Effectiveness of Using Texture Analysis in Evaluating Heterogeneity in Breast Tumor and in Predicting Tumor Aggressiveness in Breast Cancer Patients

A thesis submitted to the University of Arizona College of Medicine – Phoenix
in partial fulfillment of the requirements for the degree of Doctor of Medicine

Alix Hopp
Class of 2016

Mentor: Ronald Korn, MD, PhD

Acknowledgements

I wish to thank my mentor Dr. Ronald Korn and his collaborator Dr. Jasgit Sachdev. I would also like to thank biostatisticians Dr. Chengcheng Hu and Paul Kang for reviewing the statistics analysis in this study.

Abstract

Objective and Hypothesis

We hypothesize that tumor heterogeneity or tissue complexity, as measured by quantitative texture analysis (QTA) on mammogram, is a marker of tumor aggressiveness in breast cancer patients.

Methods

Tumor heterogeneity was assessed using QTA on digital mammograms of 64 patients with invasive ductal carcinoma (IDC). QTA generates six different values – Mean, standard deviation (SD), mean positive pixel value (MPPV), entropy, kurtosis, and skewness. Tumor aggressiveness was assessed using patients' Oncotype DX® Recurrence Score (RS), a proven genomic assay score that correlates with the rate of remote breast cancer recurrence. RS and hormonal receptor status - estrogen receptor (ER) and progesterone receptor (PR) - were collected from pathology reports. Data were analyzed using statistical tools including Spearman rank correlation, linear regression, and logistic regression.

Results

Linear regression analysis showed that QTA parameter, SD, was a good predictor of RS ($F=6.89$, $p=0.0108$, $R^2=0.0870$) at SSF=0.4. When PR status was included as a predictor, PR status and QTA parameter Skewness-Diff, achieved linear model of greater fit ($F=15.302$, $p<0.0001$, $R^2=0.2988$) at SSF=1. Among PR+ patients, Skewness-Diff was a good linear predictor of RS ($F=9.36$, $p=0.0034$, $R^2=0.1320$) at SSF=0.8.

Logistic regression analysis showed that QTA parameters were good predictors of high risk RS probability, using different cutoffs of 30 and 25 for high risk RS; these QTA parameters were Entropy-Diff for RS>30 ($chi^2=10.98$, $p=0.0009$, AUC=0.8424, SE=0.0717) and Mean-Total for RS>25 ($chi^2=9.98$, $p=0.0016$, AUC=0.7437, SE=0.0612). When PR status was included, logistic models of higher log-likelihood chi^2 were found with SD-Diff for RS>30 ($chi^2=18.69$, $p=0.0001$, AUC=0.9409, SE=0.0322), and with Mean-Total for RS>25 ($chi^2=25.56$, $p<0.0001$, AUC=0.8443,

SE=0.0591). For PR+ patients, good predictors were SD-Diff for RS>30 (chi^2=6.87, p=0.0087, AUC=0.9212, SE=0.0515), and MPP-Diff and Skewness-Diff for RS>25 (chi^2=16.17, p=0.0003, AUC=0.9103, SE=0.0482).

Significance

Quantitative measurement of breast cancer tumor heterogeneity using QTA on digital mammograms may be used as predictors of RS and can potentially allow a non-invasive and cost-effective way to quickly assess the likelihood of RS and high risk RS.

Table of Contents

List of Figures and Tables

Introduction/Significance

Breast cancer is a major worldwide health problem. In 2008, 1.4 million women were diagnosed with breast cancer worldwide and about 459,000 deaths were reported [1]. In the United States, breast cancer is the second most common cancer among American women, second only to skin cancer. Breast cancer is also the second leading cause of cancer death of American women, exceeded only by lung cancer. Approximately 1 in 8 US women will develop invasive breast cancer in their lifetime [2].

Heterogeneity of cancer tumors

In spite of an increasing incidence of breast cancer, mortality due to breast cancer has been steadily declining in major developed countries. However, there still remains much work to be done to understand and treat breast cancer. Currently there is no gold standard therapy to treat all breast cancer tumors. This complexity in treatment is due to heterogeneity of breast tumors, which according to recent studies, is the result of the Darwinian evolution of tumor, whereby the natural and founder mutations in individual tumor cells, subjected to differential selective pressure generated by the microenvironment and therapeutic interven tions, create various evolutionary pathways and result in the observed heterogeneity in cancer tumors [3]. Tumor heterogeneity is demonstrated by differences in genomic, proteonomic, and metabolic expressions of the tumor cells, as well as differences in the interaction between the tumor cells and the surrounding environment.

The recognition of heterogeneity in cancer tumor has led to targeted therapy and management efforts, which utilize specific predictive and prognostic molecular signals to guide personalized therapy to treat cancer [4]. Many current cancer research efforts are also focused on understanding molecular signals in tumors, for better cancer diagnosis, prognosis, therapy selection, and measurement of treatment response.

Imaging and cancer

Radiologic imaging, including Mammography (which includes Digital Mammography, Computer-Aided Detection, and Breast Tomosynthesis), Computed Tomography (CT), Magnetic Resonance

Imaging (MRI), ultrasound, and Positron Emission Tomography (PET), is the foundation in cancer screening, staging, and detection of treatment responses. Imaging is obtained in almost all patients with cancer diagnosis in order to measure response to treatment. The most commonly used marker for measuring clinical response is the tumor size. If there is a complete eradication of tumor, then it is considered a *complete response*. If there is a 30% or greater decrease in tumor size then it is a *partial response*. If there is a 20% or greater increase in tumor size then it is considered a *progressive disease*.

However, reduction in tumor size may not be the best measurement in treatment response. Residual tumor cells may still contain viable malignancy; furthermore, post-therapy response of the tumors, such as inflammation, fibrosis, or necrosis, may be interpreted as lack of therapy response [5]. Hence there is an active search for biologic markers that can serve as better prognostic and predictive agents for breast cancer. Gonzalez-Angulo et al [6] define a prognostic factor to be "a measurable clinical or biological characteristic associated with a disease-free or overall survival period in the absence of adjuvant therapy", whereas a predictive factor is "any measurable characteristic associated with a response or lack of a response to a specific treatment."

Biomarkers

The paradigm breast cancer biomarkers are estrogen receptors (ER) and progesterone receptors (PR), which are nuclear hormone receptors. Breast cancer tumors that tested positive for ER and/or PR expression receive remarkable benefits from endocrine therapy such as tamoxifen and aromatase inhibitors (AI), regardless of the age of the patient; whereas breast cancer tumors that are negative for ER and/or PR receive little or no benefits from the same therapy [7,8]. Another important breast cancer biomarker is HER2 (human epidermal growth factor receptor 2). HER2 protein overexpression or gene amplification is referred to as HER2+. HER2+ breast cancer can benefit from trastuzumab or lapatinib therapy.

Recent studies show that breast cancer can be typed based on its tumor source in the terminal duct: Luminal and Basilar. Tumors derived from the lumen tend to be ER+ and/or PR+, have a

prevalence of 60% of all breast cancer, and tend to respond to hormonal therapy. Basilar breast cancer tends to be HER2+, and is more aggressive than Luminal type. Luminal type is further subdivided into sub-types A and B, where Luminal A is less aggressive and is more responsive to hormonal therapy, whereas Luminal B is more aggressive and may require additional chemotherapy.

A proliferative index that is currently used clinically with breast cancer is Oncotype DX®, for both invasive carcinoma and DCIS (ductal carcinoma in situ). The test for invasive carcinoma is the Oncotype DX®, which is a genomic assay that analyzes the gene expression of a panel of 21 genes. The assay yields an Oncotype DX® Assay Recurrence Score (RS) - from 0 to 100 - that correlates to the likelihood of distant recurrence and likelihood of chemotherapy benefits, for women with newly diagnosed, stage I or II, node-negative, and ER-positive breast cancer, or postmenopausal women with newly diagnosed node-positive and hormone-receptor-positive (either ER+, PR+, or both) breast cancer [9,10,11,12,13]. This assay is included in American Society of Clinical Oncology (ASCO) and National Comprehensive Cancer Network (NCCN) guidelines. The test for DCIS is the Oncotype DX® for DCIS, which will not be utilized in this study.

Texture Analysis

Advanced imaging techniques using dynamic contrast enhanced MRI or CT, diffusion weighted MRI, and PET/CT, ultrasound, and tomosynthesis with mammography are increasingly used to measure tumor heterogeneity [14,15]. Small or pilot studies on non-small cell lung cancer [16,17], colorectal cancer [18,19,22], non-Hodgkins lymphoma [23], and breast cancer [24], have shown a promising correlation between texture analysis results and tumor heterogeneity. Texture analysis has also been used to characterize histological breast cancer types [15,25], screen high-risk breast cancer patients [26], and to monitor tumor response to therapy [23,24,27,28,29]. Advanced imaging techniques employ specialized software programs to generate quantitative data from the radiologic images in order to measure treatment response. These software programs are capable of generating a profile of characteristics and patterns within a given tumor in order to predict its response to treatment. Based on these previous

studies in which tumor texture is used to characterize biological phenotype (e.g. hypoxia, low pH, altered vascular perfusion, vascular shunting, etc.), we therefore hypothesize that tumor texture can be used to predict treatment response and to aid in prognosis.

The process of quantifying textural patterns and signals in radiologic images is called *textural analysis*. A major advantage of textural analysis is that it does not require invasive procedure such as biopsy nor even additional radiologic imaging; it is applied on routine radiologic images that are already indicated for in cancer treatment and monitoring.

Research Materials & Methods

This is a retrospective and exploratory study using archived digital mammograms and pathology reports of breast cancer patients. Approval granted by the Scottsdale Healthcare Institutional Review Board (IRB) was obtained under the expedited review process.

Patients

Eligible subjects were patients who were diagnosed with breast cancer at SMIL or the Scottsdale Healthcare hospital and who also had digital mammography performed at Scottsdale Medical Imaging (SMIL) between September of 2009 and December of 2012.

