Georgia State University

ScholarWorks @ Georgia State University

Biology Dissertations

Department of Biology

5-6-2019

Machine Learning Approaches to Predict Recurrence of Aggressive Tumors

Sergey Klimov

Follow this and additional works at: https://scholarworks.gsu.edu/biology_diss

Recommended Citation

Klimov, Sergey, "Machine Learning Approaches to Predict Recurrence of Aggressive Tumors." Dissertation, Georgia State University, 2019.
https://scholarworks.gsu.edu/biology_diss/219

This Dissertation is brought to you for free and open access by the Department of Biology at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Biology Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

MACHINE LEARNING APPROACHES TO PREDICT RECURRENCE OF AGGRESSIVE TUMORS

by

SERGEY KLIMOV

Under the Direction of Ritu Aneja, PhD

ABSTRACT

Cancer recurrence is the major cause of cancer mortality. Despite tremendous research efforts, there is a dearth of biomarkers that reliably predict risk of cancer recurrence. Currently available biomarkers and tools in the clinic have limited usefulness to accurately identify patients with a higher risk of recurrence. Consequently, cancer patients suffer either from under- or over-treatment.

Recent advances in machine learning and image analysis have facilitated development of techniques that translate digital images of tumors into rich source of new data. Leveraging these computational advances, my work addresses the unmet need to find risk-predictive biomarkers for Triple Negative Breast Cancer (TNBC), Ductal Carcinoma in-situ (DCIS), and Pancreatic Neuroendocrine Tumors (PanNETs). I have developed unique, clinically facile, models that determine the risk of recurrence, either local, invasive, or metastatic in these

tumors. All models employ hematoxylin and eosin (H&E) stained digitized images of patient tumor samples as the primary source of data. The TNBC (n=322) models identified unique signatures from a panel of 133 protein biomarkers, relevant to breast cancer, to predict site of metastasis (brain, lung, liver, or bone) for TNBC patients. Even our least significant model (bone metastasis) offered superior prognostic value than clinopathological variables (Hazard Ratio [HR] of 5.123 vs. 1.397 $p<0.05$). A second model predicted 10-year recurrence risk, in women with DCIS treated with breast conserving surgery, by identifying prognostically relevant features of tumor architecture from digitized H&E slides (n=344), using a novel two-step classification approach. In the validation cohort, our DCIS model provided a significantly higher HR (6.39) versus any clinopathological marker ($p<0.05$). The third model is a deep-learning based, multi-label (annotation followed by metastasis association), whole slide image analysis pipeline (n=90) that identified a PanNET high risk group with over an 8x higher risk of metastasis (versus the low risk group $p<0.05$), regardless of cofounding clinical variables. These machine-learning based models may guide treatment decisions and demonstrate proof-of-principle that computational pathology has tremendous clinical utility.

INDEX WORDS: Cancer biomarker, Deep learning, Machine learning, Digital image analysis, Recurrence prediction,

MACHINE LEARNING APPROACHES TO PREDICT RECURRENCE OF AGGRESSIVE TUMORS

by

SERGEY KLIMOV

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2019

Copyright by

Sergey Klimov

2019

MACHINE LEARNING APPROACHES TO PREDICT RECURRENCE OF AGGRESSIVE TUMORS

by

SERGEY KLIMOV

Committee Chair: Ritu Aneja

Committee: Yi Jiang

Arkadiusz Gertych

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2019

DEDICATION

This work, and my degree in its entirety, is dedicated to the people who stood by me through these long 5 years. Every quality which has made me the researcher I am, I owe to them.

My mother-strength.

My grandfather- inspiration.

My grandmothers- Warmth.

My brother-Curiosity.

Roma-Integrity.

My Uncle and Aunt-Humor.

My sisters-Hope.

Christal-My rock.

"When you arise in the morning, think of what a precious privilege it is to be alive - to breathe, to think, to enjoy, to love." - Marcus Aurelius

Acknowledgments

I could never possibly acknowledge all the people who have helped me get this far, but I will try. First, I must thank my Advisor Dr. Ritu Aneja, who somehow found potential in a directionless master's student so many years ago in Taiwan. Her support has helped me become a confident scientist. She allowed me to grow, make mistakes, and showed me how to find my own path. I would also like to thank my co-advisor from the mathematics/statistics side, Dr. Yi Jiang, who has helped me to blossom in so many multidisciplinary fields. I'd like to think her constant scientific curiosity has rubbed off on me. Finally, my last committee member, Dr. Arkadiusz Gertych, who has inspired me to pursue so much of the work in this thesis and has always made himself happily available for all my silly academic questions. Last, but certainly not least, I have to thank ALL the lab members of Dr.Aneja and Dr.Jiang, with whom I have shared days both happy and sad. I would specifically like to acknowledge Karuna Mittal, Shristi Bhattarai, Dr. Padmashree Rida, Howard Smith, Hao Chen, Xiuxiu He, Bin Zhang, Bikram Sahoo, and Timothy Fisher.

Table of Contents

List of Tables

List of Figures

1 Introduction and Literature Review

Despite tremendous research, activism, and improvements in screening and therapy, cancer still represents either the first or second leading cause of death for almost all age and sex groups in the United States [1]. However, the cause of the overwhelming majority of solid tumor deaths are not due to the primary cancer, but instead from the cancer recurrence, specifically metastasis [2]. The process of metastasis, and to a smaller intent local recurrence, is complicated and not completely understood. In no sequential order, cells have to pick up an aggressive phenotypes through epi/genetic instabilities, gain prerequisite properties such as self-renewal and motility, push the surrounding microenvironment towards a more invasion supportive role, gain the ability to enter, and survive, blood vessels (and change its phenotype to mesenchymal like within it), be able to 'home' into targets outside the blood vessel, extravasate toward them, and finally (and perhaps most complexly) have to be able to colonize, proliferate, and survive in a vastly different microenvironment [3]. Unsurprisingly, confidently predicting patients whose cancer will undergo this perfect storm, whose cancer will spread or comes back, has been an exercise in futility.

There are worryingly few prognostic markers actually used clinically in the overall landscape of cancer [4] and, paradoxically, the relative number (versus published rate) of biomarkers approved by the FDA has actually been going down [5]. While standard of care clinical variables, such as size and lymph node invasion, have a logical interpretations of metastatic risk, in reality, they are poorly prognostic [6] and ultimately can only accurately predict recurrence risk in around 30% of patient [7]. Although the potential is enormous, the major hurdle to gain clinical acceptance is to show utility in treatment options and move towards personalized healthcare [5]. Herein my research focuses on providing biomarkers towards predicting metastasis or recurrences, for three cancers with significant uncertainty in therapy decision making.

1.1 Triple Negative Breast Cancer and Site of Metastasis

Breast cancer is an extremely heterogeneous disease. A classical method to 'group' breast cancers (based on the underlying biology and treatment options) is through molecular subtypes. Classically, for breast cancer, they are often determined by the presence, or absence, of 3 key receptors: estrogen receptor (ER), progesterone receptor (PR), and hormone epidermal growth factor receptor 2 (HER-2) [8]. Distinct combinations of these receptors constitute different molecular subtypes, which often have differences in cancer biology and metastatic risk [9]. A particularly worrisome subtype is one which lacks each receptor, termed "Triple Negative Breast Cancer", or TNBC. TNBC's have extremely aggressive clinical behavior, with more than two times the risk of metastasis compared to other subtypes [10], and, importantly, to sites which are more likely to ultimately lead to death [11, 12]. If the propensity of a TNBC cancer to metastasize to a specific organ could be identified, then specialized preemptive, site-specific, therapies (such as through the use of RANKL antibodies if the cancer is primed for bone metastasis [13]) could be used to significantly improve patient outcomes.

1.2 Ductal Carcinoma in Situ and Recurrence

Ductal Carcinoma in Situ (DCIS) is a non-obligate precursor of invasive breast cancer where the malignant cells are confined to the lumen of a mammary duct by an intact outer myoepithelial layer and basement membrane [14]. Despite the similarity of their genetic profile [15], risk factors [16], morphology, and even the degree of heterogeneity [17] to invasive ductal carcinoma, untreated DCIS progresses to invasive disease only in ~40% of cases, sometimes only after decades [18-20]. However, this risk diminishes to ~10-20% after treatment [21]. This invasive recurrence is the most significant end-point for DCIS clinically, and yet it lacks accurate prognostic markers [22, 23]. While breast conserving surgery (also known as a lumpectomy) is the most common surgical treatment for DCIS [24], there is no consensus regarding the use of adjuvant radiation therapy, even for low grade or small DCIS tumors [25]. Without a clear understanding of

recurrence risk, clinicians are forced to balance, potentially significant, side effects associated with radiation [26-29] without knowing if it will provide marginal, if any, benefit towards suppressing recurrence [30, 31]. Ultimately, this results in many patients either over (i.e. unnecessary radiation) or under (i.e. omitted radiation when they will in fact recur) treated [32]. While attempts have been made to develop biomarkers to determine patient risk, and in turn adjuvant radiation decision making, they have lacked cost effectiveness [33], have been inconsistent [23], or simply not good enough [34].

1.3 Pancreatic Neuroendocrine Tumor and Metastasis

Despite overall trends showing a decrease in cancer deaths, pancreatic cancers death rates are alarmingly climbing [1]. Pancreatic neuroendocrine tumors are described as epithelial derived tumor cells with neuroendocrine differentiation. Through constant evolution, the World Health Organization (WHO) has described the two most common types, depending on differentiation and proliferation, as well-differentiated pancreatic neuroendocrine tumors (PanNETs) and poorly differentiated pancreatic neuroendocrine carcinomas (PanNECs)[35]. Although it current accounts for about 2.2 per 100,000 annual cases in the United States[36], PanNETs are the second most common type of pancreatic cancer, and its incidence is steadily rising [37]. Although the overall survival for PanNET patients is optimistic, long term follow up has shown that even small resected tumors show metastasis in up to 15% of cases [38]. Unfortunately, only Ki67 and Mitotic Index have widespread use prognostically for PanNET, with morphology, even high risk morphology, often not adopted [39]. Markers to determine metastasis risk for PanNET patients had either no control for overfitting or did not outperform clinopathological markers [40, 41]. A wide assortment of both systemic and targeted therapies could be tailored for patients management if risk could be accurately ascertained [42].

1.4 Computational Pathology

A promising new avenue of biomarker development is through the rapid developing field of computational pathology. Defined as a way to analyze complex, multidimensional, laboratory data, including molecular (e.g. protein or gene expression) and histological (e.g. IHC or H&E), and utilize it to generate predictive inferences [43]. Multiple cancer types including invasive breast carcinoma [44, 45], prostate [46], colon [47], and lung [48] have benefited from using these techniques to assist with diagnosis, tumor classification, predicting patient prognosis and even, extremely accurately, inferring H&E staining on unstained tissue [49]. Systems used to analyze computational pathology data, for predictive purposes, could be roughly grouped (and ordered in increasing accuracy) as statistical models, machine learning, and deep learning [50]. While deep learning networks have become the go-to for digital image analysis, sweeping most recent histopathology competitions[51], the ability of hand-crafted features within traditional machine learning models has provided super human results [52] and had the added benefit of interpretability. Although statistical models theoretically provide the poorest relative predictions, they provide the only commercial models, with genetic data, currently used in the clinic (such as the Oncotype [53] and MammaPrint [54] models), perhaps due to the obvious nature of the features and results.

I have applied all 3 types of models above to create 3 novel computational pathology tools/pipelines using a variety of multidimensional retrospective datatypes. The first model was a statistical one, used alongside an exhaustive protein biomarker panel, to determine risk of site-specific metastasis for TNBC patients. In this project I developed 4 separate regression models, of two protein markers each, that significantly identified patients at high risk for either bone, brain, liver, and lung metastasis. The crux of my research presented here, though, is developing biomarkers which leveraged whole slide image analysis of H&E stained surgical tissue. In both image analysis projects I take advantage of the tremendous variability of these stains within

different tissue regions [55] to develop a whole slide annotation tool. After this, I analyze the patterns and distribution of these stains (either through hand crafted features or within a convolutional neural network) within the previously annotated tissue classes. Variations in these stains could reflect changes in tissue architecture and cellular cytological features, or broadly, morphology. The morphology of cancer has been shown to correlate with different gene expression patterns, some of which promote invasive characteristics and therefore increase metastatic risk [56]. Perhaps the most readily observable morphological changes are within cancer cells nuclei, known as nuclear pleomorphism, and include increased size and atypia and chromatin clumping [39]. Identifying these changes are often used as backbones in tumor grading systems, traditional prognostic markers [57]. Through analysis of whole slide variations in these morphological changes, or a nested deep learning model to identify metastasis associated areas, I create tools with state-of-the-art prognostic value for DCIS and PanNET patients, respectively.

2 Novel immunohistochemistry-based signatures to predict metastatic site of triple-negative breast cancers

Parts of this chapter have been published verbatim in **British Journal of Cancer** as 'Novel immunohistochemistry-based signatures to predict metastatic site of triple-negative breast cancers'.

Authors Listed on the paper: Sergey Klimov, Padmashree CG Rida, Mohammed A Aleskandarany, Andrew R Green, Ian O Ellis, Emiel AM Janssen, Emad A Rakha and Ritu Aneja.

2.1 Abstract

Although distant metastasis (DM) in breast cancer (BC) is the most lethal form of recurrence and the most common underlying cause of cancer related deaths, the outcome following the development of DM is related to the site of metastasis. Triple negative BC (TNBC) is an aggressive form of BC characterised by early recurrences and high mortality. While multiple variables can be used to predict the risk of metastasis, few markers can predict the specific site of metastasis. This study aimed at identifying a biomarker signature to predict particular sites of DM in TNBC.

In this study, a clinically annotated series of 322 TNBC were immunohistochemically stained with 133 biomarkers relevant to BC, to develop multibiomarker models for predicting metastasis to the bone, liver, lung and brain. Patients who experienced metastasis to each site were compared to those who did not, by gradually filtering the biomarker set via a two-tailed t-test and Cox univariate analyses. Biomarker combinations were finally ranked based on statistical significance and evaluated in multivariable analyses.

Our final models were able to stratify TNBC patients into high risk groups that showed over 5, 8, 7, and 8 times higher risk of developing metastasis to the bone, liver, lung, and brain, respectively,

than low-risk subgroups. These models for predicting site-specific metastasis retained significance following adjustment for tumor size, patient age, and chemotherapy status.

