Dynamic contrast-enhanced MRI-based biomarkers of therapeutic response in triple-negative breast cancer

Daniel I Golden, 1 Jafi A Lipson, 1 Melinda L Telli, 2 James M Ford, 2 Daniel L Rubin 1, 3

ABSTRACT

Objective To predict the response of breast cancer patients to neoadjuvant chemotherapy (NAC) using features derived from dynamic contrast-enhanced (DCE) MRI.

Materials and methods 60 patients with triple-negative early-stage breast cancer receiving NAC were evaluated. Features assessed included clinical data, patterns of tumor response to treatment determined by DCE–MRI, MRI breast imaging-reporting and data system descriptors, and quantitative lesion kinetic texture derived from the gray-level co-occurrence matrix (GLCM). All features except for patterns of response were derived before chemotherapy; GLCM features were determined before and after chemotherapy. Treatment response was defined by the presence of residual invasive tumor and/or positive lymph nodes after chemotherapy. Statistical modeling was performed using Lasso logistic regression.

Results Pre-chemotherapy imaging features predicted all measures of response except for residual tumor. Feature sets varied in effectiveness at predicting different definitions of treatment response, but in general, pre-chemotherapy imaging features were able to predict pathological complete response with area under the curve (AUC)=0.68, residual lymph node metastases with AUC=0.84 and residual tumor with lymph node metastases with AUC=0.83. Imaging features assessed after chemotherapy yielded significantly improved model performance over those assessed before chemotherapy for predicting residual tumor, but no other outcomes.

Conclusions DCE–MRI features can be used to predict whether triple-negative breast cancer patients will respond to NAC. Models such as the ones presented could help to identify patients not likely to respond to treatment and to direct them towards alternative therapies.

BACKGROUND AND SIGNIFICANCE

Of the 200 000 annual breast cancer diagnoses in the USA, approximately 15% are ‘triple negative’, lacking expression of estrogen and progesterone receptors, and lacking overexpression of the HER2/neu oncogene.1–3 Triple-negative breast cancer (TNBC) has an aggressive clinical course marked by high rates of metastases4 and poorer disease-specific survival than hormone receptor-positive subtypes, with a 5-year survival rate of approximately 75%.1–3 Unlike hormone receptor-positive or HER2/neu-positive subtypes, there is no specific treatment regimen that targets the triple-negative subtype. Although neoadjuvant chemotherapy (NAC) improves rates of breast conservation and survival,4–6 not all patients respond well to treatment,7 with 20–30% of TNBC patients achieving pathological complete response (pCR) at surgery,8, 9 an outcome strongly correlated with survival.9–10 Biomarkers, either molecular11 or otherwise, which predict treatment response to NAC are crucial for tailoring treatment to individual patients and enabling precision medicine.

Imaging may complement molecular methods of disease subtyping. Dynamic contrast-enhanced (DCE) MRI provides quantitative information on vascularity and the tumor micro-environment through serial acquisition of images before and after intravenous contrast agent injection.12 DCE–MRI has proved invaluable in the diagnosis and management of breast cancer.13–14 Pharmacokinetic modeling15 of DCE–MRI data can be used to calculate the kinetic properties of tumors in individual image voxels.16–17 Lesion-averaged kinetic parameters inferred from DCE–MRI have been shown to correlate with response to NAC18–21 and disease-free and overall survival22–25 in breast cancer.

One important class of DCE–MRI biomarkers that has shown promise in predicting breast lesion malignancy and response to NAC involves quantitative measures of tumor heterogeneity.26–28 Tumor heterogeneity is pervasive at the biological level,29 and in recent studies heterogeneity measures derived from DCE–MRI have shown significant correlation with breast lesion malignancy.30–38 Breast cancer type,39 response to chemotherapy in non-breast cancers,40–41 and long-term survival for patients with breast cancer42 and other cancers.43 Histogram-based measures of heterogeneity have also been explored in predicting response to chemotherapy in breast cancer,44–45 although these measures do not consider inter-pixel or inter-voxel spatial relationships.