Quantitative Texture Analysis

QTA was performed by a board-certified radiologist, Ronald Korn MD PhD, using a unique software analysis platform called TexRAD, developed at the University of London, Sussex. TexRAD can assess the distribution of gray-levels, coarseness, and regularity within a tumor lesion, which is imaged on routine radiology scans, including CT, MRI, PET, mammograms and perhaps ultrasounds. TexRAD has been used to show correlation between a texture analysis parameter called mean positive pixel values (MPP) with hypoxia in non-small cell lung cancer (NSCLC) [16].

The process of performing texture analysis was as follows. DICOMM compatible radiology images were de-identified and imported onto a secure server, where images of interest were selected for analysis. Areas containing tumor lesions were then visually identified by drawing regions of interest (ROI) around the target lesions. A single button click then performed detailed measurement and transformation of the gray level intensity of each pixel in the ROI, followed by image filtration (Laplacian of Gaussian) using 5 filter levels (sizes 0.4mm, 0.6mm, 0.8mm, 1mm, and no filter), and finally a display of histogram of filtered pixel intensity vs. number of pixels. The filter, also known as SSF, is used to cluster small groups of pixels together in order to measure local density changes. The histogram was then processed to yield the following QTA parameters:

- Mean – Measures the average density within a cluster of pixels at a given SSF level, measured in Hounsfield unit (HU). HU measurements between 0-20 indicates fluid, 20-80 soft tissue, >80 calcium or metal, < 0 to -100 fat, and < -100 air. A measure of necrosis [21].
- Standard deviation (SD) – Measures the spread of density distribution in the filtered image. The natural logarithm of mean pixel density normalized to the total number of pixels. It is a measure of heterogeneity and microstructural changes in the entire ROI.
- Entropy – Measures the mean density of clustered pixels within the ROI, i.e. irregularity in the ROI. A measure of heterogeneity.
- Mean positive pixel values (MPP) – Measures the average density of positive pixel values. Low MPP has been correlated with areas of hypoxia in non-small cell lung cancer (NSCLC) [16]. A measure of hypoxia.
- Skewness – Measures the sharpness of density distribution. Positive skewness (i.e. slant of the peak to the left) may reflect angiogenesis [19].
- Kurtosis – Measures the symmetry of density distribution. A measure of angiogenesis, vascular shunting, and tumor homogeneity.

Next, a ROI of nearly identical size was drawn in the normal tissue of the same mammogram and QTA parameters in the same five SSF levels were obtained from the ROI of the normal tissue. Similar QTA processing was done for microcalcification tissue in mammograms where microcalcification lesions were visually identifiable (n=35).

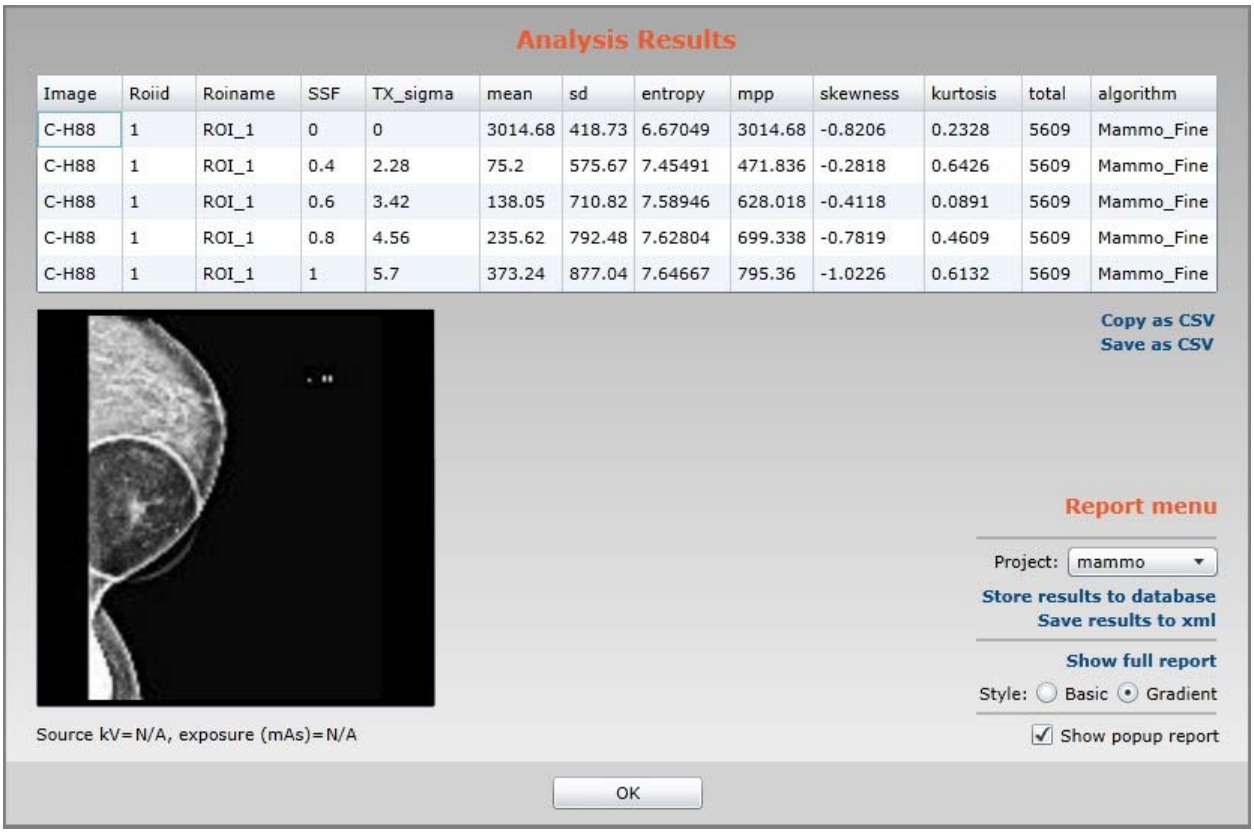

Figure 1 – Sample QTA parameter output, as generated in TexRAD after selecting a region of interest in a breast cancer mammogram and clicking a single button in TextRAD.

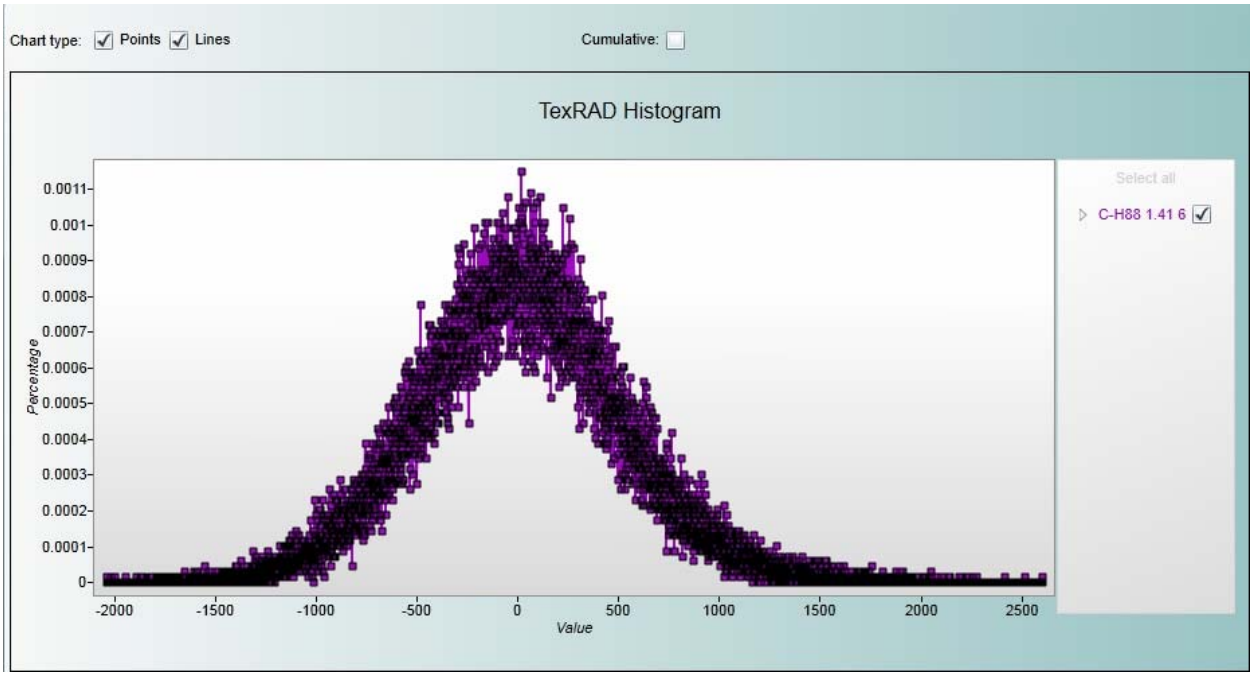

Figure 2 – Sample TexRAD histogram output for a specific filter level, generated based on the QTA parameter output seen in Figure 1. The goal of QTA is to quantify such histogram using QTA parameters Mean, Standard Deviation, Mean Positive Pixel Value, Entropy, Kurtosis, and Skewness.

The measured QTA parameters were outputted from TexRAD as a .csv file and imported to a verified Excel spreadsheet for derivation of additional QTA parameters. Three additional sets of QTA parameters are derived for each filter level, they consist of:

1) QTA parameters normalized to the size of ROI by dividing the values of primary QTA parameters (e.g. Mean, SD, Entropy, MPP, etc.) by the total number of pixels in the ROI for each filter level. They are named Mean-Total, SD-Total, Entropy-Total, MPP-Total, Skewness-Total, and Kurtosis-Total.

2) The maximal change or range in each primary parameter across the SSF levels, named Mean-Range, SD-Range, Entropy-Range, MPP-Range, Skewness-Range, Kurtosis-Range. Note that these values are the same for all SSF levels of a given tissue type (tumor or normal tissues) in the same subject.