Our novel IHC-based biomarkers signatures, when assessed in primary TNBC tumors, enable prediction of specific sites of metastasis, and potentially unravel biomarkers previously unknown in site tropism.

2.2 Introduction

Breast cancer (BC) is the most common cancer and the second leading cause of cancer related deaths among North American women [U.S. Cancer Statistics Working 58], and presents the largest overall cancer threat for women worldwide [59]. The vast majority of breast cancer deaths result from dissemination of cancer to distant metastatic sites [60]. BC, unlike prostate or sarcomas [61], shows significantly more organ variation in metastasis [62], making it very challenging to employ site-specific surveillance/preventative measures.

The molecular subtypes of BC have been shown to have very different underlying biology and distinct metastasis patterns [9].Triple negative BC (TNBC) is a subtype of BC characterized by an absence of the estrogen receptor (ER), progesterone receptor (PR), and HER2 protein over-expression. TNBCs account for around 16% of invasive BCs [63] and are considered one of the most clinically aggressive subtypes, with over twice the risk of distant metastasis relative to other molecular subtypes [10]. Compared to other subtypes, TNBCs also show much higher frequencies of metastasis to the brain and lung - sites associated with higher mortality compared to bone and other sites [11, 12]. Predicting TNBC's propensity for metastasis to those specific sites may allow preventive therapy and enable active surveillance to significantly improve outcomes.

The metastatic cascade is a multi-step process consisting of growth, vascularization, detachment, invasion, evasion of host defenses and survival in circulation, extravasation, and finally the ability to grow in the new organs' microenvironment [3, 64]. Few cells are successfully able to

accomplish all of the steps, and require specific biological properties both for general metastasis (e.g. factors which trigger EMT) and site-specific metastasis (e.g. breaching the blood-brain-barrier to colonize the brain)[3, 61, 65] . The similarity of genetic profiles between the primary and metastatic site tumors seems to suggest that many of the properties required for successful metastasis are developed early in the primary tumor cells well prior to the onset of metastasis [66]. Therefore, the identification of the specific biomarker profile of a primary tumor that is primed to metastasize to a specific site would enable the development of preventative and surveillance strategies tailored specifically to that particular site.

Being able to predict site of metastasis has very tangible evidence of improving patient survival. For instance, Denosumab [67], a RANKL antibody, and Bonefos [13], an oral bisphosphonate clodronate, have shown significant effect in reducing bone specific metastasis in clinical trials. However, both are currently only recommended for patients who already show evidence of bone metastasis [68, 69]. For brain metastases, where the blood-brain-barrier makes targeting tumor areas very difficult [70], there has been success pre-clinically, using Vorinostat [71], a histone deactylase inhibitor, and in clinical trials via sorafenib [72], a kinase inhibitor. For lung metastasis there has been success of inhibiting metastasis by blocking specific lung guiding molecules, S100A8 and S100A9 [73]. Finally for liver, where COX-2 expression is increased, using etodolac markedly decreased invasive properties [74]. Thus, if the site of metastasis could be identified in advance, active surveillance and the use of preemptive therapies could be implemented [70].

Studies involving genomic data have attempted to identify signatures for metastatic tropism. Multiple studies using microarray data have characterized gene expression profiles of BC that preferentially metastasized to lung or bone in mice [75, 76]. A retrospective study of transcriptomic data enabled the identification of a 6-gene prognostic classifier which could significantly discriminate BC patients who developed distant metastasis to lung [77]. Although the benefit of using genetic data is quite obvious, it also has pitfalls, the biggest of which is the lack of strong

correlation between (a) gene expression and protein levels, and (b) protein levels and protein activity levels, the latter of which can be extensively modified post-translationally. Additionally, while the price of sequencing a genome is decreasing exponentially, the price may still be prohibitive in a clinical setting [78].

In order to investigate protein signatures (in primary tumors) that are predictive of potential metastasis to specific anatomical sites, we evaluated a well-characterised cohort of clinically annotated TNBC with a long-term follow-up, utilizing the immunohistochemical expression of 133 biomarkers with relevance to BC progression and metastasis. By taking into account the protein localization (nuclear and/or cytoplasmic), the staining intensity, percentage of cells expressing the biomarker, and standard clinicopathologic variables, we investigated over 400 variables to produce the most relevant statistical models. In this paper we describe a method for step-wise filtering that yielded robust predictive models for four distinct sites of TNBC metastasis: bones, liver, lung, and brain.

2.3 Materials and Methods

2.3.1 Study Population

This study was based on a well-characterised series of primary operable invasive breast carcinoma cases (TNM stage I-IIIA) diagnosed in Nottingham between 1989 and 1998 (*N*=1944) of which 322 were classified as TNBC [i.e. 0% IHC staining of PR, ER, and HER2 0/1+ IHC staining or 2+ FISH non-amplified] (**Table 2.1**). Patients' clinical history and tumor characteristics, information on therapy, tumour recurrence and survival are described in previous publications [12, 79-83]. Data related to outcome including information on the development, site and time of DM and mortality were collected prospectively. Patients were treated according to a uniform protocol based on the Nottingham Prognostic Index (NPI) groups [84], ER and menopausal status. A systemic cyclophosphamide-methotrexate-5-fluorouracil (CMF) chemotherapy regimen was used if the patient was ER-negative provided the patient was considered fit enough to withstand this regimen. None of the patients received neoadjuvant or anti-HER2 targeted therapy.

Table 2.1: Patient Characteristics

Table S1. TNBC cohort clinical characteristics.	
Baseline characteristic	**Total (N = 322)**
Patient age	
Median Age (range), years	50 (25 - 71)
Age <50, n (%)	157 (48.8)
Age>=50, n (%)	157 (48.8)
Menopausal Status, n (%)	
Pre	164 (50.9)
Post	154 (47.8)
Peri	2 (0.6)
Nottingham Grade, n (%)	
1	7 (2.2)
2	19 (5.9)
3	295 (91.6)
Tubularity Grade, n (%)	
1	3 (0.9)
2	42 (13.4)
3	270 (83.9)
Mitotic Grade, n (%)	
1	14 (4.3)
2	25 (7.8)
3	276 (85.7)
Nuclear Grade, n (%)	
1	1 (0.3)
2	15 (4.7)
3	297 (92.2)
Chemotherapy, n (%)	
No Therapy	160 (49.7)
Classic CMF	135 (41.9)
Tumor Size	
Median Tumor Size (range), cm	2.2 (0.2 - 8.0)
Size <2.0, n (%)	113 (34.4)
Size >=2.5, n (%)	201 (60.5)
Nottingham Prognostic Index	
Median NPI (range)	4.5 (1.3 - 7.6)
Good Prognostic Index, n (%)	15 (4.5)
Medium Prognostic Index, n (%)	220 (66.3)
Poor Prognostic Index, n (%)	79 (23.8)
LN Stage	
1	206 (64.0)
2	81 (25.2)
3	34 (10.6)
Last follow up status	
Alive	174 (54.0)
Died -Cause Unknown	7 (2.2)
Died -Breast Cancer	109 (33.9)
Died -Other Causes	31 (9.6)

This study included 133 IHC-based biomarkers (**Table 2.2**) of clinical and biological relevance to BC [12, 79-83, 85-90] . During the follow-up period (243 months) 197 patients (61.2%) remained disease-free while 111 (34.5%) developed DM. Ethical approval was granted by Nottingham Research Ethics Committee 2 under the title 'Development of a molecular genetic classification of breast cancer' (C202313) and by The North West 7 Research Ethics Committee- Greater Manchester Central (10/H1008/72).

Table 2.2: Full list of biomarkers used in study. Over 300 total variables were extracted when taking into account protein cellular location (Nuclei vs. Cytoplasm) and type of quantification (intensity, percentage, and Hscore).

Immunity-related markers	**EGFR family members**	**Small secreted proteins**
CD3	EGFR	TFF1
CD8	**Growth factors/Tyrosine-protein kinases**	TFF3
CD20	c-erb-B3	**Transcription related proteins**
CD68	c-erb-B4	STAT1
FOXP3	Fibroblast growth factor receptor	STAT3
Cell cycle-associated, proliferation,	**Hormone receptors and ER-related**	SOX10
and apoptosis-related proteins	AGTR1	ADA3
Aurora A	Estrogen Receptor Beta 1	VGLL1
BCL2	GR	Nuclear receptor co-activator 3
CCNB1	**Hypoxia- and immunity-related and other markers**	hMOF
CDK7	hypoxia-inducible factor 1	ID4
C-Myc	IL-17	**Transport proteins**
Cyclin D1	interferon gamma receptor 1	CD71
H3mitosis	mannose receptor C type	OATP2
MAT1	**Cell Adhesion Molecules**	**Tumor suppressor genes**
Ki67	alpha v integrin b6	BARD1
MORF	E-cadherin	BRCA1
p16	EpCAM	BRCA2
p21	Keratin 23	CHK2
pAKTs473	N-Cadherin	FHIT
PARP1 Cleaved	P-cadherin	p53
PIAS1	**Luminal-associated cytokeratins**	PTEN
PIASgamma	Cytokeratin 7/8	**Basal Cytokeratins**
Retinoblastoma	Cytokeratin 18	CK5/6
TGFBeta1	Cytokeratin 19	CK14
TK1	**Nuclear Rceptor Superfamily member**	CK17
TOP2A	Androgen Receptor	**Proposed cancer stem cell marker**
UBC9	LRH1	CD24
DNA damage response	VDR (Vitamin D Receptor)	CD44
APE1	RAR□	**Other**
ATM	ROR Gamma	MAGE3
ATR	**Oncogenic**	PPARA
Dicer	HDAC1	PPARBeta
DNA PK	HDAC2	ST8SIA6
Drosha	Histone Acp53K382	SUV
gammaH2AX	Lactate dehydrogenase 5	NHERF1
KU70/KU80	MTA1	KPNA2
Nucleophosmin	PIK3CA	GMPR2
RAD51	TWIST2	Cathepsin D
SMC6L1	**Other basal/myoepithelial-associated markers**	Chromogranin A
Tip60	caveolin 1	Mucin 1
XPD	FABP7	Mucin 2
CHK1	Smooth muscle actin	STAC2

2.3.2 Biomarker preparation

Breast cancer tissue microarrays (TMA) were prepared and immunohistochemically stained as previously described [12, 79-83]. Positive and negative controls for each marker included in this study were used according to the supplier's data sheet. Two cores were evaluated from each tumor and only staining of the invasive malignant cells was considered. Each core was scored

individually and the mean of the two readings was calculated. Immunohistochemical scoring was performed in a blind fashion.

2.3.3 Selected Biomarker IHC details

For all model selected antibodies; heat induced retrieval of antigen epitopes was carried out, when required, by microwave treatment of slides in 1 liter of 10 mM sodium citrate buffer (pH 6.0) for 10 minutes at high power (800 W), followed by 10 minutes at low power (600 W). All primary antibodies were incubated for 1 hour (60 min) according to the manufacturer recommendation. Secondary detection: was performed using the Novolink Kit-polymer detection system (Leica, Newcastle, UK), except for N-Cadherin which used ABC. The following primary antibodies were used: MTA1 (Ab84136), KPNA2 (Ab84440), N-Cadherin (Sigma-C3865), XPD (Ab111596), RARα (Active Motif No: 39971), PARP1 (A6.4.12), BRCA2 (Ab110967).

2.3.4 Statistical analysis

Statistical analysis was carried out with SAS 9.4 ® software and Matlab version 9.2.0.556344 (R2017a). Patients were first grouped according to the site of distant metastasis or to a “no metastasis” group. If a patient had multiple metastases, that patient would be included in all the relevant groups based on the sites of their multiple metastases. Differences between clinicopathological proportions were determined using χ^2 test. Differences between continuous clinicopatholgical variables were evaluated via a 2-tailed t-test

2.3.5 Biomarker feature Selection

Due to variation of the number of biomarkers which each patient in our dataset was stained for (coefficient of variation = 0.36, not shown) and the difference in number of cases with informative data for each stained biomarker, we chose to select 2 biomarkers for each distant metastasis model. This allowed us to preserve substantial n numbers and to keep the models clinically facile. Biomarker selection (**Figure 2.1A**) was done using three progressive significance tests for each site. First, 2-tailed t-tests were performed between patients in whom distant metastasis occurred to that site versus patients who remained distant metastasis-free. This is to test for significant

baseline differences between all biomarkers. Significantly different (p-value <0.05) biomarkers were displayed as waterfall plots, with the height of each bar representing the average difference between expression of that biomarker in the “site-specific metastasis” group versus the “metastasis-free” group. Further selection was done through logistic regression, with “yes versus no” binary responses using all the biomarkers, one by one, as predictors. This selection appeared to be more stringent, as much fewer biomarkers were shown to be significant. Variables that were selected both by the t-test and the logistic regression are represented by asterisks on the waterfall plots. Finally, the selected biomarkers were run through a univariate Cox Proportional Hazard models for prognostic filtering, with Wald p-values <0.05 indicating significant variables (unless no biomarkers were found using this criterion, in which case it was relaxed to 0.1). Time to site-specific metastasis was considered as the time interval from date of surgery to date of distant metastasis to that particular site. Significant prognostic biomarkers were represented via arrows on the aforementioned waterfall plots.

2.3.6 Model building

Models were built by combining all previously selected prognostic biomarkers (in pairs), with the patient’s Nottingham Prognostic Index (NPI). Each model used the Cox parameter of the respective biomarkers as weights, combined into a score, and was thresholded (**Figure 2.1B**) by using Contal’s and O’Quigley’s approach [91]. The model chosen, for each distant site studied, was the one that minimized the Cox and Wald’s p-values (**Figure 2.1C**). The NPI threshold for testing risk of metastasis to each site using our models, was determined by finding the highest NPI value which would, regardless of the values for the IHC biomarkers in the relevant risk model, not allow the patient to have a score above the risk threshold (i.e., not allow the patient to fall into the high-risk group for that particular anatomical site).

