AWARD NUMBER: W81XWH-14-1-0474

TITLE: GENOMIC DIVERSITY AND THE MICROENVIRONMENT AS DRIVERS OF PROGRESSION IN DCIS

PRINCIPAL INVESTIGATOR: CARLO MALEY

CONTRACTING ORGANIZATION: ARIZONA STATE UNIVERSITY

REPORT DATE: **Oct 2019**

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
Oct 2019	Annual	09/30/2018 - 09/29/2019

4. TITLE AND SUBTITLE
Genomic Diversity and the Microenvironment as Drivers of Progression in DCIS

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-14-1-0474

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Carlo Maley

E-Mail: shelley.hwang@dm.duke.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Arizona State University
ORSPA
PO Box 856011
Tempe, AZ 85287

8. PERFORMING ORGANIZATION REPORT

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT The project is designed to test whether genetic and/or tumor environmental heterogeneity is a driving force in progression of breast DCIS. Our project, a collaboration between Duke and ASU, has made substantial progress on all 4 aims and we met our 60 month milestones. Primary achievements for 60 months are:
1) Continued Case and control identification 52 Pure DCIS & 48 adjacent DCIS with invasion) through extensive database and searching at Duke 2) Deep and comprehensive full exome sequencing for 100 cases from 30-160ng of DNA isolated from archival FFPE specimens, 3) Comparison of analytic methods to characterize somatic mutations from this full exome sequencing, 4) Application of sequencing data for copy number assessment 5) Development of dual immune-staining on DCIS lesions using 7 pairs of antibodies, 6) Imaging analysis of these stains, including quantitative analysis, 7) Identification of upstaged DCIS cases for the radiology aim, 8) Development of image analysis methods for digital mammograms, 9) Validation Aim (4) approval of the Duke IRB/ TBCRC038 protocol at 12 sites, including DOD approval to initiate collection of DCIS that either did or did not progress to invasive cancer, 10) Full integration of team members over the past year via frequent conferencing, face to face meetings, and constant communication. This multi-disciplinary progress puts our group into an ideal position to fully implement the aims of the project and reach our year 5 goals.

15. SUBJECT TERMS
DCIS, intra-tumor heterogeneity, genetic diversity, phenotypic diversity, somatic evolution, microenvironment, mammographic biomarkers

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					USAMRMC
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
U	U	U	UU	20	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

1. INTRODUCTION

Ductal carcinoma in situ (DCIS) of the breast is an increasingly common diagnosis that is related to aggressive screening patterns (mammography). This "pre-invasive" lesion may progress to invasive cancer, but does so at a relatively low frequency. Nonetheless, it is commonly treated with extensive surgery, radiation, and hormonal therapy even though most of these lesions would never progress to invasive cancer. Thus, there is a pressing clinical need to stratify the risk of DCIS tumors into those in need of intervention and those that can be safely monitored without intervention. Our project is designed to address this need by characterizing the evolvability of DCIS, detecting those that have a high likelihood of evolving to malignancy versus those that are likely to remain indolent.

2. KEYWORDS

DCIS, cancer progression, intra-tumor heterogeneity, genetic diversity, phenotypic diversity, somatic evolution, microenvironment, mammographic biomarkers

3. ACCOMPLISHMENTS

What were the major goals of the project?

Aim 1. *Determine whether genetic diversity of DCIS is greater in DCIS with adjacent invasive disease compared to DCIS without progression.* Diversity measures must be derived from geographically distinct areas of tumor. Genetic divergence of the DCIS component of tumors will be measured based on exome sequencing and SNP arrays run on two separate regions of the tumor, as well as normal tissue, in patients with DCIS either with or without adjacent invasion to determine the association between genetic diversity and progression to malignancy. Genetic diversity will be measured by the genetic divergence between the tumor samples, that is, the proportion of the genome that differs between the two samples from the same tumor.

60 Month Milestones:

- Protocol preparation, IRB submission and approval: **Completed** (Duke eIRB Pro00054515, initial Duke approval, 5/27/2014 and renewed for the current year), DOD IRB approval in place.

- Case identification and tissue block selection: Through a variety of available databases, we identified a large number of cases and controls with tissue available in the Duke Pathology archives. Each potential case and control requires extensive chart and pathology review in order to determine final eligibility and usability. For example, there is sufficient amount of the DCIS lesion (>2mm size) for isolation and DCIS is not too close to invasive cancer (it extends outside the invasive component). There must be two blocks with DCIS present that are >0.8cm apart. To date we have identified **100** cases, with pathology review.

- Sectioning of tissue blocks: **Completed**. New sections from candidate paraffin blocks are cut, stained to include one H&E at the beginning and end of each set and then reviewed by

the study pathologist. Remaining sections from candidate blocks (containing a sufficient amount of the DCIS lesion of interest) are used for macro-dissection and subsequent DNA extraction. Additional sections were also stored for immunohistochemical (IHC) analysis of key measures of tumor and micro-environmental heterogeneity. These slides are scanned for analytic and archival purposes. This process has been fully implemented and we are completing both cases and controls in this manner.

- DNA extraction of test cases: **Completed**.

- Exome sequencing of test cases: **Completed.** We chose the Genome Center at Washington University where cutting-edge methods for producing high quality data from these FFPE specimens have been developed and refined. Over the past three years, Wash U. sequenced 30-160ng from 300 individual DNA samples derived from 100 subjects (germ line sample plus 2 DCIS containing samples). They were able to derive interpretable sequence data (minimum of 40X depth at 50% coverage) from 30-160ng of FFPE DNA with qualities summarized in Figure 1, 2, 3, 4 and 5.