3) The difference between the primary QTA parameters of the tumor and normal tissues, named Mean-Diff, SD-Diff, Entropy-Diff, MPP-Diff, Entropy-Diff, Skewness-Diff, and Kurtosis-Diff. Note that these values are the same for tumor and normal tissues at the same SSF level in the same subject.

Table 1 - Primary and Derived QTA Parameters

Primary QTA Parameters	**Primary QTA Parameters Normalized To Total Number of Pixels (Derived)**	**Range of Primary Parameters Across SSF Levels (Derived)**	**Difference Between Primary QTA Parameters Between Tumor and Normal Tissues (Derived)**
Mean	Mean / Total	$Mean_{high} – Mean_{low}$	$Mean_{tumor} – Mean_{normal}$
SD	SD / Total	$SD_{high} – SD_{low}$	$SD_{tumor} – SD_{normal}$
Entropy	Entropy / Total	$Entropy_{high}$ - $Entropy_{low}$	$Entropy_{tumor}$ - $Entropy_{normal}$
MPP	MPP / Total	MPP_{high} - MPP_{low}	MPP_{tumor} - MPP_{normal}
Skewness	Skewness / Total	$Skewness_{high}$ - $Skewness_{low}$	$Skewness_{tumor}$ - $Skewness_{normal}$
Kurtosis	Kurtosis / Total	$Kurtosis_{high}$ - $Kurtosis_{low}$	$Kurtosis_{tumor}$ - $Kurtosis_{normal}$

The derivation was done automatically by Excel functions and Stata program to eliminate manual calculation mistakes.

Tumor Aggressiveness

To assess tumor aggressiveness, we used the patients' Oncotype DX® Recurrence Score (RS). This score is the result of the 21-Gene Oncotype DX Breast Cancer Assay, a genomic assay clinically validated and recommended by major guidelines for all node-negative, ER+ breast cancer patients. The score directly correlates with the 10-year distant recurrence rate, with score > 30 considered high risk, score between 18 and 30 inclusive considered intermediate risk, and score < 18 considered low risk. A more recent, alternative risk criterion considers score > 25 as high risk and score < 11 as low risk [31]. RS was collected from the Oncotype pathology reports stored in Scottsdale Healthcare ChartMAXX database. In addition, ER and PR status, if available, were also collected from the biopsy or surgery pathology reports stored in SMIL RIS system or Scottsdale Healthcare portal, respectively.

Statistical Analysis

All statistical analysis was carried out using STATA/IC version 13 on a PC.

It is known that many biological variables do not meet the underlying assumptions of parametric statistical tests (e.g. linear regression). For example, biological data often are not normally distributed or the variances of the residuals are not homogeneous, in other words, they do not exhibit normality or homoscedasicity. Standard data transformation, such as log or square-root transformation, preserves the monotonicity of the data, while simultaneously improving the normality and the homoscedasicity of the variables; the effect is often clear especially when there is a large number of observations. Therefore, in order to examine such data using parametric statistical analysis, transformation of the data is often an important prerequisite. In this study, log (base-10) and square-root transformation were both utilized to improve normality and homoscedasicity of the data for linear regression analysis. Untransformed data was used in the logistic regression analysis, because logistic regression does not assume normal and homoscedastic data.

Table 2 - Data Transformation of Select QTA Parameters

QTA Parameter	Transformed QTA Parameter
Mean	*log*(Mean + 400)
Kurtosis	*log*(Kurtosis + 2)
Mean-Total	*Sqrt*(Mean-Total)
SD-Total	*log*(SD-Total)
Entropy-Total	*log*(Entropy-Total)
MPP-Total	*log*(MPP-Total)
Kurtosis-Total	*Sqrt*(Kurtosis-Total)
SD-Range	*Sqrt*(SD-Range)
Skewness-Range	*log*(Skewness-Range)
Kurtosis-Range	*log*(Kurtosis-Range)

First, we assessed whether QTA parameters can predict RS by performing robust multiple linear regression using RS as the dependent variable or outcome, and QTA parameters and age as independent variables or predictors. A linear model was considered statistically significant if the t-statistics of every predictor in the model is > 2 (i.e. greater than 2 standard deviations) and its associated p-value is <= 0.05. The fitness of the linear model was determined by examining *F*-statistics and the associated p-value: The best-fit linear model has the highest *F*-statistics and the lowest associated p-value. As recommended by Stock and Watson [30], robust algorithm for linear regression was used to control for heteroskedasticity, an assumption of linear regression that means that the variance of the residuals should be constant.

For each SSF level, the best-fit linear regression model of RS was found by using backward step-wise elimination, which consisted of two steps: First, Spearman rank correlation was performed between RS and each of QTA parameters with significance level of 0.20, allowing for the identification of any QTA parameters that may combine to have significance contribution in the final model. Next, linear regression was performed iteratively, dropping the predictor with the biggest p-value greater than 0.05 each time, until all predictors have p-value less or equal to 0.05. This analysis was done three times, first time including PR status as a predictor, second time without PR status, and third time examining PR+ patients only. The rationale for the first two analyses was that PR status may not be known at the time of QTA. The third analysis allowed us to see if QTA parameters can predict RS in PR+ population. All best-fit linear models were found with a custom program written in Stata 13 programming language that implemented the backward step-wise elimination algorithm as outlined; the automation minimized manual errors and allowed for reproduction of models.

The resulting linear models were examined with regression diagnostic tools, including check for normalized residuals via STATA's distributional diagnostic plots (e.g. *kdensity*, *pnorm*, *qnorm*) and the Shapiro-Wilk Test, check for severe outliers with the Interquartile Range Test, check for collinearity via Variance Inflation Factors and condition index tests, check for homoscedasticity via graph of residuals vs. predicted values, check for model specification via specification link

test, and check for omitted variables via the Ramsey (1969) regression specification-error test (RESET).

Second, we assessed whether QTA parameters can predict high risk RS (both RS > 30 and RS > 25) by performing multiple logistic regression using high RS as the outcome (1 if RS>30 or 25, 0 otherwise), and QTA parameters and age as predictors, also employing backward step-wise elimination. This analysis was done three times, first time including PR status as a predictor, second time without PR status, and third time examining PR+ patients only. The rationales for the three times logistic regression analysis is the same as those for linear regression analysis.

Following each logistic regression analysis, the log likelihood chi^2 value of the model was checked to see if the model as a whole was statistically significant. Hosmer and Lemeshow's goodness of fit test was employed to examine the fit of the logistic models, meaning how well the model is able to predict the outcome. A link test was also performed to detect model specification error, which could mean either that logit function (used in logistic regression) was not the correct function to use, or that the relationship between the logit of the outcome and the predictors was not linear. Collinearity of the predictors was also checked using STATA's collinearity test. Finally, ROC analysis was performed on the best-fit model to assess the discrimination of the model, which is how well the model distinguishes patients who have high risk RS from those have non-high risk RS. The ROC area under curve (AUC) was recorded.

We considered the problem of multiple comparisons in our analysis, since multiple filters were used on the same ROIs. One possible solution was to skip certain filter level(s), and another was to use more stringent significance level, namely instead of the standard 0.05, use 0.05/5=0.01. However, given the significance levels in the models found – many less than 0.01, and the fact that this is an exploratory study, we decided not to skip any filter level in our analysis.

Results

142 patients were identified based on the eligibility requirement, but 78 were eliminated because of one or more of the following reasons: Had microcalcification lesion only, no tumor lesion that was visually identifiable on the mammograms, no verifiable Oncotype DX® Recurrence Score (RS), or had multiple tumor lesions with different hormonal status and/or RS. After exclusion, we had 64 patients remaining. One of them (subject ID 066) had 2 tumor lesions, both lesions had identical hormonal status and RS, therefore statistical analysis was performed on 65 data points. For each patient, QTA was performed on a single tumor lesion at SSF levels 0.4, 0.6, 0.8, 1, and 0 (SSF=0 means no filter) with the exception of subject ID 066, who had 2 tumor lesions.

The mean age at the time of diagnosis was 61 (standard deviation=11 years) and ranged from 36 to 83. The patients' ER and PR status are shown in

Table 3.

Table 3 - Tabulation of ER and PR Status of Patients with RS

	PR		
ER	**Positive**	**Negative**	**Total**
Positive	56	7	63
Negative	1	0	1
Total	57	7	64

The RS had a mean of 20 (standard deviation=11) and ranged from 4 to 65. As mentioned previously, the standard thresholds for risk stratification are RS < 18 for low risk, 18 <= RS <= 30 for intermediate risk, RS > 30 for high risk [30]. However, recent studies have attempted to elucidate the tumor aggressiveness of patients in the intermediate risk RS group (18 <= RS <= 30) by stratifying risk group with a different set of thresholds: RS < 11 for low risk, 11 <= RS <= 25 for intermediate risk, and RS > 25 for high risk [31].

Table 4 and

Table 5 show the breakdown of the 64 patients into the three risk groups, based on the two definitions.

Table 6 lists the mean, standard deviation, and range of the primary QTA parameters in tumor tissues at SSF levels 0.4 to 1. Figure 3 shows the average values of the QTA parameters as the filter (SSF) changes from 0.4 mm to 1.0 mm and also when there is no filter.