These studies show that measures of spatial heterogeneity have unrealized potential in predicting the response to chemotherapy for breast cancer patients. One popular quantitative method of assessing lesion heterogeneity is by the gray-level co-occurrence matrix (GLCM),46 which is a method of quantifying the inter-pixel spatial relationships of an image. Features derived from the GLCM have been used successfully to predict breast lesion malignancy,11 12 34–38 to predict treatment response in limb sarcoma,31 and to predict certain breast imaging-reporting and data system (BI-RADS) terms, namely the presence of heterogeneous enhancement.38

In this exploratory study, we attempt to predict the response to NAC in early-stage TNBC patients using quantitative measures of spatial heterogeneity, measured by the GLCM, from DCE–MRI-derived lesion kinetic maps. To put our results into context, we also consider features derived from clinical data as well as from the semantic DCE–MRI BI-RADS47 and semantic DCE–MRI patterns of response.

(derived empirically from this dataset). BI-RADS descriptors in particular are known to be useful for characterizing breast cancer lesion malignancy, and they are usually assessed as part of the standard of care. A successful model to predict treatment response would allow therapy to be avoided in patients who would not benefit, thereby either informing the use of other therapies, or suggesting enrolment in clinical trials of novel therapies.

MATERIALS AND METHODS

Patient cohort
With institutional review board approval, de-identified patient data were drawn from a pre-existing, multi-institutional, prospective clinical trial of treatment-naive patients with early-stage triple-negative or BRCA mutation-associated (ie, showing defective homologous recombination of DNA) breast cancer. All patients received four or six cycles of platinum-based NAC. Of 93 enrolled patients, 75 patients had completed their enrolment in the trial at the time of our study and were considered for our analysis. Of these 75 patients, 15 patients were excluded from this study because they either had BRCA mutation-associated breast cancer but not TNBC, they had ended the trial early (eg, due to disease progression), they had more than one candidate primary lesion (eg, bilateral or multicentric disease), or outcome information was not available. The remaining 60 patients (table 1) were included in our analyses. Some patients were excluded from some analyses due to missing, incomplete, or poor-quality pre-chemotherapy or post-chemotherapy MRI.

For each patient, standard-of-care DCE–MRI imaging was performed before the initiation of chemotherapy but usually following the initial lesion biopsy. All imaging data used in this study were acquired between April 2009 and July 2012. All patients included in this imaging study underwent treatment (but not necessarily imaging) at our institution; patients enrolled in the trial who received treatment at other institutions were not included.

DCE–MRI imaging was performed at 25 different imaging centers, with a variety of different pulse sequences. The choice of scanner, sequence, and selected parameters within that sequence were selected according to each institution’s standard-of-care protocols. This large variation in scanning protocols initially appears as a disadvantage when grouping our patients together into a single cohort. However, the variation in scanning protocols has the advantage of highlighting the real-world applicability of our analysis method, as a successful model in this case will prove the robustness of our features to variations in DCE–MRI protocol.

Approximately 40% of pre-chemotherapy scans and 60% of post-chemotherapy scans were acquired at our institution. Dynamic sequences at our institution were typically either Spiral (a multisection, k-space-trajectory, gradient-echo sequence), or ‘differential sub-sampling with Cartesian ordering’ (DISCO, a three-dimensional, variable density pseudo-random k-space segmentation scheme). For each of these sequences, the period between acquisitions of the full image volume (ie, the time resolution) is approximately 10 s. For Spiral sequences, high-spatial-resolution three-dimensional spectral–spatial excitation magnetization transfer images were acquired before and after injection of the contrast agent. All scans at our institution had at least 19 time points, often with a gap between ‘wash-in’ and ‘wash-out’ series of between 6 and 12 min. At other institutions, the most common pulse sequence was GE VIBRANT (although the name of the pulse sequence could often not easily be determined), and time resolution ranged from 1 to 5 min (median 2 min, 46 s). All dynamic sequences had at least three time points (median five time points).

