MR images and not on the parametric maps. Nie *et al* [48] used images 1 min after contrast arrival and Prescott *et al* [49] used the images at the plateau phase at 75 s. Although it seems efficient to use only one time point from the dynamic series, the question remains whether or not specific information is lost. Both the pharmacokinetic and the heuristic model DCE-MRI parameters integrate information from several time points of the time-contrast or the time intensity curve. If the heterogeneity in a plateau phase image contains relevant information for diagnostic [48] or treatment outcome prediction [49], DCE-MRI could be confined to one late enhancement MRI. In this study, we preferred the use of both pharmacokinetic and model-free-based parameters based on the contrast time or intensity time curve.

In conclusion, in a relatively small group of patients, we established a high correlation between the parametric maps created using the pharmacokinetic and the heuristic model-based approach. In addition, we demonstrated the potential usefulness of features derived from DCE-MRI parametric maps (irrespective of the exact nature of the parametric map) to study tumour response to chemotherapy. Using this method, *in vivo* tumour mass can be monitored by serial MRI studies during the whole course of chemotherapy. When correlated to histopathological findings, this method might be clinically useful in understanding pathophysiologic changes, predicting tumour response and guiding therapeutic approach.

Conclusions

This study investigated whether heterogeneity of contrast uptake parameters from routinely acquired DCE-MRI scans, irrespective of the modelling approach, can be used to predict treatment outcome and monitor treatment effect. Results show that there is a considerable correlation in the parametric maps between pharmacokinetic modelling and the heuristic model-based approach. Therefore, it is not surprising that the heterogeneity measures derived from the parametric maps of both approaches show similar results in monitoring and predicting treatment response.

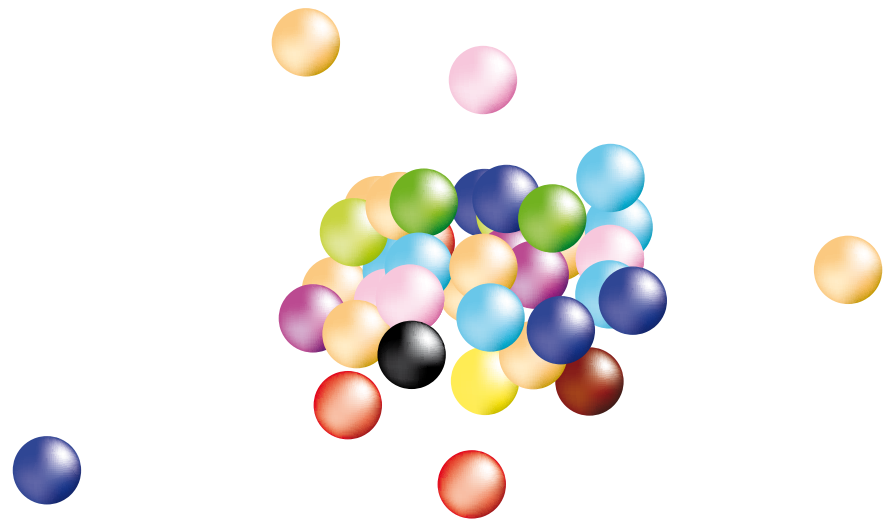
Our results suggest that specific heterogeneity features, computed from both pharmacokinetic and heuristic model-based parametric maps, show potential as a biomarker for monitoring tumour response.

References

1. Padhani, A.R., *Dynamic contrast-enhanced MRI in clinical oncology: current status and future directions.* J Magn Reson Imaging, 2002. 16(4): p. 407-22.
2. O'Connor, J.P., et al., *DCE-MRI biomarkers in the clinical evaluation of antiangiogenic and vascular disrupting agents.* Br J Cancer, 2007. 96(2): p. 189-95.
3. Zweifel, M. and A.R. Padhani, *Perfusion MRI in the early clinical development of antivascular drugs: decorations or decision making tools?* Eur J Nucl Med Mol Imaging, 2010. 37(Suppl 1): p. S164-82.
4. Galbraith, S.M., et al., *Combretastatin A4 phosphate has tumor antivascular activity in rat and man as demonstrated by dynamic magnetic resonance imaging.* J Clin Oncol, 2003. 21(15): p. 2831-42.
5. Morgan, B., et al., *A simple, reproducible method for monitoring the treatment of tumours using dynamic contrast-enhanced MR imaging.* Br J Cancer, 2006. 94(10): p. 1420-7.
6. Yu, H.J., et al., *MRI measurements of tumor size and pharmacokinetic parameters as early predictors of response in breast cancer patients undergoing neoadjuvant anthracycline chemotherapy.* J Magn Reson Imaging, 2007. 26(3): p. 615-23.
7. Nagashima, T., et al., *Dynamic enhanced MRI predicts chemosensitivity in breast cancer patients.* Eur J Radiol, 2006. 60(2): p. 270-4.
8. Pickles, M.D., et al., *Prognostic value of pre-treatment DCE-MRI parameters in predicting disease free and overall survival for breast cancer patients undergoing neoadjuvant chemotherapy.* Eur J Radiol, 2009. 71(3): p. 498-505.
9. van Laarhoven, H.W., et al., *Prediction of chemotherapeutic response of colorectal liver metastases with dynamic gadolinium-DTPA-enhanced MRI and localized 19F MRS pharmacokinetic studies of 5-fluorouracil.* NMR Biomed, 2007. 20(2): p. 128-40.
10. Giesel, F.L., et al., *Dynamic contrast-enhanced MRI of malignant pleural mesothelioma: a feasibility study of noninvasive assessment, therapeutic follow-up, and possible predictor of improved outcome.* Chest, 2006. 129(6): p. 1570-6.
11. Lavini, C., et al., *Pixel-by-pixel analysis of DCE MRI curve patterns and an illustration of its application to the imaging of the musculoskeletal system.* Magn Reson Imaging, 2007. 25(5): p. 604-12.
12. Marusyk, A. and K. Polyak, *Tumor heterogeneity: causes and consequences.* Biochim Biophys Acta, 2010. 1805(1): p. 105-17.
13. Parker, G.J., et al., *Probing tumor microvascularity by measurement, analysis and display of contrast agent uptake kinetics.* J Magn Reson Imaging, 1997. 7(3): p. 564-74.
14. Hayes, C., A.R. Padhani, and M.O. Leach, *Assessing changes in tumour vascular function using dynamic contrast-enhanced magnetic resonance imaging.* NMR Biomed, 2002. 15(2): p. 154-63.
15. van Rijswijk, C.S., et al., *Dynamic contrast-enhanced MR imaging in monitoring response to isolated limb perfusion in high-grade soft tissue sarcoma: initial results.* Eur Radiol, 2003. 13(8): p. 1849-58.
16. Leach, M.O., et al., *The assessment of antiangiogenic and antivascular therapies in early-stage clinical trials using magnetic resonance imaging: issues and recommendations.* Br J Cancer, 2005. 92(9): p. 1599-610.
17. Robinson, S.P., et al., *Tumour dose response to the antivascular agent ZD6126 assessed by magnetic resonance imaging.* Br J Cancer, 2003. 88(10): p. 1592-7.
18. Gaustad, J.V., et al., *Dynamic contrast-enhanced magnetic resonance imaging of human melanoma xenografts with necrotic regions.* J Magn Reson Imaging, 2007. 26(1): p. 133-43.
19. Mullamitha, S.A., et al., *Phase I evaluation of a fully human anti-alphav integrin monoclonal antibody (CNTO 95) in patients with advanced solid tumors.* Clin Cancer Res, 2007. 13(7): p. 2128-35.
20. Issa, B., D.L. Buckley, and L.W. Turnbull, *Heterogeneity analysis of Gd-DTPA uptake: improvement in breast lesion differentiation.* J Comput Assist Tomogr, 1999. 23(4): p. 615-21.
21. Emblem, K.E., et al., *Glioma grading by using histogram analysis of blood volume heterogeneity from MR-derived cerebral blood volume maps.* Radiology, 2008. 247(3): p. 808-17.

22. Mills, S.J., et al., *Do cerebral blood volume and contrast transfer coefficient predict prognosis in human glioma?* AJNR Am J Neuroradiol, 2006. 27(4): p. 853-8.
23. Mayr, N.A., et al., *Pixel analysis of MR perfusion imaging in predicting radiation therapy outcome in cervical cancer.* J Magn Reson Imaging, 2000. 12(6): p. 1027-33.
24. Checkley, D., et al., *Dynamic contrast-enhanced MRI of vascular changes induced by the VEGF-signalling inhibitor ZD4190 in human tumour xenografts.* Magn Reson Imaging, 2003. 21(5): p. 475-82.
25. Li, K.L., et al., *Heterogeneity in the angiogenic response of a BT474 human breast cancer to a novel vascular endothelial growth factor-receptor tyrosine kinase inhibitor: assessment by voxel analysis of dynamic contrast-enhanced MRI.* J Magn Reson Imaging, 2005. 22(4): p. 511-9.
26. Rose, C.J., et al., *Quantifying spatial heterogeneity in dynamic contrast-enhanced MRI parameter maps.* Magn Reson Med, 2009. 62(2): p. 488-99.
27. Karahaliou, A., et al., *Assessing heterogeneity of lesion enhancement kinetics in dynamic contrast-enhanced MRI for breast cancer diagnosis.* Br J Radiol, 2010. 83(988): p. 296-309.
28. Alic, L., et al. *Quantification of Heterogeneity in Dynamic Contrast Enhancement MRI Data for Tumor Treatment Assessment*. 2006. IEEE International Symposium on Biomedical Imaging: From Nano to Macro. Arlington, Virginia.
29. Haralick, R.M., K. Shanmugam, and J. Dinstein, *Textural features for image classification.* IEEE Trans Syst Man Cybern, 1973. 6: p. 12.
30. Peleg, S., et al., *Multiple Resolution Texture Analysis and Classification.* IEEE Transactions on pattern analysis and machine intelligence, 1984. 6: p. 518-523.
31. Mandelbrot, B.B., *The fractal geometry of nature*. Updated and augm. ed. 1983, New York: W.H. Freeman. 468 p.
32. Falconer, K., *Fractal Geometry: Mathematical Foundations and Applications*. 1990: John Wiley & Sons. 310.
33. Lynch, J.A., D.J. Hawkes, and J.C. Buckland-Wright, *Analysis of texture in macroradiographs of osteoarthritic knees using the fractal signature.* Phys Med Biol, 1991. 36(6): p. 709-22.
34. Veenland, J.F., et al., *Estimation of fractal dimension in radiographs.* Med Phys, 1996. 23(4): p. 585-94.
35. Veenland, J.F., J.L. Grashuis, and E.S. Gelsema, *Texture analysis in radiographs: the influence of modulation transfer function and noise on the discriminative ability of texture features.* Med Phys, 1998. 25(6): p. 922-36.
36. Grunhagen, D.J., et al., *Outcome and prognostic factor analysis of 217 consecutive isolated limb perfusions with tumor necrosis factor-alpha and melphalan for limb-threatening soft tissue sarcoma.* Cancer, 2006. 106(8): p. 1776-84.
37. Creech, O., Jr., et al., *Chemotherapy of cancer: regional perfusion utilizing an extracorporeal circuit.* Ann Surg, 1958. 148(4): p. 616-32.
38. Eggermont, A.M., et al., *Isolated limb perfusion with tumor necrosis factor and melphalan for limb salvage in 186 patients with locally advanced soft tissue extremity sarcomas. The cumulative multicenter European experience.* Ann Surg, 1996. 224(6): p. 756-64; discussion 764-5.
39. Grunhagen, D.J., et al., *One hundred consecutive isolated limb perfusions with TNF-alpha and melphalan in melanoma patients with multiple in-transit metastases.* Ann Surg, 2004. 240(6): p. 939-47; discussion 947-8.
40. Olieman, A.F., et al., *Hyperthermic isolated limb perfusion with tumor necrosis factor alpha, interferon gamma, and melphalan for locally advanced nonmelanoma skin tumors of the extremities: a multicenter study.* Arch Surg, 1999. 134(3): p. 303-7.
41. Gustafson, D.E. and W.C. Kessel. *Fuzzy clustering with a fuzzy covariance matrix.* in *IEEE Conference on Decision and Control*. 1979.
42. Tofts, P.S. and A.G. Kermode, *Measurement of the blood-brain barrier permeability and leakage space using dynamic MR imaging. 1. Fundamental concepts.* Magn Reson Med, 1991. 17(2): p. 357-67.
43. Hittmair, K., et al., *Method for the quantitative assessment of contrast agent uptake in dynamic contrast-enhanced MRI.* Magn Reson Med, 1994. 31(5): p. 567-71.
44. Duda, R.O., P.E. Hart, and D.G. Stork, *Pattern Classification*. 2001: John Wily & Sons. Inc..

45. Jackson, A., et al., *Imaging tumor vascular heterogeneity and angiogenesis using dynamic contrast-enhanced magnetic resonance imaging.* Clin Cancer Res, 2007. 13(12): p. 3449-59.
46. Ah-See, M.L., et al., *Early changes in functional dynamic magnetic resonance imaging predict for pathologic response to neoadjuvant chemotherapy in primary breast cancer.* Clin Cancer Res, 2008. 14(20): p. 6580-9.
47. Kozlowski, P., et al., *Combined prostate diffusion tensor imaging and dynamic contrast enhanced MRI at 3T--quantitative correlation with biopsy.* Magn Reson Imaging, 2010. 28(5): p. 621-8.
48. Nie, K., et al., *Quantitative analysis of lesion morphology and texture features for diagnostic prediction in breast MRI.* Acad Radiol, 2008. 15(12): p. 1513-25.
49. Prescott, J.W., et al., *Temporal Analysis of Tumor Heterogeneity and Volume for Cervical Cancer Treatment Outcome Prediction: Preliminary Evaluation.* J Digit Imaging, 2010. 23(3): p. 342-357.



Chapter 5

Regional heterogeneity changes in DCE-MRI as response to isolates limb perfusion in experimental soft-tissue sarcoma

This chapter is based uppon:
L Alić, M van Vliet, PA Wielopolski, TLM ten Hagen, CF van Dijke, WJ Niessen, JF Veenland. Regional heterogeneity changes in DCE-MRI as response to isolated limb perfusion in experimental soft-tissue sarcomas. *Contrast Media Mol Imaging. 2013, 8(4): 340-349.*

Abstract

Experimental evidence supports an association between heterogeneity in tumour perfusion and response to chemotherapy/radiotherapy, disease progression and malignancy. Therefore, changes in tumour perfusion may be used to assess early effects of tumour treatment. However, evaluating changes in tumour perfusion during treatment is complicated by extensive changes in tumour type, size, shape, and appearance. Therefore, this study assesses the regional heterogeneity of tumours by dynamic contrast-enhanced MRI (DCE-MRI) and evaluates changes in response to isolated limb perfusion (ILP) with TNF-α and melphalan. Data were acquired in an experimental cancer model, using a macromolecular contrast medium (MMCM), albumin-$(Gd\text{-}DTPA)_{45}$. Small fragments of BN 175 (a soft-tissue sarcoma) were implanted in eight brown Norway (BN) rats. MRI of five drug treated and three sham treated rats was performed at baseline and 1 h after ILP-intervention. Properly co-registered baseline and follow-up DCE-MRI were used to estimate the volume transfer constant (K^{trans}) pharmacokinetic maps. The regional heterogeneity was estimated in 16 tumour sectors and presented in cumulative map-volume histograms. On average, ILP treated tumours showed a decrease in regional heterogeneity on the histograms. This study shows that heterogenic changes in regional tumour perfusion, estimated using DCE-MRI pharmacokinetic maps, can be measured and used to assess the short-term effects of a potentially curative treatment on the tumour microvasculature in an experimental soft-tissue sarcoma model.

Introduction

Increased use of current and development of novel anti-angiogenic agents for cancer treatment poses new challenges to medical imaging technologies. In contrast to the systemic effect of traditional cytotoxic drugs, these new agents selectively target the endothelial cells of tumour neovasculature [1-3]. Although often costly, these treatments are potentially effective and have become a priority for oncologists and pharmaceutical companies.