Table 4 - Tabulation of Standard RS Risk Group and PR Status

	PR		
RS Risk Group	**Positive**	**Negative**	**Total**
Low Risk (RS < 18)	35	0	35
Intermediate Risk (18 <= RS <= 30)	19	3	22
High Risk (RS > 30)	3	4	7
Total	57	7	64

Table 5 - Tabulation of Alternative RS Risk Group and PR Status

	PR		
RS Risk Group	**Positive**	**Negative**	**Total**
Low Risk (RS < 11)	9	0	9
Intermediate Risk (11 <= RS <= 25)	42	1	43
High Risk (RS > 25)	6	6	12
Total	57	7	64

Table 6 – The mean, standard deviation, and range of the primary QTA parameters at SSF=0.4, 0.6, 0.8, 1, and 0 (No filter)

SSF	QTA Parameter	Sample Size (n)	Mean	Standard Deviation	Minimum	Maximum
0.4	Mean	65	68.60708	55.85314	-12.01	316.61
0.4	SD	65	634.9731	154.9119	348.56	1031.47
0.4	Entropy	65	7.653379	0.2358882	7.17405	8.04565
0.4	MPP	65	517.3695	131.9921	272.202	838.151
0.4	Skewness	65	-0.09662	0.3005977	-1.7556	0.4905
0.4	Kurtosis	65	0.5355985	1.097498	-0.5714	7.7133
0.6	Mean	65	149.7208	109.6911	-16.94	570.49
0.6	SD	65	765.0869	211.1919	404.45	1497.25
0.6	Entropy	65	7.796314	0.2568659	7.25682	8.26074
0.6	MPP	65	654.8278	192.9854	335.024	1274.12
0.6	Skewness	65	-0.1088385	0.335995	-1.5016	0.7663
0.6	Kurtosis	65	0.4318538	0.9044054	-0.5396	4.7585
0.8	Mean	65	256.7908	169.9689	-14.52	785.95
0.8	SD	65	865.8097	261.5368	461.5	1750.13
0.8	Entropy	65	7.885341	0.2633471	7.30915	8.43365
0.8	MPP	65	787.6868	258.1746	403.857	1750.13
0.8	Skewness	65	-0.1185754	0.393803	-1.1378	0.9672
0.8	Kurtosis	65	0.3651415	0.987174	-0.8932	4.2965
1	Mean	65	385.6463	235.56	2.22	1003.88
1	SD	65	944.4589	307.0153	492.44	2137.81
1	Entropy	65	7.945265	0.2764926	7.32828	8.5702
1	MPP	65	920.5114	323.4779	430.198	2141.74
1	Skewness	65	-0.1165585	0.4346557	-1.4324	0.9026
1	Kurtosis	65	0.2624369	1.140967	-0.9769	5.7072
0	Mean	65	2534.86	429.3455	1385.75	3387.36
0	SD	65	444.3408	150.9218	243.61	938.4
0	Entropy	65	6.571937	0.3008233	6.07411	7.33031
0	MPP	65	2534.86	429.3455	1385.75	3387.36
0	Skewness	65	-0.4918954	0.4406189	-1.7215	0.3882
0	Kurtosis	65	0.22204	0.858151	-1.1182	3.2106

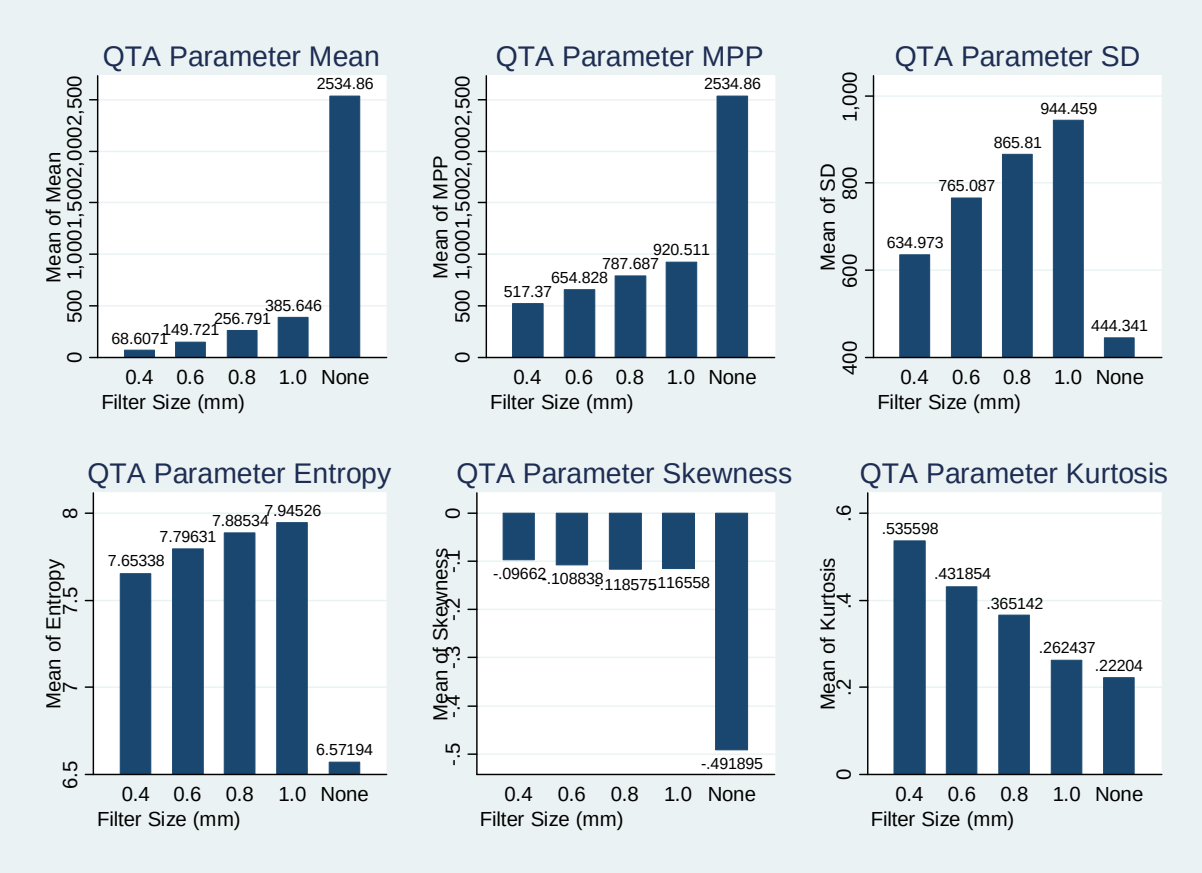

Figure 3 - Average of primary QTA parameters at filter levels 0.4-1.0 mm and no filter.

Differences in QTA Parameter Mean Across Different Risk Groups

We observed that at each SSF level, the average values of QTA parameters differed among the three RS risk groups, demonstrated in Table 7 and Figure 4 through Figure 9. Among the QTA parameters, Mean showed the most statistically significant difference between the high and non-high risk groups (i.e. low and intermediate risk groups). Highest *t* statistic was observed at SSF=0.8, where QTA parameter Mean of the tumor tissue was lower in the high risk RS group than in the low or intermediate risk RS group (Intermediate vs. high risk group: t=3.1756, p=0.0044. Low vs. high risk group: t=4.2251, p=0.0002), see Figure 4 and

Table 7. This relationship between RS and Mean was observed at all SSF levels except for SSF=0 (i.e. no filter). Furthermore, this relationship was not observed in normal tissue, see Figure 10.

Table 7 - t Test Comparison of QTA Parameter Mean in different risk groups in tumor tissue. Values in red boldface font highlight the statistically significant differences. High risk = Score > 30. Mid risk = Score between 18 and 30. Low risk = Score < 18.

ssf	Risk Group	n	Mean	SE	SD	*t*	p-value	Diff	Diff SE
0.4	High	7	26.35	7.809821	20.66284	.	.	.	.
0.4	Mid vs. high	22	62.89091	8.97	42.07303	**3.0724**	**0.0057**	36.54091	11.89345
0.4	Low vs. high	36	80.31694	10.6071	63.64258	**4.0971**	**0.0003**	53.96694	13.17208
0.4	Low vs. Mid	.	.	.	.	1.2544	0.2149	17.42604	13.89141
0.6	High	7	63.52	16.08923	42.56809	.	.	.	.
0.6	Mid vs. high	22	140.6391	18.55678	87.03902	**3.1400**	**0.0048**	77.11909	24.56048
0.6	Low vs. high	36	172.0319	20.47731	122.8638	**4.1668**	**0.0003**	108.5119	26.04195
0.6	Low vs. Mid	.	.	.	.	1.1360	0.2609	31.39285	27.63466
0.8	High	7	119.5986	25.72049	68.05003	.	.	.	.
0.8	Mid vs. high	22	245.3786	30.12085	141.2793	**3.1756**	**0.0044**	125.7801	39.60819
0.8	Low vs. high	36	290.4411	31.20084	187.2051	**4.2251**	**0.0002**	170.8425	40.43558
0.8	Low vs. Mid	.	.	.	.	1.0391	0.3035	45.06247	43.36771
1	High	7	193.9643	36.86359	97.53188	.	.	.	.
1	Mid vs. high	22	372.5791	42.63327	199.9677	**3.1691**	**0.0045**	178.6148	56.36062

1	Low vs. high	36	430.9033	42.90795	257.4477	**4.1885**	**0.0003**	236.939	56.56869
1	Low vs. Mid	.	.	.	.	0.9642	0.3393	58.32424	60.48709
0	High	7	2661.97	130.1412	344.3212	.	.	.	.
0	Mid vs. high	22	2516.122	92.67463	434.6826	-0.9129	0.3783	-145.8477	159.7664
0	Low vs. high	36	2521.595	74.48862	446.9317	-0.9361	0.3705	-140.375	149.9509
0	Low vs. Mid	.	.	.	.	0.0460	0.9635	5.472727	118.8997