The contrast protocol at our institution consisted of 0.1 mmol of Magnevist (Gd-DTPA; Schering AG, Berlin, Germany) per kilogram of patient body weight injected at a rate of 2 mL/s; for Spiral scans, injection was begun 40 s after imaging began, and for DISCO scans, injection was begun at the beginning of the dynamic sequence, following the acquisition of several pre-contrast images. In general, the contrast protocols could not be determined for other institutions. For our quantitative analyses, when a contrast protocol was unknown, we assumed that the contrast agent was 0.1 mmol/kg body weight of Magnevist administered at a rate of 2 mL/s, injected midway between the pre-contrast and first post-contrast images. In practice, these assumptions about contrast agent protocol are not likely to affect our results because our investigation focuses on intra-tumor heterogeneity and the contrast protocol is invariant within a given tumor.

Clinical lesion features
Our initial feature set consisted of clinical features. All features from table 1 were included, including germline BRCA mutation.

<table>
<thead>
<tr>
<th>Table 1 Cohort information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characteristic</td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>SD</td>
</tr>
<tr>
<td>AJCC clinical stage</td>
</tr>
<tr>
<td>IA</td>
</tr>
<tr>
<td>IIA</td>
</tr>
<tr>
<td>IIB</td>
</tr>
<tr>
<td>III</td>
</tr>
<tr>
<td>AJCC grade</td>
</tr>
<tr>
<td>I</td>
</tr>
<tr>
<td>II</td>
</tr>
<tr>
<td>III</td>
</tr>
<tr>
<td>ER status (&gt;5%)</td>
</tr>
<tr>
<td>Positive</td>
</tr>
<tr>
<td>Negative</td>
</tr>
<tr>
<td>PR status (&gt;5%)</td>
</tr>
<tr>
<td>Positive</td>
</tr>
<tr>
<td>Negative</td>
</tr>
<tr>
<td>HER2 status</td>
</tr>
<tr>
<td>Positive</td>
</tr>
<tr>
<td>Negative</td>
</tr>
<tr>
<td>Germline BRCA1 status</td>
</tr>
<tr>
<td>Mutant</td>
</tr>
<tr>
<td>Wild type</td>
</tr>
<tr>
<td>Germline BRCA2 status</td>
</tr>
<tr>
<td>Mutant</td>
</tr>
<tr>
<td>Wild type</td>
</tr>
<tr>
<td>Variant of unknown significance</td>
</tr>
<tr>
<td>Number of treatment cycles</td>
</tr>
<tr>
<td>4 Cycles</td>
</tr>
<tr>
<td>6 Cycles</td>
</tr>
<tr>
<td>Ki67%</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>SD</td>
</tr>
<tr>
<td># Unknown</td>
</tr>
</tbody>
</table>

AJCC, American Joint Committee on Cancer; ER, estrogen receptor; PR, progesterone receptor.
status (obtained for all patients on a research basis) as well as T and N categories per American Joint Committee on Cancer (AJCC) TNM breast cancer staging. Per protocol eligibility, all patients had stage I–IIIA breast cancer with an index lesion 1 cm or greater by breast MRI. Categorical features were converted into a feature vector of binary dummy variables, with one dummy variable (‘present’ or ‘absent’, encoded as 1 or 0) per term. Ki67% (a protein whose increased expression is associated with poorer prognosis)⁴⁴ and patient age were included as continuous variables.

Semantic morphological features
Two different semantic morphological feature sets were determined for each patient. The first set consisted of descriptors from the MRI BI-RADS lexicon (standard-of-care morphological descriptions of lesion shape, margins, etc.) from the mass and non-mass-like enhancement categories. For each case, a board-certified, subspecialty fellowship-trained breast radiologist (JAL) described the morphology of the dominant lesion using the standard MRI BI-RADS lexicon, and these terms were converted into a feature vector of binary dummy variables.

We also defined another feature set consisting of five semantic ‘patterns of response’ for each lesion. These patterns, determined by the same radiologist, represent the change in appearance of the lesion on MRI before and after chemotherapy. These patterns were defined empirically from this image set, and they were not based on any previously published patterns. Categories include ‘progression’ (the lesion became larger or grew into the chest wall or skin), ‘no change,’ ‘regression with fragmentation’ (the lesion appeared to ‘break apart’ into two or more lesions with a decrease in net volume), ‘regression without fragmentation’ (the lesion decreased in size but did not fragment) and ‘resolution’ (the lesion was no longer visible) (figure 1). Response categories were represented as binary dummy variables in a feature vector. Like other features discussed here, our goal was to determine whether these MRI-based patterns of response could be used as biomarkers to model our gold standard pathological definitions of treatment response (discussed later).