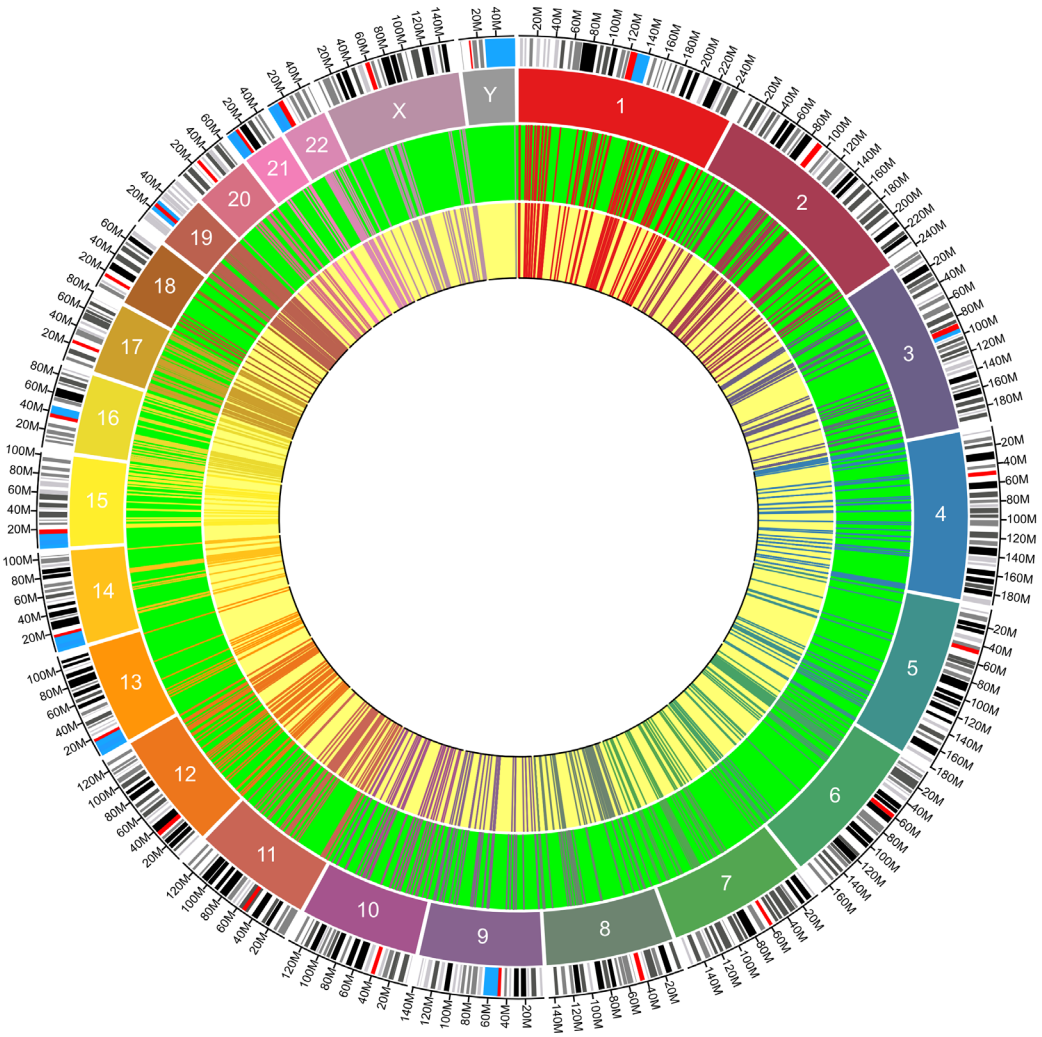

Figure 1: Exomic variants in Outer Track is genome; Middle Track is pure DCIS and Inner Track is adjacent DCIS.

- Development of a pipeline for identification of somatic genetic alterations: **Completed.** In order to assess and minimize artefacts induced by the FFPE procedure and the small amounts of DNA obtained from FFPE samples we developed a strategy based on 28 sequencing technical replicates. We used a 5-fold cross-validation procedure. We partitioned the patients into 5 complementary subsets. The patients were randomly assigned to the groups but each group is composed of 5 samples and each sample has a different amount of DNA: 20, 40, 60, 80, 100 ng. One subset (training set) was taken as hold out and evaluated against the rest of the patients (training data set). 5 rounds of cross-validation were performed using different partitions (Figure 2). Although our pipeline has been completed and is fully functional, we continue to work to improve it. These improvements have been statistically significant as seen in Figure 3, Improved SNV Bioinformatics Pipeline (Wilcoxon signed-rank test, p=0.008).

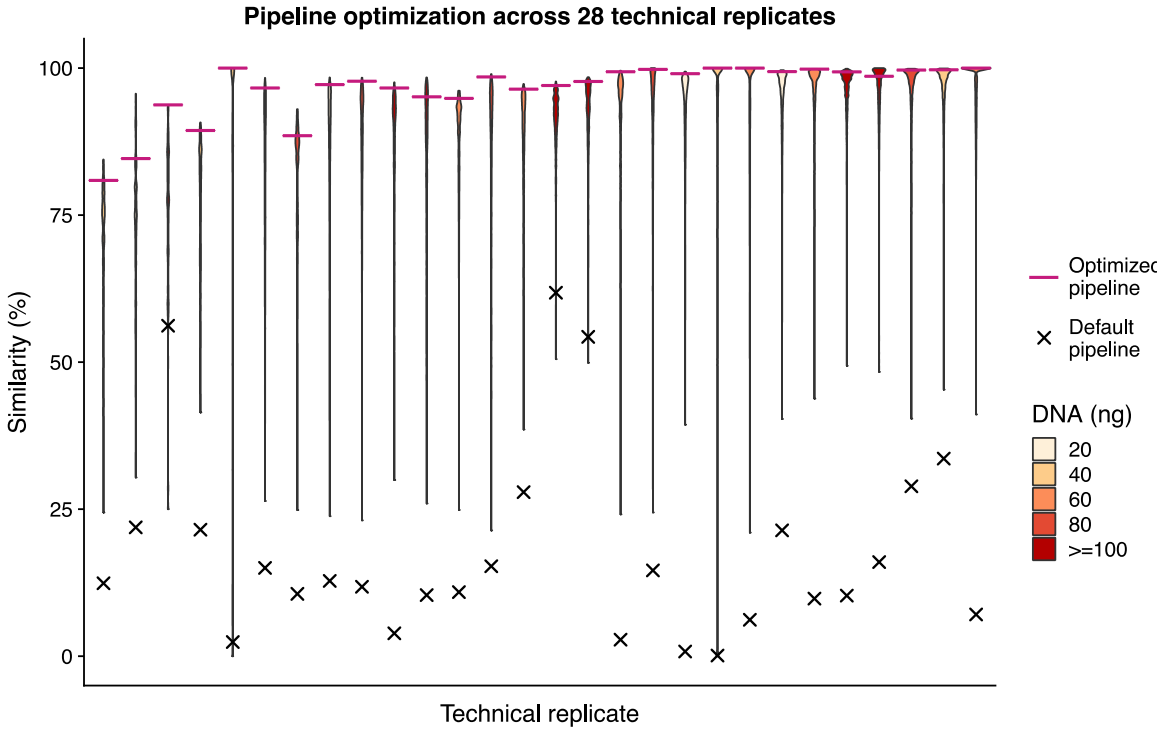

Figure 2: We developed an empirical method for optimizing the analysis algorithm through the comparison of sequences of the same DNA sample. We used a combination of filtering parameters on 28 different DNA samples extracted from 25 different patients. The same DNA analyzed twice with the same methodology should give the same results and detect the same somatic mutations in both analyses, ideally scoring 100% similarity. Of course, technical noise in the sequencing process interferes with achieving that ideal. After parameter optimization the similarity between the technical replicates was 96.8% ± 0.04 SD in average (x= similarity before optimization; — = similarity after optimization; colors indicate the amount (ng) of DNA used as template).

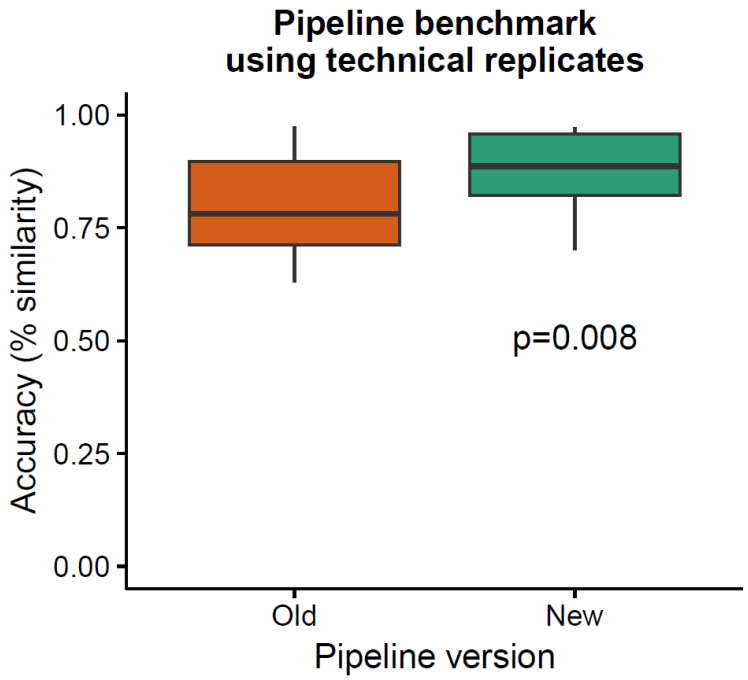

Figure 3: Improved SNV Bioinformatics Pipeline. We achieved a significantly better concordance between technical replicates of exon sequenced samples using our adjusted mutation calling pipeline.