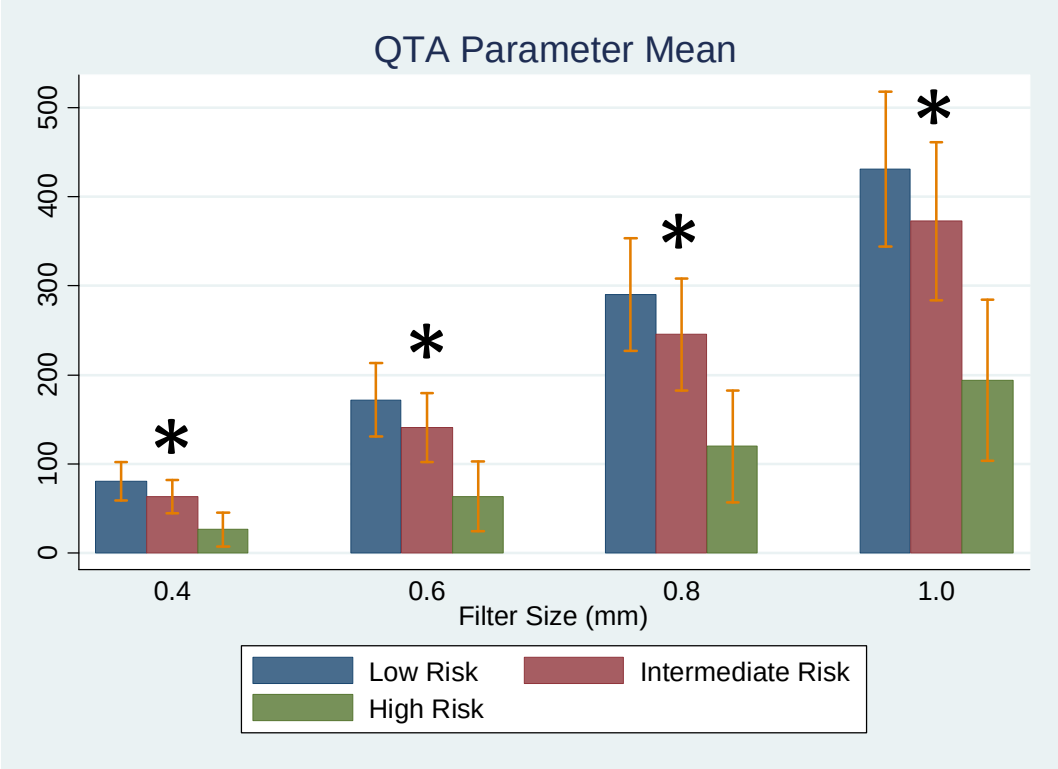

Figure 4 – Mean and standard deviation of QTA parameter Mean, at different RS risk groups (<18, 18-30, and >30) and filter levels (0.4-1.0 mm). Asterisk indicates statistically significant difference with p<0.05.

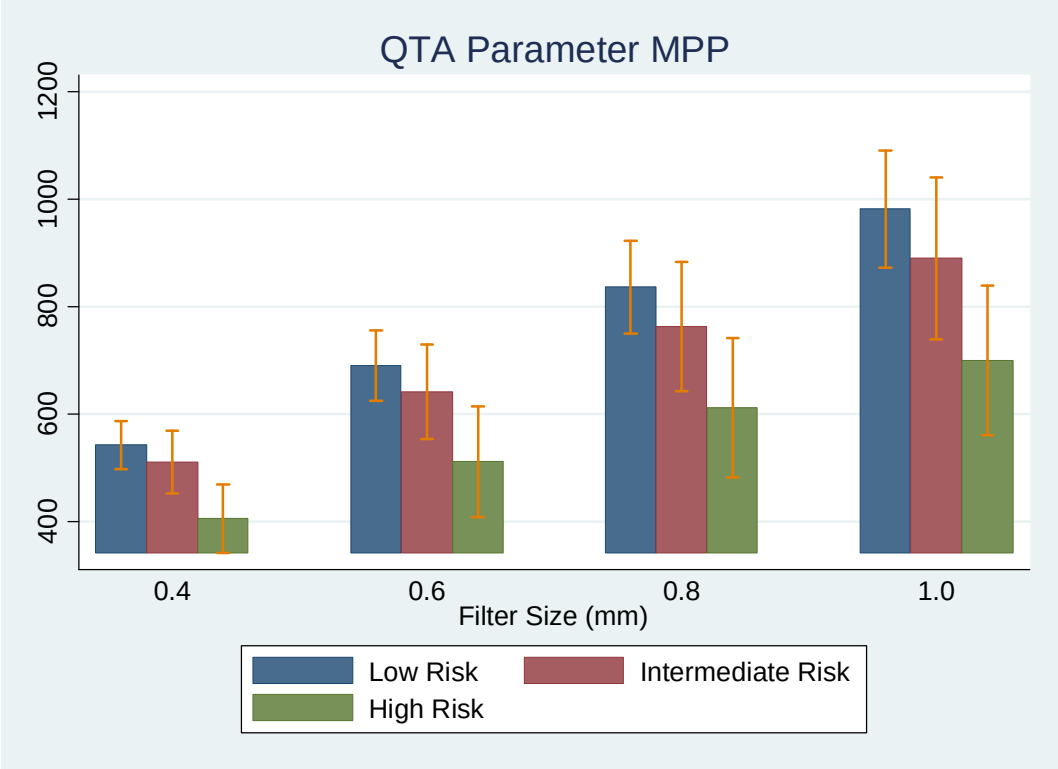

Figure 5 - Mean and standard deviation of QTA parameter MPP, at different RS risk groups (<18, 18-30 and >30) and filter levels (0.4-1.0 mm).

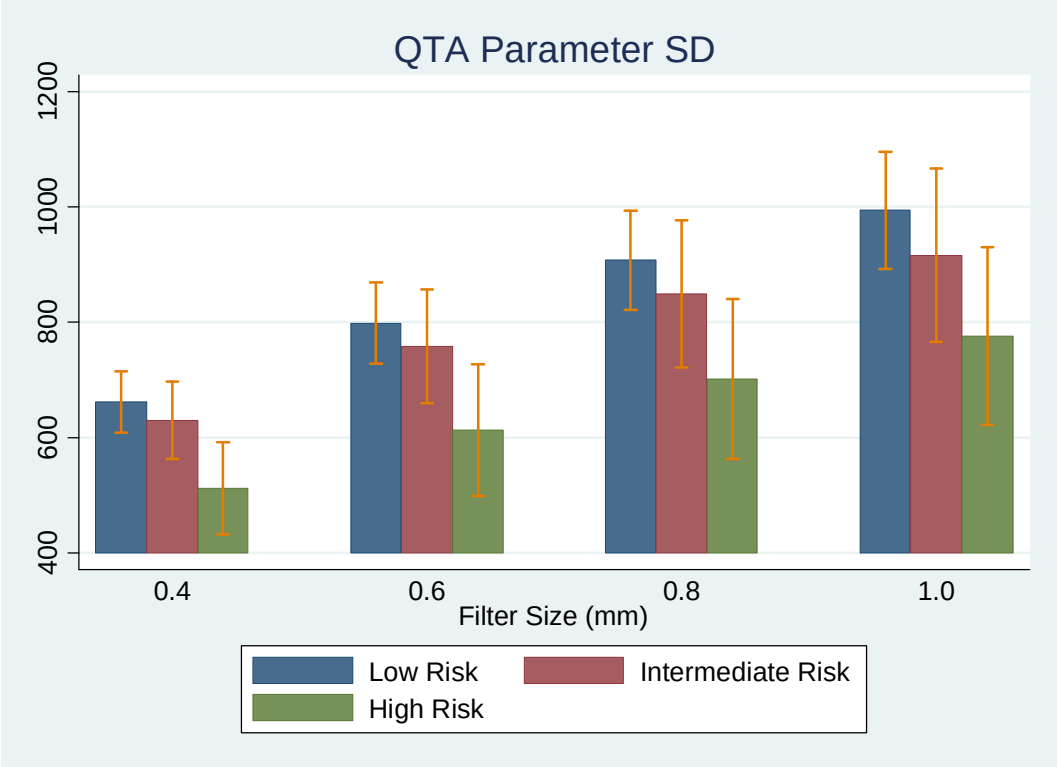

Figure 6 - Mean and standard deviation of QTA parameter SD, at different RS risk groups (<18, 18-30, and >30) and filter levels (0.4-1.0 mm).

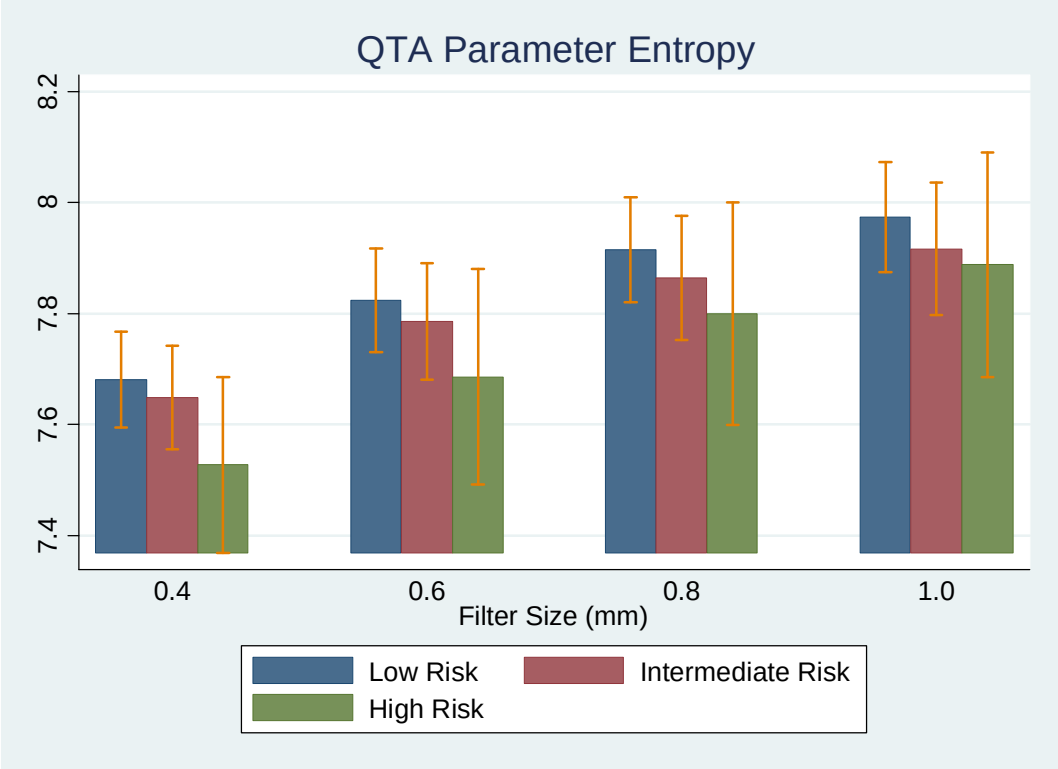

Figure 7 - Mean and standard deviation of QTA parameter Entropy, at different RS risk groups (<18, 18-30, and >30) and filter levels (0.4-1.0 mm).