Low local concentration of anti-tumour agents and dose-limiting toxicity are often the cause of failure of chemotherapy. In isolated limb perfusion (ILP) local drug concentrations are increased while systemic exposure to the drug is minimal [4]. By utilizing ILP, tumours are exposed to concentrations up to 15-20 times higher than the maximum tolerated dose of tumour necrosis factor alpha (TNF-α) without major side-effects. TNF-α also has an indirect effect on endothelial cells and tumour-associated vasculature [5-7]. By combining TNF-α with a cytostatic drug (e.g. melphalan) an immediate TNF-mediated tumour-selective enhanced drug uptake can be realised [8, 9]. This effect depends on the vascularisation of the tumour and is followed by a tumour-selective permeabilisation and eradication of the tumour-associated vasculature [10, 11].

Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) enables to non-invasively characterise tumours, and has emerged as a valuable tool for monitoring pathophysiologic changes in various aspects of tumour vascular structure and function [12, 13]. DCE-MRI with macromolecular contrast media (MMCM) depicts microvessel permeability [14] and has been shown in experimental models to express tumour perfusion solely due to malfunctioning tumour vessels [5, 15].

Experimental evidence shows the association of tumour vascular heterogeneity with disease progression and malignancy [3]. Heterogeneity of tumour blood supply results in the formation of hypoxic voids associated with oxidative stress, promotion of survival factors and genomic instability [16]. This indicates the importance of tumour vascular heterogeneity for disease mechanism and possibly will lead to the discovery of novel treatment strategies. A heterogeneous blood supply will also affect treatment response due to poor delivery of chemotherapeutic agents to areas of low vascularity. The importance of neovascular heterogeneity for tumour treatment is well documented [3, 17, 18]. Although the use of tumour heterogeneity as a biomarker for treatment response has been evaluated [19-21], these studies summarise tumour heterogeneity across the whole tumour and thereby fail to identify regional changes.

An alternative approach is to deal with intratumoural heterogeneity at the sub-tumour level. This method partitions a tumour into volumes of interest (VOI) based on certain tumour properties, such as identifying regions with bulky enhancement, i.e., the 'hot-spot'. This has been performed either manually [22-24] or by thresholding a parameter value or a fit statistic [25, 26]. Another approach, inspired by the strong peripheral tumour enhancement common in solid tumours, involves tumour segmentation into concentric bands [5, 27, 28 , 29, 30]. However, the number and size of those bands is rather arbitrary, and these studies do not account for the heterogeneity within such a concentric band.

The present preclinical study evaluates the potential of heterogeneity, in DCE-MRI with MMCM, as a biomarker for the early effect of a combined cancer treatment with TNF-α and melphalan in an experimental model of soft-tissue sarcoma. The changes in temporal pattern of signal enhancement due to MMCM are expected to reflect changes due to treatment. Following change in the tumour over time requires properly co-registered baseline to follow-up pairs, which is complicated by extensive changes in tumour size, tumour shape and DCE-MRI appearance. This makes it difficult to analyse tumours on a complete voxel-wise level. Therefore, we assessed the regional changes in DCE-MRI parametric maps which were co-registered to facilitate such an analysis. To quantify tumour heterogeneity, we introduce the cumulative map-volume (CMV) histogram. Although haemorrhagic necrosis and destruction of the endothelial cells will develop in time after TNF-α administration, we hypothesise that the effect of ILP-intervention will be evident 1 h after initiation.

Material and methods

Animal and tumour model

Eight male inbred BN strain rats (Harlan-CPB, Austerlitz, the Netherlands), with a mean body weight of 300 g were implanted subcutaneously, in the hind limb, with 4 mm fragments of the syngenic BN175 soft-tissue sarcoma. The animals were inspected daily for tumour growth and general appearance. The tumours were imaged by MRI when they reached approximately 15 mm in diameter. Prior to MRI, the animals were randomly assigned to the drug treated group (five) or sham treated group (three). They were anesthetized with subcutaneous injection of 150 µl of a 1:1 (v:v) mixture of ketamine (Alfasan, Woerden, the Netherlands) and Xylazine (Bayer AG, Leverkusen, Germany). A 25-gauge butterfly cannula was inserted into a tail vein for injection of contrast medium. The study was conducted with the approval of the local Committee for Animal Research.

Magnetic resonance contrast media

Albumin-(Gd-DTPA)$_{45}$ is a water-soluble macromolecular contract medium (MMCM) with a molecular weight of 92 kDa which corresponds to about 45 molecules of Gd-DTPA covalently bound to each albumin molecule. It was synthesized following the method of Ogan et al. [31]. This contrast agent has a distribution volume of 0.05 l/kg and a plasma half-life of 3 h in rats. Albumin-(Gd-DTPA)$_{45}$ was injected at a dose of 0.03 mMGd/kg. The size of the MMCM molecule prevents leakage through properly matured vessels outside the tumour, resulting in a prolonged period of intravascular retention of the contrast agent. Only newly formed, hyper-permeable tumour vessels contribute to contrast extravasation, When evaluated with DCE-MRI, MMCM demonstrates *in vivo* imaging properties that correlate with histological features of angiogenesis [32].

Magnetic resonance imaging

MRI was performed using a clinical 1.5T MRI scanner (Signa CVi, GE Healthcare, Milwaukee, WI, USA). A custom-made dedicated single-loop surface coil with an internal diameter of 1.7 cm was constructed specifically to create a high signal-to-noise ratio. The imaging protocol consisted of two consecutive sequences, a calibration and a dynamic scan. To estimate the $T_{1,0}$ of blood prior to contrast injection, a flow compensated gradient echo inversion recovery echo planar imaging (IR GE-EPI) sequence [33] with 25 inversion time (TI) values (start:step:end = 100:50:1200 ms) was performed using a TR/TE 15000/24.3 ms, matrix of 128x128 with an in-plane resolution of 0.55 x 0.55 mm and slice thickness of 0.8 mm covering the entire tumour (30 slices, scan time 11 minutes). Dynamic images were acquired using a 3D T1-weighted RF Spoiled Gradient Echo sequence with imaging parameters TR/TE 12.37/2.95 ms, flip angle of 18° and a readout bandwidth of 81 Hz/pixel. The matrix selected was 256x256x64 with a field-of-view (FOV) of 30x26 mm^2 leading to true voxel resolution of 0.12 x 0.14 x 0.50 mm^3. The temporal resolution was 3.1 min per 3D volume. The first 10 time points were acquired at a temporal resolution of 3.1 min. To sample the time frame efficiently, after the first 10 acquisitions, time gaps of 15 min were applied and after each gap 2 series were acquired. The total scan was acquired in approximately 90 min covering 16 time points.

Experimental protocol

MRI was performed at baseline and 1 h after isolated limb perfusion (ILP) [4]. ILP-intervention allows the delivery of high doses of cytostatic drugs by isolating the tumour-bearing limb from the systemic circulation. Briefly, the animals were anesthetized with ketamine and xylazine, and 50 units heparin were injected intravenously. The femoral artery and vein of anesthetized rats were cannulated with silastic tubing. Collaterals were occluded by groin tourniquet and perfusion was started after the

tourniquet was tightened. An oxygenation reservoir and a roller pump were included in the circuit. The animals were randomly assigned to either the drug treatment or the control group and perfused with 5 ml Haemaccel (Behring Pharma, Amsterdam, the Netherlands) with haemoglobin of 0.9 mmol/l.50g. In the drug treatment group, 50 µg *TNF-α* (Boehringer, Ingelheim, Germany) and 40 µg melphalan (Alkeran, Wellcome, Beckenham, UK) were added to the perfusate. In the control group no drugs were added to the perfusate. Perfusion was maintained for 30 min at a flow rate of 1.8 ml min^{-1}, and finalised by washout with 5 ml oxygenated Haemaccel. To restore physiological limb perfusion, the femoral artery was decannulated and sutured. During the ILP-intervention and MRI imaging, the rat's temperature was maintained at 38-39°C by means of a warm water mattress.

Motion correction and co-registration

Spatial correspondence between MRI sequences is not always guaranteed; in particular, baseline and follow-up scans may not be aligned. In addition, the long MRI acquisition time complicates the co-registration problems. To perform a proper quantitative analysis the spatial relation between the sequences needs to be established. Therefore, we co-register the images before and after the intervention. Figure 1 shows the registration procedure. First, the intra-sequence motion correction was performed for both sequences separately. Second, the separate sequences were co-registered. Third, the late-stage enhanced image for follow-up was registered to the baseline image to ensure the spatial correspondence before and after treatment. All registration steps were performed using Elastix [34], first rigidly and then as an affine transformation.

Pharmacokinetic modelling

The tumour contrast agent transport through the capillary plasma compartment was modelled by the Tofts model [35, 36]. The pharmacokinetic analysis of DCE-MRI requires estimates of the contrast concentration, C(t). To estimate C(t), a flow compensated inversion recovery gradient echo planar imaging (IR-GE EPI) sequence was acquired prior to contrast injection. For each animal, longitudinal relaxation time ($T_{1,0}$) was estimated voxelwise by a nonlinear least-square fit of the acquired signal intensities $S(T_{IN})$ against inversion times [37].

$$S(T_{IN}) = c* (1- 2e^{-T_{IN}/T_{1,0}}) \quad \text{Eq. 1}$$

To account for the incomplete clearance of the contrast agent after the intervention due to the prolonged retention of MMCM in the intravascular space, pre-contrast $T_{1,0}$ is estimated at follow-up acquisitions as well. The estimated $T_{1,0}$ was used in calculation of plasma contrast concentration over time, $C_p(t)$, and the tumour contrast concentration over time, $C_t(t)$. For both concentration curves, the molar concentration over time is defined by Eq. 2 [38].

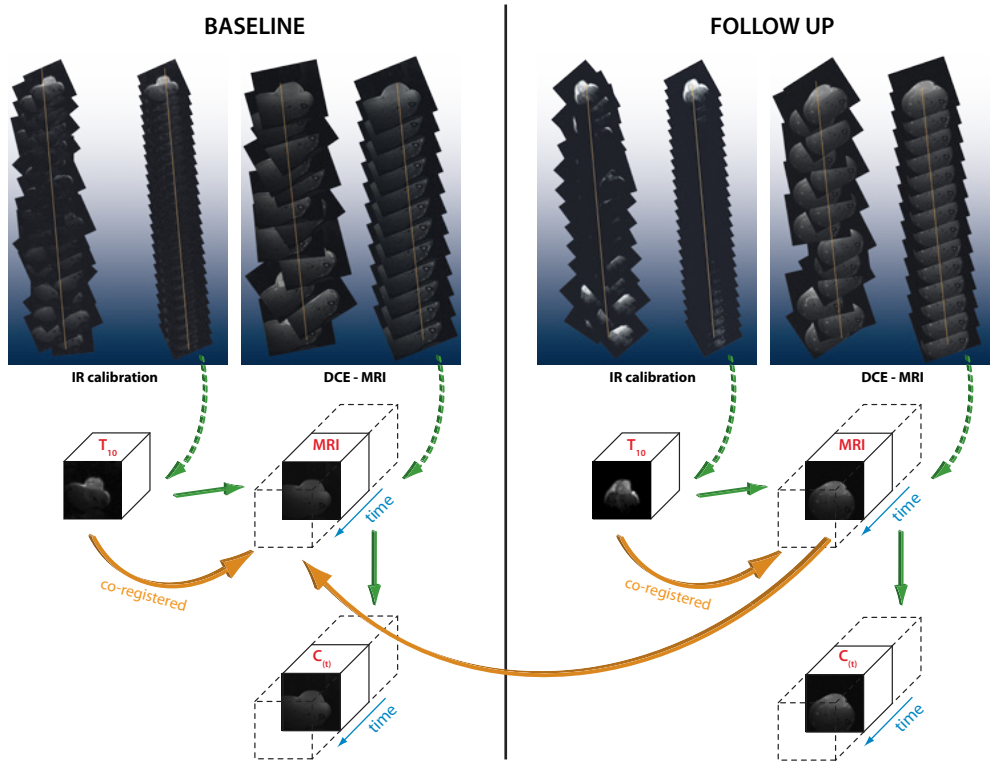


Figure 1. Overview of the registration steps at baseline (left-hand side) and follow-up (right-hand side) performed previous to pharmacokinetic analysis.

$$C(t) = \frac{1}{r_1} * (R_{1,0}(t) - R_{1,0}(0)) \quad \text{Eq. 2}$$

where $R_{1,0}$ (=1/$T_{1,0}$) and $R_{1,t}$ (=1/R1,t) represent the longitudinal relaxation rate of the tissue prior and at different time points after the contrast agent injection; r_1 is the T_1 relaxivity of $Albumin\text{-}(Gd\text{-}DTPA)_{45}$ at 1.5T (0.273 $M^{-1}sec^{-1}$ [31]).

Expressing the enhancement of MRI signal intensity, for a GRE sequence [39], after the contrast agent injection and separating parameter A offers the molar concentration of the contrast agent over time (Eq. 3).

$$A = \frac{S(t)}{S0} * \frac{\left(1-e^{-T_R/T_{1,0}}\right)}{\left(1-\cos\alpha e^{-T_R/T_{1,0}}\right)}$$

$$C(t) = \frac{1}{r_1} * \left[\frac{1}{T_R} * \ln\left(\frac{1-A\cos\alpha}{1-A}\right) - \frac{1}{T_{1,0}}\right] \quad \text{Eq.3}$$

For the plasma contrast concentration, often referred to as a vascular input function (VIF), we outlined manually a major tumour-feeding vessel in the latest contrast concentration image by using a volume of interest (VOI). By applying this VOI to all time points and spatial averaging, we created a contrast plasma concentration curve, $C_p(t)$. Subsequently, by assuming a bi-exponentially decaying VIF [40], we fitted $C_p(t)$ using the following function:

$$C_p(t) = a_1 e^{-m_1 t} + a_2 e^{-m_2 t} \quad \text{Eq. 4}$$

To compute K^{trans} (µl/sec), the estimated parameters a_1, a_2, m_1 and m_2 were used. The tumour contrast concentration curve $C_t(t)$ was fitted voxel-wise, by a nonlinear least-square fit, using Eq. 5.

$$C_t(t) = DK^{trans} \sum_{n=1}^{2} \frac{a_i}{m_i - k_{ep}} \left(e^{-m_i t} + e^{-k_{ep} t}\right) \quad \text{Eq. 5}$$

The fitting procedure resulted in a voxel-wise K^{trans} value and a coefficient of determination r^2. This coefficient summarizes the resemblance between observed values and the values expected under the Tofts model [38, 39]. The inherent image noise will mask low signal enhancement, resulting in low r^2. We excluded all voxels with $r^2<0.75$ from further analysis, since these voxels do not follow the Tofts model properly. Voxels with $r^2>0.75$ were identified as enhancing. The volume transfer constant, i.e., K^{trans}, was presented as parametric maps before and after the intervention by ILP. DCE-MRI data analysis was performed using in-house developed software based on MATLAB (MathWorks, Natick, MA, USA). All steps were performed voxel-wise after motion correction and co-registration.

Tumour features

The whole tumour was outlined manually by an experienced radiologist (CFvD) in the final DCE-MRI image. Using the whole tumour mask, we generated automatically the following masks (Figure 2): tumour rim, tumour core, tumour periphery. The tumour rim was defined as the peripheral 10% of the tumour volume. The tumour core was defined as the 50% of the tumour diameter covering the tumour core. The tumour periphery was defined as the outer 50% of the tumour. Tumour core and periphery were automatically each subdivided into 8 tumour sectors. Figure 2 shows the position of the various VOIs within a tumour. For whole tumour, tumour rim, tumour core, tumour periphery, and for all 16 tumour sectors we calculated first-order statistics (e.g., mean, median, standard deviation) of K^{trans}.