- Calculation of genetic diversity scores for Aim 1: **Completed**. The percent of mutations different between two regions of the same tumor (called genetic divergence) is slightly (but not statistically significant) higher in DCIS adjacent to invasive disease than Pure DCIS (Figure 4, Exonic variants, Protein-coding variants, and Protein-altering variants). We found mutated genes in all patients. Current analysis of genetic diversity suggests that the genetic variability in DCIS adjacent samples was accumulated in the early phase of cancer development and then maintained during the subsequent tumor expansion. DCIS adjacent to invasive disease had slightly more mutations than Pure DCIS samples, but the difference was not statistically significant and could not be used for prognosis (Figure 5).

  We further analyzed the mutated genes to evaluate the molecular processes or signaling pathways that are deregulated based on Reactome (https://reactome.org) and DAVID (https://david.ncifcrf.gov) gene functional analysis. Both pure DCIS and DCIS adjacent to invasive disease have a statistically significant enrichment of immune-related pathways. The impairment of molecular mechanisms involved in immune molecular mechanisms could allow cancer cells to escape immune surveillance and cancer cells harboring these mutations could be to be positively selected in the tumor. DCIS adjacent to invasive disease showed an enrichment in taste of G-protein coupled receptors, TAS2Rs (Fold enrichment=7.4, p=0.031 after Benjamini correction, DAVID analysis). TAS2Rs expression is capable of inhibiting tumorigenicity. Thus, their impairment could contribute to DCIS progression. Mutations in TAS2Rs are candidates for validation as prognostic factors in Aim 4.

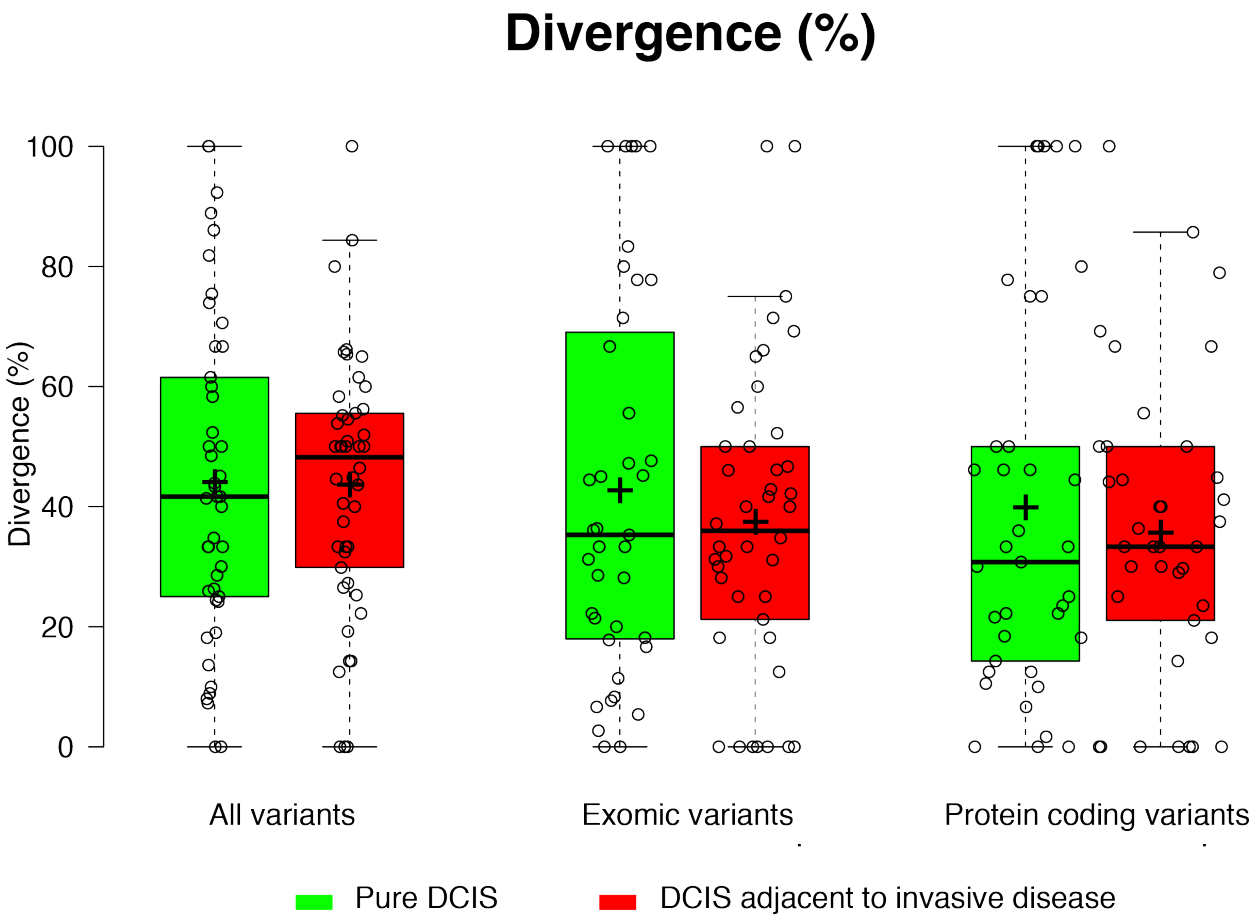

Figure 4: Exonic variants in Pure DCIS vs. DCIS adjacent to invasive disease comparing (% Divergence) the two regions sequenced from each case. Variants are categorized based on their protein coding consequences. There is no statistical difference between the two groups for any of the comparisons. For all variants, though the median is higher in the DCIS adjacent to the invasive cancers, the mean divergences (marked with +) are almost identical: pure DCIS=44.09±6.57 s.e.m., DCIS adjacent to invasive disease =43.61±6.43 s.e.m.).