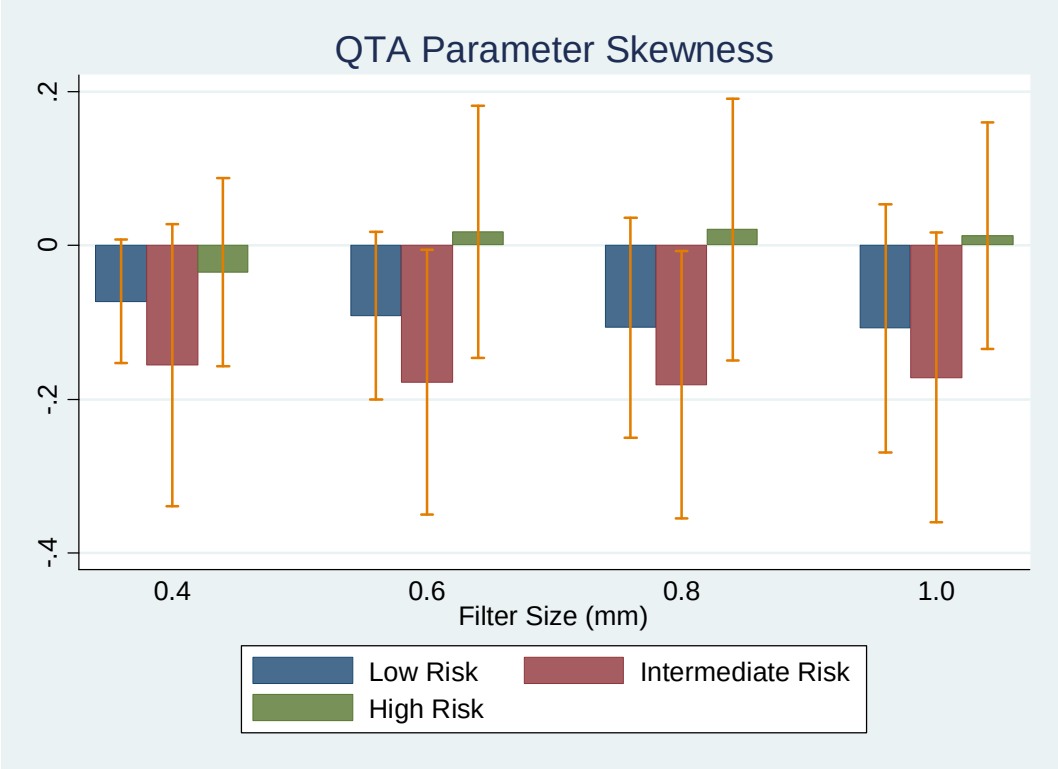

Figure 8 - Mean and standard deviation of QTA parameter Skewness, at different RS risk groups (<18, 18-30, and >30) and filter levels (0.4-1.0 mm).

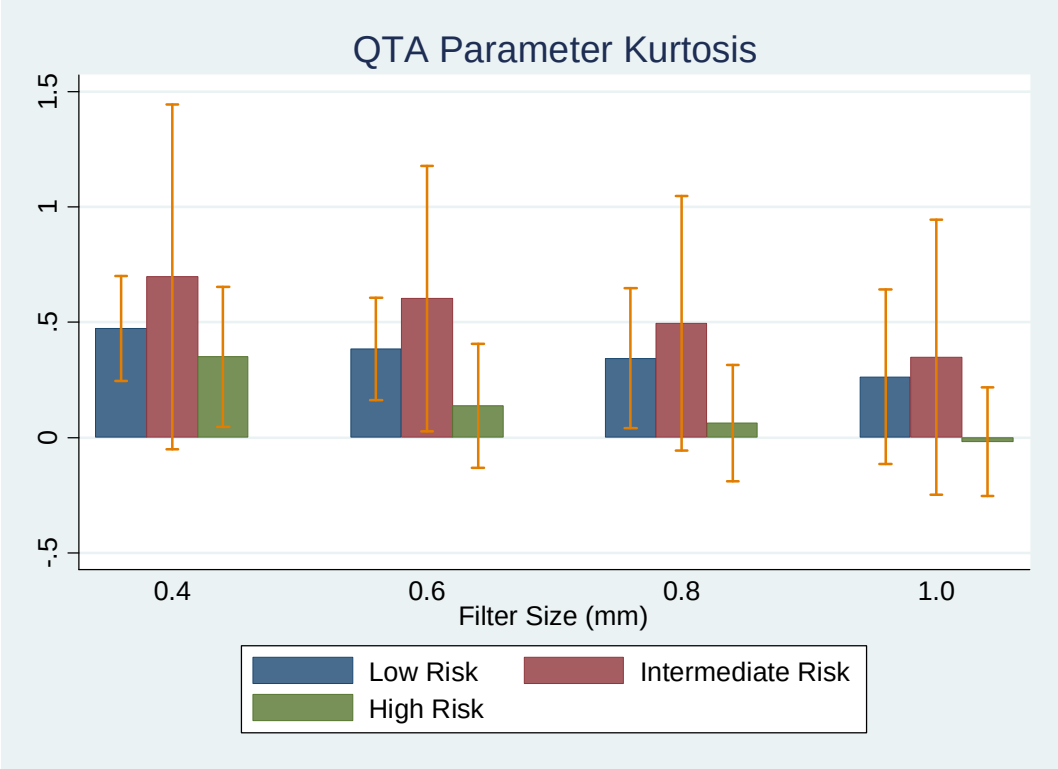

Figure 9 - Mean and standard deviation of QTA parameter Kurtosis, at different RS risk groups (<18, 18-30, and >30) and filter levels (0.4-1.0 mm).

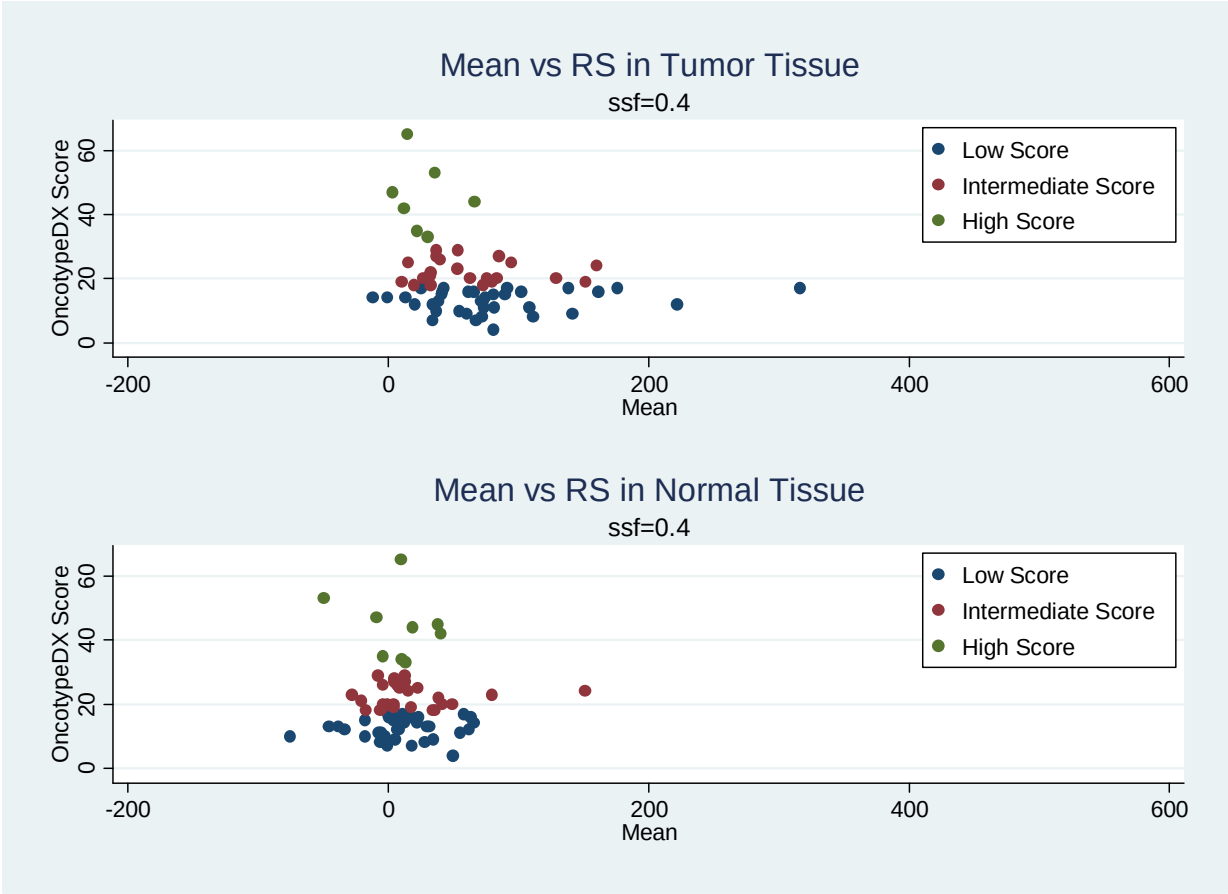

Figure 10 – Raw data of RS vs. QTA Parameter Mean in Tumor vs. Normal Tissues.