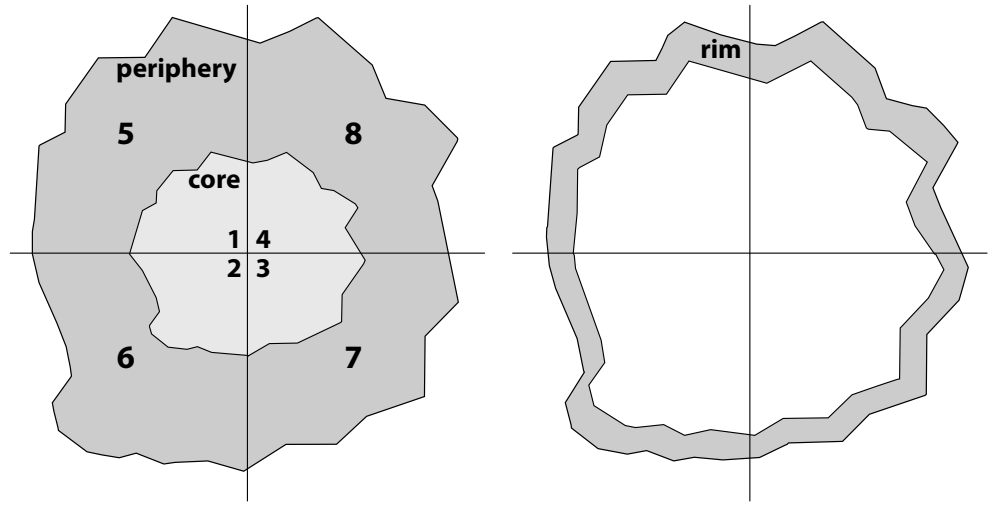


Figure 2. The tumour regions of interest (left-hand side) showing tumour periphery, tumour core and first eight tumour sectors. The tumour rim is shown on the right-hand side.

Cumulative map-volume histogram

Cumulative dose-volume histograms (DVH) have been used in radiotherapy to graphically summarize three-dimensional (3D) dose-distribution in two-dimensional graphs [41]. This concept has been expanded to cumulative dose-[functioning]-mass histogram (D[F]MH) [42, 43]. To describe the 3D map distribution of a pharmacokinetic map, we propose a cumulative map-volume (CMV) histogram. Besides the origin of the information that it summarizes, the CMV histogram is identical to D[F]MH. The pharmacokinetic map is first divided into 50 equal bins (rather an arbitrary number). As we are primarily interested in the enhancing fraction, only the voxels

belonging to this fraction were used for further analysis. The CMV histogram plots K^{trans} on the x-axis and percentage volume of the tumour (with $K^{trans} \geq$ that specific value) on the y-axis. The CMV always slopes from north-west to south-east and is usually a sigmoid shaped function. The slope of this function reflects the distribution of the K^{trans}, and was estimated (by a least squares fitting) for three categories of K^{trans} values ($0<K^{trans}\leq1$), ($1<K^{trans}\leq2$), and ($2<K^{trans}\leq5$). The distribution of all estimated slopes was captured per tumour as mean and standard deviation over all 16 sectors. The variance (or standard deviation) of slopes, between the regions, is a measure for tumour heterogeneity. To report the group trends for both distribution parameters, we computed the mean and standard deviation over all animals belonging to a specific category (treated, sham, before treatment and after treatment).

Statistical analysis

Statistical analysis was performed using SPSS (SPSS for Windows, Version 17.0, SPSS Inc., Chicago, IL, USA). Two categories were defined: treated and sham. The Mann–Whitney U-test was used to assess differences between the two response categories based on the K^{trans} parametric map acquired before and after treatment. For the follow-up examination, the Wilcoxon signed rank test was used to assess the differences between the two response categories based on the K^{trans} parametric maps. A *p*-value of <0.05 was considered statistically significant.

Results

When implanted subcutaneously in the hind limb, the fragments of the syngenic BN175 soft-tissue sarcoma grew into tumours that appeared as masses composed of several lobes. Figure 3 shows the features of a lobular carcinoma with viable tumour parts (A) and less viable tumour parts (B). Tumour growth rates mirrored those previously published [5, 8], with an average post inoculation growth period of two weeks to reach 15 mm in diameter.

Co-registered baseline and follow-up DCE-MRI were used to estimate the pharmacokinetic maps of the volume transfer constant (K^{trans}). To estimate K^{trans} (µl/sec) accurately, we assessed all components at baseline and follow-up separately. The voxelwise tumour native longitudinal relaxation time ($T_{1,0}$) was estimated prior to the administration of contrast agent. The vascular input function (VIF) was estimated after the administration of contrast agent. To quantify the effect of tumour treatment, we evaluated the slope of the cumulative map-volume (CMV) histogram for the 16 tumour regions, only considering the enhancing fraction. DCE-MRI derived maps from a tumour before (top row) and 1 h after the intervention (bottom row) are shown in Figure 4 for ILP-intervention with TNF-α and melphalan and in Figure 5 for sham intervention. Panels A and D (for both figures) show a maximum intensity projection (MIP) of the late-stage enhanced image with arrows illustrating a VIF VOI position. Consecutive columns (in both Figure 4 and Figure 5) represent $T_{1,0}$ map before contrast agent administration (B and E), and K^{trans} map (C and F).

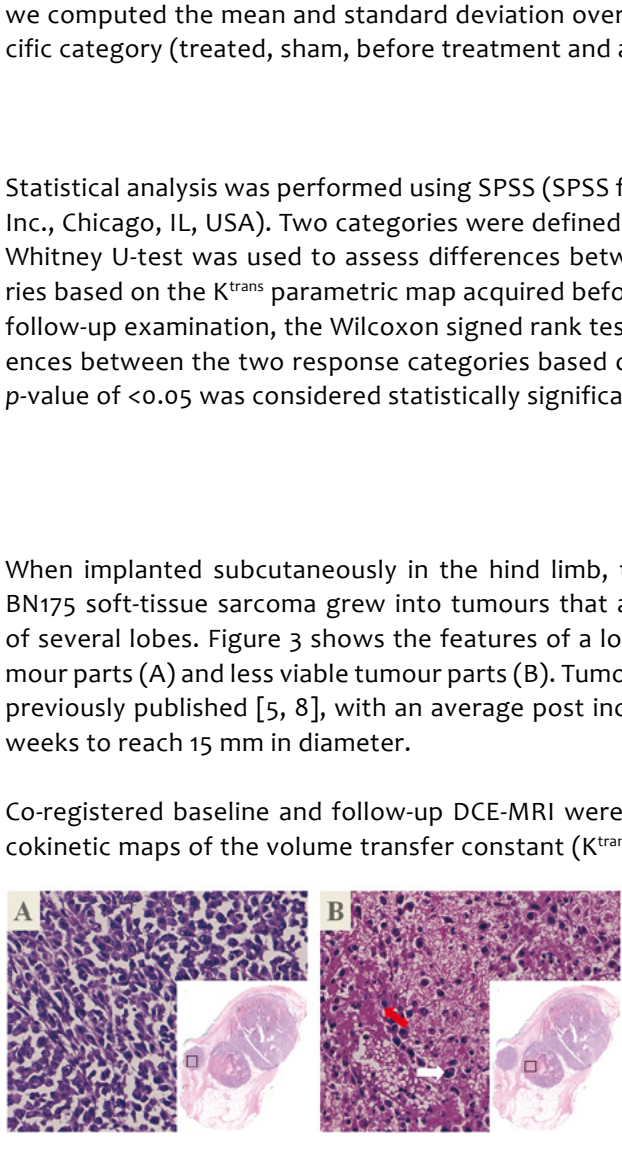


Figure 3. Morphology of syngenic BN175 soft-tissue sarcoma showing viable parts (A) and less viable parts (B) with good developed nuclei (white arrow) and pyknotic nuclei (red arrow).

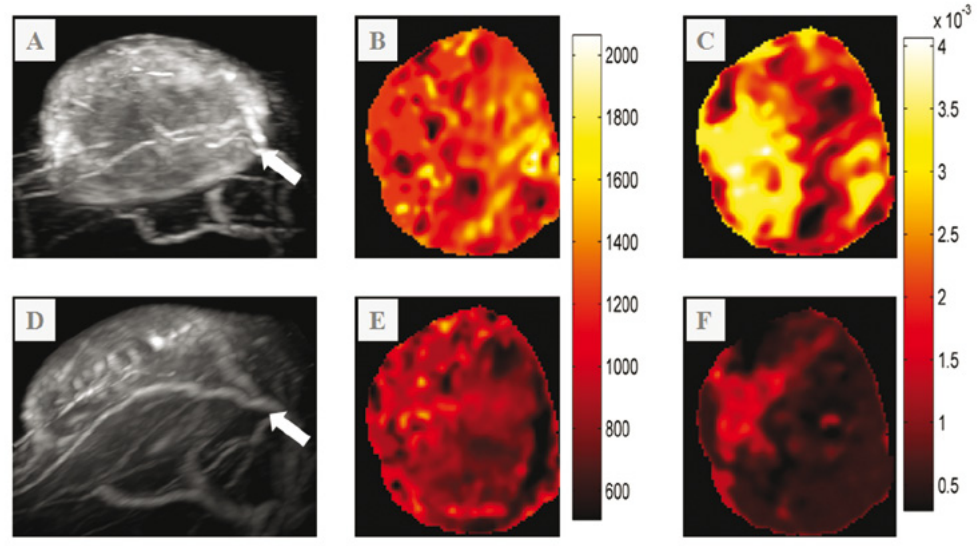


Figure 4. A tumor before (top row) and after (bottom row) ILP-intervention with TNF-α and melphalan. First column (A and D) shows a maximum intensity projection (MIP) of the late stage enhanced image with arrows illustrating position of a VIF VOI. Consecutive columns represent T10 map before contrast agent administration (B and E), and K^{trans} map (C and F).

Longitudinal relaxation rate $R_{1,0}$

For the VIF VOI, the average $T_{1,0} = 1/R_{1,0}$ at baseline was equal to 1.44±0.13 s. The average $T_{1,0}$ after the ILP-intervention was 1.98±0.12 s for animals treated with TNF-α and melphalan, and 1.30±0.06 s for animals treated with sham perfusion. For all fits, r^2-values were within the range of 0.61-0.99 (mean r^2 = 0.88±0.14). These results are similar to previous measurements in rat blood [5, 44].

For the tumour VOI the average $T_{1,0} = 1/ R_{1,0}$ at baseline was 1.16±0.2 s. The $T_{1,0}$ after the ILP-intervention was 0.72±0.12 s for animals treated with TNF-α and melphalan, and 1.07±0.15 s for animals after sham perfusion. Figure 4 shows the $T_{1,0}$ tumour map at baseline (B) and at follow-up (E) for an animal after ILP-intervention with TNF-α and melphalan. Figure 5 shows the $T_{1,0}$ tumour map at baseline (B) and at follow-up (E) for an animal after sham intervention.

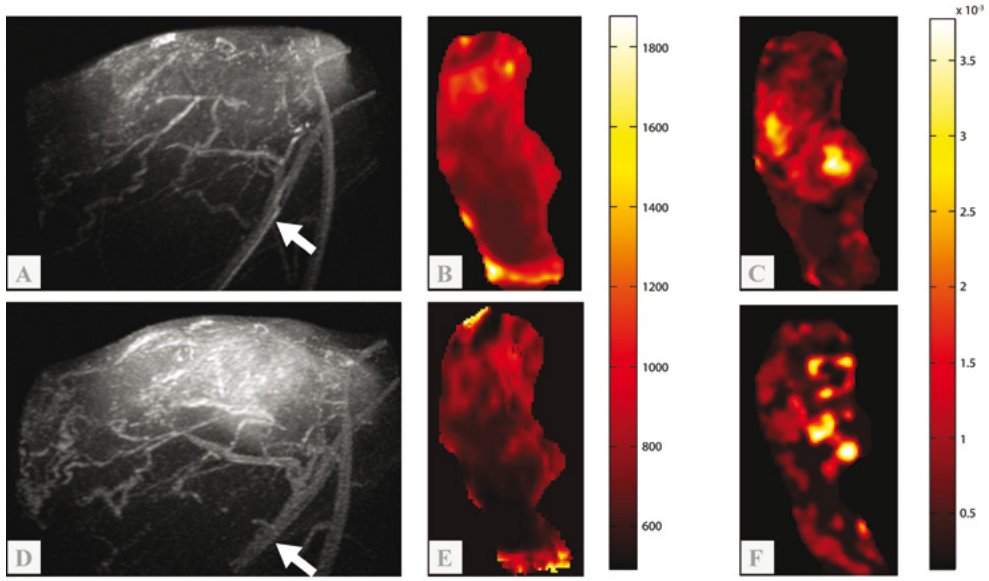


Figure 5. A tumor before (top row) and after (bottom row) sham-intervention with TNF-α and melphalan. First column (A and D) shows a maximum intensity projection (MIP) of the late stage enhanced image with arrows illustrating position of a VIF VOI. Consecutive columns represent $T_{1,0}$ map before contrast agent administration (B and E), and K^{trans} map (C and F).

Pharmacokinetics

Regardless the treatment category (i.e. treated or control), the intervention significantly decreases the enhancing fraction of the tumour from 78.02 ± 11.46% at baseline to 42.71 ± 23.32% at follow-up. There was no significant difference, in enhancing fraction, at baseline and at follow-up between the treatment categories or between the regions. Similar effects were observed in the tumour core (79.24 ± 15.04 at baseline and 40.76 ±26.87 at follow-up), the tumour periphery (77.43 ± 11.28 at baseline and 43.12 ± 22.63 at follow-up) and the tumour rim (73.53 ± 9.63 at baseline and 48.09 ± 18.66 at follow-up).

Table 1 presents the results for the mean estimated K^{trans} for different VOIs, before and after intervention. Figure 4 (ILP-intervention with TNF-α and melphalan) and

Figure 5 (sham intervention) show the K^{trans} tumour map before (C) and after (E) the intervention. The Mann-Whitney U-test was used to assess significant differences in parametric maps between the response categories (treated and sham) as measured before and after the intervention separately.

Table 1. Mean K^{trans} before and after isolated limb perfusion intervention for both treatment categories (TNF-α and melphalan, and sham). *Significantly different from baseline value, and §significantly different from control group. Statistical significance was set at p<0.05.

K^{trans} – mean	Treated		Sham	
	Baseline	Follow-up	Baseline	Follow-up
Tumour whole	1.59 ± 0.43	0.62 ± 0.17*,§	1.84 ± 0.98	1.85 ± 0.94
Tumour rim	1.53 ± 0.37	0.61 ± 0.17*,§	1.66 ± 0.88	1.96 ± 0.90
Tumour core	1.56 ± 0.48	0.62 ± 0.17*,§	1.94 ± 0.87	1.69 ± 1.16
Tumour periphery	1.60 ± 0.41	0.62 ± 0.17*,§	1.80 ± 1.00	1.91 ± 0.89

For all VOIs, the mean K^{trans} at baseline showed no significant difference between treatment and sham intervention. On the other hand, the mean K^{trans} at follow-up showed a significant difference between treatment and sham intervention (p=0.036). In addition, the difference in parametric maps at baseline and follow-up was assessed using the Wilcoxon signed rank test for the treatment and sham groups. The mean K^{trans} at follow-up was significantly different from baseline for treatment intervention (p=0.04), but not significantly different for sham intervention. For animals treated with TNF-α and melphalan, K^{trans} decreased in all VOIs. Figure 6 shows the box plots of K^{trans} at baseline and follow-up for both categories: i.e., treated (TNF-α and melphalan) and sham.

Sector results

Figure 7 clearly shows the distinction between animals treated with TNF-α and melphalan, and animals treated with sham perfusion. This scatter plot shows the mean K^{trans} per sector, at baseline and at follow-up, for all tumours. The mean was computed for the voxels belonging to the enhancing fraction. For the TNF-α and melphalan treated animals, a clear decrease in K^{trans} variance due to treatment. For the sham treated animals, the variance seems to increase. The average cumulative map-volume (CMV) histogram of the K^{trans} parametric map is plotted in Figure 8 for TNF-α and melphalan-treated animals (blue), and for sham-treated animals (red). Each of the 16 sectors was used to create 16 CMV histograms which were averaged afterwards. The dotted lines represent the standard deviation pooled over all sectors and all rats.