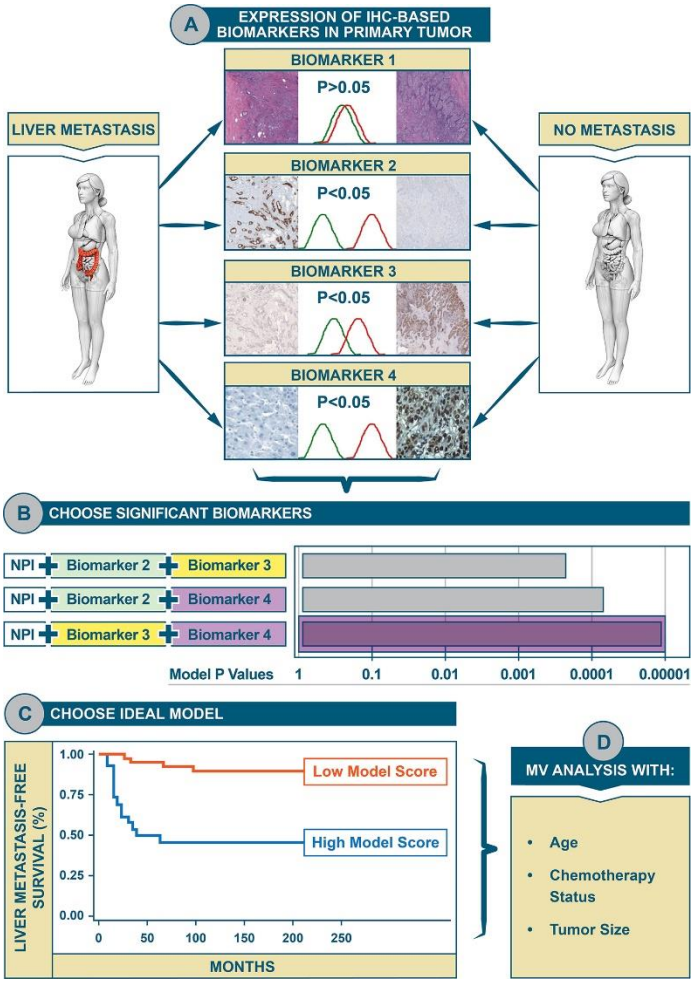


Figure 2.1: Schematic depicting sequence of steps leading to development of a model that predicts site-specific metastasis in TNBC. Briefly, a two tailed t-test was used to compare the biomarker profile for each patient who developed a site-specific metastasis versus every patient who did not have any metastasis. The biomarkers that showed significant differences in expression were then compared prognostically, with a continuous univariate Cox model, for site-specific metastasis hazard. Those significant variables which had a p-value <0.1 were then all tested with each other to identify the best combination, alongside NPI

To evaluate if their ability to predict risk of site-specific metastasis was robust regardless of the nature of model used, the selected biomarkers were also evaluated using two different machine learning algorithms: a support vector machine (SVM) and an Ensemble tree-based method. Hyperparameters for both types of models were found using Bayesian optimization, through maximization of the 'expected-improvement-plus'[92, 93] over 60 iterations (**Table 2.3 and Table 2.4**). The following parameters were optimized for the SVM algorithm: Box Constraint, Kernel

Scale, Kernel type, Polynomial order (if polynomial kernel), and feature standardization. For the Ensemble tree algorithm, the following hyperparameters were optimized: Ensemble method (Bagging, GentleBoost, LogitBoost, AdaBoost, RUSBoost'), maximum number of branch nodes, minimum number of leaf nodes, and the split criteria. Both methods were also built with/without empirical prior dataset probabilities for site-specific metastasis. The optimized hyperparameters for each model are detailed in the online Supplementary Data.

Table 2.3: Hyperparameters identified for each distant metastasis site's Ensemble-based model through Bayesian optimization via maximization of the 'expected-improvement-plus' over 60 iterations.

	Ensemble Method	**Max Branches Nodes**	**Min Leaf Nodes**	**Split Criteria**	**Priori**
Bone	RUSBoost	9	10	deviance	Uniform
Lung	Bag	10	9	Gini's diversity index	Uniform
Liver	AdaBoostM1	3	3	Gini's diversity index	Uniform
Brain	AdaBoostM1	1	4	deviance	Uniform

Table 2.4:Hyperparameters identified for each distant metastasis site's SVM-based model through Bayesian optimization via maximization of the 'expected-improvement-plus' over 60 iterations.

	Box Constraint	**Kernel Scale**	**Kernel Type**	**Polynomial Order**	**Standardization**	**Priori**
Bone	0.001004734	1.389753067	gaussian	-	TRUE	Uniform
Lung	0.011543653	-	polynomial	3	TRUE	Uniform
Liver	0.005632345	-	linear	-	TRUE	Uniform
Brain	870.3162264	18.03512966	gaussian	-	TRUE	Uniform

2.3.7 Model Validation

All models (namely, our combined and then thresholded model, the optimized SVM, and the optimized Ensemble) to each site, were 5-fold cross-validated for survival risk evaluation. Kaplan-Meier survival curves were created by combining the 5 testing sets and then used to confirm significance and rank models. The comparison metric used to compare the cross validated models was the Akaike Information Criterion (AIC), a measure of fit. The model which granted the lowest AIC per site, was considered the optimal model for that site. Multivariate analysis was also performed to control for the effects of chemotherapy, tumor size, and age.

2.4 Results

The ability of clinical variables to predict distant metastasis [94] and specifically in TNBCs [95] is well documented, with common features, as for instance tumor size, and nodal stage providing significant prognostic ability. Our data corroborate these findings by showing tumor size (HR = 0.002), age ($p < 0.048$), and NPI ($p < 0.0001$) as having significant univariate impact on distant metastasis-free survival. However, a comparison of the distribution of these clinical factors for specific metastasis sites (**Table 2.5**) showed no difference in mean values of these variables or in the proportions of patients in each group.

Table 2.5: Clinical characteristics of TNBC patients in this study categorized according to site of distant metastasis. Proportions/means significance noted via either chi-square/ANOVA p-values.

Baseline characteristic	Total (N = 322)	Bone metastasis (n = 47)	Liver metastasis (n = 29)	Brain metastasis (n = 29)	Lung metastasis (n = 34)	χ^2 P-value
Patient age						
Median Age (range), years	50 (25 - 71)	50 (28 - 71)	48 (28 - 70)	47 (31 - 68)	48 (28 - 69)	
Age <50, n(%)	157 (48.8)	23 (48.9)	15 (51.7)	17 (58.6)	17 (50.0)	0.8759
Age>=50, n(%)	157 (48.8)	24 (51.1)	14 (48.3)	12 (41.4)	16 (47.1)	
Menopausal Status, n (%)						
Pre	164 (50.9)	23 (48.9)	16 (55.2)	18 (62.1)	17 (50.0)	
Post	154 (47.8)	22 (46.8)	11 (37.9)	11 (37.9)	15 (44.1)	0.9109
Peri	2 (0.6)	1 (2.1)	1 (3.4)	0 (0.0)	1 (2.9)	
Nottingham Grade, n (%)						
1	7 (2.2)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
2	19 (5.9)	4 (8.5)	2 (6.9)	0 (0.0)	2 (5.9)	0.4651
3	295 (91.6)	42 (89.4)	27 (93.1)	29 (100.0)	32 (94.1)	
Chemotherapy, n (%)						
No Therapy	160 (49.7)	21 (44.7)	11 (37.9)	9 (31.0)	16 (47.1)	0.8231
Classic CMF	135 (41.9)	20 (42.6)	13 (44.8)	13 (44.8)	13 (38.2)	
Median Tumor Size (range), cm	2.2 (0.2 - 8.0)	2.2 (0.6 - 7.0)	2.2 (1.2 - 6.5)	2.3 (1.3 - 6.0)	2.2 (1.1 - 6.5)	0.9862
Median NPI (range)	4.5 (1.3 - 7.6)	3.8 (3.1 - 6.9)	5 (3.2 - 7.3)	4.9 (4.3 - 6.8)	4.7 (3.4 - 7.3)	0.7488

We also observed that chemotherapy did not affect recurrence patterns (**Table 2.6**). This led us to investigate if any of our biomarker models could provide the required specificity of being both prognostically relevant and unique to specific distant metastasis sites.

Table 2.6: Recurrence pattern to specific sites depending on therapy regime. Proportions significance noted via chi-square p-values.

	% post CMF	% no chemotherapy	*P*-value
Bone Metastasis			
No	85.19	86.79	0.6918
Yes	14.81	13.21	
Liver Metastasis			
No	90.37	93.08	0.3975
Yes	9.63	6.92	
Lung Metastasis			
No	90.37	89.94	0.9012
Yes	9.63	10.06	
Brain Metastasis			
No	90.37	94.34	0.1974
Yes	9.63	5.66	

2.4.1 Bone metastasis

Among the protein biomarkers available in our TNBC dataset, those whose expression was significantly different in patients who developed bone metastasis (**Fig. 2.2A**), included several that were overexpressed (blue lines) or highly underexpressed (red lines) in the primary tumor. The 8 biomarkers which are eligible for inclusion into the final model, based on univariate prognostic significance, are indicated. **Fig. 2.2B** shows the results of the parameter selection, with the lowest p-value ($p<0.0001$) obtained combining the MTA1 nuclear H-score, KNPA2 nuclear percentage, in addition to NPI. NPI was included in all our models as a stand-in for a

"generalized risk of metastasis" as high-NPI patients have a higher risk of distant metastasis compared to low-NPI (Metastasis HR = 1.6, $p<0.001$).

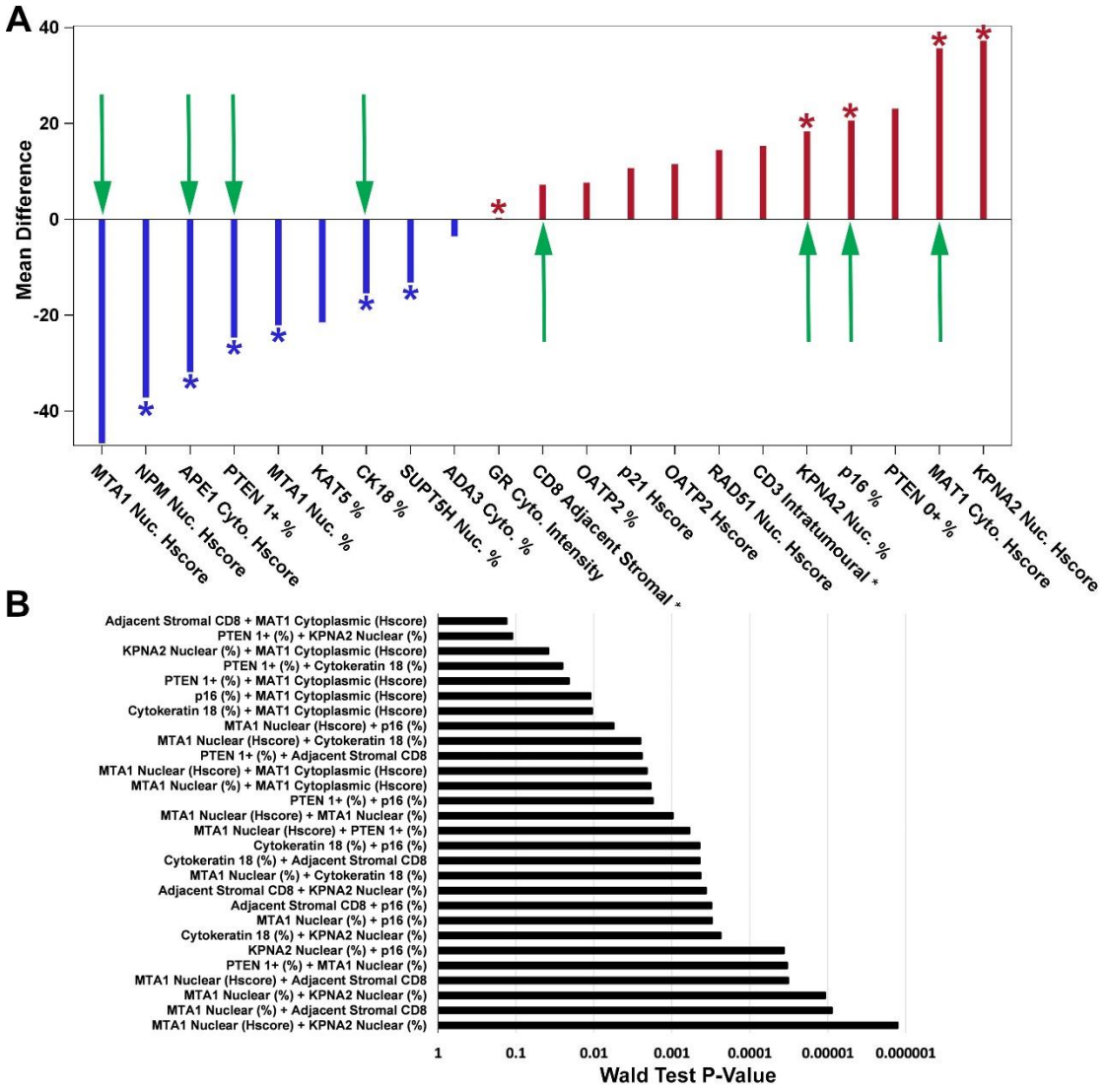


Figure 2.2: A) Waterfall plot showing significantly different biomarkers (as determined via a t-test), when comparing patients who had bone metastasis versus patients with no metastasis. The y-axis represents difference between mean biomarker levels in patients who had no metastasis and patients who had bone metastasis. Blue lines represent significantly higher mean expression of biomarker among patients who had metastasis, while red represents the opposite. Asterisks indicate significance in logistic regression. Green arrows indicate significance in Cox regression for time until bone specific metastasis. B) Ranked list of biomarker combinations (alongside NPI) based on their Cox regression p-values for bone metastasis.

This model, detailed below, enables us to identify patients who have a 5 times higher risk of developing metastasis to bones (**Fig. 2.3A**) and stayed significant after cross validation. Multivariate analysis (**Table 2.7**) confirmed the prognostic value of our model by having it independently associated with bone metastasis risk ($p < 0.0001$) following adjustment for age, chemotherapy status, and tumor size.