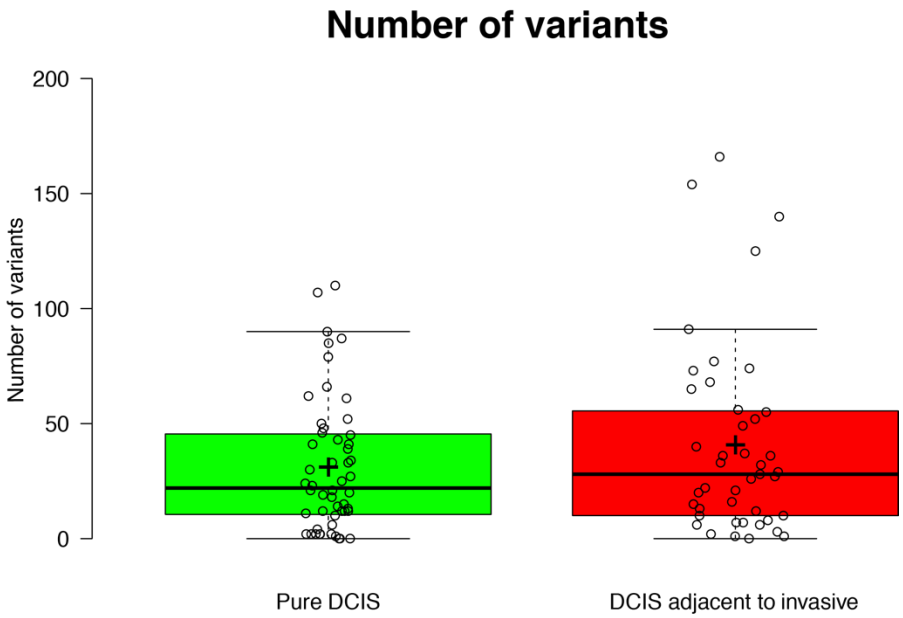

Figure 5: Mutational burden. The number of mutations in DCIS adjacent to invasive patients (n=47) is slightly higher than in Pure DCIS (n=52) but it is not statistically significant. The two population variances differ (F-test, p=0.02). Mean (+): pure DCIS=31.04±4.02 s.e.m., DCIS adjacent to invasive disease =41.90±6.39 s.e.m.).

Preliminary Conclusions: The degree of genetic divergence between regions does not seem to distinguish cases of pure DCIS from DCIS that is adjacent to invasive disease. Nor does the overall mutational burden. This suggests that the evolutionary dynamics do not significantly differ between DCIS that will or will not progress. Whether or not DCIS progresses may be a matter of which particular mutations they acquire (and differences in the ecologies – see Aim 2). This hypothesis is supported by the fact that DCIS adjacent to invasive disease accumulate statistically significantly more mutations in TAS2R genes, which may be acting as critical tumor suppressors in breast neoplastic progression.

Table 1: Cohort Demographics

	Pure DCIS (n=55)	Adjacent DCIS (n=61)
Age (mean)	56.9	57.5
Race		
White	32	42
Black	20	15
Other	3	4
Tumor Size (mean, cm)	3.5	4.6
Nodal Status		

Positive	0	27	
Negative	55	34	
Grade		*DCIS*	*Invasive*
1	1	1	11
2	24	22	27
3	30	37	22
Surgery			
Lumpectomy	24	38	
Mastectomy	19	23	
Estrogen Receptor			
Positive	39	42	
Negative	9	19	
Equivocal	0	0	
Progesterone Receptor			
Positive	33	38	
Negative	12	23	
Equivocal	3	0	
HER2 Status			
Positive	0	14	
Negative	0	45	

Aim 2. Determine whether phenotypic diversity of DCIS and the tumor microenvironment (TME) is greater in DCIS with adjacent IDC compared to DCIS without IDC. Since genomics is not the sole driver of tumor behavior, we will phenotypically characterize DCIS and its microenvironment including markers of hypoxia, migration, proliferation, matrix organization, and immune signaling in the same samples used in Aim 1. We will compute microenvironmental divergence to determine if specific components of the TME, or the divergence between TMEs from the same tumor, differs between DCIS with and DCIS without adjacent IDC.

In the past 12 months, we have analyzed our phenotypic diversity markers on a total of 85 cases (43 pure DCIS, 42 mixed invasive/DCIS, Table 1). To evaluate these elements, we have used a detailed expert scoring that captures the distribution of intensity of staining. This allows us to fully evaluate heterogeneity between regions of the cancer following the original study design and the genetic analyses.

This work is still in progress, both staining the last few cases and expert quantitation of the staining. However, we have performed an interim analysis of the data to date specifically with respect to heterogeneity in the DCIS component. We calculated earth mover distance, Manhattan distance, and Euclidean distance between the two areas that are genetically defined by exome sequencing (Aim 1). Similar results were obtained using the three computational methods for defining distance. Results shows that there is greater heterogeneity between regions in pure DCIS samples

for GLUT1 staining (glucose uptake) and FOXP3 (T-regulatory cells = immune suppression), but that DCIS adjacent to invasive disease had more heterogeneity between regions for CA9 (hypoxia), ALDH in the stroma ("ALDHS" stem cell marker) and COL15 (basement membrane, invasion). There is also evidence of an overall reduction in staining for CA9 and ALDH in the stroma of the DCIS samples that are adjacent to invasive disease. Reduction of ALDH expression in breast cancer stroma has been associated with poor survival. These are our most promising markers for validation in Aim 4.

We hypothesize that local tumor ecology for individual DCIS creates differential selective forces and ultimately influences its potential for progression to invasive cancers. To characterize the local ecological features for each DCIS component within the tissue, we first designed a deep learning pipeline for automated detection and simultaneous segmentation of DCIS. Comparison of multiple cutting-edge convolutional neural networks including SSD, faster RCNN, showed that MIMOnet was the most accurate in identifying and delineating individual DCIS.

To further explore the ecological features immediately adjacent to each duct, we used the topological context to investigate whether deep learning extracted useful image features from carcinoma in situ to learn the difference in biology between cases with DCIS adjacent to invasive cancers versus cases with pure DCIS (Figure 6).

Spatial tessellation centered at each DCIS created the boundary in which local ecology can be studied. Subsequently, a deep learning method was used to classify single cells into lymphocytes, epithelial, stroma cells and others.

These developments in methodologies enable us to quantify the spatial relationship between lymphocytes and DCIS. Our preliminary results indicate that, while pure DCIS cases have overall more lymphocytes, the lymphocytes in adjacent cases tend to co-localize with DCIS ($p<10^{-8}$), suggesting a more inflamed ecology locally to DCIS in tissue adjacent to invasive breast cancer. This is one of our most promising measures for prognosis that we will evaluate in the full cohort for Aim 2 and then validate in Aim 4.