Image pre-processing and lesion region of interest selection
For each patient, a representative two-dimensional slice containing the primary lesion was selected on the pre-chemotherapy and post-contrast DCE-MRI image with the help of the radiologist mentioned above. Representative slices were chosen that contained the largest tumor area with minimal artifacts (eg, due to biopsy markers). Pre and post-chemotherapy slices were not chosen to align precisely to the same position in a given patient because tumors often shrank non-uniformly, which would complicate alignment, and representative slices were by definition representative, which should obviate the need to choose one representative slice over another. All processing was performed on data from the representative slice using built-in functions and custom code from Matlab 2012a (MathWorks, Inc., Natick, Massachusetts, USA).

To examine heterogeneity at the pixel level, it was necessary to register the representative slices spatially across time within a given scan for each patient. Rigid registration was performed manually by choosing the values in a transformation matrix for each image. Registration quality was assessed by playing back a movie of the sequence and ensuring there were no ‘jumps’ in image landmarks, particularly near the lesion.

An empirical kinetic map based on three time points (eg, Degani et al)⁵⁵ was created from the representative slice. The three time points were defined as: (1) the known or assumed time of contrast injection; (2) either 110 s after the contrast injection, or the first post-contrast image, whichever was later; and (3) the time of the last image in the series that was no more than 20 min after contrast injection. This choice of time points did a good job of capturing the rise (‘wash-in’) and subsequent fall (‘wash-out’) of contrast amplitude in all of our dynamic imaging sequences (figure 2A). To display the empirical kinetic map, we used a hue saturation value color map, with ‘hue’ set proportional to wash-out, saturation set to 1, and ‘value’ proportional to wash-in. Therefore, regions of the image that do not enhance are nearly black, and those that enhance are colored by the rate of wash-out (figure 2B).

This empirical map provided strong contrast between the lesion and the surrounding tissue and allowed us to draw a
region of interest (ROI) confidently around the primary lesion. Although a radiologist assisted with locating the general region of the primary tumor, the actual ROI in the quantitative analysis were drawn without assistance from a radiologist. For non-mass-like lesions, ROI were drawn around the entire area of the lesion. ROI were drawn in this manner both on the pre-chemotherapy and post-chemotherapy data, when available. For patients with no visible residual tumor in their post-chemotherapy MRI, an ROI was drawn around non-enhancing tissue near where the tumor had previously been identified. All lesions’ ROI were at least 22 mm², with a mean pre-chemotherapy size of 594 mm² and a mean post-chemotherapy size of 212 mm².

Quantitative lesion features

Our quantitative lesion analysis is focused on assessing lesion heterogeneity by the texture of lesion kinetic maps. Our intuition was that heterogeneity in kinetic maps would reflect heterogeneity in the physiology of the tumors. Two classes of kinetic parameters were considered: empirical parameters and modeled pharmacokinetic parameters. Empirical parameters consisted of the three time points rates of wash-in, wash-out, and the area under the curve (AUC) between time points 1 and 3, inclusive. Pharmacokinetic parameters consisted of the transfer constant ($K_{\text{trans}}$), the rate constant ($k_e$) and the extravascular extracellular volume fraction ($v_e = K_{\text{trans}}/k_e$). These parameters were estimated for each voxel using a two-compartment pharmacokinetic model\textsuperscript{15} using the arterial input function from Tofts and Kermode\textsuperscript{56} based on data from Weinmann et al.\textsuperscript{57} $K_{\text{trans}}$ and $k_e$ values were estimated using the non-linear least-squares curve fitting function ‘lsqcurvefit’ in Matlab 2012a, and both parameters were constrained to be between 0.1 and 2.5/minute; $v_e$ was determined from the fitted $K_{\text{trans}}$ and $k_e$ values. All available time points for each patient were utilized in estimating the pharmacokinetic parameters. Because $T_{10}$ maps were not available for our data, we made the naive assumption that $T_{10}$ is equal to 1 s for all pixels. Probably because of this assumption, $v_e$ tended to take on non-physical values greater than 1. However, as our texture analysis examines intra-tumor heterogeneity, the absolute truth of our estimated parameters is not important as long as the analysis is consistent throughout the tumor. Example kinetic maps and modeled curves are shown in figure 3.