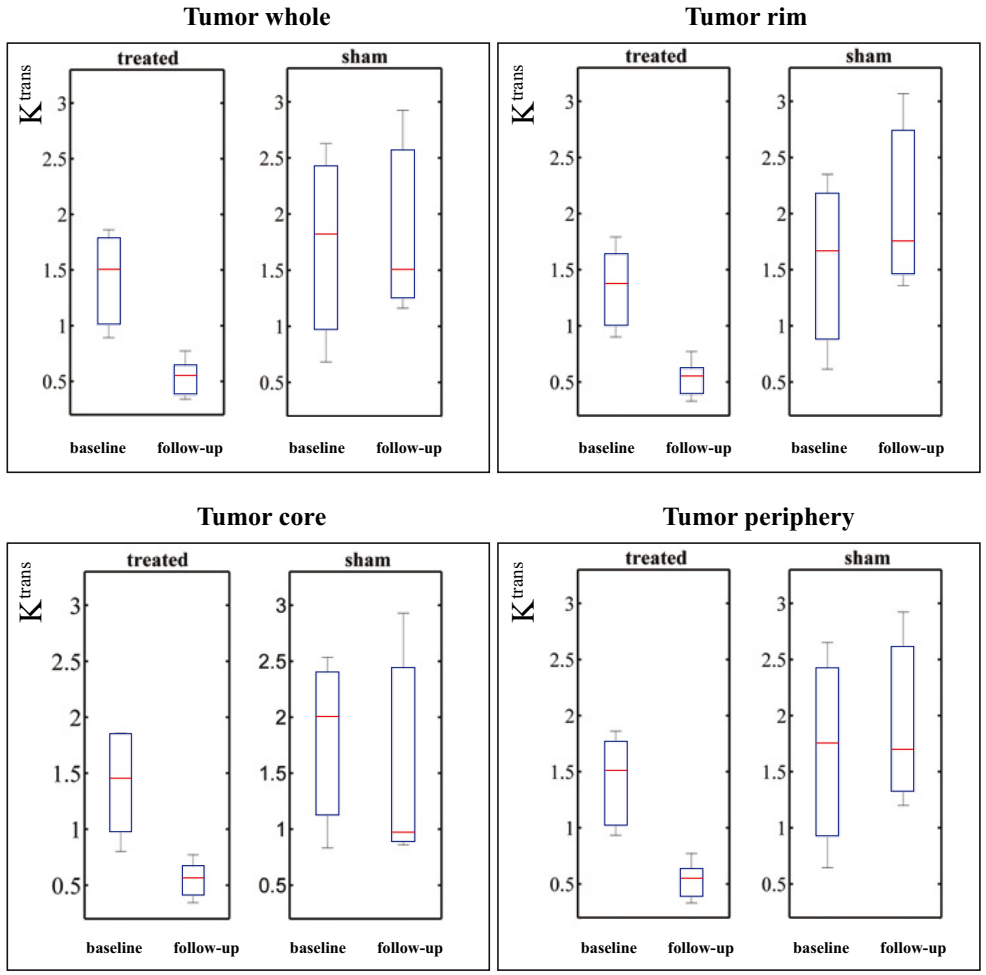


Figure 6. The box plots of K^{trans} at baseline and follow-up examinations for both categories: i.e., treated (TNF-α and melphalan) and sham. Treated group showed statistically significant difference between baseline and follow-up (Wilcoxon signed rank test, $p<0.05$) or all four VOIs: i.e., tumor whole, tumor rim, tumor core and tumor periphery.

The CMV histogram values (Table 2) were fitted linearly as a function of K^{trans}, in a least squares sense, for three K^{trans} categories. The Mann-Whitney U-test was used to assess differences in the slope of the CMV histogram between the response categories (treated and sham) as measured before and after the intervention separately. At baseline, the slopes showed no significant difference between treatment and sham intervention. At follow-up, on the other hand, the slopes for all K^{trans} categories showed significant differences between treatment and sham intervention. For the

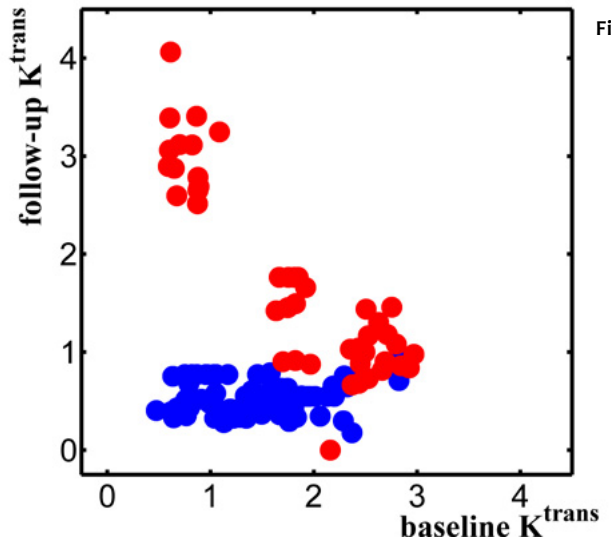


Figure 7. Scatter plot showing mean K^{trans} per sector (estimated for the voxels belonging to the enhancing fraction) at baseline and follow-up for both categories: TNF-α and melphalan treated (blue) and sham reated (red).

follow-up examination, the difference in the slope of the CMV histogram at baseline and follow-up was assessed using the Wilcoxon signed rank test for the two separate groups, i.e. treatment and sham. For all K^{trans} categories, the slopes of the CMV histogram were significantly different from baseline for treatment intervention but not significantly different for sham intervention. The variance in the slopes over the regions is a measure for tumour heterogeneity. The standard deviation between the

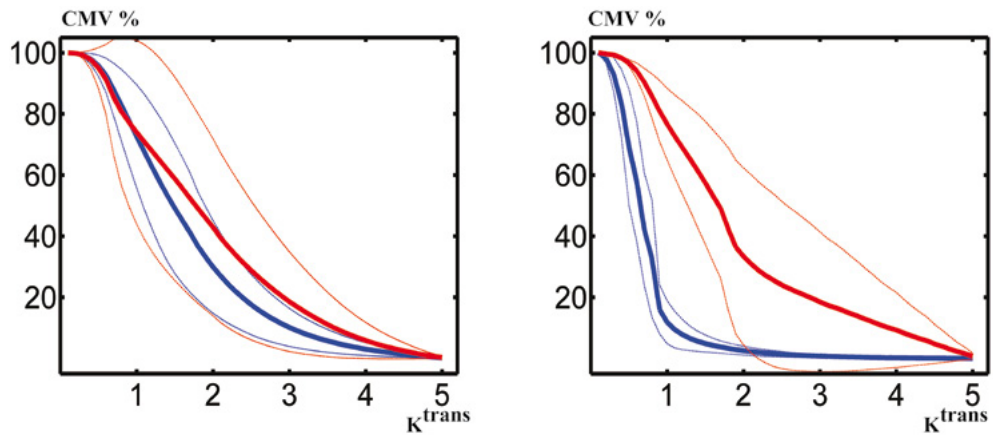


Figure 8. Cumulative map-volume (CMV) histogram representing the 3D-map distribution of K^{trans} for the whole tumor enhancing voxels as averaged over 16 predefined sectors. Blue line demonstrates averaged CMV histogram (solid line) and its deviation (dotted line) for all treated animals and all 16 sectors. Red line demonstrates the same information as blue line but for all sham animals. On the left the results at baseline and on the right the results at follow-up.

Table 2. The first order statistics for the slope of the cumulative map-volume (CMV) histogram over the 16 different sectors, before and after isolated limb perfusion intervention for both categories (TNF-α and melphalan treated, and sham treated). *Significantly different from baseline value, and §significantly different from control group. Statistical significance was set at $p<0.05$.

CMV histogram slope	**Treated**		**Sham**	
Mean (sectors)	Baseline	Follow-up	Baseline	Follow-up
($0<K^{trans}<1$)	- 33.70 ± 25.45	-109.57 ± 9.37*,§	-30.75 ± 45.33	-33.29 ± 26.23
($1<K^{trans}<2$)	-39.76 ± 8.20	-6.14 ± 5.37*,§	-30.86 ± 13.92	-37.76 ± 21.02
($2<K^{trans}<5$)	-8.52 ± 5.79	-0.66 ± 0.64*,§	-13.49 ± 11.74	-9.99 ± 12.87
Stddev (sectors)	Baseline	Follow-up	Baseline	Follow-up
($0<K^{trans}<1$)	16.13 ±10.28	13.55 ± 6.92	7.98 ± 8.80	16.78 ±10.43
($1<K^{trans}<2$)	12.04 ± 4.65	4.83 ± 3.69*,§	9.64 ± 4.59	15.74 ± 5.83
($2<K^{trans}<5$)	4.26 ± 3.40	0.66 ±0.56*	4.09 ± 3.40	2.61 ± 2.28

sectors, for the K^{trans} values ranging 1-5 µl/sec, decreases due to treatment (Table 2). This is also illustrated in Figure 6: the variance between the sectors for the drug-treated rats decreased after treatment whereas the variance of the sham-treated rats is similar to the variance before treatment. This implies the treatment induced homogenization of all tumour sectors with respect to K^{trans} values higher than 1 µl/sec.

Discussion and conclusions

Experimental evidence shows that, for several treatment alternatives, perfusion heterogeneity is associated with cancer treatment outcome [3, 29]. The present preclinical study evaluates the short-term effects of ILP-intervention, i.e., 1 h after the intervention. This allows to assess the almost immediate effects of ILP on the microvasculature. Data from this pilot study indicate that the effects of ILP treatment can be successfully monitored by DCE-MRI 1 h after the intervention: in the treatment group the volume transfer constant (K^{trans}) decreased significantly after treatment with TNF-α and melphalan, whereas in the sham group no decrease in K^{trans} was observed.

To assess tumour heterogeneity we separated the tumour in several different VOIs., i.e., tumour rim, tumour core, tumour periphery. Tumor core and periphery were

each subdivided, using three orthogonal planes, into 8 sectors. The 16 sectors, as such, have no pathophysiological meaning; they are a tool to study the heterogeneity in the tumour. The tumour rim was used since the rim is, generally, the most vascularised region of a solid tumour [5].

The tumour VOIs are selected after careful co-registration of pharmacokinetic maps, before and after the intervention. Therefore, we were able to pinpoint the region-based tumour treatment changes. To neatly summarize the gray value distribution of the estimated pharmacokinetic map, we exploited the CMV histogram by estimating its slope for three different K^{trans} categories, and for all 16 regions. The variance in the slopes over the regions is a measure for tumour heterogeneity. Our results show that the heterogeneity between the sectors for the K^{trans} values, ranging 1-5 µl/sec, decreases due to treatment. This implies that the treatment induces tumour homogenization with respect to K^{trans} values higher than 1 µl/sec. This regional analysis could be a helpful tool in evaluating local treatment effect. When part of the tumour 'escapes' from treatment, this can be detrimental for the whole treatment effect. With a regional analysis this 'escape' could be spotted within hours after start of treatment.

The ILP-intervention is generally applied for the treatment of limb threatening sarcomas and melanomas. During treatment, the limb is isolated from the systemic circulation. After perfusion the physiological limb perfusion is restored. The results show that, due to the intervention, in both the drug-treated and the sham group, the enhancing fraction is decreased. The strength of this intervention is a combined effect of tumour necrosis factor alpha (TNF-α) and an alkylating agent (melphalan) intended to kill tumour cells. TNF-α augments the drug levels, within the tumour, via an increase of the melphalan concentration by six times [8, 45, 46]. This strong effect increases the tumour cell eradication. TNF-α increases the vascular permeability of the tumour vessels already early during the perfusion while the quiescent vasculature of the healthy tissue is not affected. TNF-α may also induce hemodynamic changes. Both agents induce a cascade of effects in the tumour microenvironment. However, which combination of biological effects cause the decrease in K^{trans} is not clear. For instance, vessel disintegration or increased interstitial pressure could play a role.

For an accurate K^{trans} computation, a reliable $T_{1,0}$ estimation is a prerequisite. *In vivo* estimation of blood $T_{1,0}$, and thereby also VIF, is a non-trivial problem as it is susceptible for motion, system imperfections, in-flow effects and partial volume effects. To estimate K^{trans} as accurately as possible, we accounted for these problems as follows. The inoculation location, the hind limb, was chosen in order to minimize tumor movement during scanning. To estimate $T_{1,0}$, we selected a flow compensated inversion recovery gradient echo planar imaging sequence (IR-GE EPI). The IR technique

is relatively immune for the system imperfections such as RF power and slice imperfections [47]. With GE EPI, a shorter acquisition time was possible compared to standard IR scans (GE or SE). We have selected a VOI far from a big artery with the consequence that the remaining blood flow in this VOI is slow enough to prohibit signal dephasing. The geometric distortions, due to GE-EPI readout, were corrected for in the carefully performed co-registration procedure.

Even though a limited number of animals was examined, the data from this study support the hypothesis that the tumour heterogeneity patterns in the volume transfer constant as measured with CMV are an appropriate measure of tumour changes in response to ILP-intervention. Therefore, the CMVs at can serve as non-invasive early outcome predictors in treatment monitoring and may guide therapy adaptation.

In summary, the present preclinical study suggests that heterogeneity in DCE-MRI pharmacokinetic maps can be used as a potential biomarker to quantify short-term effects of ILP-intervention on tumour microvasculature in an experimental model of soft-tissue sarcoma. This provides reliable assessment of tumour treatment effect within 1 h after the ILP-intervention. Further assessment of treatment effects requires comparative studies of ILP over different time points using registered DCE-MRI pharmacokinetic maps to histological sections.

Acknowledgements

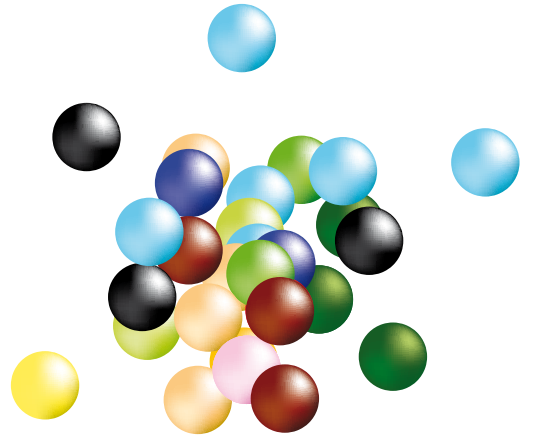
Albumin-(Gd-DTPA)$_{45}$ was generously provided by Dr. RC Brasch, Center of Pharmaceutical and Molecular Imaging, Department of Radiology, University of California San Francisco, CA, USA.