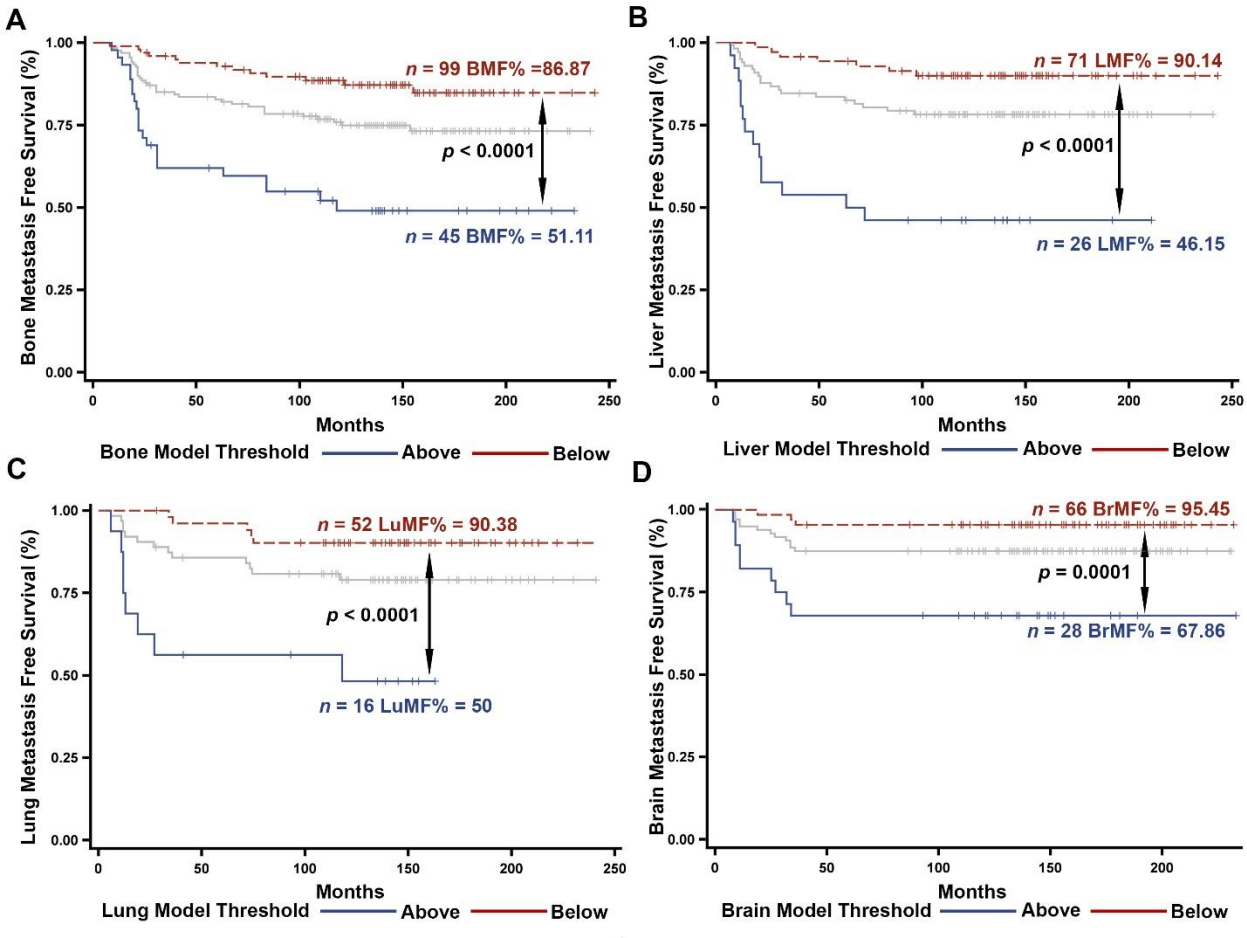


Figure 2.3: Kaplan-Meier survival curves showing patient stratification via our survival-based models for A) Bone (BMF = Breast Metastasis Free), B) Liver (LMF = Liver Metastasis Free), C) Lung (LuMF = Lung Metastasis Free), and D) Brain sites (BrMF = Brain Metastasis Free). All significances are measured via the log-rank test. Light grey lines represent baseline survival for the patients before stratification by the respective site-specific metastasis predictive models.

$$Bone\ Metastasis\ Score = (0.27) * MTA1\ Nuclear\ Hscore - (1.26) * KPNA2\ Nuclear\ \% + (43.49) * NPI$$

$$If\ Bone\ Metastasis\ Score \geq 196, then\ high\ risk\ of\ bone\ metastasis$$

We also compared the performance of this model in the patient subgroup that received adjuvant CMF chemotherapy versus the subgroup that received no adjuvant chemotherapy, to determine if the model's prognostic value was affected by therapy. Results showed that the model for predicting bone-specific metastasis-maintained significance regardless of whether chemotherapy was administered or not (**Table 2.8**).

Table 2.7: Univariate and multivariate Cox regression analysis of common clinicopathological variables and IHC models affecting distant metastasis risk.

Table 1. Cox regression analysis of common clinopathological variables and IHC models							
	variables	Univariate Analysis			Multivariate Analysis		
		Hazard Ratio	95% Confidence interval	*P*-value	Hazard Ratio	95% Confidence	*P*-value
Bone							
Age of diagnosis		1.022	0.997 - 1.048	0.0813	0.998	0.959- 1.038	0.9028
Chemotherapy	CMF vs. None	0.801	0.452 - 1.419	0.4471	0.741	0.301 - 1.824	0.5148
Tumor Size	Per CM	1.397	1.001 - 1.951	0.0496*	1.355	0.862 - 2.132	0.188
Risk Model	High vs. Low	5.123	2.572 - 10.201	<.0001*	4.939	2.281 - 10.692	<.0001*
Liver							
Age of diagnosis		1.002	0.971 - 1.034	0.9042	0.969	0.925 - 1.016	0.1957
Chemotherapy	None vs. CMF	1.202	0.570 - 2.536	0.6294	0.57	0.185 - 1.757	0.3275
Tumor Size	Per CM	1.688	1.171 - 2.432	0.005*	1.134	0.592 - 2.172	0.7047
Risk Model	High vs. Low	8.039	3.230 - 20.005	<.0001*	9.156	3.376 - 24.837	<.0001*
Lung							
Age of diagnosis		1.004	0.973 - 1.037	0.7878	0.972	0.902 - 1.047	0.4486
Chemotherapy	None vs. CMF	0.808	0.402 - 1.626	0.5507	0.686	0.156 - 3.016	0.618
Tumor Size	Per CM	1.761	1.221 - 2.540	0.0025*	1.735	0.712 - 4.226	0.225
Risk Model	High vs. Low	7.661	2.491 - 23.564	0.0004*	6.306	1.567 - 25.374	0.0095*
Brain							
Age of diagnosis		1.005	0.971 - 1.040	0.7718	0.943	0.867 - 1.026	0.1721
Chemotherapy	None vs. CMF	1.261	0.572 - 2.778	0.565	0.035	0.003 - 0.390	0.0064*
Tumor Size	Per CM	1.717	1.129 - 2.610	0.0115*	4.632	1.519 - 14.132	0.0071*
Risk Model	High vs. Low	8.506	2.299 - 31.471	0.0013*	36.362	4.276 - 309.228	0.001*

Table 2.8: Prognostic value of the survival-based model for each site, depending on adjuvant therapy received (univariate Cox regression).

	Hazard Ratio	95% Confidence interval	*P*-value
Bone Model			
Classic CMF (N = 73)	5.003	1.777 - 14.085	0.0023*
No Chemotherapy (N = 66)	5.362	1.871 - 15.266	0.0018*
Liver Model			
Classic CMF (N= 54)	6.147	1.790 - 21.112	0.0039*
No Chemotherapy (N = 38)	12.533	2.509 - 82.598	0.0007*
Lung Model			
Classic CMF (N = 39)	9.122	1.832 - 45.412	0.0069*
No Chemotherapy (N = 27)	6.468	0.661 - 63.249	0.1086
Brain Model			
Classic CMF (N = 49)	40.307	0.331 - 3.653	0.2903
No Chemotherapy (N = 41)	14.955	2.687 - 83.241	0.002*
*P<0.05.			

Interestingly, the cross-validated AICs showed that this survival-based model slightly outperformed the SVM and Ensemble-based models (**Table 2.9**). Notably, all models tested yielded statistically significant stratification.

Table 2.9: AIC values for each model were obtained from survival analyses of combined 5-fold cross-validated test sets and used as a comparison matrix. The model which grants the lowest AIC, per site, is considered the optimal model for that site. Log-rank p-value

	Without Covariates	Survival Model	Optimized SVM	Optimized Ensemble
Bone	334.58	**319.73**	321.533	323.431
Lung	106.661	**101.639**	102.399	101.764
Liver	187.201	180.305	**173.041**	184.067
Brain	107.641	98.478	**94.977**	96.003

2.4.2 Liver metastasis

For patients with liver metastases, we observed that the majority of differentially expressed biomarkers ($p<0.05$; 29 vs. 9) were underexpressed in the patient subgroup with liver metastases compared to metastasis-free patients (**Fig. 2.4A**). Furthermore, we found that the underexpression of majority of these biomarkers (7 vs. 1), was statistically significant in univariate analyses. The combination that yielded the lowest p-value ($p <0.0001$) involved N-cadherin H score, the cytoplasmic intensity of xeroderma pigmentosum complementation group D (XPD), and NPI (**Fig. 2.4B**). The model shown below can stratify patients into a high-risk group that shows ~8x higher risk of liver metastasis (**Fig. 2.3B**) and retained significance after cross validation. Multivariate analysis indicated that this model is contributing predictive information for liver-specific metastasis independently of other factors (**Table 2.8**).

$$Liver\ Metastasis\ Score = (0.61) * (N-cadherin\ Hscore) - (108) * XPD\ cytoplasmic\ intensity + (90) * NPI$$

$$If\ Liver\ Metastasis\ Score \geq 436, then\ high\ risk\ of\ liver\ metastasis$$

As with the bone model, the survival-based model for lung retained significance regardless of chemotherapy (**Table 2.8**) and performed marginally better than machine learning approaches (**Table 2.9**).

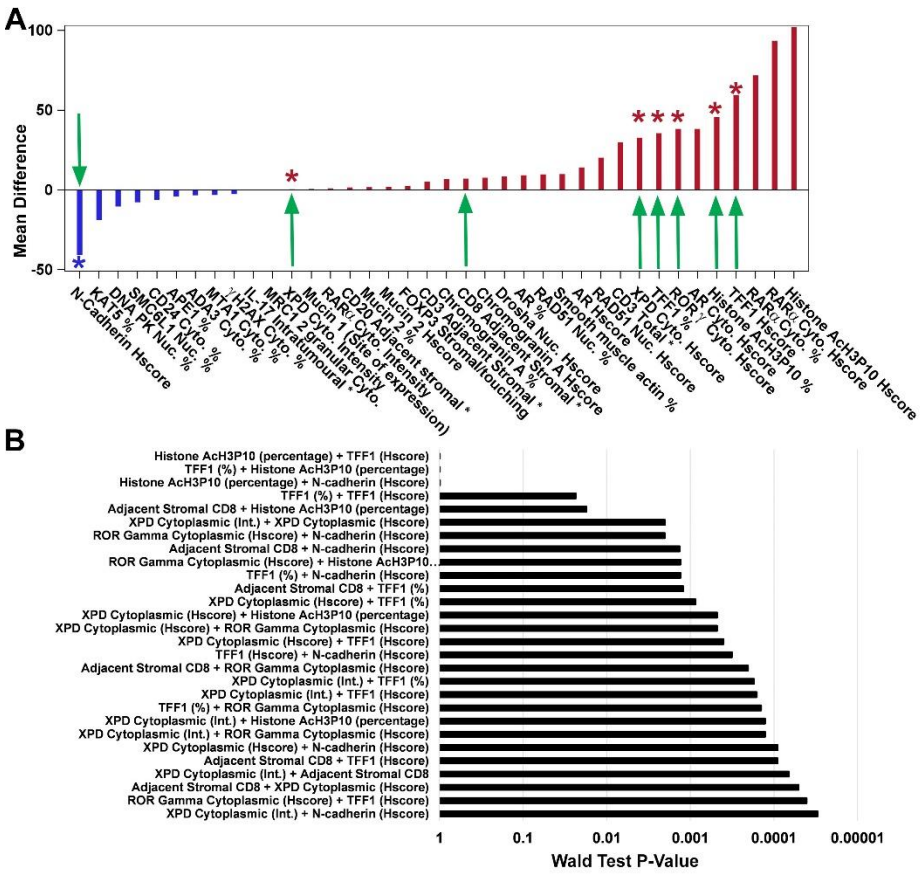


Figure 2.4: A) Waterfall plot showing significantly different biomarkers (as determined via a t-test), when comparing patients who had liver metastasis versus patients with no metastasis. The y-axis represents difference between mean biomarker expression levels in patients who had no metastasis and patients who had liver metastasis. Blue lines represent significantly higher mean expression of biomarker among patients who had metastasis, while red represents the opposite. Asterisks indicate significance in logistic regression. Green arrows indicate significance in Cox regression for time until liver specific metastasis. B) Ranked list of biomarker combinations (alongside NPI) based on their Cox regression p-values for liver metastasis.