We used the GLCM\textsuperscript{46} to quantify lesion heterogeneity from the kinetic maps. Of the original 14 scalar GLCM parameters, we used the following four: the ‘angular second moment’ (energy), ‘contrast’, ‘correlation’, and ‘inverse difference moment’ (homogeneity). This subset of four parameters was chosen because it is built in to the ‘graycoprops’ function in Matlab 2012a. When applying the GLCM to our images, we included only pixels within the lesion ROI. We quantified the kinetic maps amplitudes to eight gray levels between the first and 99th percentiles of their full range and resized them using the Lanczos-3 kernel\textsuperscript{58} to a common resolution of 1.5 mm ($2.25$ mm$^2$/pixel, which is the coarsest resolution of the images used in this study. We generated the GLCM using the target pixel’s eight neighborhood; the resulting GLCM is symmetric. Given four GLCM properties and six kinetic maps (three empirical and three pharmacokinetic), we derived 24 GLCM properties for each lesion. For completeness, we also included the six lesion-averaged kinetic parameters. Finally, we included the lesion area within the ROI, giving a total of 31 GLCM-related features defined for each patient.

Modeling procedure

We define Boolean ‘response to treatment’ in four similar ways, based on histopathology of the resected tumor bed after therapy:
1. pCR
2. Residual invasive breast tumor (possibly with residual lymph node metastases)
3. Residual lymph node metastases (possibly with residual tumor)
4. Residual invasive breast tumor and lymph nodes metastases

pCR is defined as the absence of residual invasive carcinoma and lymph node metastases after therapy (in-situ carcinoma may be present). The presence of residual invasive breast tumor or residual lymph nodes metastases are both associated with poorer prognosis.\textsuperscript{59} These definitions of response are distributed well over our cohort (figure 4).

We established several different primary and secondary feature sets as follows:
1. Clinical/pathological/genetic features
   a. All clinical features (29 features, 46 patients)
   b. All clinical features except for Ki67 (28 features, 60 patients)

Figure 2  (A) Dynamic sequence showing lesion-averaged voxel intensity versus time, along with images at three time point time points. The window level does not change between images. (B) A three time point empirical kinetic map derived from the lesion dynamic sequence, colored by wash-in (value) and wash-out (hue). The lesion shown is the pre-therapy lesion from figure 1A. Access the article online to view this figure in colour.
2. Semantic morphological features
   a. BI-RADS (12 features, 59 patients)
   b. Patterns of response (five features, 57 patients)

3. Quantitative texture (GLCM) features
   a. Pre-chemotherapy (31 features, 53 patients)
   b. Post-chemotherapy (31 features, 46 patients)
   c. Pre and post-chemotherapy (31 features, 46 patients)

4. Combinations
   a. Pre-chemotherapy and BI-RADS (43 features, 53 patients)

Note that the number of features listed includes each dummy variable separately, so not all features are strictly independent (ie, of four clinical dummy variables representing AJCC clinical stage, namely, IA, IIA, IIB or IIIA, one and only one of those variables was set to one for each patient). We separated Ki67 from the other clinical features because not all patients had Ki67 assessed (table 1) and we wanted to avoid excluding those patients when assessing the other clinical features. Of these feature sets, only post-chemotherapy GLCM and patterns of response use information from after chemotherapy; the remaining feature sets use only pre-chemotherapy information.