References

1. Jain, R.K., et al., *Biomarkers of response and resistance to antiangiogenic therapy.* Nat Rev Clin Oncol, 2009. 6(6): p. 327-38.
2. Duda, D.G., M. Ancukiewicz, and R.K. Jain, *Biomarkers of antiangiogenic therapy: how do we move from candidate biomarkers to valid biomarkers?* J Clin Oncol, 2010. 28(2): p. 183-5.
3. Yang, X. and M.V. Knopp, *Quantifying tumor vascular heterogeneity with dynamic contrast-enhanced magnetic resonance imaging: a review.* J Biomed Biotechnol, 2011. 2011: p. 732848.
4. Manusama, E.R., et al., *Isolated limb perfusion with TNF alpha and melphalan in a rat osteosarcoma model: a new anti-tumour approach.* Eur J Surg Oncol, 1996. 22(2): p. 152-7.
5. Preda, A., et al., *Dynamic contrast-enhanced MRI using macromolecular contrast media for monitoring the response to isolated limb perfusion in experimental soft-tissue sarcomas.* MAGMA, 2004. 17(3-6): p. 296-302.
6. Seynhaeve, A.L., A.M. Eggermont, and T.L. ten Hagen, *TNF and manipulation of the tumor cell-stromal interface: "ways to make chemotherapy effective".* Front Biosci, 2008. 13: p. 3034-45.
7. Hoving, S., et al., *Synergistic antitumor response of interleukin 2 with melphalan in isolated limb perfusion in soft tissue sarcoma-bearing rats.* Cancer Res, 2005. 65(10): p. 4300-8.
8. van der Veen, A.H., et al., *TNF-alpha augments intratumoural concentrations of doxorubicin in TNF-alpha-based isolated limb perfusion in rat sarcoma models and enhances anti-tumour effects.* Br J Cancer, 2000. 82(4): p. 973-80.
9. de Wilt, J.H., et al., *Prerequisites for effective isolated limb perfusion using tumour necrosis factor alpha and melphalan in rats.* Br J Cancer, 1999. 80(1-2): p. 161-6.
10. Eggermont, A.M., et al., *Isolated limb perfusion with tumor necrosis factor and melphalan for limb salvage in 186 patients with locally advanced soft tissue extremity sarcomas. The cumulative multicenter European experience.* Ann Surg, 1996. 224(6): p. 756-64; discussion 764-5.
11. van Etten, B., et al., *Degree of tumour vascularity correlates with drug accumulation and tumour response upon TNF-alpha-based isolated hepatic perfusion.* Br J Cancer, 2003. 88(2): p. 314-9.
12. Padhani, A.R., *Functional MRI for anticancer therapy assessment.* Eur J Cancer, 2002. 38(16): p. 2116-27.
13. Zweifel, M. and A.R. Padhani, *Perfusion MRI in the early clinical development of antivascular drugs: decorations or decision making tools?* Eur J Nucl Med Mol Imaging, 2010. 37 Suppl 1: p. S164-82.
14. Padhani, A.R., *MRI for assessing antivascular cancer treatments.* Br J Radiol, 2003. 76 Spec No 1: p. S60-80.
15. Daldrup-Link, H.E. and R.C. Brasch, *Macromolecular contrast agents for MR mammography: current status.* Eur Radiol, 2003. 13(2): p. 354-65.
16. Nelson, D.A., et al., *Hypoxia and defective apoptosis drive genomic instability and tumorigenesis.* Genes Dev, 2004. 18(17): p. 2095-107.
17. Jain, R.K., *Normalization of tumor vasculature: an emerging concept in antiangiogenic therapy.* Science, 2005. 307(5706): p. 58-62.
18. Seynhaeve, A.L., et al., *Tumor necrosis factor alpha mediates homogeneous distribution of liposomes in murine melanoma that contributes to a better tumor response.* Cancer Res, 2007. 67(19): p. 9455-62.
19. Alic, L., et al. *Quantification of Heterogeneity in Dynamic Contrast Enhancement MRI Data for Tumor Treatment Assessment*. 2006. IEEE International Symposium on Biomedical Imaging: From Nano to Macro.Arlington, Virginia.
20. Rose, C.J., et al., *Quantifying spatial heterogeneity in dynamic contrast-enhanced MRI parameter maps.* Magn Reson Med, 2009. 62(2): p. 488-99.
21. Alic, L., et al., *Heterogeneity in DCE-MRI parametric maps: a biomarker for treatment response?* Phys Med Biol, 2011. 56(6): p. 1601-16.

22. Pickles, M.D., et al., *Prognostic value of pre-treatment DCE-MRI parameters in predicting disease free and overall survival for breast cancer patients undergoing neoadjuvant chemotherapy.* Eur J Radiol, 2009. 71(3): p. 498-505.
23. Nagashima, T., et al., *Dynamic enhanced MRI predicts chemosensitivity in breast cancer patients.* Eur J Radiol, 2006. 60(2): p. 270-4.
24. van Rijswijk, C.S., et al., *Dynamic contrast-enhanced MR imaging in monitoring response to isolated limb perfusion in high-grade soft tissue sarcoma: initial results.* Eur Radiol, 2003. 13(8): p. 1849-58.
25. Robinson, S.P., et al., *Tumour dose response to the antivascular agent ZD6126 assessed by magnetic resonance imaging.* Br J Cancer, 2003. 88(10): p. 1592-7.
26. Gaustad, J.V., et al., *Dynamic contrast-enhanced magnetic resonance imaging of human melanoma xenografts with necrotic regions.* J Magn Reson Imaging, 2007. 26(1): p. 133-43.
27. Checkley, D., et al., *Use of dynamic contrast-enhanced MRI to evaluate acute treatment with ZD6474, a VEGF signalling inhibitor, in PC-3 prostate tumours.* Br J Cancer, 2003. 89(10): p. 1889-95.
28. Benjaminsen, I.C., et al., *Assessment of tumor blood perfusion by high-resolution dynamic contrast-enhanced MRI: a preclinical study of human melanoma xenografts.* Magn Reson Med, 2004. 52(2): p. 269-76.
29. Benjaminsen, I.C., et al., *Assessment of extravascular extracellular space fraction in human melanoma xenografts by DCE-MRI and kinetic modeling.* Magn Reson Imaging, 2008. 26(2): p. 160-70.
30. Jia, G., et al., *Colorectal liver metastases: contrast agent diffusion coefficient for quantification of contrast enhancement heterogeneity at MR imaging.* Radiology, 2008. 248(3): p. 901-9.
31. Ogan, M.D., et al., *Albumin labeled with Gd-DTPA. An intravascular contrast-enhancing agent for magnetic resonance blood pool imaging: preparation and characterization.* Invest Radiol, 1987. 22(8): p. 665-71.
32. Barrett, T., et al., *Macromolecular MRI contrast agents for imaging tumor angiogenesis.* Eur J Radiol, 2006. 60(3): p. 353-66.
33. Guo, J.Y., et al., *Improved accuracy and consistency in T1 measurement of flowing blood by using inversion recovery GE-EPI.* Med Phys, 2005. 32(4): p. 1083-93.
34. Klein, S., et al., *elastix: a toolbox for intensity-based medical image registration.* IEEE Trans Med Imaging, 2010. 29(1): p. 196-205.
35. Tofts, P.S., *Modeling tracer kinetics in dynamic Gd-DTPA MR imaging.* J Magn Reson Imaging, 1997. 7(1): p. 91-101.
36. Tofts, P.S., et al., *Estimating kinetic parameters from dynamic contrast-enhanced T(1)-weighted MRI of a diffusable tracer: standardized quantities and symbols.* J Magn Reson Imaging, 1999. 10(3): p. 223-32.
37. Barral, J.K., et al., *A robust methodology for in vivo T1 mapping.* Magn Reson Med, 2010. 64(4): p. 1057-67.
38. Bhujwalla, Z.M., D. Artemov, and J. Glockner, *Tumor angiogenesis, vascularization, and contrast-enhanced magnetic resonance imaging.* Top Magn Reson Imaging, 1999. 10(2): p. 92-103.
39. Liney, G., *MRI from A to Z : a definitive guide for medical professionals*. 2nd ed. 2010, Dordrecht ; New York: Springer. 385 p.
40. Weinmann, H.J., M. Laniado, and W. Mutzel, *Pharmacokinetics of GdDTPA/dimeglumine after intravenous injection into healthy volunteers.* Physiol Chem Phys Med NMR, 1984. 16(2): p. 167-72.
41. Shipley, W.U., et al., *Proton radiation as boost therapy for localized prostatic carcinoma.* JAMA, 1979. 241(18): p. 1912-5.
42. Butler, L.E., et al., *Dosimetric benefits of respiratory gating: a preliminary study.* J Appl Clin Med Phys, 2004. 5(1): p. 16-24.
43. Nioutsikou, E., et al., *Reconsidering the definition of a dose-volume histogram.* Phys Med Biol, 2005. 50(11): p. L17-19.
44. Schwickert, H.C., et al., *Quantification of liver blood volume: comparison of ultra short TI inversion recovery echo planar imaging (ULSTIR-EPI), with dynamic 3D-gradient recalled echo imaging.* Magn Reson Med, 1995. 34(6): p. 845-52.

45. ten Hagen, T.L., A.M. Eggermont, and F.J. Lejeune, *TNF is here to stay--revisited.* Trends Immunol, 2001. 22(3): p. 127-9.
46. de Wilt, J.H., et al., *Tumour necrosis factor alpha increases melphalan concentration in tumour tissue after isolated limb perfusion.* Br J Cancer, 2000. 82(5): p. 1000-3.
47. Jackson, A., D. Buckley, and G.J.M. Parker, *Dynamic contrast-enhanced magnetic resonance imaging in oncology*. Medical radiology. 2003, Berlin ; New York: Springer. p.



Chapter 6

Summary and general discussion

Any intelligent fool can make things bigger and more complex. It takes a touch of genius – and a lot of courage to move in the opposite direction.

Albert Einstein

Introduction

Different types of cancer have large inter-tumour heterogeneity, for example, in their origin, site, genetic expression, and pathologic appearance. Molecular studies have also revealed that heterogeneity can exist within one tumour. Subtypes of the same cancer differ in terms of both biology and response to treatment, causing regional variations in cell death, metabolic activity, proliferation, and vascular structure. Biologically a tumour is a complex system in which distinct populations of cancer cells interact in a competitive manner. The more aggressive tumour populations, e.g. those that proliferate faster, have a higher neo-angiogenesis level or are less sensitive to treatment, will suppress the less aggressive populations. There is also clinical evidence that recurrent tumours are often more malignant than the primary tumour: the more aggressive populations have survived. In this respect, visualisation and quantification of tumour heterogeneity is a useful tool in grading, differentiation, monitoring response to tumour treatment, and predicting outcome. Tumour heterogeneity, quantified from radiological images, has the potential to become an important biomarker for grading, differentiation, monitoring treatment response and predicting outcome.

The aim of this thesis is to develop and evaluate tumour heterogeneity quantification techniques and to investigate their importance for tumour treatment monitoring and outcome prediction. In particular, this thesis focuses on the following questions:

- Do MR imaging data reveal the tumour heterogeneity?
- Which analysis methods are used to quantify tumour heterogeneity for diagnostic and/or treatment purposes, and what is the reported performance of these methods?
- Is tumour heterogeneity in Dynamic Contrast Enhanced (DCE)-MRI, as quantified with texture analysis methods, sensitive to changes due to therapy, and can patient outcome be predicted?

Summary

Chapter 1 provides a general introduction to this thesis. The epidemiology of cancer, the biological background of intra-tumour heterogeneity, radiological imaging methods and the methods available for the quantification of heterogeneity are briefly discussed, and the main aim of the thesis is outlined.

Chapter 2 presents a methodology for obtaining an accurate 3D relation between high resolution *in vivo* T2*-w MRI and the corresponding 3D histology of the tumour tissue. The key features of the methodology are: 1) standardized acquisition and processing, 2) use of an intermediate *ex vivo* MRI, 3) use of a reference cutting plane, 4) dense histological sampling, 5) elastic registration, and 6) use of complete 3D data sets. The methodology consists of two separate registration steps, both exploiting a three-step strategy of gradually increasing degrees of freedom (rigid, affine, and elastic transformation). These two registration steps involve *in vivo* MRI to *ex vivo* MRI registration, and *ex vivo* MRI to histology registration. The resulting accuracy was assessed by two independent observers and was on average 0.7 mm, between *in vivo* MRI and 3D reconstructed histology. This accuracy corresponds (on average) with 30-50 cells and is similar to the inter-observer variation of the manual annotations. We have shown that, based on T2*-w MRI signal intensity, automatic identification of necrotic tissue is feasible. However, the separation of two other tissue types, i.e. haemorrhagic and viable tissue, was not possible. For the separation of these tissue types, other MRI sequences are needed. This work is a first step in MRI tumour characterization. Now that spatial correspondence between *in vivo* MRI and 3D H&E histology has been established, the extension to multi-spectral MR and multi-stained histological sections is the next logical step. The 3D correspondence of tumor histology and *in vivo* MRI enables extraction of MRI signatures for histologically defined regions.

Chapter 3 provides a systematic review of the literature on tumour heterogeneity quantification methods computed from radiological images for grading, differentiation, outcome prediction and tumour-response monitoring. The heterogeneity analysis methods are divided into four categories, i.e., non-spatial methods (NSM), spatial grey level methods (SGLM), fractal analysis (FA) methods, and filter and transforms (F&T). The reported performance of the heterogeneity features in terms of measures such as accuracy, sensitivity, specificity and AUC, or by statistical tests reporting significance levels, was compared. We found that, among the 8,956 unique studies identified, 192 studies reported heterogeneity as a biomarker in cancer imaging. Since 2009, the number of publications reporting the quantification of tumour heterogeneity has increased. Up until 2006 most heterogeneity papers were based on ultrasound (US), whereas after 2007 the number of studies using MRI has increased. The NSM and the SGLM are the two methods most frequently used during that period. Most papers focus on heterogeneity quantification for differentiation, grading, or outcome prediction has recently increased. The performance of the heterogeneity features was mostly (68%) evaluated by classification experiments reporting performance measures such as accuracy, sensitivity, specificity and AUC. The other papers used only statistical tests (30%) or did not perform a quantitative study (2%). Almost 98% of the studies reported positive findings. The reported AUC ranged from 0.5 to 1 with a median of 0.89. No relation was found between the performance measures and the imaging method or quantification method used. Of all the classification publications, 41% did not report the use of cross-validation as a technique to limit the effect of overfitting on the available data. A negative correlation was found between the tumour-feature ratio and the AUC, which might be caused by overfitting. Publications reporting significance testing often did not perform a correction of the significance levels for multiple comparisons. For 36% of the papers a significant decrease in the number of significant features was observed after the Holm-Bonferroni correction. Only 12% of all included studies had a prospective study design. Although the use of retrospectively collected data is necessary to develop, test and evaluate heterogeneity as a biomarker for tumour grading, differentiation, outcome prediction and treatment response monitoring the real test is to evaluate the performance of the developed features in a prospective study design. Moreover, in most retrospective studies the performance of the heterogeneity feature is evaluated without taking into account other available clinical predictive information. While the researcher may be interested in the performance of the feature itself, the clinician is interested in the additional value of the feature on top of the already available clinical features. To enable the translation of imaging biomarkers from the research stage to clinical practice, future research should focus on prospective studies to investigate the additional value of the proposed heterogeneity biomarker on top of the clinically established markers.

Chapter 4 uses the DCE-MRI images of sarcoma patients undergoing isolated limb perfusion (ILP) with TNF-α and melphalan, to evaluate promising imaging-biomarkers: enhancing fraction, SGLM and FA. The potential of these imaging biomarkers to monitor tumour changes and predict tumour response to treatment was evaluated in 18 patients. Using routinely acquired DCE-MRI scans, this study investigates the potential of SGLM and FA to measure the treatment-induced tumour changes for different pharmacokinetic modelling approaches. The correlation between pharmacokinetic and different heuristic parametric maps, averaged over all tumours, was estimated by Spearman's rank correlation coefficient and ranged between 0.44 and 0.61. The monitoring study demonstrated that, regardless of the origin of the estimated parametric map (pharmacokinetic or heuristic model-based), the enhancing fraction and coherence showed a significant difference between baseline and follow-up acquisitions for the group that responded to therapy. When evaluating the enhancing fraction and coherence on the pre-treatment scan, it was found that

both features differentiate between outcome categories. It seems that patients with a large viable tumour fraction and a high coherence respond well to therapy, whereas patients with large necrotic areas and a low coherence show less response to treatment. Moreover, we demonstrated that heterogeneity measures derived from DCE-MRI parametric maps (irrespective of the exact nature of the parametric map) were related to tumour response to chemotherapy. Therefore, we can conclude that DCE-MRI has potential for *in vivo* monitoring of the tumour during chemotherapy. When correlated to histopathological findings, this method may be clinically useful in understanding pathophysiological changes, predicting tumour response and guiding the therapeutic approach.

Chapter 5 investigates the regional heterogeneity changes in DCE-MRI as response to ILP in an experimental soft-tissue sarcoma model. DCE-MRI of drug-treated and sham-treated rats was performed at baseline and 1 h after ILP intervention. The enhancing data were acquired using a macromolecular contrast medium (MMCM), albumin-(Gd-DTPA)$_{45.}$ To pinpoint the regional changes accurately, the DCE-MRI at baseline and follow-up were spatially registered. To assess the regional heterogeneity, tumours were divided into 16 sectors. For each sector the cumulative map-volume (CMV) histogram of K^{trans} was computed and the variance in its slope was used as a measure for tumour heterogeneity. The results indicate that the heterogeneity between sectors decreases due to treatment. This implies that the treatment induces tumour homogenization with respect to K^{trans} demonstrating the potential of regional analysis for evaluating local treatment effects of ILP intervention. When part of the tumour 'escapes' from treatment, this can be detrimental for the whole treatment effect. Using a regional analysis this 'escape' could potentially be spotted within hours after the start of treatment. Therefore, CMVs may serve as non-invasive biomarker of early treatment effects and may guide therapy adaptation.