2.4.3 Lung metastasis

Unlike for liver, multiple IHC biomarkers, such as Fascin-1, Id1, and Id3 have been reported to mediate lung colonization in invasive BC including TNBCs [96, 97]. Unlike the previously mentioned proteins (where overexpression was correlated with lung metastasis) our model filtering (**Fig. 2.5**) led to the selection of two biomarkers that accorded a favorable prognosis when expressed at a high level. Combining TFF1 and RARa, as shown below,

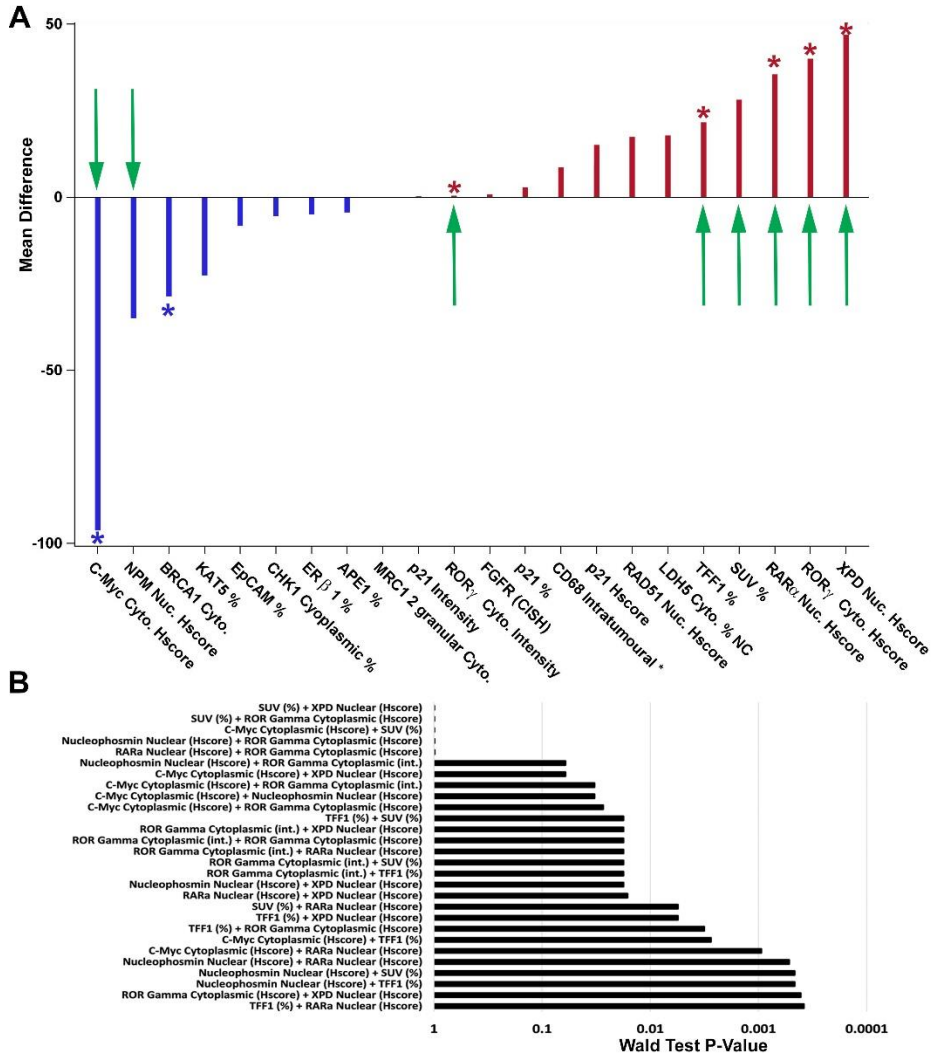


Figure 2.5: A) Waterfall plot showing significantly different biomarkers (as determined via a t-test), when comparing patients who had lung metastasis versus patients with no metastasis. The y-axis represents difference between mean biomarker expression levels in patients who had no metastasis and patients who had lung metastasis. Blue lines represent significantly higher mean expression of biomarker among patients who had metastasis, while red represents the opposite. Asterisks indicate significance in logistic regression. Green arrows indicate significance in Cox regression for time until lung specific metastasis. B) Ranked list of biomarker combinations (alongside NPI) based on their Cox regression p-values for lung metastasis.

produced a high-risk group which had over a 7 times higher risk of developing lung metastasis (**Figure 2.3C**). Additionally, this model retained its significance in cross validation and multivariable analysis, independent of other factors (**Table 2.8**).

$$Lung\ Metastasis\ Score = (-1.25) * TFF1\ \% - (0.82) * RARa\ nuclear\ Hscore + (115) * NPI$$

$$If\ Lung\ Metastasis\ Score \geq 436, then\ high\ risk\ of\ lung\ metastasis$$

Unlike the previous two site-specific models, using an SVM to predict lung metastasis produced a marginally superior AIC, and thus fit (**Table 2.9**), although all models retained significant stratification. Also, while the model showed powerful prognostic ability among CMF-treated patients, it lost significance in the patient subgroup that did not receive CMF (**Table 2.8**, p = 0.1); this was likely due to the low number of metastatic events and metastasis-free patients in that patient subgroup (4 and 27, respectively).

2.4.4 Brain metastasis

Although brain metastasis only accounts for around 10-16% of all breast metastasis sites [98], and is a relatively longer process due to the blood-brain barrier [99], it results in a very poor survival and a dramatic reduction in quality of life [100]. The current paucity of biomarkers with the ability to predict metastasis to the brain [101], coupled with lack of an effective targeted treatment [102] demonstrate that this is an area of urgent and unmet clinical need for BC patients. Recently, though, αB-crystallin, a chaperone protein predominantly expressed in brain metastasis, has shown promise as a TNBC site-specific IHC biomarker [103, 104]. In our dataset, we found only a few biomarkers that (a) showed significantly different expression between patients with metastasis to the brain and those with no metastases, and (b) had prognostic value in univariate analyses (**Figure 2.6A**). Post-hoc survival analysis using a non-optimized (minimized Wald p-value) biomarker combination for brain metastasis patients yielded a very imbalanced high-risk group that included only 2 patients (not shown). We therefore combined the biomarkers whose combination had the second-best p-value (**Figure 2.6B**) to develop the model shown below. With this model, high-risk patients possessed more than a 7x higher risk of brain metastasis (**Figure 2.3D**). This effect was maintained in multivariate analysis (**Table 2.8**) and cross validation.

$$Brain\ Metastasis\ Score = (0.77) * Parp1\ Nuclear\ Hscore + (0.87) * BRCA2\ cytoplasmic\ Hscore + (79) * NPI$$

$$If\ Brain\ Metastasis\ Score \geq 568, then\ high\ risk\ of\ brain\ metastasis$$

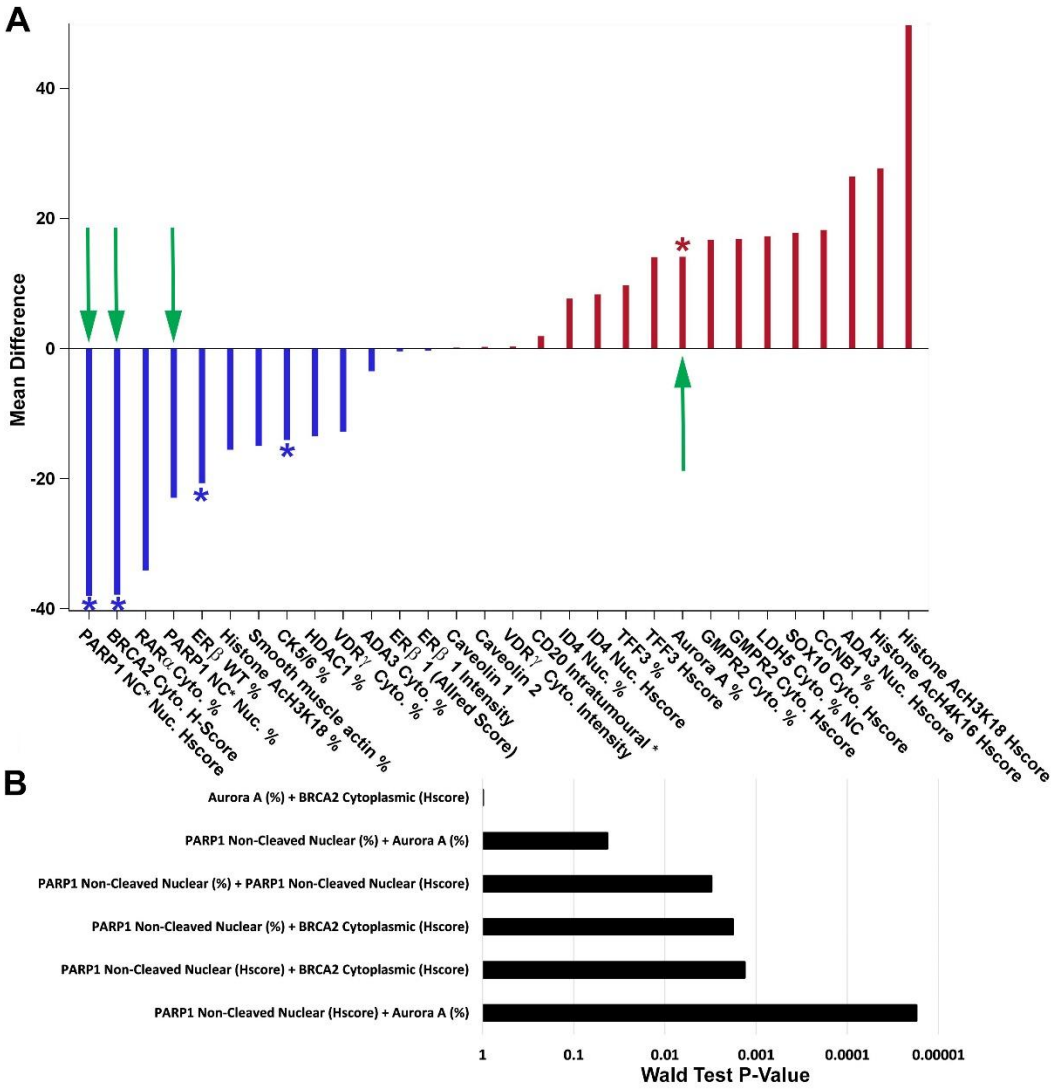


Figure 2.6: A) Waterfall plot showing different biomarkers (as determined via a t-test), when comparing patients who had brain metastasis versus patients with no metastasis. The y-axis represents difference between mean biomarker expression levels in patients who had no metastasis and patients who had brain metastasis. Blue lines represent significantly higher mean expression of biomarker among patients who had metastasis, while red represents the opposite. Asterisks indicate significance in logistic regression. Green arrows indicate significance in Cox regression for time until brain specific metastasis. B) Ranked list of biomarker combinations (alongside NPI) based on their Cox regression p-values for brain metastasis.

The prognostic value appears to result from significant stratification of the untreated patients (**Table 2.8**), as only 3 treated patients, who were stained for both markers, had distant metastasis to the brain. Patient prognosis, while significant with our model, was predicted slightly better using an SVM (**Table 2.9**).

We then addressed the question of whether every TNBC patient in the clinic should be prescribed the test for our panel of 8 IHC-based biomarkers that are able to foretell risk of metastasis to specific sites. Interestingly, we found that the vast majority of TNBC patients in this dataset had an NPI>4 regardless of whether they experienced metastasis or not (**Figure 2.7**), and thus would require testing for all 8 biomarkers. We confirmed elevated NPI among TNBCs in a second independent dataset (**Figure 2.7B**). These data suggest that with the exception of a very small proportion of TNBC patients whose NPI is below 4, the majority of TNBCs may require testing for all 8 biomarkers to determine risk of metastasis to these sites in the future.

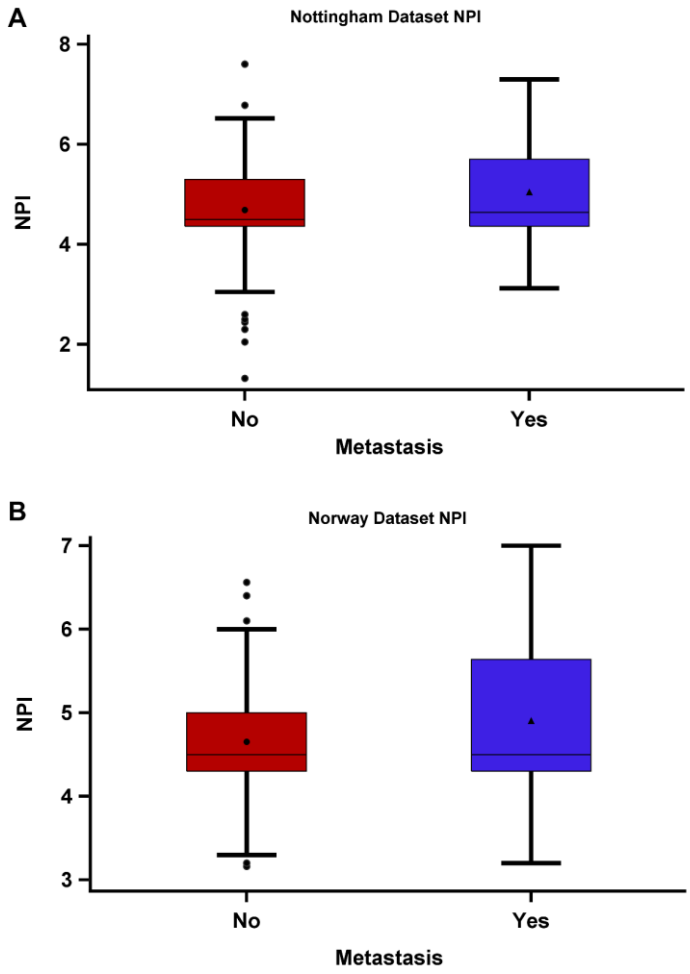


Figure 2.7: Mean NPI value for patients who have developed distant metastasis and those without metastases for the A) Nottingham University Hospital, and B) Stavanger University Hospital (in Norway) cohorts. T-tests for both hospitals presented a p-value > 0.05 when comparing mean NPI between patients who developed distant metastasis versus those who did not.

2.5 Discussion

BC patients with DM have a median survival of only 2-3 years [105]. Even more worrisome is the fact that both the time until DM and survival after metastasis is greatly reduced for TNBCs, especially among those with residual disease after neoadjuvant treatment [106, 107]. However, metastasis to different sites is associated with distinct survival times after metastasis with some metastatic sites associated with poorer outcomes compared to others. Therefore, predicting DM before it occurs and identifying the potential sites of metastasis would have a significant impact in management of TNBC.

Previous studies investigating biomarkers predictive of the site of DM in BC have mainly utilized either global gene expression data using high-throughput techniques such as microarrays and next generation sequencing or single proteins using IHC [108-110]. No studies have investigated DM using large groups of protein biomarkers in primary TNBC tumor samples. Using our novel models, we are able to introduce a clinically-facile IHC biomarker panel that can identify high-risk subgroups among TNBCs, with at least a 5x increased risk of site-specific metastasis. The strength of the current study stems from (a) the large number of cases in our TNBC series, (b) their long-term follow-up and detailed clinical annotation, (b) the unique, and, to the best of our knowledge, largest IHC biomarker dataset available for this cohort, and (c) the comprehensive analytical approaches.