Feature selection and modeling was performed by logistic regression with the Lasso algorithm using the 'lassoglm' function in Matlab 2012a, which operates as follows. Initially, for any given set of features (eg, clinical, BI-RADS, GLCM, etc.), all features are included as ‘candidate features’. As the regularization parameter, \( \lambda \), is increased, more feature coefficients are set to zero, yielding simpler models. The optimal feature subset is chosen as that which results from the largest value of \( \lambda \) (simplest model), which yields a regression model whose 10-fold cross-validated deviance is within one SE of the minimum model deviance over all values of \( \lambda \) (this is the ‘one SE’ rule, eg, section 7.10.1 of Hastie et al). Model performance was assessed using cross-validated receiver operating characteristic (ROC) curves. Because of the inherent randomness in choosing subsets during 10-fold cross-validation, 20 Monte Carlo repetitions were run, and the resulting ROC curves were averaged to assess the model performance.

RESULTS
Figure 5 shows the results of various models to predict treatment response. Model ROC AUC is shown with a blue triangle, and the 95% CI over all 20 Monte Carlo repetitions is indicated with a black line. The perfect model would have AUC=1, and a ‘random chance’ model would have AUC=0.5. We consider results with AUC<0.6 to be failed models and are labeled as ‘junk’ (in fact, results with 0.5<AUC<0.6 do perform better than random chance, but not well enough to have clinical value). A discussion of the individual features selected for each mode appears in supplementary appendix A (available online only).

We note several major results from figure 5. Overall, more feature sets were successful at predicting residual lymph node metastases (‘residual nodes’ in figure 5) or residual tumor and nodes than were successful at predicting pCR or residual tumor; this is despite that fact that all imaging features were derived from the primary tumor and not lymph nodes. Pre-chemotherapy GLCM features (GLCM Pre) were successful on their own at predicting pCR and residual nodes and they outperformed all feature sets at predicting residual nodes (except for the combination of GLCM Pre and BI-RADS, which had the same performance). GLCM features assessed after therapy (GLCM Post) were more successful than those assessed.

Figure 3  Pharmacokinetic maps for a single lesion, and estimated pharmacokinetic curves for three pixels. The lesion shown is the pre-therapy lesion from figure 1A. Note that values of \( V_e > 1 \) are unphysical, but they probably appear due to uncertainties in model inputs, particularly \( T_10 \). Access the article online to view this figure in colour.

Figure 4  Proportion of patients with residual invasive breast tumor, residual lymph nodes metastases, or both, after therapy. Access the article online to view this figure in colour.
before therapy (GLCM Pre) for all models except for those to predict residual nodes. At least one imaging feature set outperformed clinical features for every definition of response (although, for predicting residual tumor, no pre-chemotherapy feature sets were successful). Combinations of features (GLCM Pre and GLCM Post and GLCM Pre and BI-RADS) did not significantly outperform their individual constituent feature sets.

Patterns of response were not successful at predicting any outcome measures. If we define any pattern of response other than ‘resolution’ as a positive test result, and residual tumor by pathology as a positive condition result, we see that the patterns of response probably failed in the lasso model due to a high false positive rate (tumor was seen on MRI but not by pathology, figure 6). The sensitivity and specificity were 0.90 and 0.38, respectively.

DISCUSSION
Because of the wide variation in TNBC patients’ response to chemotherapy, predicting treatment is critical. The strength of DCE-MRI as a whole-tumor, minimally invasive, functional imaging method is unparalleled, and the results of this study strongly suggest that DCE-MRI-derived biomarkers, particularly those involving tumor kinetic texture, have a bright future in the management of breast cancer. For each definition of treatment response considered here, at least some feature sets were successful at predicting treatment response. Imaging features were more effective than clinical features, with the exception that in predicting residual tumor, the successful texture-based imaging features were assessed after chemotherapy.

The fact that patients received DCE-MRI scans at many different imaging centers, with many different (and often unknown) imaging protocols, initially appears to be a caveat in interpreting these results. However, we believe that this imaging protocol variation is instead indicative of the strength of our modeling results. The fact that we were able to model treatment response under these conditions indicates that our chosen features (especially BI-RADS and GLCM features) are robust to variations in both the implemented imaging sequence and...
uncertainties about the contrast protocol. Therefore, these imaging features may be used to predict treatment response without requiring standardization of DCE–MRI protocols across different institutions.