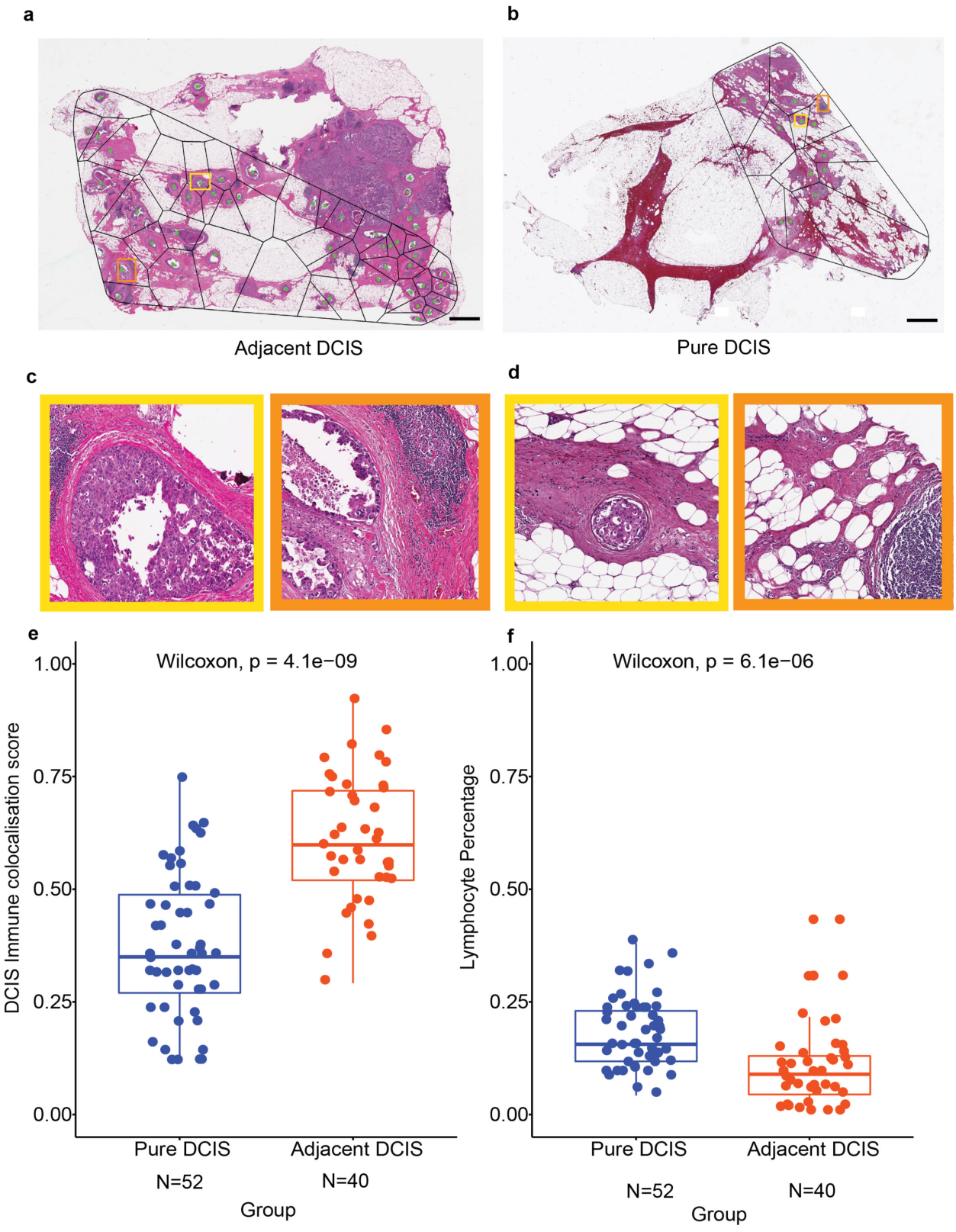
a
b
Adjacent DCIS
Pure DCIS
c
d
e
Wilcoxon, p = 4.1e−09
DCIS Immune colocalisation score
1.00
0.75
0.50
0.25
0.00
Pure DCIS
Adjacent DCIS
N=52
N=40
Group
f
Wilcoxon, p = 6.1e−06
Lymphocyte Percentage
1.00
0.75
0.50
0.25
0.00
Pure DCIS
Adjacent DCIS
N=52
N=40
Group

Figure 6: Comparison of TIL distribution pattern local to DCIS ducts in adjacent versus pure DCIS cases. (a) Voronoi tessellation of adjacent DCIS excluding invasive components (b) Voronoi tessellation of Pure DCIS. Scale bar represents 100□m. (c) Representative DCIS region enclosed within the voronoi of adjacent DCIS. (d) Representative DCIS region enclosed within the voronoi of Pure DCIS. (e) Boxplot illustrating the difference in DCIS immune colocalisation score calculated using the Morisita index. It was computed by associating individual DCIS duct with the surrounding lymphocyte within the voronoi region; a high score indicates the spatial colocalisation of lymphocytes and DCIS ducts. Each point corresponds to a WSI image, 52 WSI from n=40 patients in the pure DCIS and 40 WSI from n=25 patients in the adjacent DCIS group. (f) Box plot illustrating the difference in overall lymphocyte percentage in all cells for WSIs of pure DCIS and adjacent DCIS cases (after exclusion of invasive tumor regions), using only single-cell classifications.

60 Month Milestones:

- IHC staining of candidate markers on all cases **Close to completion (80%)**
- Expert scoring of all markers on all cases (50% **Completed)**
- Data analysis using distance metrics to determine which markers demonstrate significant heterogeneity that distinguishes pure DCIS from mixed DCIS/invasive cases **Completed**
- Stain Aim 4 cases with the most promising markers of tumor and microenvironmental heterogeneity
- Scan IHC and H&E stained slides for Automated image analysis (AIA) **Completed**
- Training and validation of AIA for the identification and enumeration of cell types (epithelial, stromal, lymphocytes, blood vessels). Computer algorithms are trained by expert identification of cell types (study pathologist, Allison Hall). Accuracy of the computer identification is evaluated by comparison back to the expert scoring. Apply methods for quantitative image analysis
- Test computer vision methods for measuring nuclear size as a surrogate for tumor grade

Aim 3. *Create and test a computational learning algorithm to compare mammographic characteristics and diversity measures in pure DCIS compared to DCIS with IDC.* A weighted computational algorithm using mammographic features of lesional and stromal characteristics as well as heterogeneity measures derived from Aims 1 and 2 will be constructed. The tool will be designed to allow for radiologic discrimination between good and poor prognosis DCIS, and will be evaluated in a validation set.

60 Month Milestones:

- We published in *IEEE Transactions in Biomedical Engineering* the study to improve the prediction of pure DCIS (negative) versus upstaged (positive) cases by leveraging the

adjunctive roles of two related classes, Atypical Ductal Hyperplasia (ADH) and Invasive Ductal Carcinoma (IDC). This study compared four different predictive models based on a suite of over 100 computer vision or "radiomics" features derived from the mammography images.

- We developed a new, deep learning model for segmentation of microcalcifications in mammograms. Our previous studies for this project were all based on computer vision algorithms, which are subjectively designed to fit a certain data set. Such "handcrafted" approaches do not generalize well to other data sets, such as mammograms from a different manufacturer. The new deep learning model is a U-Net that was trained only on magnification views from a single manufacturer (GE), but has so far demonstrated robust performance on not only magnification but also full-field digital mammograms, as well as being able to generalize between the two major manufacturers (GE and Hologic). An example is shown below in Fig 7.