Bone is the most studied BC DM site, with multiple steps of the metastasis cascade elucidated in substantial detail [111, 112]. Alongside this molecular knowledge, multiple bone metastasis-specific biomarkers have been proposed. For example, Winczura et al. [113] have found that a reduction of osteopontin was consistently observed in patients who developed bone metastasis while Mihai et al. [114] have found that the calcium-sensing receptor (CaR) was commonly expressed in breast tumors which metastasized to the bone. While our models uncovered some proteins previously known to be associated with metastasis, it also uncovered several proteins that generally have not been studied in the context of BC tropism to specific metastatic sites/tissues, or have not been implicated directly in regulating metastasis. For example, upregulation of the high risk biomarker MTA1 in our bone metastasis model, is seen in several aggressive cancers [115] and has been linked to bone metastasis from prostate cancer [116]. In BC, MTA1 upregulation has been shown to promote lung-specific metastasis in mice [117]. By contrast, we found that underexpression of the karyopherin, KPNA2, is associated with development of bone metastasis. KPNA2 has not been implicated in site specificity of metastasis; in fact, its overexpression was correlated with poorer recurrence-free overall survival in BC [118,

119]. More importantly, KPNA2 expression in patients with no metastases and patients with metastasis to sites other than bone, was higher than in patients with bone metastasis. These results suggest that TNBC patients (a) have high baseline expression of KPNA2 [90], and (b) this high expression preferentially selects for all the other metastatic sites (seen in **Figure 2.8**, along with the other site specific biomarker comparisons).

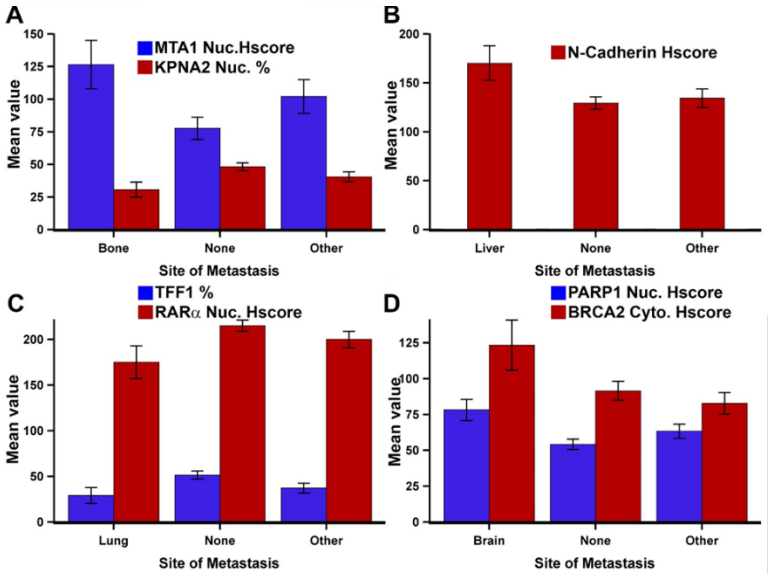


Figure 2.8: Box plots comparing selected model biomarkers expression between that site, all other sites (combined), and with patients who experienced no metastasis. Note that for liver the second biomarker used was the cytoplasmic intensity value of XPD, a categorical variable; thus this biomarker is not represented in these box plots. All biomarkers were significantly different when comparing patients who have had metastasis to the specific site versus patients combined in the two other groups

Although liver is one of the most common sites of metastasis for BC patients [60] there is little research into potential liver metastasis-specific IHC biomarkers. Interestingly, the biomarkers that were differentially expressed in patients with liver metastases showed a strong tendency to be overexpressed in patients with liver metastases. The best model included N-cadherin, whose upregulation has been associated with pro-migratory phenotypes [120], and is thus believed to contribute to the general risk of metastasis. There is also evidence suggesting a preference for tropism to liver in BCs that overexpress N-cadherin [121, 122], although the mechanistic underpinning of that preference is yet to be uncovered. The other biomarker selected, XPD (also

known as ERCC2), has no previously published evidence of being involved in BC metastasis. In fact, most studies focus on the association between mutations in this gene and an increased risk of developing BC [123], or specifically TNBC [124].

We also found that the roles reported for some of the metastasis biomarkers in our models appear to differ between ER-positive and TNBC patients. For instance, in ER-positive BC cohorts, gene expression studies showed *TFF1* to be very highly overexpressed in patients who had bone metastasis versus metastasis to another site [125]. However, IHC data showed no significant difference between patients who developed bone metastasis and those with no metastasis [126]. There are also conflicting reports regarding the impact of TFF1 overexpression on BC prognosis with some studies suggesting that it may play an oncogenic role [127] while others indicate an association between its overexpression and a favorable prognosis [128]. In our TNBC cohort, reduced TFF1 expression was associated with high risk of lung metastasis. However, studies in ER-positive BC suggest that high TFF1 levels could promote lung metastasis via TFF1's role in enhancing chemotaxis [129]. Another protein that has not previously studied with regard to promoting metastasis to any specific site is RARα. Our data suggest that TNBC patients who experience lung metastasis underexpress nuclear RARα. In ER-positive BC, the presence of ER both correlates with the number of RARα receptors and the ability of ER to inhibit cell growth in concert with RARα [130]. In TNBCs that underexpress RARα, it is plausible that the brakes on proliferation are lifted; however, the molecular basis of the propensity of these low-RARα TNBCs to metastasize to the lung is currently unclear and merits further study.

An important feature that cancer cells require to metastasize to the lung, the ability to extravasation through non-fenestrated capillaries, is also vital for brain metastasis. In fact, multiple genetic similarities were shown between cells primed to metastasize to the brain and to the lungs, such as COX2 [131]. In the brain metastasis model we derived, we observed an unexpected combination of overexpressed biomarkers that were not significant for metastasis to any other

site. A previous study of IHC biomarkers had shown that an increase in both PARP1 and nuclear BRCA2 expression is associated with a stark decrease in both OS and RFS, separately and when combined [132]. By contrast though BRCA2 in our brain metastasis model was cytoplasmic; more studies are required to clarify the functions of cytoplasmic BRCA2 [133].

The aim of this retrospective study was exploratory, to identify IHC-based biomarkers which held statistical significance in predicting TNBC metastasis to specific sites. Our study highlights the importance of evaluating protein subcellular localization and identification of such “phenotypic’ biomarkers, as subcellular localization can profoundly influence biological activities and prognostic significance of protein biomarkers. In fact, we found that in the majority of the cases, the nuclear-localized or cytoplasmic pools of the proteins in our signatures held prognostic significance while the overall levels did not. This finding emphasizes a key limitation of gene-expression-based signatures where robust gene expression-based signatures would be limited to the subset of proteins whose cellular activities are directly proportional to mRNA expression levels.

It is also noteworthy that in our dataset, among patients with metastases to multiple sites, the exact order of metastases is unknown, and each metastasis was treated independently even though it is possible that some of these metastases may have arisen from other earlier metastases rather than from the primary tumor. In closing, our novel multi-parametric prognostic models allow for very significant identification of patients with TNBC who will experience distant metastasis to a specific site. Design of a cost-effective, clinically-facile IHC-based battery of tests to predict the most likely site of metastasis for TNBCs would (a) enable early detection of metastases through increased surveillance, (b) allow use of preventative therapy to prevent disease progression, and (c) improve outcomes for TNBCs.

Copyright by
Sergey Klimov
2019

3 A Whole Slide Image Based Machine Learning Approach to Predict Ductal Carcinoma in Situ (DCIS) Recurrence Risk

Parts of this chapter have been submitted to **Breast Cancer Research** as 'A Whole Slide Image Based Machine Learning Approach to Predict Ductal Carcinoma in Situ (DCIS) Recurrence Risk'.

Authors Listed on the paper: Sergey Klimov, Islam M. Miligy, Arkadiusz Gertych, Yi Jiang, Michael S. Toss, Padmashree Rida, Ian O. Ellis, Andrew Green, Uma Krishnamurti, Emad A. Rakha, and Ritu Aneja1

3.1 Abstract

Breast ductal carcinoma in situ (DCIS) represent approximately 20% of screen-detected breast cancers. The overall risk for DCIS patients treated with breast conserving surgery stems almost exclusively from local recurrence. Although a mastectomy or adjuvant radiation can reduce recurrence risk, there are significant concerns regarding patient over-/under-treatment. Current clinicopathological markers are insufficient to accurately assess recurrence risk. To address this issue, we developed a novel machine learning (ML) pipeline to predict risk of ipsilateral recurrence using digitized whole slide images (WSI) and clinicopathologic long-term outcome data from a retrospectively collected cohort of DCIS patients (n=344) treated with lumpectomy at Nottingham University Hospital, UK.

The patient cohort was split case-wise into training: (n=159, 37 with recurrence); and validation (n=185, 32 with recurrence) sets. Sections from primary tumors were stained with H&E, then digitized and analyzed by the pipeline. In the first step, a classifier trained manually by pathologists was applied to digital slides to annotate areas of stroma, normal/benign ducts, cancer ducts, dense lymphocyte region, and blood vessels. In the second step a recurrence risk classifier was trained on 8 select architectural and spatial organization tissue features from the annotated areas to predict recurrence risk.

The recurrence classifier significantly predicted the 10-year recurrence risk in the training [hazard ratio (HR)=11.6; 95% confidence interval or CI: 5.3–25.3, p<0.0001; concordance index=0.77 (95% CI: 0.69–0.85)] and independent validation [HR=6.39 (95% CI: 3.0–13.8), p<0.0001; concordance index=0.69 (95% CI: 0.59–0.78)] cohorts. Our tool outperformed clinicopathological variables in predicting overall local, invasive, and DCIS recurrences (p<0.0001), had a superior concordance index compared to all clinicopathological variables, and identified patients that might benefit from additional therapy (validation cohort p=0.0006). Our machine learning-based model fills an unmet clinical need for accurately predicting recurrence risk for lumpectomy treated DCIS patients.

3.2 Introduction

The incidence of ductal carcinoma in situ (DCIS) has rapidly risen over the past few decades [1] and is estimated to affect over 1 million US women by 2020 [134]. Despite the excellent overall survival of DCIS patients [135, 136], over-treatment is a considerable concern [137]; which results mainly from the inability of standard clinicopathologic factors to accurately identify a low risk group unlikely to recur [138, 139].

One of the goals of DCIS treatment is to curb local recurrence, especially invasive recurrence. Common histopathological factors such as age at diagnosis, DCIS growth pattern, tumor size, margin status, nuclear grade, presence of comedo necrosis [21, 140], and combinations of the aforementioned (such as in the Van Nuys Prognostic Index or in prognostic nomograms) [141, 142] have been shown to have limited value in predicting recurrence. Efforts to introduce new DCIS molecular prognostic variables have not offered consistent results [23] nor were they found to be significantly prognostic tools [34]. Additionally, transcriptomic models have restrictive requirements [143], are not cost-effective [33], lack significant “genetic patterns leading to invasive disease” signatures [144], and do not take into account the tumor stromal

microenvironment. Thus, there is an unmet clinical need for novel tools to improve recurrence risk stratification of DCIS [145].

With the advent of technology able to process data in a high throughput manner, computational pathology has shown promise as a valuable prognostic tool. By integrating image analysis, data generation, and medical statistics, computational pathology enables a high-level quantitative tissue analysis [43, 146]. Although relatively new, computational pathology has already shown marked success in assisting with diagnosis, tumor classification, and predicting patient prognosis in a variety of cancer types [44-48, 147]. Whole slide quantitative image analysis pipelines have demonstrated significant discriminatory success not only using features stemming from pixel (stain) intensities [148, 149], but also morphometric features and texture [150, 151]. For predicting DCIS recurrence, various scales of these image features have been studied using H&E-stained tissue, such as through quantifying image features of comedo necrosis within ducts [152]. At the cellular level, chromatin distribution, long considered a computationally quantifiable feature of cancer cells [153], has also been used to predict DCIS recurrence [154, 155] and was shown to outperform its pathological analog, nuclear grade [156]. However, these results focus on a narrow range of very specific characteristics of the DCIS and discard the rich information that could potentially be derived from consideration of other architectural features (e.g., surrounding stromal, blood vessel-related) within the sample.

Human eye limitations and lack of concordance between pathologist's impact DCIS grading in clinical practice. Notably, the breadth of DCIS grading is limited to a single (high grade) duct, and oftentimes histopathologic features are grouped into qualitative categories instead of capturing and analyzing more granular data derived from quantitative features. This simplification overlooks (a) the prognostic value of surrounding microenvironment [157-159] and even alterations in non-cancerous epithelial cells [160], and (b) the tremendous intra-tumor heterogeneity, which cannot be categorized in a fundamentally meaningful way [161]. Our current study evaluates whether

quantitatively analyzing the whole slide, dubbed whole slide image analysis (WSI) [162] has prognostic and predictive value with respect to recurrence prediction for DCIS.

In the retrospective study presented herein, we developed a machine learning-based image analysis pipeline, identified prognostically relevant features obtained from the texture of H&E slides [55], and designed a novel two-step classification approach to predict 10-year recurrence risk in DCIS patients treated with breast conservative surgery (BCS) (**Fig 3.1**).