Appendix A presents a brief review of the automatic registration approach to transform different datasets into one coordinate system. In this thesis, the registration is used to correct for different deformations of *ex vivo* tumours with respect to the original *in vivo* shape (Chapter 2), to compensate for motion artefacts (Chapter 5), and to pinpoint the region-based tumour treatment changes between baseline and follow-up scans (Chapter 5).

Discussion and future prospective

Quantification of tumour heterogeneity in diagnostic images has the potential to improve tumour grading, differentiation, outcome prediction and response monitoring. The challenge is to identify, per tumour type, the imaging modality in conjunction with the analysis method that is suitable for clinical application.

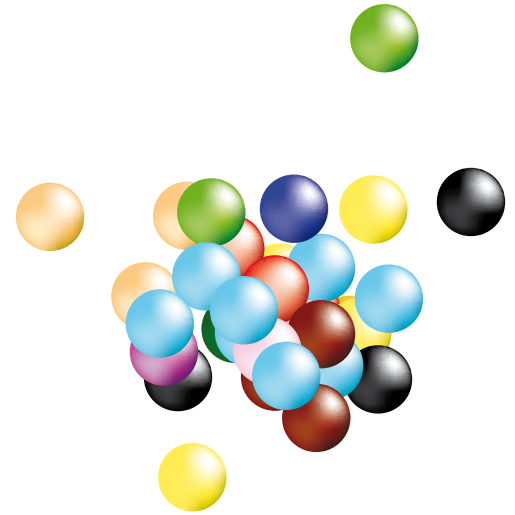
This thesis contributes a number of possibilities to address this challenge. First, we developed a method to establish an accurate 3D-relation between *in vivo* MRI and histological sections. This method will enable the verification of MRI-based signatures by histologically confirmed regions. Although this study demonstrated the feasibility of automatic identification of parts of the tumour, full characterisation of the whole tumour was not possible using only one MRI sequence. We hypothesize that different imaging sequences and, possibly, different imaging modalities will be required for this purpose.

Second, we provide evidence that heterogeneity in DCE-MRI pharmacokinetic maps is a promising technique for tumour diagnosis and evaluation of therapeutic response to anti-angiogenic drugs. The preclinical and clinical studies included in this thesis indicate that the treatment effect may be monitored and predicted by heterogeneity measures derived from parametric maps, which are obtained from non-model-based statistical features such as co-occurrence matrices. In addition, we show the potential of regional tumour heterogeneity patterns in DCE-MRI pharmacokinetic maps, as measured with CMV, to measure short-term tumour changes in response to chemotherapy.

To develop and implement tumour heterogeneity as a biomarker for tumour grading, differentiation, response monitoring and outcome prediction in clinical practice, a careful evaluation strategy is required. Generally, the publications using heterogeneity features only evaluate the value of the heterogeneity features, without taking into account currently accepted clinical features such as mean signal intensity, tumour size, grade or border regularity. Other studies report only the combined classification performance of heterogeneity and clinical features. A large number of publications even use the mean signal intensity as a feature to estimate tumour heterogeneity, even though this is clearly not a heterogeneity measure (i.e., mean signal intensity does not measure intra-tumour heterogeneity). Based on these types of studies, it is not possible to evaluate the added value of heterogeneity to currently accepted clinical features (e.g. mean signal intensity, tumour size, grade or border regularity). Since the quantification of heterogeneity is usually more complex and computationally more costly than computing the mean intensity, the benefit of the added effort to characterise heterogeneity should be justified. Many studies support enhancing fraction (EF) or non-enhancing fraction (NEF) of a tumour volume as an important feature for treatment monitoring and outcome prediction. The relation between EF/NEF and heterogeneity features needs to be further investigated. In addition, most of the studies lack a balanced study design in terms of the requirements for a pattern recognition study. Generally, a large number of features is used compared to the number of validated tumours, which introduces data overfitting. Also, cross validation is often not performed during the estimation of classification power. Moreover, in most studies the performance of the heterogeneity feature is

evaluated without taking into account currently accepted clinical-features. Future studies need a careful design that takes into account the requirements from pattern recognition: i.e. a balanced number of subjects and features, cross-validation, independent test dataset, and a prospective study design, and focus on the added value of the proposed heterogeneity biomarker on top of the clinical established markers. In satisfying these requirements, the real performance of heterogeneity features can be assessed more reliably.

In conclusion, monitoring of tumour treatment and prediction of outcome using heterogeneity biomarkers could have widespread clinical applications. These imaging biomarkers could be used to optimize treatment for the individual patient. Currently, however, most research is still in the proof-of-concept stage. For clinical acceptance and applicability, prospective studies with an appropriate study design are required.



Appendix A

Registration summary

This chapter is based uppon:
L Alić, JC Haeck, K Bol, S Klein, ST Van Tiel, PA Wielopolski, M Bijster, M Bernsen, M de Jong, WJ Niessen, JF Veenland. Facilitating tumor functional assessment by spatially relating 3D tumor histology and *in vivo* MRI: Image registration approach

Medical image registration (also referred to as fusion, superposition, matching or alignment) considers the process of transforming different data sets into one coordinate system to achieve biological, anatomical or functional correspondence. This transformation is necessary to compare (or integrate) the information from different modalities, from baseline to follow-up scans, from pre to post contrast scans, from pre to post treatment scans, and between different subjects. Thereby, it reveals additional information not apparent in the separate images. A comprehensive review on the subject is given by [1, 2]. The choices regarding implementation of the different registration components can have considerable impact on the registration results. All registrations in this paper were performed using Elastix [Figure 1. The basic components of the registration framework containing two input images, a transform, a metric, an interpolator and an optimizer (adopted from Ibanez et al. [4]).1 and are described in more detail in this section.

One image, called the *moving image* I_M, is deformed to fit the other image, called the *fixed image* I_F. More formally, registration is the problem of finding a *coordinate transformation T* that aligns $I_M(T)$ spatially with I_F. The quality of alignment is defined by a cost function $C(T ; I_F , I_M)$. The optimal coordinate transformation is estimated by minimizing the cost function with respect to *T*, usually by means of an iterative optimization method embedded in a hierarchical (multi-resolution) scheme.

Mathematically, the registration is defined as an optimization problem where the cost function *C* is minimized with respect to *T*, eq. 1.

$$\hat{\mu} = \arg\min_{\mu} C(T_\mu; I_F, I_M) \qquad \text{Eq.1}$$

where the subscript **μ** indicates the transform parameterization, and contains the transformation parameters. The minimization problem (eq. S1) is solved with an iterative optimization method, usually in a multi-resolution setting.

Cost Function. The cost function *C* measures the similarity between the images. We have used mutual information (MI) [5] as it only assumes a statistical relation between image intensities, and is therefore suitable for registration of images acquired by different modalities [6, 7]. The MI was parameterized using a 32-bin image intensity histogram.

Optimizer. To optimize the metric criterion with respect to the transform parameters, an iterative optimization procedure is employed. We used an adaptive stochastic gradient descent algorithm [8].

Transform. For the coordinate transformation *T*, different deformations with varying degrees of freedom are available. In order of increasing transformation flexibility: the rigid, the affine, the deformable (non-rigid, or elastic) transformations. After proper rough initialization, achieved by reference plane orientation, we used a three-step strategy of gradually increasing degrees of freedom, starting with rigid registration, followed by affine registration, and finalized by non-rigid refinement. The displacement field was parameterized using a third order B-spline model.

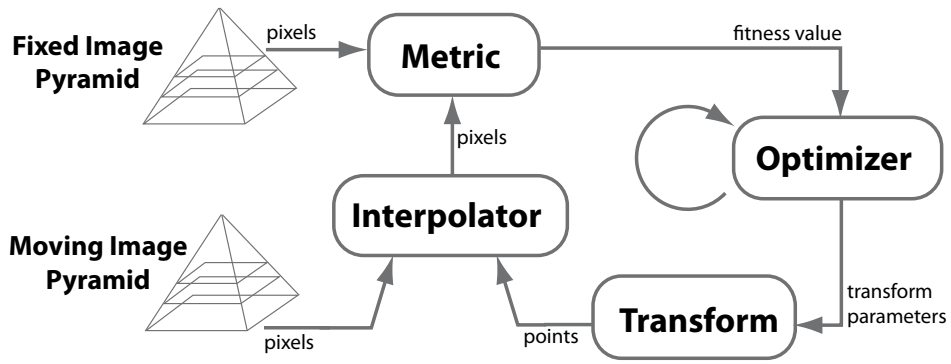


Figure 1. The basic components of the registration framework containing two input images, a transform, a metric, an interpolator and an optimizer (adopted from Ibanez et al. [4]).

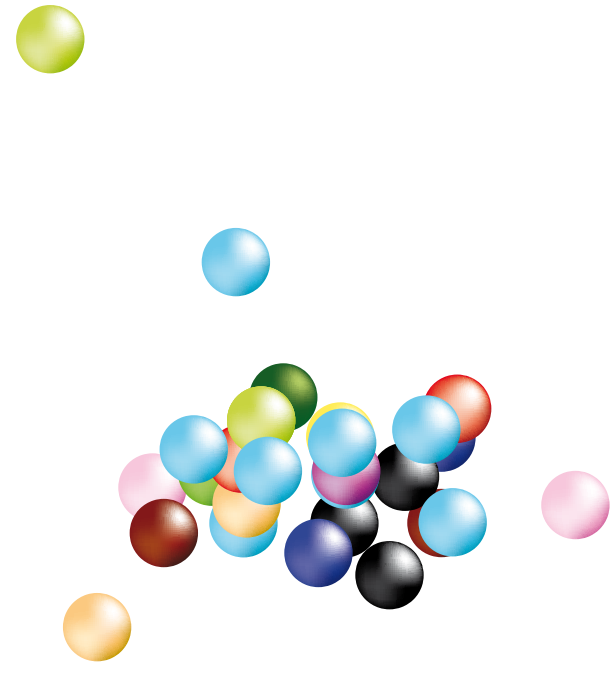
Sampling Strategies. To compute the cost function *C*, a set of samples needs to be selected. The most straightforward strategy is to use all voxels from the fixed image, which has the obvious disadvantage that it is time consuming for large images. A common methodology is to use a subset of voxels, selected on a uniform grid, or sampled randomly. Another strategy is to pick only those points that are located on striking image features, such as edges. In our application, we used a subset of voxels sampled randomly.

Interpolation. For computation of the cost function, the moving image I_M needs to be evaluated at non voxel positions which requires the intensity interpolation. Several interpolation methods (varying in quality and speed) have been proposed and compared by Pluim et al. [9]. In this paper we have used a third order B-spline interpolation.

Optimization. Multi-resolution (hierarchical) strategies are an important aspect of image registration [2, 10]. Such coarse-to-fine schemes in general improve registration accuracy and increase robustness by eliminating local minima of the cost function at coarser scales. The basic idea is to perform the first registration at a coarse scale with down sampled images. The spatial mapping determined at the coarse level is then used to initialize registration at the next finer scale. This process is repeated until it reaches the finest scale.

References

1. Modersitzki, J., *Numerical Methods for Image Registration*. 2003: Oxford University Press.
2. Hill, D.L., et al., *Medical image registration.* Phys Med Biol, 2001. 46(3): p. R1-45.
3. Klein, S., et al., *elastix: a toolbox for intensity-based medical image registration.* IEEE Trans Med Imaging, 2010. 29(1): p. 196-205.
4. Ibanez, L., et al., *The ITK Software Guide*. 2005: Kitware, Inc.
5. Rueckert, D., et al., *Nonrigid registration using free-form deformations: application to breast MR images.* IEEE Trans Med Imaging, 1999. 18(8): p. 20.
6. Maes, F., et al., *Multimodality image registration by maximization of mutual information.* IEEE Trans Med Imaging, 1997. 16(2): p. 187-198.
7. Thevenaz, P. and M. Unser, *Optimization of mutual information for multiresolution image registration.* IEEE Trans Image Process, 2000. 9(12): p. 2083-2099.
8. Klein, S., M. Staring, and J.P. Pluim, *Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines.* IEEE Trans Image Process, 2007. 16(12): p. 2879-2890.
9. Pluim, J.P.W., J.B.A. Maintz, and M.A. Viergever, *Interpolation artefacts in mutual information-based image registration.* Computer Vision and Image Understanding, 2000. 77(2): p. 211-232.
10. Lester, H.A., S.R, *A survey of hierarchical non-linear medical image registration.* Pattern Recognition, 1999. 32(1): p. 21.



Publications

List of publications
(in bold those related to this thesis)

Papers in International Journals

Alić L, Niessen WJ, and Veenland JV, Quantification of Heterogeneity as a Biomarker in Tumour Imaging: A Systematic Review. *2013. Submitted.*

Alić L, van Vliet M, Wielopolski PA, ten Hagen TLM, van Dijke CF, Niessen WJ, Veenland JF. Regional heterogeneity changes in DCE-MRI as response to isolated limb perfusion in experimental soft-tissue sarcomas. *Contrast Media and Molecular Imaging, 2013. 8(4): 340-349.*

Alić L, Haeck JC, Bol K, Klein S, van Tiel ST, Wielopolski PA, Bijster M, Bernsen M, de Jong M, Niessen WJ, Veenland JF. Facilitating tumor functional assessment by spatially relating 3D tumor histology and in vivo MRI: image registration approach. *PLoS One, 2011. 6(8): e22835.*

Alić L, van Vliet M, van Dijke CF, Eggermont AMM, Veenland JF, Niessen WJ. Heterogeneity in DCE-MRI parametric maps: a biomarker for treatment response? *Phys Med Biol, 2011. 56(6): 1601-16.*

Lourens MS, Alić L, van den Berg B, Verbraak AFM, Bogaaard JM. Estimation of expiratory time constants via fuzzy clustering. *J Clin Monit Comput, 2002. 17(1): 15-22.*

Babuska R, Alić L, Lourens MS, Verbraak AFM, and Bogaard JM. Estimation of respiratory parameters via fuzzy clustering. *Artif Intell Med, 2001. 21(1-3): 91-105.*

Mathura KR, Alić L, van Deventer S, and Ince C. The rectal microcirculation observed during inflammatory bowel disease using OPS imaging. *Journal of Vascular Research, 2000. S1.*

Book Chapters

Mathura KR, Alić L, and Ince C, eds. *I*nitial clinical experience with OPS imaging for observation of the human microcirculation. *Yearbook of Intensive Care and Emergency Medicine, ed. J.L. Vincent. 2001, Springer-Verlag: Berlin. 725.*

Papers in Conference Proceedings

Schols RM, Wieringa FP, Alić L, Bouvy ND, and Stassen LPS. In vivo wide-band reflectance spectrometry during thyroid and parathyroid surgery. *ECBO 2013: European Conferences on Biomedical Optics, 2013. Munich, Germany.*

Bol K, Haeck JC, Alić L, Bernsen M, de Jong M, Niessen WJ, and Veenland JF. Quantification of DCE-MRI: A validation of three teahniques with 3D-histology. ***IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2012. Barcelona, Spain.***

Bol K, Haeck JC, Alić L, Bernsen M, de Jong M, Niessen WJ, and Veenland JF. Developing a tool for the validation of quantitative DCE-MRI. ***SPIE Medical Imaging, 2011. Orlando, USA.***

Alić L, Haeck JC, Klein S, Bol K, van Tiel ST, Wielopolski PA, Bijster M, Niessen WJ, Bernsen M, Veenland JF, and de Jong M. Multi-modal image registration: matching MRI with histology. ***SPIE Medical Imaging, 2010. San Diego, USA.***

Alić L, Veenland JF, van Vliet M, van Dijke CF, Eggermont AMM, and Niessen WJ. Quantification of Heterogeneity in Dynamic Contrast Enhancement MRI Data for Tumor Treatment Assessment. ***IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2006. Arlington, Virginia.***

Abstracts in Conference Proceedings

Schols RM, Alic L, Beets GL, Breukink SO, Wieringa FP, and Stassen LPS. In vivo optical characterization of critical tissues during colorectal surgery. *EAES: 21st Interna-tinal Congres of EAES, 2013. Vienna, Austria.*