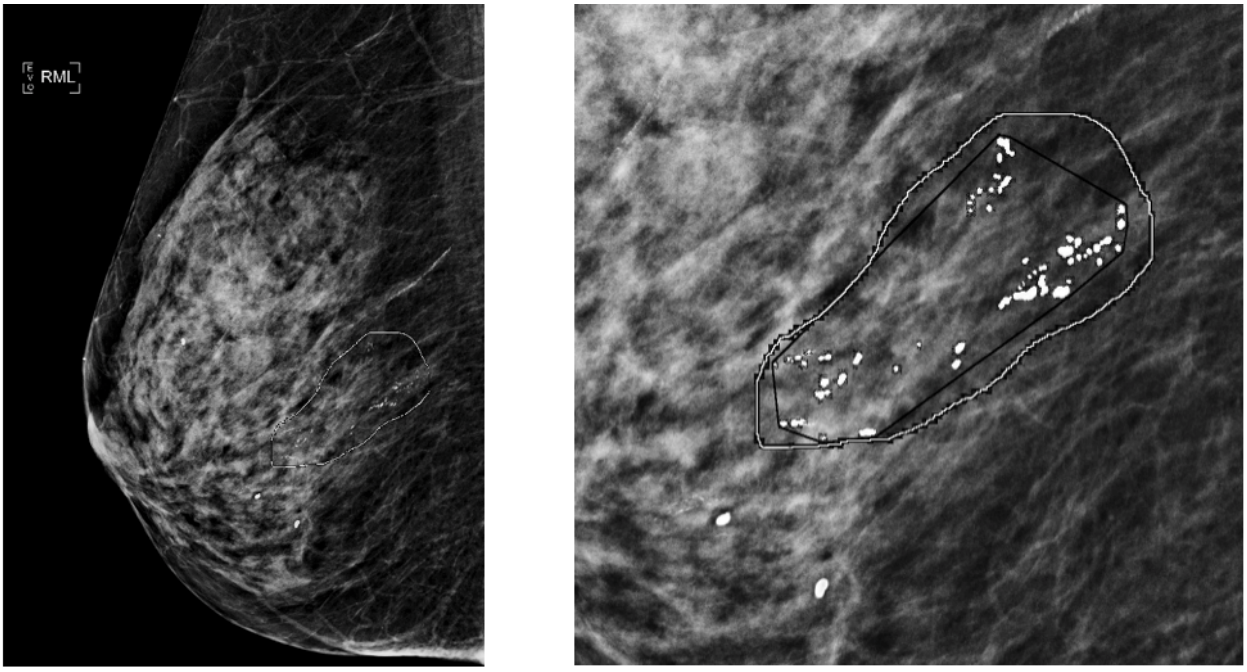

Figure 7: Example of microcalcification segmentation based on a U-Net deep learning model. Left: Mammogram with DCIS lesion annotated by radiologist. Right: Close-up of showing radiologist annotation as a white boundary, individual microcalcifications as white objects, and automated lesion boundary as a convex black polygon.

Dataset Version	Total Numbers	Pure	Upstaging	Upstaging Rate
Initial Study	137	105	32	23.4%
New Training Testing Split				
Training (59.54+-10.23)	400	334	66	16.5%
Testing (59.30+-10.97)	218	184	34	15.6%
ALL	**618**	**518**	**100**	**16.2%**

Table 2: Overview of radiology image data collection. The top portion (gray) shows the 137 cases used during the first 4 years of this project. We now have 618 total cases, which have been split in preparation for the final training and testing phase of the project.

- All studies in previous years were based on an initial 137 mammography cases while we focused our efforts on collecting all cases from our institution. We have concluded this effort now with 618 total cases (including the previous subset of 137 cases). We have randomly selected a cohort of 400 cases for training and validation, which have been matched for age and upstage rate. This data is approximately 3 times the size of our previous data, and will allow us to create more stable versions of the 4 types of logistic regression models recently published in the IEEE paper as well as to create new machine learning models using random forests.
- We have also investigated the potential to leverage radiologist interpretations of DCIS. We published a study in *Radiology* to model the growth dynamics of DCIS vs. benign lesions. We also conducted a large reader study involving ten radiologists each interpreting 150 DCIS cases for the task of predicting DCIS. Interestingly, radiologists performed comparably to our radiomics-based models with average AUC of 0.62. Their individual ROC curves and AUCs are shown in Fig 8. We conducted a follow-up reader study, where the same radiologists underwent a training process based on their collective experiences, defined new clinical criteria to predict upstaging, then applied those criteria to try to improve their performance on a new set of DCIS cases. That data is now undergoing statistical analysis.

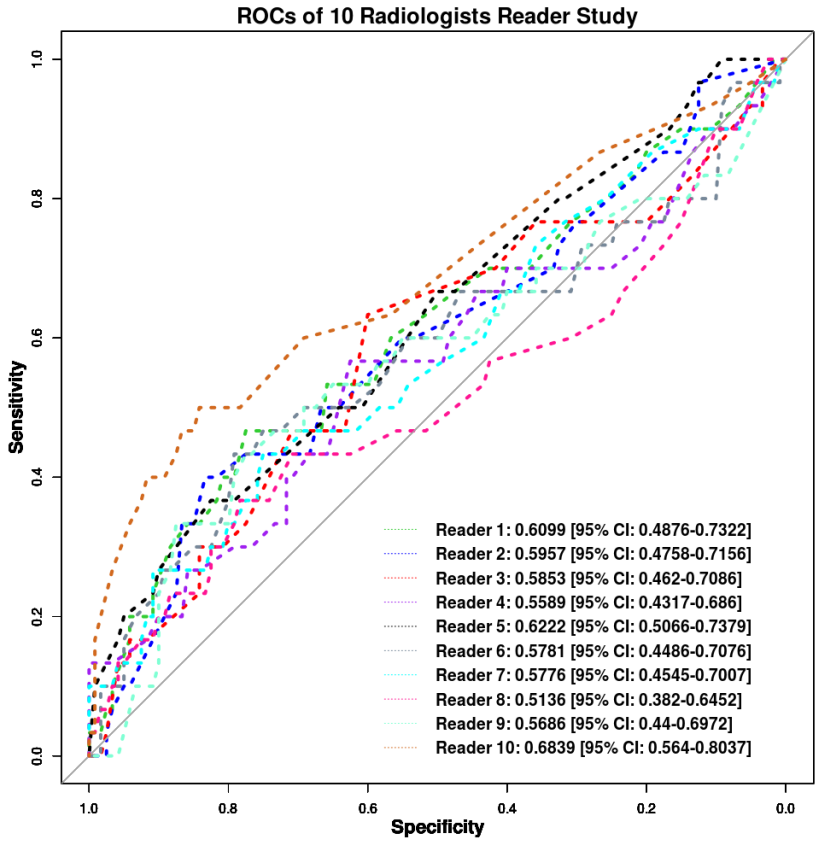

Figure 8: ROC results from radiologist reader study. Ten radiologists each interpreted 150 DCIS cases to predict upstaging. Average AUC was 0.620 (95% CI: 0.489-0.751).

- Future plans: We will also finish the collection of physician-derived features, including radiological, pathological, and clinical data. We will create new models based only on those physician-derived features, and also combine them with the imaging features for an hybrid models that use all available data. We will select the top few logistic regression and random forest models. After selecting a few best models, we will conduct independent testing on the reserved 218 cases. This will conclude the work for this project.

Aim 4. *Test the predictive performance of the best diversity measures in an independent validation set of pure DCIS with and without subsequent invasive recurrence.* Genotypic and phenotypic measures of diversity derived from Aims 1-2 will be applied to an independent case-control, longitudinal, tissue bank of DCIS with and without invasive recurrence to validate their utility. The Duke IRB approved protocol has been approved at 12 sites. For the next budget year, we will continue to accrue cases of pure DCIS that are long term disease free or recurred with invasive cancer. Slides are being shipped to Duke for macrodissection for DNA analysis and for immunodetection of phenotypic heterogeneity.

60 Month Milestones: This aim will be carried out after aims 1-3 are complete. We obtained approval to obtain these specimens through the Translational Breast Cancer Research Consortium (TBCRC) and Duke IRB approval. We have identified 12 high volume academic medical center members of the consortium who obtained regulatory approval, DOD approval and MTA's.