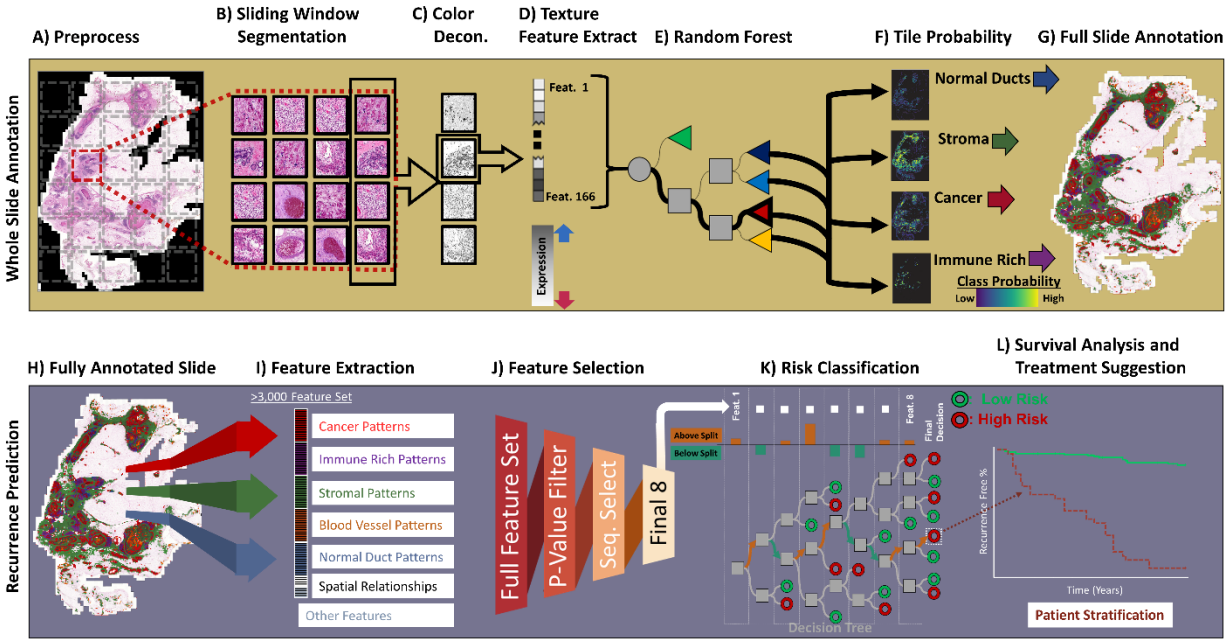


Figure 3.1: Two-step WSI method for stratifying DCIS patients based on their recurrence risk. The first step in this pipeline automatically annotates the patient's whole surgical H&E slides into prognostically informative tissue classes. For this automated annotation, the patient's whole virtual slide is (A) preprocessed through whole slide color normalization and down-sampling followed by (B) a sliding window, over the whole slide, which extracts non-overlapping image tiles which are then (C) color deconvoluted to yield the hematoxylin image from which (D) values for 166 texture features are extracted. These features are then (E) input into a random forest annotation classifier which (F) outputs a probability of each tile belonging to a specific class (malignant ducts of DCIS, surrounding breast parenchyma / ducts, blood vessels, and stromal regions with and without dense immune infiltration [immune cells occupying at least 50% of the tile area]) which are combined to produce (G) a whole slide annotation. The second step extracts tissue architecture features and features of spatial relationship between these tissue classes, from the previously annotated slides, and compiles them into what serves as the 'full slide' feature set. For prediction of DCIS recurrence risk, (H) each annotation is analyzed through (I) feature distributions, spatial features which compare distances between different classes, and other features such as region confidence. (J) The final (optimized) feature list ,

alongside the patients follow-up (recurrence) data as the labels, is used to train a (K) random forest recurrence risk classifier to predict (L) high- versus low-risk of recurrence and allows for the recommendation of optimal therapy.

3.3 Methods

3.3.1 Study population

The study population was obtained from patients diagnosed at Nottingham City Hospital (DCIS case series), spanning the period from 1989 to 2012. The training cohort comprised slides from 159 patients (127 of whom had multiple tumor blocks yielding a total of 335 slides); these slides were used for the model development **(Table 3.1)** and training. A further 185 patients (9 of whom had multiple slides, yielding a total of 199 slides) comprised an independent validation cohort for the recurrence risk classifier **(Table 3.1)**. Patients included in this study were exclusively those presenting with pure DCIS (without any invasive component/tumor in the primary biopsy whether ductal, lobular, or any special type), without bilateral disease, and treated with BCS, rather than mastectomy. The DCIS classification was initially identified through pathological records and further verified through a review of slides by 2 pathologists. Details on clinicopathological variables including size, tumor grade (classified according to the three-tier nuclear grading system [163]), comedo necrosis (defined as presence of central acellular necrosis with nuclear debris), and final margins], demographic information, and follow-up data/recurrence status were retrospectively obtained from patient medical records and validated by pathologists. Post BCS, patients at Nottingham were screened once a year until their 5^{th} year, after which they were followed up every 3 years. Recurrence-free survival (RFS) was calculated from the date of pathologic diagnosis until the first ipsilateral breast local recurrence or last follow-up. Local recurrence (either invasive or DCIS) was considered as an event. Cases with contralateral recurrences, or those who developed a second lower grade tumor, were treated as censored at the time of development to avoid mixing recurrences with new primaries.

Table 3.1: Patient characteristics. Descriptive data detailing the training and validation cohort's clinicopathological variables. The cutoff point for positive margins was 2 mm. In the training cohort the tumor size of 3 cases was not known and a patient was missing data for margin status and grade.

Baseline characteristic	Training Cohort (N = 159)	Validation Cohort (N = 185)
Patient age		
Median Age (range), years	57 (30 - 83)	59 (36 - 77)
Age <50, n (%)	26 (16.3)	23 (12.4)
Age>=50, n (%)	133 (83.7)	162 (87.6)
Menopausal Status, n (%)		
Pre	31 (19.5)	29 (15.7)
Post	128 (80.5)	156 (84.3)
Presentation, n (%)		
Screening	85 (53.5)	120 (64.9)
Symptomatic	74 (46.5)	65 (35.1)
Comedo Necrosis, n (%)		
No	60 (37.7)	34 (18.4)
Yes	99 (62.3)	151 (81.6)
Radiation, n (%)		
No	117 (73.6)	145 (78.4)
Yes	42 (26.4)	40(21.6)
Grade, n (%)		
1	25 (15.8)	0 (0.0)
2	24 (15.2)	0 (0.0)
3	109 (69.0)	185 (100.0)
Margins, n (%)		
Negative	154 (97.5)	183 (98.9)
Positive	4 (2.5)	2 (1.1)
Tumor Size		
Median Tumor Size (range), cm	1.7 (0.1-14.5)	1.7 (0.2-12.0)
Size <2.0, n (%)	88 (56.4)	101 (55.6)
Size >=2.5, n (%)	68 (43.6)	84 (45.4)
Survival status, n (%)		
Alive	109 (68.6)	159 (86.0)
Dead	50 (31.4)	26 (14.0)
Recurrence status, n (%)		
Recurrence free	122 (76.7)	153 (82.7)
Recurred	37 (23.3)	32 (17.3)

3.3.2 Tumor Slide Selection

All diagnostic slides, from the lumpectomy surgical sample, for each patient were pathologist-reviewed (IMM and MST) and the best representative (to ensure presence of adequate tumor tissue for analysis, morphological variation, and to confirm the pure DCIS diagnosis) formalin-

fixed paraffin-embedded (FFPE) tumor blocks (donor) for each patient's specimen were retrieved and included in the study. A fresh full-face section of 4 µm thickness was cut from each selected block, stained with H&E to standardize the consistency of staining quality, and again pathologist reviewed. Slide scanning was performed with a slide scanner at 40× magnification (0.24 µm/pixel) (PANNORAMIC 250 FLASH III, 3DHISTECH). Slides were reviewed for image quality and those with out of focus areas re-scanned and those with folded over tissues removed from analysis.

3.3.3 Scanning Options:

Automatic scanner mode was selected with Optovar position Pos10_1.6_1 (The Panoramic 250 Flash III, 3DHISTECH, Hungary) for good quality images scanning option (JPEG: 80, 24bit depth, and 8 bit per color channel). Flash mode was selected with 6 focus distance in field of view single layer using stitching mode without Bright-field compensation.

3.3.4 Automated Full Slide Annotation

OpenSlide software [164] allowed for 4x down-sampling of the full slides for computational feasibility. A simple graphical user interface (GUI) was developed to manually select and extract 50x50 pixel, pathologist-identified, "ground truth" image tiles from our training cohort, for training our annotation classifier to identify stroma, benign epithelial ducts (including normal breast parenchyma elements, epithelial hyperplasia and other non-malignant epithelial changes), cancerous ducts, stromal regions with dense immune infiltration (immune cells occupying at least 50% of the tile area), and blood vessels (**Fig 3.2**).

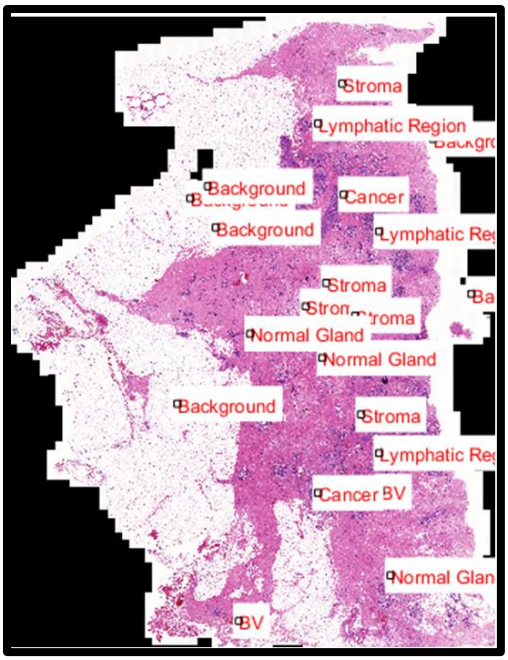


Figure 3.2: An example of the Graphical User Interface (GUI) developed to allow for ground truth annotations used for classifier training. Through this interface, a user will select regions representative of each class, from which the program will apply feature extraction from multiple 50x50 pixel windows within that region

Regions which fell outside these classes (such as areas of fat), or slide areas that were non-tissue, were given a background classification. An effort was made to select non-mixed-class (mutually exclusive) ground truth regions, which were completely surrounded by the pathologists' manual annotation, with occasional edge cases (such as intersections of classes) being labeled by the predominant class in the image tile. Each 50x50 pixel image tile used was color normalized to a standard H&E staining distribution [165] to account for specimen and staining variability, and to improve classifier performance [166]. The normalized image tiles were then color deconvoluted [167] into separate hematoxylin and eosin channels through an optical density matrix which contains the relative absorbance of each stain in the RGB color channel (**Table 3.2**). A total of 166 texture features **(Table 3.3)** were extracted from the deconvoluted hematoxylin (nuclear stain) channel for training the random forest annotation classifier.

Table 3.2: Optical density matrix. This matrix is used to deconvolute RGB H&E images into greyscales of each layer whose intensity correlated with stain absorbance.

R	G	B	
0.644	0.717	0.267	Hematoxylin
0.093	0.954	0.283	Eosin
0.636	0.001	0.772	Zero Matrix

Table 3.3: Breakdown of textural features extracted and used in region annotation.

Textural Feature Type	No. of Features	Source
Entropy	1	[168]
Gray-Level Co-occurrence Matrix (GLCO)	16	[169]
Gray-Level Run Length (GRLRL)	44	[170, 171]
Segmentation-based Fractal Texture Analysis (STFA)	45	[172, 173]
Gabor wavelet filters	60	[174-177]
Sum:	166	

To reduce same slide bias, testing of the classification ability was performed on a slide-based leave-one-out cross-validation. Each held out set of image tiles used for testing was composed of (pathologist-annotated) ground truth regions from single individual slides, such that the test fold always consisted of extracted image tiles from a slide which was not used in training. The classifier was re-trained with increasing tile N numbers in the training sets, until the cross-validated test set accuracy leveled off. To take into account the rotational invariance of the data (all of the image tiles have the same label regardless of the angle), and increase the size of the dataset, without decreasing the quality [178], we augmented the training image tiles by 4-fold, by performing diagonal flipping, 90° rotation, and the combination of the two, on all training tiles. Tissue features extracted from the augmented set of image tiles were used to train a random forest classifier [179] for tissue annotation on the slide class (development depicted in **Fig 3.3A**). The output of this random forest was the probability of the input image tile belonging to each of the five classes with the final assigned annotation determined by the highest probability.

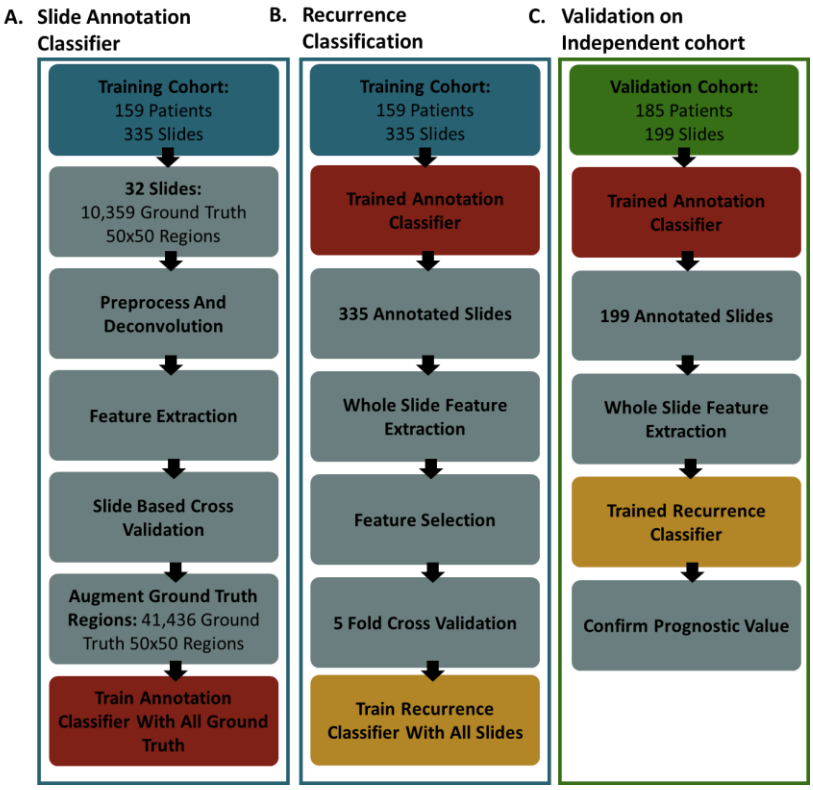


Figure 3.3: Summary of the methodology for model development. (A) The slide annotation classifier was developed using a random selection of slides within the training cohort. The ground truth regions were preprocessed and color deconvoluted so that texture features could be extracted from the hematoxylin distributions. Five-fold cross validation was performed to determine the model's classification ability after which the training set was augmented through rotation and transposition of ground truth regions and input into the final annotation classifier (red box fill). (B) To develop the recurrence classifier the training slides were first annotated through the trained annotation classifier (red box fill). The fully class-annotated slides had whole slide features extracted and selected to identify the set of features that differed most significantly between patients who recurred and recurrence-free patients. The performance of these features within a classifier was determined through 5-fold cross validation, and the full training cohort was used to train a recurrence classifier (gold box fill). (C) The prognostic value of the pipeline was confirmed on a validation cohort. Both the previously-trained annotation classifier and recurrence classifier were applied towards the patient samples in this validation cohort, and the resulting stratification of patients was evaluated.