Schols RM, Wieringa FP, Alić L, Bouvy ND, and Stassen LPS. In vivo wide-band reflectance spectrometry during thyroid and parathyroid surgery. *SAGES: Aanual meeting Society of American Gastrointestinal and Endoscopic Surgeons, 2013. Baltimore, USA.*

Schols RM, Wieringa FP, Alić L, Bouvy ND, and Stassen LPS. Difuse optical reflectance spectrometry in thyroid and parathyroid surgery. *EAES: 21st Internatinal Congres of EAES, 2013. Vienna, Austria.*

Schols RM, Wieringa FP, Alić L, Bouvy ND, and Stassen LPS. Visible and infrared diffuse reflectance spectrometry during thyroid and parathyroid surgery. *Wetenschapsmiddag MUMC, 2013. Maastricht, The Netherlands.*

Bol K, Haeck JC, Alić L, Niessen WJ, de Jong M, Bernsen MD, and Veenland JF. Validation of DCE-MRI parametric maps: developing a tool using 3D-histology. ***ESMRMB Congress: 28th Annual Scientific Meeting. 2011. Leipzig, Germany.***

Alić L, van Vliet M, van Dijke CF, Eggermont AMM, Niessen WJ, and Veenland JF. Quantifying DCE-MRI for response prediction of cancer treated by ILP. ***EMIM: 6th European Molecular Imaging Meeting, 2011. Leiden, The Netherlands.***

Bol K, Haeck JC, Alić L, Bernsen M, de Jong M, Niessen WJ, and Veenland JF. Methodology for the validation of quantitative DCE-MRI with 3D-histology. ***EMIM: 6th European Molecular Imaging Meeting, 2011. Leiden, The Netherlands.***

Alić L, van Vliet M, van Dijke CF, Eggermont AMM, Veenland JF, and Niessen WJ. Heterogeneity in DCE-MRI parametric maps: a biomarker for treatment response? ***ECR: European Congress of Radiology, 2011. Vienna, Austria.***

Bol K, Haeck JC, Alić L, Bernsen M, de Jong M, Niessen WJ and Veenland JF. Validation of DCE-MRI parametric maps: developing a tool using 3D-histology. ***ECR: European Congress of Radiology, 2011. Vienna, Austria.***

Bol K, Haeck JC, Alić L, Bernsen M, de Jong M, Niessen WJ, and Veenland JF. Validation of DCE-MRI parametric maps: developing a tool using 3D-histology. ***BME: 3th Duch Bio-Medical Engineering Conference, 2011. Egmond aan Zee, The Netherlands.***

Alić L, Haeck JC, Klein S, Bol K, Niessen WJ, Bijster M, Bernsen M, de Jong M, and Veenland JF. Validation of MRI by accurate maching with histology. ***BME: 3th Duch Bio-Medical Engineering Conference, 2011. Egmond aan Zee, The Netherlands.***

Alić L, Haeck JC, Klein S, Bol K, Bijster M, Niessen WJ, Bernsen M, Veenland JF, and de Jong M. Correlation of MRI with histology using non rigid registration. ***ECR: European Congress of Radiology, 2010. Vienna, Austria.***

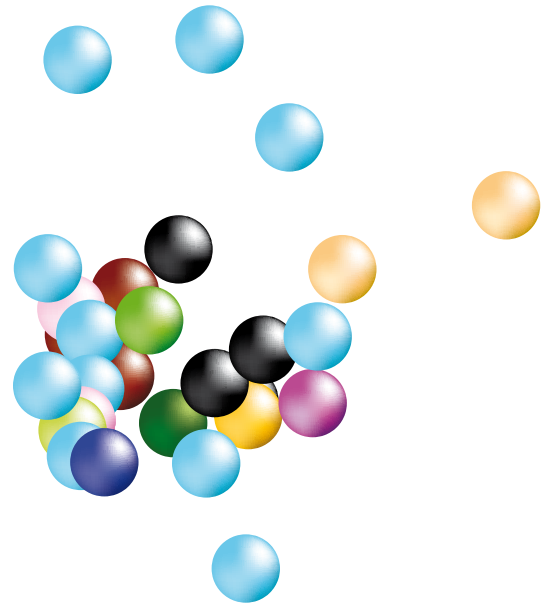
Veenland JF, Haeck JC, Alić L, van Tiel ST, Klein S, Wielopolski PA, Niessen WJ, Bernsen M, and de Jong M. Correlation of MRI with histology: a method based on non-rigid registration. ***ESMRMB: 26th Annual Scientific Meeting, 2009. Antalya, Turkey.***

Alić L, Haeck JC, van Tiel ST, Wielopolski PA, Bijster M, Klein S, Niessen WJ, Krestin GP, Krenning EP, de Jong M, Bernsen M, and Veenland JF. Matching MRI with histology: does elastic registration help? ***EMIM: 4th European Molecular Imaging Meeting, 2009. Barcelona, Spain.***

Haeck JC, Alić L, van Tiel ST, Wielopolski PA, Bijster M, Veenland JF, Bernsen M, and de Jong M. How to match tumor structures and morphological information? ***EMIM: 4th European Molecular Imaging Meeting, 2009. Barcelona, Spain.***

van Vliet M, Veenland JF, Wielopolski PA, ten Hagen TLM, van Dijke CF, Preda A, Alić L, Grünhagen DJ, Eggermont AMM, and Kerstin CP. Tumor vessel visualization for treatment monitoring: how useful is MRA? *RSNA: 91st Scientific Assembly and Annual Meeting of the Radiological Society of North America, 2005. Chicago, USA.*

Mathura KR, Alić L, van Deventer S, and Ince C. The rectal microcirculation observed during inflammatory bowel disease using OPS imaging. *21st Conference on Microcirculation, 2000. Stockholm, Sweden.*



Samenvatting

Verschillende soorten kanker vertonen een grote verscheidenheid in hun oorsprong, locatie, genetische expressie en pathologische verschijningsvorm. Dankzij moleculaire studies weet men nu dat ook binnenin de tumor verscheidenheid aanwezig is. Onderdelen van dezelfde tumor tonen verschillen, zowel biologisch als in de reactie op behandeling. Biologisch gezien is een tumor een complex systeem waarin verschillende populaties van kankercellen op een concurrerende wijze op elkaar inwerken. Agressievere tumorpopulaties, bijvoorbeeld populaties die zich sneller vermenigvuldigen, of een hoger neo-angiogenetisch potentieel hebben of minder gevoelig zijn voor behandeling, onderdrukken de minder agressieve populaties. Klinische studies tonen aan dat terugkerende tumoren vaak kwaadaardiger zijn dan de oorspronkelijke tumor, de meer agressieve kankercellen hebben de overhand gekregen. In dit opzicht is visualisering en kwantificering van de heterogeniteit binnen een tumor een belangrijk instrument in het graderen, het differentiëren, het volgen van behandeleffecten het voorspellen van de uiteindelijke uitkomst.

Het doel van deze thesis is enerzijds ontwikkeling en evaluatie van kwantitatieve technieken ten behoeve van de bepaling van tumorheterogeniteit, en anderzijds analyse van het belang van deze technieken voor het volgen van het behandeleffect en het voorspellen van de uiteindelijke uitkomst. Deze thesis richt zich voornamelijk op de volgende vragen:

- Kunnen MRI-beelden de tumorheterogeniteit weergeven?
- Welke analysemethoden worden gebruikt voor kwantificatie van tumorheterogeniteit ten behoeve van diagnose en/ofbehandeling en wat zijn de gerapporteerde prestaties van deze methoden?
- Is tumorheterogeniteit in Dynamisch Contrast Enhanced (DCE)-MRI-beelden, zoals gekwantificeerd met textuur-analysemethoden, gevoelig voor veranderingen als gevolg van therapie,en kan de uitkomstvoor de patiënt worden voorspeld?

Hoofdstuk 1 bevat een algemene inleiding van het onderzoek. De epidemiologie van kanker, de biologische achtergrond van intra-tumor heterogeniteit, de beeldvormende technieken in de oncologie en de beschikbare methoden voor kwantificatie van heterogeniteit binnen een tumor worden kort besproken en het hoofddoel van het proefschrift wordt geschetst.

Hoofdstuk 2 introduceert een methodologie voor het verkrijgen van een nauwkeurige 3D relatie tussen in vivo MRI-beelden met een hoge resolutie en de bijbehorende 3D-histologische datasets van het tumorweefsel. De belangrijkste kenmerken van dezemethodologie zijn: 1) de gestandaardiseerde data-acquisitie en verwerking, 2) het gebruik van ex vivo MRI-beelden, 3) het gebruik van een referentiesnijvlak, 4) het gebruik van histologiebemonstering met een hoge dichtheid, 5) het gebruik van

elastische registratie en 6) het gebruik van volledige 3D datasets. De methode bestaat uit twee aparte registratiestappen, beide met een geleidelijk toenemend aantal vrijheidsgraden (rigide, affine en elastische). Deze twee registratiestappen koppelen in vivo MRI-beelden aan ex vivo MRI-beelden en ex vivo MRI-beeldenaan histologische beelden. De resulterende nauwkeurigheid werd geëvalueerd door twee onafhankelijke waarnemers en was gemiddeld 0.7 mm tussen de in vivo MRI-beeldenen de gereconstrueerde 3D-histologische beelden. Deze nauwkeurigheid komt overeen met ongeveer 30 tot 50 cellen en is vergelijkbaar met de inter-waarnemer nauwkeurigheid. Automatische identificatie van necrotisch weefsel op basis van MRI intensiteit lijkt mogelijk. De identificatie van twee andere weefselsoorten, hemorragischen vitaal weefsel, was echter niet mogelijk met behulp van een T2*-W MRI sequentie. Voor de scheiding van deze weefseltypes zijn andere MRI-sequenties nodig. Dit werk is een eerste stap in het karakteriseren van tumoren met behulp van MRI-beelden. Nu de ruimtelijke correspondentie tussen de in vivo MRI-beelden en de 3D-H&E histologische data is vastgesteld, is de uitbreiding naar multi-spectrale MRI en meerdere histologie kleuringen een logische volgende stap. De 3D correspondentie van tumor histologie met in vivo MRI-beeldenmaakt de extractie van MRI-karakteristieken voor histologisch gedefinieerde regio's mogelijk.

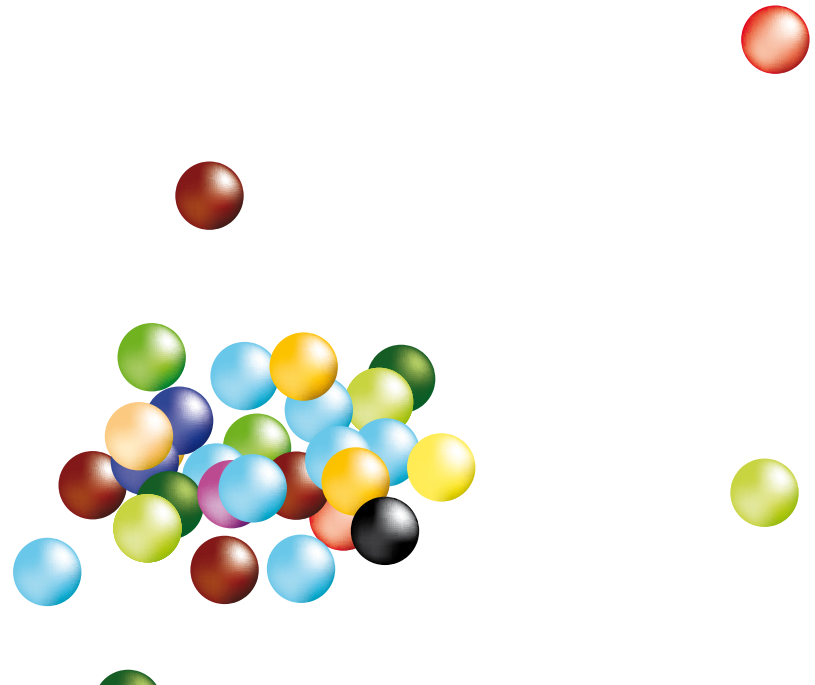
Hoofdstuk 3 biedt een systematisch literatuuroverzicht van de methoden voor kwantificatie van heterogeniteit voor tumor gradering, tumor differentiatie, uitkomstvoorspelling en het monitoren van behandeleffecten in oncologische beelden. De analysemethoden werden ingedeeld in vier categorieën: niet-ruimtelijke methoden (NSM), ruimtelijke grijswaarde methoden (SGLM), fractale analyse (FA) methoden, en filter en transformaties (F&T) methoden. De prestaties van de heterogeniteitskenmerken, gerapporteerd met behulp van maten zoals sensitiviteit, specificiteit, nauwkeurigheid en AUC, of met behulp van statistische testen, werden vergeleken. Van de 8956 unieke studies, rapporteerden 192 heterogeniteit als een biomarker voor oncologische belden. Sinds 2009 groeit het aantal studies dat over de kwantificatie van heterogeniteit in oncologische beelden rapporteert. Tot 2006 waren de meeste publicaties gebaseerd op echografie, na 2007 is een duidelijke toenamete zien van het aantal publicaties dat gebruik maakt van MRI. De meest gebruikte methoden in de hele periode zijn de NSM en SGLM. De meeste publicaties waren gericht op tumor gradering, differentiatie of voorspellen van de uitkomst. De prestaties van de heterogeniteitkenmerken werdenvooral gemeten met behulp van classificatie experimenten (68%) en statistische testen (30%). In 2% van de publicaties werd er geen kwantitatief onderzoek uitgevoerd. Bijna 98% van de studies vermelden positieve bevindingen. De AUC varieerde tussen de 0.5 en 1 met een mediaan van 0.89. Er is geen relatie gevonden tussen de verschillende prestatiematen en de bijbehorende beeldvormende modaliteit noch met de analysemethode. Van alle classificatie-studies gebruikte 41% geen cross-validatie als techniek om het effect van 'inleren' op de beschikbare data te voorkomen. Er werd een negatieve correlatie gevonden tussen de tumor-kenmerk ratio en de AUC. Het vermoeden bestaat dat deze negatieve correlatie veroorzaakt zou kunnen zijn door het 'inleren'. In een groot deel van de publicaties die testen op significantie, is geen correctie voor multiple tests uitgevoerd. In 36% van deze studies werd een significante afname van het aantal significante kenmerken geconstateerd na een Holm-Bonferroni correctie. Niet meer dan 12% van de geincludeerde studies hadden een prospectief studie-ontwerp. Hoewel het gebruik van retrospectief verkregen data noodzakelijk is voor het ontwikkelen, testen en evalueren van heterogeniteitsmaten als biomarker voor graderen, differentiëren, monitoren van behandeleffecten en voorspellen van uitkomsten, de werkelijke test is om de prestatie van de ontwikkelde kenmerken te bepalen in een prospectieve studie. Bovendien worden in de meeste retrospectief uitgevoerde studies de uitkomsten geëvalueerd zonder rekening te houden met andere beschikbare klinische informatie. Waar de onderzoeker voornamelijk geïnteresseerd kan zijn in de prestatie van het kenmerk op zich, is de arts geïnteresseerd in de toegevoegde waarde van het kenmerk in combinatie met de al beschikbare informatie. Om de heterogeniteitskenmerken uit hetonderzoekslab de klinische praktijk in te helpen, zouden onderzoekers zich moetenmeer moeten gaan richten op prospectieve studies voor de evaluatie van de toegevoegde waarde van de voorgestelde heterogeniteit biomarkerin aanvulling op de beschikbare klinische kenmerken.