The REDCap database is used for data entry online and slide inventory control. To date we have obtained cases from 11 of the 12 sites. We have received 93 cohort 0, 68 cohort 1 and 57 cohort 2 cases to date. Overall, this aspect of the project is adhering to our proposed timeline and should achieve its accrual and analysis goals.

Below is the list of approved centers participating in this study and accrual to date.

Table 3: Multicenter Site Update for Aim 4

Site Name	PI	Cohort 0	Cohort 1	Cohort 2
Baylor	Julie Nangia	**0**	**2**	**3**
Chicago	Rita Nanda	**0**	**0**	**3**
DFCI	Tari King	**0**	**6**	**2**
Duke	ES Shelley Hwang	**71**	**26**	**17**
Indiana	Anna Maria Storniolo	**0**	**1**	**1**
Mayo	Fergus Couch	**5**	**5**	**7**
MDACC	Joanna Lee	**0**	**1**	**3**
Montefiore	Bryan Harmon	**10**	**13**	**6**
Pittsburgh	Priscilla McAuliffe			
UNC	Kristalyn Gallagher	**5**	**3**	**2**
UWashington	Mark Kilgore	**0**	**6**	**4**
UPENN	Angela DeMichele	**0**	**4**	**4**
UAB	Gabrielle Rocque	**2**	**1**	**5**
Total		**93**	**68**	**57**

What was accomplished under these goals?

Our primary goals have been met including, most importantly, identifying the most efficient method of sequence generation from small amounts of fixed DNA. We have acquired more radiology imaging data sets and established the computer vision algorithms for their analysis. Further, based on our databases, we are confident of accruing sufficient cases and controls at Duke to fulfill the Aim 1 and 2 goals of the project. Overall, we are in excellent position to complete the proposed work in the project period along the time line that was provided.

What opportunities for training and professional development has the project provided?

We hired several new post-doctoral fellows in the previous year to continue expanding our analysis. Priya Narayananhas and Faranak Sobhani acquired new skills in deep learning methods and attended a conference on breast cancer diagnosis.

How were the results disseminated to communities of interest?

We had a DCIS abstract based on aims 1 and 2 presented at the San Antonio Breast Cancer Symposium in December 2018.
Rui Hou was accepted for a talk at SPIE Medical Imaging 2018.

What do you plan to do during the next reporting period to accomplish the goals?

Aim 1: We have completed the identification, extraction and DNA sequencing for cases and controls (n=100). DNA extracted from these slides has been exome sequenced. Returned data from these assays is being analyzed using our current pipeline in order to scale up from the pilot study to a study with a bigger sample size, which will allow us to get more insights from the data. Moreover, we will investigate the biological meaning of the most common variants of the two different tumor types. These data will be prepared for publication during this period. In addition, we are preparing a manuscript to detail our methodologic approaches to sequence analysis focused on the technical replicates and the pipeline developed from these samples.

Aim 2:
We will complete the dual IHC staining on the remaining cases, as they come off line after pathology review. We will refine methods for agnostic computer scoring of IHC stains. These methods will be implemented on all images. Further, we will develop computer vision methods to measure nuclear size of the epithelial component. These methods have been developed by Dr. Yuan's team and are in testing phase. All cases will be analyzed for this parameter by Dr. Narayanan and Dr. Sobhani. .

Aim 3:
We will complete another paper describing the final results of the transfer learning of deep features. We will complete the analysis of the forced labeling study to improve classification by addition of neighboring classes, and submit that as an additional paper. We will then perform the majority of the final modeling studies using all cases from our institution, as well as begin to analyze cases from other institutions.

Aim 4:
This multicenter validation arm of the project is set up through the Translational Breast Cancer Research Consortium (TBCRC), a collaborative group set up to conduct innovative and high-impact breast cancer clinical trials.

The validation protocol has been approved by both the TBCRC and the Duke IRB (3/18/2016). Twelve (13 including Duke) external sites have obtained local IRB approval. Sites have both IRB as well as DOD approval and completed an SIV call training session with key personnel from each site,

We will finalize the collection of cases from sites. We have 218 cases in the RedCap database from all sites. We currently participate in monthly calls with TBCRC participating sites (12) where clinical coordinators, from all active TBCRC studies, provide updates and questions are addressed. These cases will be analyzed using the genetic, informatic, and phenotypic approaches developed in Aims 1 and 2. These data will constitute the validation of the results from the first two aims and will be prepared for publication.

4. IMPACT

Successful completion of this project will lead to a variety of biomarkers (genetic, IHC and radiographic) to distinguish high risk from low risk DCIS. This would reduce patient suffering and conserve clinical resources for the women with low risk DCIS, and focus management efforts and clinical resources on women with high risk disease, potentially justifying the risks of interventions. As the project is in its initial stages, these important impacts await in the future.

What was the impact on the development of the principal discipline(s) of the project?

We continue to advance the field's understanding of DCIS progression and the impact of tumor heterogeneity on the fate of DCIS. The final deliverables of this proposal will impact how DCIS is regarded both by the scientific and clinical communities.

What was the impact on other disciplines?

We have contributed to emerging knowledge regarding the digital radiographic characteristics of DCIS and continue to extend the applications for machine learning in breast cancer. We are one of the most active teams in the field, as evidenced by numerous publications and invited talks.

What was the impact on technology transfer?

Nothing to report.

What was the impact on society beyond science and technology?

Nothing to report.

5. CHANGES/PROBLEMS

Changes in approach and reasons for change

There have been no changes in approach.

Actual or anticipated problems or delays and actions or plans to resolve them

So far the problems that have emerged have been primarily technical. Full exome sequencing from small amounts of FFPE tissue has been the primary challenge, and is now proceeding smoothly at Wash U.

Changes that had a significant impact on expenditures

None

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents

None

Significant changes in use or care of human subjects

None

Significant changes in use or care of vertebrate animals.

Not applicable.

Significant changes in use of biohazards and/or select agents

Not applicable

6. PRODUCTS

Publications

1. J Hou R, Mazurowski MA, Grimm LJ, Marks JR, King LM, **Maley CC, Hwang ES**, **Lo JY**. Prediction of Upstaged Ductal Carcinoma in situ Using Forced Labeling and Domain Adaptation. IEEE Trans Biomed Eng. 2019. Epub 2019/09/11. doi: 10.1109/TBME.2019.2940195. PubMed PMID: 31502960.

2. Grimm LJ, Miller MM, Thomas SM, Liu Y, **Lo JY, Hwang ES**, Hyslop T, Ryser MD, "Growth Dynamics of Mammographic Calcifications: Differentiating Ductal Carcinoma in Situ from Benign Breast Disease," Radiology, 182599 (2019). Epub 2019/05/22.

3. Hou R, Ren Y, Grimm LJ, Mazurowski MA, Marks JR, King L, **Maley CC, Hwang ES**, **Lo JY**, "Malignant microcalcification clusters detection using unsupervised deep autoencoders," Proc. SPIE 10950, Medical Imaging 2019: Computer-Aided Diagnosis, 109502Q (13 March 2019).

4. Walther, V., Hiley, C.T., Shibata, D., Swanton, C., Turner, P.E., and **Maley, C.C.**: Can oncology recapitulate paleontology? Lessons from species extinctions. Nature Reviews Clinical Oncology, 12:273-285, 2015. doi:10.1038/nrclinonc.2015.12. Published. Acknowledged federal support.
5. Caulin, A.F., **Maley, C.C.:** Solutions to Peto's Paradox Revealed by Mathematical Modeling and Cross-Species Cancer Gene Analysis. Philosophical Transactions of the Royal Society of London B, 370 (1673):20140222. Published. Acknowledged federal support.