Because post-chemotherapy GLCM features were, by definition, assessed after the completion of the chemotherapy regimen, they are not themselves useful to redirect treatment regimens for patients who are not likely to respond. However, the favorable performance of these features to predict residual tumor is suggestive that GLCM features assessed during treatment may also be predictive of response. Further studies will be required to test this hypothesis.

An important result of our analysis is that, for different definitions of response, one or the other of BI-RADS or quantitative texture features (the GLCM) tended to be significantly more successful. This result indicates that morphological lesion characteristics and quantitative texture-based lesion properties are complimentary and both are relevant depending on the desired definition of treatment response.

A disappointing result was that the radiologist-determined ‘patterns of response’ were not useful at predicting any of our response categories, due to a high false positive rate. In this study, for many patients with no residual disease by pathology, small regions of enhancement could nonetheless be seen in the DCE–MRI images in the vicinity of the original lesion. This is consistent with previous work that showed that high false positive rates (low specificity) are systemic in the use of DCE–MRI to estimate residual disease after chemotherapy.62–64

One limitation of this study was that texture features were assessed in two dimensions instead of in three dimensions. We expect that extending this work to three dimensions may yield improved model accuracy, as has proved true when classifying breast lesion malignancy.35 In addition, the many different DCE–MRI imaging sequences and unknown contrast protocols used probably reduced the quality of some of our results; greater predictive power may be obtained if uniform imaging sequence and contrast protocol were used for all patients. However, the strength of our results in their current form suggests that our imaging features are somewhat invariant to the imaging details; this is important in real-world applications of our modeling approach, in which there will probably be variation in the DCE–MRI techniques across centers.

Models that use pre-chemotherapy imaging features like the ones shown here could be used in the future to help triage TNBC patients to standard NAC, or, if that regimen is predicted to be unsuccessful, to clinical trials of experimental treatments. Our methods of quantifying the cancer tumor phenotype in order to predict treatment response will probably prove useful not only in TNBC, but also in other breast cancers and in solid cancers of other organ systems.

CONCLUSION
In this study, we showed that semantic medical morphological DCE–MRI features and quantitative kinetic texture features to characterize lesion heterogeneity can be used to predict whether triple-negative patients will respond to NAC. Models of this form could be used to inform recommendations for likely non-responders to enable alternative therapies to be chosen. These models may also be useful in predicting treatment response in cancers beyond TNBC.

Correction notice This paper has been corrected since it was published Online First. The order of author names has changed. Daniel L Rubin has changed from the second author to the last author.

Acknowledgements The authors would like to thank Dr Nicholas Hughes for providing pharmacokinetic modeling software, Katie Planey for helping to download and de-identify DCE–MRI data, Pei-Jen Chang and Elizabeth Schackmann for help with the clinical database, and Dr Bruce Daniel for helpful discussions.

Contributors DIL helped conceive the project, performed the analysis and interpretation of the data, and drafted the manuscript and figures; he is the content guarantor. JAL determined BI-RADS descriptors and MRI ‘patterns of response’, aided in logical interpretation of images, and edited the manuscript. MLT and JMF led the clinical trial from which this study derives patient data, helped to define the motivation for the project, and edited the manuscript. DLR helped conceive the project, helped with interpretation of the data and edited the manuscript.

Funding This work was supported by NIH grants T32 CA009695 and U01 CA142555.

Competing interests All authors have completed the ICMJE uniform disclosure form at http://www.icmje.org/coi_disclosure.pdf and declare: support for the submitted work from the NIH (DIL, DLR), Bipar Sciences (JAL, MLT, JMF) and PERCOG (MLT), personal or institutional financial relationships with the following organizations in the past 36 months: Sanofi Aventis, Herbert and Shatterwhite, Facing Our Risk of Cancer Empowered (MLT), Bristol-Myers Squibb, The Breast Cancer Research Foundation, USA Department of Defense (JMF), American Society of Clinical Oncology, Susan G. Komen for the Cure (MLT, JMF), California Breast Cancer Research Program (DLR).

Ethics approval The study received ethics approval from the Stanford University Institutional Review Board.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES
Research and applications