In Hoofdstuk 4 worden dynamische contrast MRI data (DCE-MRI) van sarcoom patiënten, behandeld met TNF-α en melphalan, gebruikt om veel belovende heterogeniteitsmethoden te evalueren: de aankleurende fractie, SGLM en FA. Het vermogen van deze beeldgebaseerde biomarkers om behandelingseffecten te volgen en de uitkomst te voorspellen werd geëvalueerd met data van 18 patiënten. Met behulp van routinematig vervaardigde DCE-MRI scans, wordt in dit hoofdstuk onderzocht of SGLM en FA in staat zijn om veranderingen veroorzaakt door de therapie te meten wanneer gebruik gemaaktwordt van verschillende farmaco-kinetische modelleringsmethoden. De correlatie tussen de farmaco-kinetische en heuristische kenmerken, is bepaald met de Spearman's rangorde correlatiecoefficient. Deze coëfficiënt varieerde, gemiddeld over alle tumoren, tussen de 0.44 en 0.61. Uit de resultaten is gebleken dat, ongeacht de oorsprong van de geschatte parametrische map (farmaco-kinetisch of heuristisch), de aankleurende fractie en de coherentie, een SGLM-kenmerk, in staat waren om een significant verschil te meten tussen de follow-up scans en de uitgangs-scans voor de groep patiënten die reageerde op de therapie. Met deze kenmerken konden ook, wanneer ze bepaald werden op de uitgangs-scan, de twee groepen (de groep die reageerde op de therapie en de groep die niet reageerde) van elkaar onderscheiden worden. Het lijkt erop dat de patiënten met een grote vitale tumorfractie en een hoge coherentie goed reageren op de behandeling. Daarnaast reageren de patiënten met grote necrotische gebieden en een lage coherentie veel slechter op de behandeling. In deze studie werd aangetoond dat heterogeniteits kenmerken afgeleid van DCE-MRI beelden, onafhankelijk van de pharmaco-

kinetische analysemethode, gerelateerd zijn aan de response van de tumor op de therapie. We kunnen hieruit concluderen dat DCE-MRI beelden geschikt kunnen zijn om in vivo het effect van therapie op de tumor te volgen. In combinatie met histopathologische bevindingen, zou deze methode klinisch bruikbaar kunnen zijn voor de begripsvorming van de pathofysiologische veranderingen die optreden gedurende de behandeling, voor het bijsturen van de therapie en voor het voorspellen van de tumorrespons.

In Hoofdstuk 5 worden deveranderingen onderzocht in de regionale heterogeniteit, gemeten in DCE-MRI, als gevolg van behandeling met Isolated Limb Perfusion (ILP) met behulp van een sarcoom-diermodel. DCE-MRI werd geacquireerd voor en 1 uur na de ILP. Als contrastmiddel, is er gebruik gemaakt van een macromoleculair contrastmiddel, albumin-(Gd-DTPA)$_{45}$. Om regionale verschillen te kunnen beoordelen is er beeldregistratie uitgevoerd tussen de beelden voor en na behandeling. Om de heterogeniteit in de tumor te bepalen, zijn alle tumor volumes verdeeld in 16 sectoren waarover het cumulatieve mapvolume (CMV) van K^{trans} is uitgerekend. De variantie in de hellingshoek tussen de CMV curves van de verschillende sectoren is bepaald als maat voor de heterogeniteit. De resultaten geven aan dat de heterogeniteit van K^{trans} tussen de sectoren afneemt als gevolg van behandeling. Dit impliceert dat de behandeling een homogenisering van de tumor induceert, en geeft het potentieel aan van regionale analyse om een lokale behandelingseffect van ILP te evalueren. Wanneer een deel van de tumor aan de behandeling 'ontsnapt', kan dat funest zijn voor het totale behandelingseffect. Door gebruik te maken van deze regionale analyse zou deze 'ontsnapping' mogelijk binnen uren na aanvang van therapie opgemerkt kunnen worden. Resumerend, CMV's zouden kunnen fungeren als niet invasieve biomarkers van vroege behandeleffecten en dus gebruikt kunnen worden om tot therapie-aanpassingen te besluiten.

Appendix A geeft een kort overzicht van de beeldregistratietechnieken om verschillende data sets naar één coördinatensysteem te transformeren. Deze technieken zijn in dit proefschrift gebruikt om de verschillende vervormingen te corrigeren van de ex vivo tumorenten opzichte van de in vivo tumoren, zoals afgebeeld met MRI (Hoofdstuk 2). In Hoofdstuk 5 zijn deze technieken gebruikt voor bewegingscorrectie en voor het matchen van baseline en follow-up scans.



PhD portfolio

Name PhD Student	Lejla Alić
Erasmus MC Departments	Radiology & Medical Informatics
Graduate school	ASCI & Molecular Medicine
PhD period	2005-2010
Promotor	Prof.dr. WJ Niessen

General academic scills	**year**	**ECTS**[1]
Biomedical Scientific English Writing (MolMed)	2010	4
Research Management (MolMed)	2009	1
Grant proposal writing (MolMed)	2009	0.5
SPSS (MolMed)	2009	1
LaTeX (MolMed)	2009	1
Good clinical practice	2006	0.3

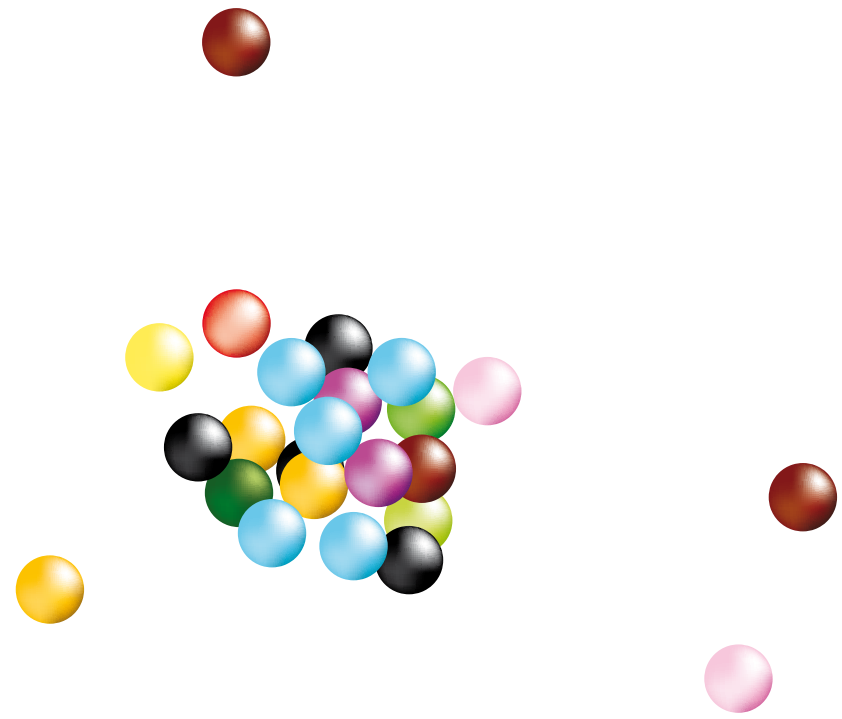
In-depth courses	**year**	**ECTS**
Measuring features (ASCI)	2008	4
Knowledge Driven Image Segmentation (ASCI)	2005	4
Mathematical morphology(ASCI)	2005	4

International Conferences		
ECR: European Congress of Radiology, Vienna.	2011	1
EMIM: 6th European Molecular Imaging Meeting, Leiden.	2011	1
ECR: European Congress of Radiology, Vienna.	2010	1
SPIE Medical Imaging, San Diego.	2010	1
ESMRMB: 26th European Society for Magnetic Resonance in Medicine and Biology, Antalya.	2009	1
EMIM: 4th European Molecular Imaging Meeting, Barcelona.	2009	1
IEEE International Symposium on Biomedical Imaging, Arlington.	2006	1

1 European Credit Transfer System (1 ECTS equals 28h of study)

Seminars, symposia, meetings and workshops		
BME: 3th Duch Bio-Medical Engineering Conference, Egmond aan Zee	2011	1
Photoshop & Illustrator workshop(MolMed)	2010	0.3
Medical Informatics days	2009	1
PK modeling workshops	2009	1
3rd Animal Imaging Workshop by AMIE (MolMed)	2009	1.4
ANGIONET	2007	1
Basic and Translational Oncology (MolMed)	2007	1.8
Medical Imaging Symposium for PhD students	2006	1
Medical Imaging symposium for PhD students, Rotterdam	2006	1
MICCAI: 9th International Conference on Medical Image Computing and Computer Assisted Intervention. Challenges in Clinical Oncology workshop	2006	1
Medical Informatics days	2005	1

Other		
Reviewing for Journal of Magnetic Resonance Imaging, International Journal of Computer Assisted Radiology and Surgery, Journal of the American Medical Informatics Association, International Conference on Medical Image Computing and Computer Assisted Intervention.	2009-	4
Research seminars at BIGR	2008-	1
Introduction MevisLab, MeVis, Bremen.	2006	0.5
Total		**43.8**



Acknowledgements

Many are accountable for this thesis (and research behind it) and the preparation of its public defense. As I am aware of the danger of forgetting people, I would like to start with expressing my most sincere gratitude to everybody who contributed to it in any way. However, there are certain people I wish to acknowledge in more detail.

First of all professor Niessen (beste Wiro), I am honored to have you as my supervisor. I always have valued your promptness in reading my work all over again. Despite your tight schedule, you have always been very creative in making time for a quick meeting, even though that sometimes include Skyping from another part of the world.

I also would like to express my sincere gratitude to Dr Veenland. Beste Jifke, I am grateful for you as my co-supervisor, however I would like to express my appreciation to your family first. Jan, Hendrik, Yme, and Afke, thank you for all weekends and evenings you have had to miss Jifke during the long discussions we sometimes had, and during the preparation of the manuscripts. I suppose the quantity of cookies you get on weekly basis could go down by now. Dear Jifke, thank you for numerous and valuable scientific advice, incredible responsiveness and energetic support that helped to improve the quality of this research.

This thesis would never be a fact without support and assistance of many people during the planning and execution of numerous experiments. Thanks to the departments of Experimental Surgical Oncology (Timo, Sandra, Linda), Radiology (Marion, Joost, Monique), Nuclear medicine (Marion, Magda, Harald), Pathology (Peter, Bart). Thanks are also due to those who preceded my research and those who have continued my research, both represented in my two paranymphs during the defense: Marion van Vliet and Karin Bol.

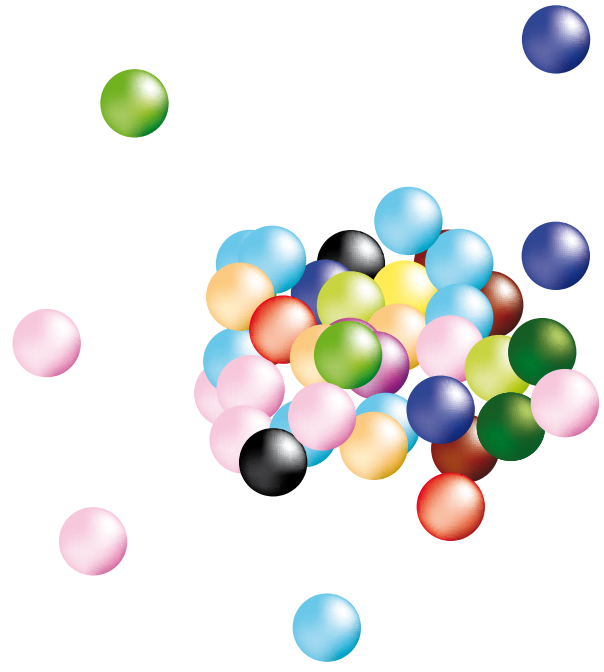
I also want to thank Ton for having beeen so helpful in creating many nice figures, and Laraine for helping me juggling between UK and USA English and improving my writing skills. The two secretaries (Petra and Desiree), they provided invaluable practical support and managed to get me organized. I also would like to thank all people I shared a room with during those years I spend at the Erasmus MC: Henri, Jifke, Oleh, Ihor, Rashindra, Danijela, Azadeh, Reinhardt, and Edward.

I would like to express my gratitude to prof.dr. Sir Mike Brady for inspiring me during my stay at Robotics Research Laboratory in Oxford, UK. I am also grateful to prof.dr. Alison Noble for sorting out all paper work before I started there, even though she was on maternity live at the time. To all lab members and friends in UK, I had a great stay in Oxford.

During my working experience outside the university, I enjoined company of a number of nice people. At i-Optics I discovered ways the applications in medical image processing fit in a commercial setting. I would like to thank my colleagues at Topcon for teaching me some valuable commercial lessons. Erik, Pierre, Toru-san, I think with pleasure about my time at Topcon.

I would like to express my sincere gratitude to my TNO colleques for their guidance and encouragement. It felt almost surreal to come back and walk same corridors after almost ten years after I left in the first place. As there are many colleagues, to many to mention individually here, I wish to acknowledge the kindness of all of you.

I wish to thank my parents for things one can receive only from parents, things generally taken for granted but which for many turn out to be true only in theory. I also wish to show appreciation to my family and friends for their understanding I couldn't be on yet another family event. I promise to do better now. My two lovely sisters (Enisa and Dinka): each time I see you, I feel we are not spending enough time together. Last and most important my own household (I can't even say 'little family' as it has grown to a respectable size of five) ... Renato, Ben, Om and Hannah, all my love is for you.



Curriculum vitae

http://nl.linkedin.com/in/lejlaalic

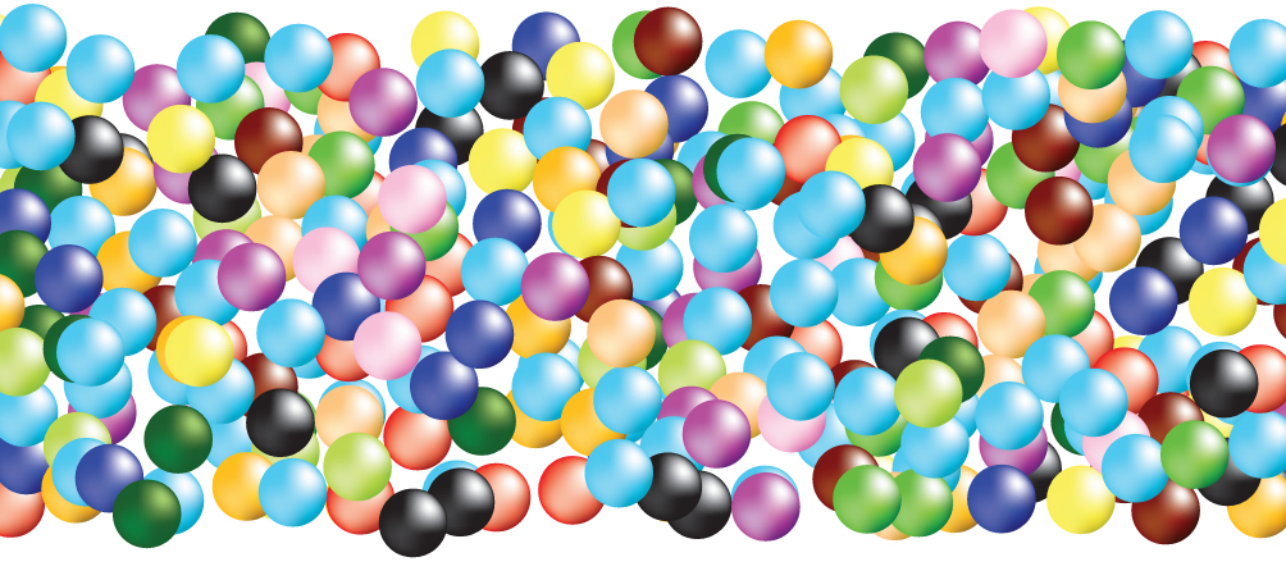
https://twitter.com/#!/Lejla_Alic

LejlaResearch@gmail.com

Lela Alic (MSc, 2001) studied Electrical Engineering at Delft University of Technology, the Netherlands. Part of the MSc program was carried out at Leiden University Medical Center, the Netherlands (department of Radiology, Division of Image Processing -LKEB); and Academic Medical Center university of Amsterdam (department of Experimental Anesthesiology).

In 2001 she joined the TNO-FEL laboratory at The Hague where she worked on image analysis of ground penetrating radar data. In 2004 she started working as a research assistant at BIGR where she was awarded with a NWO grant. From Mei 2006 to August 2007 she visited the Robotics Research Group at the Department of Engineering Science of University of Oxford, UK. Lejla co-organized the MICCAI 2006 workshop: 'Medical Image Processing: The Challenges in Clinical Oncology'.

Currently she holds a scientist position at TNO (department of Intelligence Imaging) where she works on medical imaging applications. She is involved in hyperspectral data analysis within 'van het Hoff' program on medical photonics. Her research interests include many aspects of image processing, image analysis and pattern recognition, and their applications in oncology and molecular imaging.



ISBN: 978-90-8891-630-4