6. Aktipis, C.A., Boddy, A.M., Jansen, G., Hibner, U., Hochberg, M.E., **Maley, C.C.**, Wilkinson, G.S.: Cancer across the tree of life: Cooperation and cheating in multicellularity. Philosophical Transactions of the Royal Society of London B, 370 (1673):20140219. Published. Acknowledged federal support.

7. Noemi Andor, Trevor A. Graham, Marnix Jansen, Li C. Xia, C. Athena Aktipis, Claudia Petritsch, Hanlee P. Ji, **Carlo C. Maley**: Pan-cancer analysis of the extent and consequences of intra-tumor heterogeneity. Published. Nature Medicine 22:105-13, 2016. Acknowledged federal support.

8. **Carlo C. Maley**, Konrad Koelble, Rachael Natrajan, Athena Aktipis and Yinyin Yuan: An ecological measure of immune-cancer colocalization as a prognostic factor for breast cancer. Breast Cancer Research 17:1-13, 2015. Published. Acknowledged federal support.

9. Shi B, Grimm LJ, Mazurowski MA, Baker JA, Marks JR, King LM, Maley CC, Hwang ES, **Lo JY**, "Can Occult Invasive Disease in Ductal Carcinoma In Situ Be Predicted Using Computer-extracted Mammographic Features?" Academic Radiology, 24 (9), 1139-1147 (2017). PMC5557686. Published. Acknowledged federal support.

10. Shi B, Grimm LJ, Mazurowski MA, Baker JA, Marks JR, King LM, Maley CC, Hwang ES, **Lo JY**, Prediction of occult invasive disease in ductal carcinoma in situ using deep learning features, J Am Coll Radiol, accepted (2017). Acknowledged federal support

11. Shi B, Grimm LJ, Mazurowski MA, Marks JR, King LM, Maley CC, Hwang ES, **Lo JY**, Prediction of occult invasive disease in ductal carcinoma in situ using computer-extracted mammographic features, Proc. SPIE 10134, Medical Imaging 2017: Computer-Aided Diagnosis, Armato SG, Petrick NA, Eds., 101341I (2017). Published. Acknowledged federal support.

12. Shi B, Grimm LJ, Mazurowski MA, Marks JR, King LM, Maley CC, Hwang ES, **Lo JY**, "Can upstaging of ductal carcinoma in situ be predicted at biopsy by histologic and mammographic features?" Proc. SPIE 10134, Medical Imaging 2017: Computer-Aided Diagnosis, Armato SG, Petrick NA, Eds., 101342X (2017). Published. Acknowledged federal support.

13. Abegglen, L.M., Caulin, A.F., Chan, A., Lee, K., Robinson, R., Campbell, M.S., Kiso, W.K., Schmitt, D.L., Waddell, P.J., Bhaskara, S., Jensen, S.T., **Maley, C.C**.†, Schiffman , J. D.†: Potential Mechanisms for Cancer Resistance in Elephants and Comparative Cellular Response to DNA Damage in Humans. JAMA, 314:1850-1860, 2015. Published. Acknowledged federal support.

14. Li, X., Paulson, T.P., Galipeau, P.C., Sanchez, C.A., Liu, K., Kuhner, M.K., **Maley, C.C**., Self, S.G., Vaughan, T.L., Reid, B.J., Blount, P.L.,: Assessment of esophageal adenocarcinoma risk using somatic chromosome alterations in longitudinal samples in Barrett's esophagus. Cancer Prevention Research, 8:845-56, 2015. Published. Acknowledged federal support.

15. Kostadinov, R., **Maley, C.C**., Kuhner, M.K.: Bulk genotyping of biopsies can create spurious evidence for hetereogeneity in mutation content. PLoS Computational Biology, 12:e1004413, 2016. Published. Acknowledged federal support.

16. Andor, N., **Maley, C.C**., Ji, H. P. Genomic Instability in Cancer: Teetering on the Limit of Tolerance. Cancer Research 77:2179-2185, 2017. Published. Acknowledged federal support.

17. Tollis, M., Boddy, A. M., **Maley, C.C ,**. Peto's Paradox: How has evolution solved the problem of cancer prevention? BMC Biology 15:60, 2017. Published. Acknowledged federal support.

18. Fortunato, A., Boddy, A., Mallo, D., Aktipis, A., **Maley, C.C**. †, & Pepper, J. W. †: Natural Selection in Cancer Biology: From Molecular Snowflakes to Trait Hallmarks. Cold Spring Harbor Perspectives in Medicine, a029652, 2016. († = co-senior authors) Published. Acknowledged federal support.

19. **Maley, C.C**., Aktipis, A., Graham, T.A., Sottoriva, A., Boddy, A.M., Janiszewska, M., Silva, A.S., Gerlinger, M., Yuan, Y., Pienta, K.J., Anderson, K.S., Gatenby, R., Swanton, C., Posada, D., Wu, C.-I., Schiffman, J.D., Hwang, E.S., Polyak, K., Anderson, A.R.A., Brown, J.S., Greaves, M., Shibata, D.: Classifying the Evolutionary and Ecological Features of Neoplasms. Nature Reviews Cancer, Sept. 15, 2017. Published. Acknowledged federal support.

Technologies or techniques

Nothing to report

Inventions, patent applications, and/or licenses

Nothing to report

Other Products

Case report forms for Duke and outside cases and databases to efficiently capture this information

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Co-PI: Dr. Shelley Hwang (M.D., M.P.H.): Duke University (no change)
Co-PI: Dr. Carlo C. Maley (Ph.D.): Arizona State University (no change)

Co-Investigators:
Dr. Jeffrey Marks (Ph.D.): Duke University (no change)
Dr. Joseph Geradts (M.D.): Duke University (departed during year one)
Dr. Allison Hall (M.D.): Duke University, replacing Dr. Geradts.
Dr. Joseph Lo (Ph.D.): Duke University (no change)
Dr. Jay Baker (M.D.): Duke University (no change)
Dr. Yin Yin Yuan (Ph.D): Institute for Cancer Research, UK (no change)
Dr. Lars Grimm (M.D.): Duke University (no change)
Dr. Trevor Graham (Ph.D.): Barts Cancer Institute, Queen Mary University of London (no change)
Dr. C. Athena Aktipis (Ph.D.): Arizona State University (no change)
Dr. Shane Jensen (Ph.D.): University of Pennsylvania (departed during year one)

Post-Docs:
Dr. Mengyu Wang (Ph.D): Duke University (departed during year one)
Dr. Violet Kovacheva (Ph.D): Institute for Cancer Research, UK (departed during year two)
Dr. Narayanan (Ph.D): Institute for Cancer Research, UK, replacing Dr. Kovacheva.
Dr. Sobhani (Ph.D): Institute for Cancer Research, UK.
Dr. Lorraine King (Ph.D): Duke University (no change)
Dr. Bibo Shi (Ph.D): Duke University (departed during year two)
Rui Hou, ECE Ph.D. student, Duke University (no change)
Dr. Angelo Fortunato (Ph.D): Arizona State University (no change)
Dr. Diego Mallo (Ph.D): Arizona State University (no change)