Linear Models of RS

We isolated QTA parameters that showed correlation with RS risk group using Spearman rank correlation test, then we tested for linear relationship between QTA parameters and RS. The following show the results from the linear modeling.

If PR status was unknown and therefore not included, the best-fit linear model had SD as the predictor at SSF=0.4 (n=65, F=6.89, p=0.0108, R^2=0.0870). Statistically significant linear models were also found at SSF=0.6 and 0.8. The value of PR in the linear models was either zero for PR- or one for PR+.

If PR status was known and included, the best-fit linear model had PR and Skewness-Diff as statistically significant predictors at SSF=1 (n=65, F=15.30, p<0.0001, R^2=0.2988). Statistically significant linear model was also found at SSF=0.4 (n=65, F=10.74, p<0.0001, R^2=0.3224).

Among PR+ patients, linear regression showed that QTA parameter Skewness-Diff is a statistically significant predictor of *log*(RS) at SSF=0.8 and 1 (At SSF=0.8, n=58, F=9.36, p=0.0034, R^2=0.1320. At SSF=1, n=58, F=7.25, p=0.0093, R^2=0.0770.). No statistically significant models for predicting RS among PR+ patients were found at SSF=0.4, 0.6, and 0.

The linear models are summarized in

Table 8.

Table 8 - Best-fit QTA-Based Linear Model In Predicting log(RS)

PR Status	SSF	Out-come	Predictors	Linear Model	n	F	p	R^2	Condi-tion Number
Not Included	0.4	*log*(RS)	SD	*log*(RS) = -0.0004177*(SD) + 1.500248	65	6.89	0.0108	0.0870	17.8116
Included	1	*log*(RS)	PR, Skewness-Diff	*log*(RS) = -0.3759396*(PR) + 0.0785932*(Skewness-Diff) + 1.592273	65	15.30	< 0.0001	0.2988	12.3685
PR+ Only	0.8	*log*(RS)	Skewness-Diff	*log*(RS) = 0.3033226*(Skewness-Diff) + 2.83791	58	9.36	0.0034	0.1320	14.7062

All three linear models showed no severe outliers and exhibited normalized residuals; they also succeeded model specification test and omitted variables test. However, their condition numbers exceeded 10 and less than 30, which indicated that the linear models were slightly unstable but not severely so.

Logistic Models of High Risk RS

We also want to see if there is logistic relationship between QTA parameters and RS risk group. The following show the results from the logistic modeling.

Statistically significant logistic models were identified with QTA parameters as explanatory variables in predicting the probability of being in the standard high risk group (RS > 30) or in the alternative high risk group (RS > 25). These models are shown in

Table 9.

Among PR+ patients, QTA parameter, SD-Diff, was shown to be a statistically significant predictor for the probability of having high risk RS with RS > 30 (n=58, chi^2=6.87, p=0.0087, AUC=0.9212, SE=0.0515). Also among PR+ patients, QTA parameters, Skewness and SD-Diff, were shown to be statistically significant predictors for the probability of the patient being in the alternative high risk group with RS > 25 (n=58, chi^2=9.68, p=0.0079, AUC=0.8814, SE=0.0453)

Table 9 - Best-fit QTA-Based Logistic Model in Predicting High Risk RS

PR Status	Pr	Predictors	Logistic Model	SSF	n	chi^2	p	AUC	SE
Not Included	RS>30	Entropy-Diff	p=*exp*(x)/[1+*exp*(x)] where x= -5.905371*(Entropy-Diff) –2.254296	0	65	10.98	0.0009	0.8424	0.0717
Not Included	RS>25	Mean-Total	p=*exp*(x)/[1+*exp*(x)] where x = -140.8101*(Mean-Total) – 0.2993403	0.6	65	9.98	0.0016	0.7437	0.0612
Included	RS>30	PR, SD-Diff	logit (p)= -3.548941*(PR) – 0.0092257*(SD-Diff) + 0.363196	0	65	18.69	0.0001	0.9409	0.0322
Included	RS>25	PR, Mean-Total	logit(p) = -4.321735*(PR) -159.0879*(Mean-Total) + 3.430188	0.6	65	25.56	<0.0001	0.8443	0.0591
PR+ Only	RS>30	SD-Diff	p=*exp*(x)/[1+*exp*(x)] where x = -0.0124166*(SD-Diff) – 3.547383	0	58	6.87	0.0087	0.9212	0.0515
PR+ Only	RS>25	MPP-Diff, Skewness-Diff	logit(p) = -0.0095748*(MPP-Diff) + 6.487719*(Skewness-Diff) – 2.761748	0.6	58	16.17	0.0003	0.9103	0.0482

Where *exp* denotes exponential, and

$$logit(Pr) = ln\frac{Pr}{1 - Pr}$$

Pr is the probability or likelihood of having high risk RS (RS > 30 or RS > 25), and $\frac{Pr}{1-Pr}$ is the odds of having high risk RS given the QTA parameters.

Note that when there is a single quantitative explanatory variable for predicting high risk RS, the logit function can be solved directly for *Pr* in that case.

All logistic models for predicting high risk RS succeeded the link test, Hosmer and Lemeshow's goodness of fit test, and the collinearity check.

Discussion

Several recent pilot and small studies have shown promising results in using Quantitative Texture Analysis (QTA) to characterize tumor biology in non-small cell lung cancer [16,17], colorectal cancer [18,19,22], non-Hodgkins lymphoma [23], and breast cancer [24]. Based on the results from those studies, we predicted that QTA can also be used to characterize breast tumor aggressiveness. We used TexRAD software platform to perform QTA and we chose Oncotype DX® Recurrence Score (RS) to be the quantitative measure of breast tumor aggressiveness. QTA was performed on digital mammograms at 4 filter levels (SSF=0.4mm, 0.6mm, 0.8mm, 1.0mm, and 0=no filter) and yielded 6 primary QTA parameters - Mean, MPP, SD, Entropy, Skewness, and Kurtosis – and 18 secondary QTA parameters, which are derived from primary QTA parameters and included the normalization, ranges, and tumor-minus-normal tissue difference of primary QTA parameters. RS has been shown to predict remote recurrence of node-negative and ER+ breast cancer in 10 years. Per standard convention, RS is categorized into three groups, high risk (RS > 30), intermediate risk (18 >= RS >= 30), and low risk (RS < 18), whereas an increasingly used clinical convention divides RS into high risk (RS > 25), intermediate risk (11 >= RS >= 25), and low risk (RS < 11).

First, we were able to observe correlation between RS risk group and QTA parameters, specifically membership in high risk group and QTA Mean. QTA parameter Mean was lower in high risk RS group than non-high risk RS group with statistical significance at all SSF levels except when SSF=0 (see Figure 4 and

Table 7). Since QTA Mean is a measure of necrosis, it means that high risk RS correlates with higher necrosis. As a control, we observed that there was no such difference in the Means between high risk RS and non-high risk RS groups in normal tissues (see Figure 10). Non-parametric analysis confirmed that multiple QTA parameters correlated with RS.

Next, we were able to derive models for predicting RS or RS risk group using QTA parameters as predictors. However, we found that PR status was a significant predictor in addition to the QTA parameters. Introduction of PR status as a predictor allowed for better fit linear and logistic regression models. In fact, there was a statistically significant difference in the mean RS between PR+ and PR- patients in our sample of 64 patients ($t=3.5146$, $p=0.0113$): 7 PR- patients had a mean RS of 38 and 57 PR+ patients had a mean RS of 17. This 21-point difference explains why including PR status increased the goodness of fit in both linear and logistic models. However, the number of PR- patients was small in this study, and a larger PR- sample size is recommended for future studies to confirm this result.

To understand the predictive power of QTA, we identified the best-fit models when PR status was included as a predictor, when PR status was not included as a predictor, and when subjects were PR+ patients only. We derived best-fit linear models that can predict RS (see

Table 8) as well as best-fit logistic models that can predict RS high risk group (see

Table 9), all with p-value < 0.02. Below are the findings:

- When PR status is unknown, QTA parameter SD can be used to predict RS (b=-0.0004177, t=-2.63, p=0.011) in a linear model at SSF=0.4. This inverse relationship between SD and RS seemed contradictory at first, as one would expect tumor heterogeneity to increase with tumor aggressiveness. However, recall that SD represents the distribution of the density distribution, and that the Means of high risk RS group were highly clustered on the low end of Mean spectrum, unlike the wide-spread distribution of the Means of the low and intermediate risk RS groups (see Figure 10). Hence SD is expected to be lower for high risk RS (i.e. more aggressive tumor).
- When PR status known, PR status (b=-0.3759396, t=-5.28, p=0.000) and QTA parameter Skewness-Diff (b=0.0785932, t=2.63, p=0.011) can be used together to predict RS in a linear model at SSF=1. Recall that Skewness measures the symmetry of density distribution, and hence as the symmetry difference between tumor and normal tissue increases (i.e. a positive skewness delta), RS would increase based on the model. This is in accordance with the previous report that positive Skewness may reflect angiogenesis.
- Analysis also showed that Skewness-Diff may be used alone in a linear model at SSF=0.8 to predict RS among PR+ patients.
- QTA parameter Entropy Diff can be used to predict the probability of having RS > 30. The ROC AUC for this model is 0.8424 (SE=0.0717), indicating that the discrimination of the model is moderately well. We propose a cut point of Entropy Diff=0.021 (sensitivity=85.71%, specificity=67.24%, LR+=2.6165, LR-=0.2125). Entropy Diff less than 0.021 in tumor tissue indicates a high probability of RS > 30.
- QTA parameter Mean-Total can be used to predict the probability of having RS > 25. The ROC AUC is 0.7437 (SE=0.0612), indicating that the discrimination of the model is only fair. We propose a cut point of Mean-Total=0.00769932 (sensitivity=75.00%, specificity=64.15%, LR+=2.0921, LR-=0.3897); a Mean-Total less than 0.00769932 in tumor tissue indicates a high probability of RS > 25.
- When PR status is known, PR status and SD-Diff can be used together to predict the probability of having RS > 30. The ROC AUC is 0.9409 (SE=0.0322), which means the

model can distinguish patients with RS > 30 well. PR status and Mean-Total can also be used together to predict the probability of having RS > 25. The ROC AUC is 0.8443 (SE=0.0591), which means the model can distinguish patients with RS > 25 moderately well.

- Even though QTA parameter Mean showed an interesting correlation with RS risk group in the initial correlation analysis (see Figure 4), Mean is not the QTA parameter that yields the best-fit linear or logistic model using backward step-wise elimination.

Future Directions

For future studies, we recommend that a larger set of breast cancer patients with RS be analyzed to validate the QTA-based models for predicting RS and RS high risk group as found in this study.

We also recommend whole-breast QTA processing, which is running QTA over the entire breast area including normal and tumor tissues, occult or visible. In this study, we performed QTA only on breast tumor lesions that had clear margins, in other words, we excluded mammogram-occult breast tumors. Even though whole-breast QTA processing would eliminate the six secondary QTA parameters derived from the difference between tumor and normal tissues, this alternative mode of QTA processing would eliminate lesion selection bias and operator subjectivity when running QTA, it would also broaden the applicability of QTA to include mammogram-occult breast tumors, therefore it is worth exploring in a future study.

Another possible future direction is to study the correlation between QTA parameters and microcalcification lesions, which was skipped in this study due to limited sample size.

Finally, it would be interesting to see if QTA can predict hormonal receptor status, including ER and PR status. Our study was unable to do due to the small number of subjects with ER- or PR-status. This was because we chose subjects with RS, which is recommended for breast cancer patients with stage I or II, node-negative, and estrogen receptor-positive breast tumors, or post-menopausal women with node-positive, hormone receptor-positive breast cancer.

Conclusions

We conclude that tumor heterogeneity in digital mammograms, as quantified by QTA analysis, shows promising opportunity in predicting breast cancer tumor aggressiveness, as measured by Oncotype DX® Recurrence Score (RS). Since our sample size was only moderate (n=65), there was insufficient number of subjects to perform validation of the models. However, given the statistically significant results in linear prediction of RS and in logistic prediction of high risk RS, we believe that QTA offers great potential in quantifying tumor aggressiveness in a non-invasive, real-time, cost-effective way. QTA analysis and its relationship to tumor biology should be studied further to derive clinically relevant application of QTA in breast cancer risk stratification and prediction of treatment response.

References

1. Youlden DR, Cramb SM, Dunn NA, Muller JM, Pyke CM, Baade PD. The descriptive epidemiology of female breast cancer: An international comparison of screening, incidence, survival and mortality. *Cancer Epidemiol*. 2012;36(3):237-248.
2. DeSantis C, Siegel R, Bandi P, Jemal A. Breast cancer statistics, 2011. *CA Cancer J Clin*. 2011;61(6):409-418. http://onlinelibrary.wiley.com/doi/10.3322/caac.20134/full
3. Ng CK, Pemberton HN, Reis-Filho JS. Breast cancer intratumor genetic heterogeneity: Causes and implications. *Expert Rev Anticancer Ther*. 2012;12(8):1021-1032.
4. Goldhirsch A, Wood WC, Coates AS, et al. Strategies for subtypes--dealing with the diversity of breast cancer: Highlights of the st. gallen international expert consensus on the primary therapy of early breast cancer 2011. *Ann Oncol*. 2011;22(8):1736-1747. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3144634/
5. Kramer-Marek G, Gijsen M, Kiesewetter DO, et al. Potential of PET to predict the response to trastuzumab treatment in an ErbB2-positive human xenograft tumor model. *J Nucl Med*. 2012;53(4):629-637. http://jnm.snmjournals.org/content/53/4/629.long#ref-23
6. Gonzalez-Angulo AM, Morales-Vasquez F, Hortobagyi GN. Overview of resistance to systemic therapy in patients with breast cancer. *Adv Exp Med Biol*. 2007;608:1-22.
7. Hammond ME, Hayes DF, Dowsett M, et al. American society of clinical Oncology/College of american pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version). *Arch Pathol Lab Med*. 2010;134(7):e48-72. http://www.archivesofpathology.org/doi/full/10.1043/1543-2165-134.7.e48
8. Rastelli F, Crispino S. Factors predictive of response to hormone therapy in breast cancer. *Tumori*. 2008;94(3):370-383.
9. Dowsett M, Cuzick J, Wale C, et al. Prediction of Risk of Distant Recurrence Using the 21-Gene Recurrence Score in Node-Negative and Node-Positive Postmenopausal Breast Cancer Patients Treated with Anastrozole or Tamoxifen: A TransATAC Study. *J Clin Oncol*. 2010; *28(11):1829-1834.*
10. Albain K, Barlow W, Shak S, et al. Prognostic and Predictive Value of the 21-Gene Recurrence Score Assay in Postmenopausal, Node-Positive, Estrogen Receptor-Positive Breast Cancer. *Lancet Oncology*. 2010; *11(1):55-65.*
11. Paik S, Tang G, Shak S, et al. Gene Expression and Benefit of Chemotherapy in Women with Node-Negative, Estrogen Receptor-Positive Breast Cancer *J Clin Oncol*. 2006; *24(23):3726-3734.*
12. Habel LA, Shak S, Jacobs M, et al. A Population-Based Study of Tumor Gene Expression and Risk of Breast Cancer Death Among Lymph Node-Negative Patients. *Breast Cancer Res*. 2006; *R25.*
13. Paik S, Shak S, Tang G, et al. A-Multigene-Assay-to-Predict-Recurrence-of-Tamoxifen-Treated-Node-Negative-Breast-Cancer. *N Engl J Med*. 2004; *2817-2826.*
14. Chicklore S, Goh V, Siddique M, Roy A, Marsden PK, Cook GJ. Quantifying tumour heterogeneity in (18)F-FDG PET/CT imaging by texture analysis. *Eur J Nucl Med Mol Imaging*. 2012.

15. Davnall F, Yip CS, Ljungqvist G, et al. Assessment of tumor heterogeneity: An emerging imaging tool for clinical practice? *Insights Imaging*. 2012. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3505569/
16. Ganeshan B, Goh V, Mandeville HC, Ng QS, Hoskin PJ, Miles KA. Non-small cell lung cancer: Histopathologic correlates for texture parameters at CT. *Radiology*. 2013;266(1):326-336.
17. Ganeshan B, Panayiotou E, Burnand K, Dizdarevic S, Miles K. Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: A potential marker of survival. *Eur Radiol*. 2012;22(4):796-802.
18. Ganeshan B, Miles KA, Young RC, Chatwin CR. Hepatic enhancement in colorectal cancer: Texture analysis correlates with hepatic hemodynamics and patient survival. *Acad Radiol*. 2007;14(12):1520-1530.
19. Ganeshan B, Ziauddin X, Goh VJ, et al. Quantitative imaging biomarkers from PET-CT as potential correlates for angiogenesis and hypoxia in colorectal cancer. European Society of Radiology 2012, Vienna, Austria.
20. Miles KA, Ganeshan B, Griffiths MR, Young RC, Chatwin CR. Colorectal cancer: Texture analysis of portal phase hepatic CT images as a potential marker of survival. *Radiology*. 2009;250(2):444-452.
21. Miles KA, Ganeshan B, Hayball MP. CT texture analysis using the filtration-histogram method: what do the measurements mean? *Cancer Imaging*. 2013;13(3):400-406. doi:10.1102/1470-7330.2013.9045.
22. Ng F, Ganeshan B, Kozarski R, Miles KA, Goh V. 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):177-184. http://radiology.rsna.org/content/266/1/177.abstract
23. Harrison LC, Luukkaala T, Pertovaara H, et al. Non-hodgkin lymphoma response evaluation with MRI texture classification. *J Exp Clin Cancer Res*. 2009;28:87-9966-28-87. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2711966/
24. Ganeshan B, Strukowska O, Skogen K, Young R, Chatwin C, Miles K. Heterogeneity of focal breast lesions and surrounding tissue assessed by mammographic texture analysis: Preliminary evidence of an association with tumor invasion and estrogen receptor status. *Front Oncol*. 2011;1:33. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3355915/
25. Holli K, Laaperi AL, Harrison L, et al. Characterization of breast cancer types by texture analysis of magnetic resonance images. *Acad Radiol*. 2010;17(2):135-141.
26. Li H, Giger ML, Lan L, et al. Computerized analysis of mammographic parenchymal patterns on a large clinical dataset of full-field digital mammograms: Robustness study with two high-risk datasets. *J Digit Imaging*. 2012;25(5):591-598.
27. Ahmed A, Gibbs P, Pickles M, Turnbull L. Texture analysis in assessment and prediction of chemotherapy response in breast cancer. *J Magn Reson Imaging*. 2012.
28. Alic L, van Vliet M, van Dijke CF, Eggermont AM, Veenland JF, Niessen WJ. Heterogeneity in DCE-MRI parametric maps: A biomarker for treatment response? *Phys Med Biol*. 2011;56(6):1601-1616.

29. Huber S, Medl M, Vesely M, et al. Ultrasonographic tissue characterization in monitoring tumor response to neoadjuvant chemotherapy in locally advanced breast cancer (work in progress). J Ultrasound Med. 2000;19(10):677-687. http://www.jultrasoundmed.org/content/19/10/677.long
30. Sparano JA, Paik S. Development of the 21-gene assay and its application in clinical practice and clinical trials. J Clin Oncol 2008;26:721-728.
31. Sparano J, Gray R, Makower D, et al. Prospective validation of a 21-gene expression assay in breast cancer [published online September 27, 2015]. N Engl J Med. doi: 10.1056/NEJMoa1510764. http://www.nejm.org/doi/full/10.1056/NEJMoa1510764#t=article
32. Stock J, Watson M. Heteroskedasticity-robust standard errors for fixed-effects panel-data regression. *Econometrica* 2008;76: 155-174.