Full slides being processed by the WSI pipeline (i.e., slides that were not previously used for training the annotation classifier) were annotated through a grid approach wherein adjacent non-overlapping 50x50 pixel image tiles (that made up the full slide) were processed **(Fig 3.1A/B/C)**, as previously detailed for the training data, their features input into the trained random forest **(Fig**

3.1D/E), and the classified image tiles stitched together **(Fig 3.1F/G)**. Additional post processing, using neighborhood voting, was performed only for the analysis of spatial features (see next section). In this approach, the class assigned to a region was amended if the sum of all its direct neighbors' trees classifications resulted in a larger proportion vote for a different annotation (**Fig 3.4** shows an example).

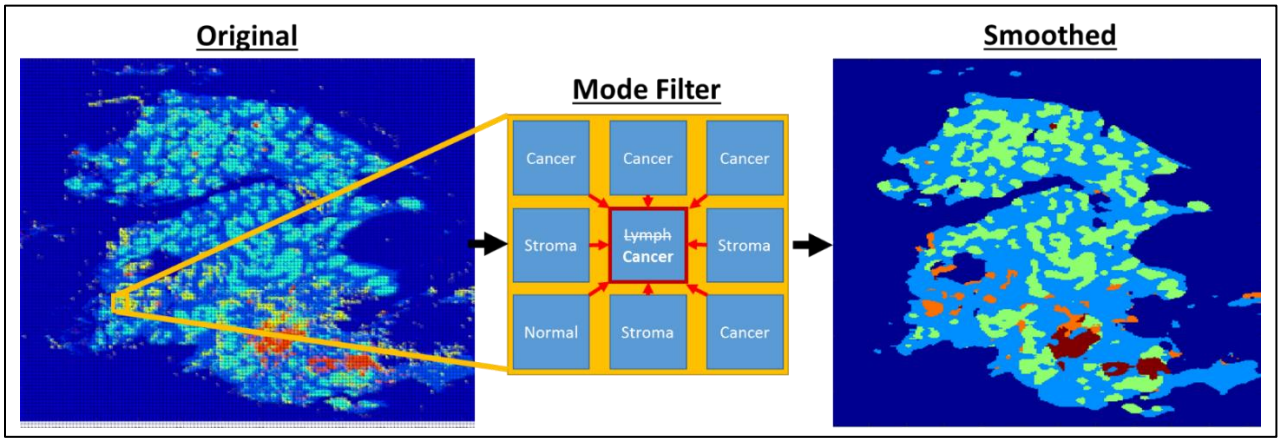


Figure 3.4: Example of region smoothing using a mode (class appearing most often) filter. In this example the middle tile was originally classified as a lymphocyte-dense region. The surrounding neighbors though, were predominantly classified as cancer; thus, the middle tile had its class changed to cancer. While this example showed the mode depending on each tile's predicted class, our model actually uses the mode of tree predictions of surrounding neighbors to adjust the middle tile classification.

3.3.5 Full-Slide Feature Optimization and Recurrence Prediction

Following automated slide annotation, a set of distinct full-slide features can be extracted **(Fig 3.1I) (Table 3.4)**. The majority (99%) of these features consist of statistical moments **(Fig 3.5)** of the 166 texture features for each annotated class and provide information on the shape of the texture features distribution for that class.

Table 3.4: Features extracted from class-annotated virtual/digital slides. The texture feature distribution statistics constitute the majority of evaluated features as they include the mean, standard deviation, skew, and kurtosis for each of the 166 textural features within each of the 5 annotated classes.

Full-Slide Feature Type	No. of Features
Distribution Statistics	3320
Spatial Distance Densities	12
Class Proportions	5
Confidence Metric	5

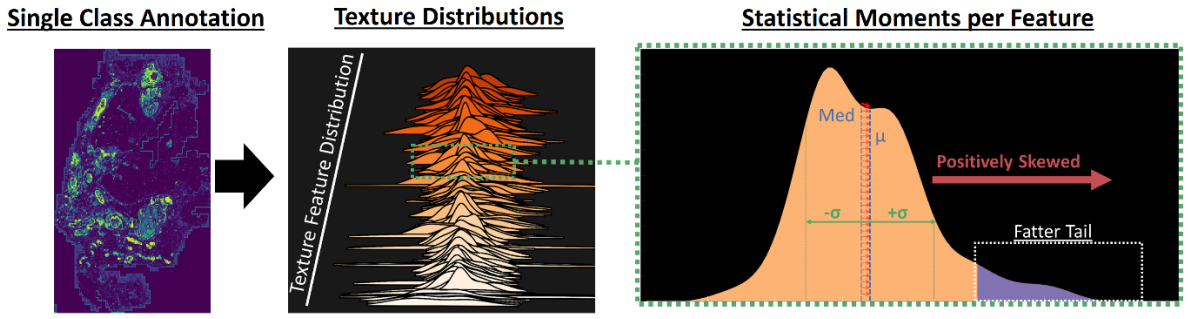


Figure 3.5: An example of the statistical moments obtained from full slide analysis. For each window for an annotated class the distribution of all texture features was computed. From each of these distributions, the mean, standard deviation, skew, and kurtosis was calculated and input as individual components of the full slide feature list.

Additionally, spatial features were derived that related the distance and size of cancer to either blood vessels or immune-rich stroma, as literature suggests that both these spatial relationships have prognostic relevance **(Fig 3.6)** [180, 181].

$$Density\ Distance = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{D_{i,j}} * A_i * A_j}{\sum A_i}$$

Figure 3.6: Density Distance Statistic. Statistic comparing the size (A) and distance (D) between all (sum) cancer (i) areas (connected regions) and either immune-rich or blood vessel (BV) areas (j), normalized (divided) by the total cancer area.

Finally, proportions of each class, such as the amount of tumor on a slide (a quantity commonly calculated in cancer staging), and average annotation confidence (calculated by averaging the number of trees which voted for each annotated class, such that low values would be given if

there was large ambiguity for any annotation on that slide) were included as features. To reduce data dimensionality and improve training time and prediction accuracy [182], a feature reduction step was performed. First, we selected a maximum follow-up time point past which a patient will be right censored and considered as a non-recurring patient. For the selected follow-up time, we filtered and sequentially selected the list of candidate features within multiple machine learning models, and using patient recurrence status as the input label, to build an optimized classifier **(Fig 3.1J)**. The performance of this final DCIS recurrence risk classifier model was then examined univariately through Kaplan-Meier curves **(Fig 3.1K/L)**.This model output a prognostic risk on a slide level, and as some patients had multiple slides (n = 127 in this cohort), a simple logic was used if a patient's slides had discordant risk classifications (i.e. a situation wherein one slide belonging to the patient classified the patient as high-risk, while another did not). In these cases, patients were given a high-risk classification **(Fig 3.7).** For comparison, we performed a separate analysis wherein we omitted these patients to test if the model performance suffered. The development of this full slide classifier is depicted in **Fig 3.3B**.

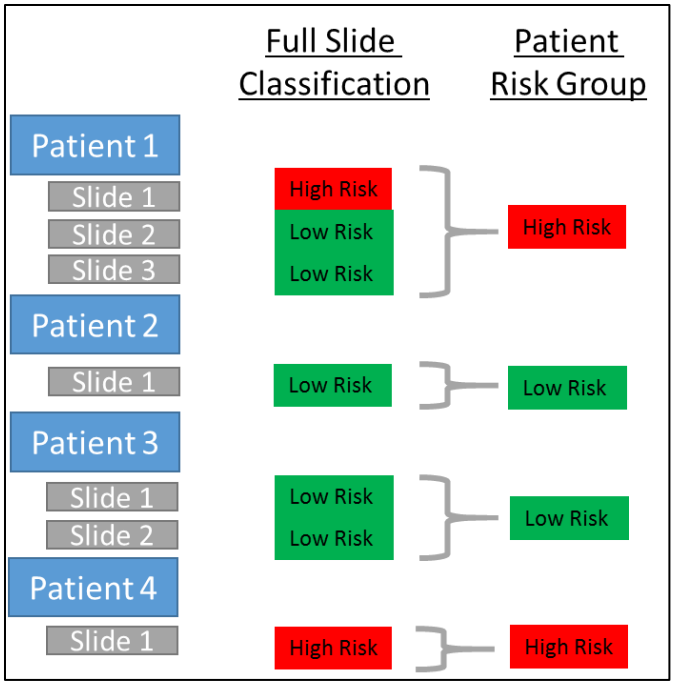


Figure 3.7: Schematic of the logic used to translate risk category of patient slides to patient risk. Patients who possessed multiple resection slides were put into a high-risk subgroup if any of their slides were classified as high-risk by the recurrence classifier.

To develop a continuous metric we utilized the selected features within a random survival forest (RSF) [183, 184] and provided each patient a 'risk score' which was equal to 1 – the RSF's output survival function for that patient for the previously selected follow-up time.

3.3.6 Time Threshold Selection:

As patients who experience recurrence after a very long follow up may possess features resembling features in patients who do not recur, it becomes imperative to select a time point that is both clinically relevant and maximizes the number of significant features that separate recurring and non-recurring groups. Therefore, t-tests were run on all of the full slide features (texture distributions, spatial features, annotation proportions, and the confidence metric) between recurrence-free and recurring (at a specified time point) patients, starting at a follow-up period of 5 years, as most patients recur within 10 years of diagnosis [185, 186]. To identify the temporal change in significant features, the same process was performed for every additional year of

follow-up until a maximum follow-up period of 25 years. The maximum follow-up time selected for our study was the one which provided the greatest number of significant features between patients who recurred by that time versus those that did not.

3.3.7 Feature and Machine Learning Model Selection:

The full features set was first filtered to those that were significantly different (t-test p-value < 0.05) between slides of patients who recurred versus those that did not. The retained features were further evaluated by sequential forward feature selection with random forest, k-nearest neighbor, and support vector machine classifiers (**Fig 3.1J**) with the goal of identifying a classifier and a subset of features that together best predict the DCIS risk recurrence. The retained features were sequentially added one by one to the training of a classifier, and the resulting classifier's performance was measured through the misclassification rate observed upon 5-fold cross validation. Features which minimized the misclassification rate the most were retained. The process of adding features was continued until there was no further improvement in classifier's performance. The selected features alongside the classifier which provided the best cross validated accuracy and HR was selected for the final DCIS recurrence risk prediction model.

3.3.8 Comparison of recurrence classifier accuracy with or without inclusion of standard clinicopathologic variables

To evaluate if our final model provides an advantage over DCIS recurrence risk prediction using available clinicopathologic parameters (comedo necrosis, size, grade, surgical margins, and patients age), we (a) performed multivariable Cox proportional hazard regression analysis using these clinicopathologic variables as covariates, and (b) concatenated the clinicopathologic variables to the 8 (optimized) features in our model, and assessed the performance of this expanded machine learning model, and the importance of each variable to the overall prediction accuracy of this model, via a variable permutation approach.

3.3.9 Prediction of DCIS recurrence risk in the context of different adjuvant therapies

We then evaluated our final model's ability to predict DCIS recurrence risk among patients who (a) were diagnosed as having high-grade DCIS (due to the clinical relevance), (b) were treated with BCS alone, and (c) received adjuvant radiotherapy after BCS. The risk of invasive recurrence was also analyzed within the classified patient risk groups.

3.3.10 Recurrence Classifier Validation

To validate the recurrence classifier's significant prognostic ability, we applied it to a second independent cohort of BCS-treated patients diagnosed with pure DCIS. The final feature-selected recurrence risk classifier model and pipeline, as previously trained for both annotation and recurrence classification, was used on 199 slides (of 185 patients, which were not included in the training cohort). The patients predicted by the model to be in the high-risk subgroup were compared with patients predicted to be in the low recurrence risk subgroup through survival analysis (Kaplan-Meier and Cox regression) of their 10-year recurrence outcomes (**Fig 3.3C**).

3.3.11 Statistical Analysis

Statistical analysis was carried out with SAS 9.4 software (Cary, NC, USA), MATLAB R2017b (Natick, MA, USA), the Python programming language (Python Software Foundation, https://www.python.org/), and R (R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/). Significance of texture feature differences between annotated classes, were analyzed with an Analysis of Variance (ANOVA) with a post-hoc Tukey-Kramer procedure. T-tests used during the initial stage of feature selection were 2-tailed. The accuracy metric was calculated as the sum of true positives and true negatives divided by the total observations. Accuracy for the training recurrence classifiers were ascertained through the average of 100 repeated 5-fold cross validation, with confusion matrices chosen from the combined testing folds of one of the repeats. When analyzing invasive or DCIS recurrence separately, patients who experienced DCIS or invasive recurrence, respectively, were treated as censored. For the training cohort, both the Kaplan-Meier survival analysis and the subsequent multivariate analyses were

performed on the 5-fold cross validated data with risk classification groups taken from the cross validated test sets [187] and significance determined using the log-rank test and Wald chi-square test respectively. Comparisons between clinicopathological proportions of training/testing versus the validation cohort was carried out through a chi-square test. Multivariate analysis was controlled for comedo necrosis, size, grade, age, and the surgical margin status. Model fit was compared with the through the Akaike Information Criterion (AIC) [188], a measure of goodness of fit/efficiency within the Cox regression statistical model. The lower the AIC value the better the likelihood. Model discrimination ability was analyzed through the Harrell's c-statistic [189] using a SAS macro [190]. Feature importance within the RF model that included standard clinicopathologic variables concatenated with the features in our recurrence classifier, was determined through 100 iterations of out-of-bag variable permutations in which the average increase in prediction error, for each variable whose value was permuted, was calculated for out-of-bag observations [179]. For fitting and optimizing survival forests, the R package 'randomForestSRC' [191] was used. When necessary, dichotomization of continuous features was performed by identifying an optimal outcome-based threshold [192]. To facilitate visualization of hazard ratios for continuous variables, z-score transformation of features was used.

3.4 Results

3.4.1 Traditional clinicopathological factors have limited DCIS recurrence risk predictive ability

The major clinicopathological characteristics for the cohorts of DCIS patients used to train and validate our model, are shown in **Table 3.1**. For the training cohort, While the recurrence rate was low (23%), the majority (84%) of recurrences occurred within the first 10 years of follow-up **(Fig 3.8)**.

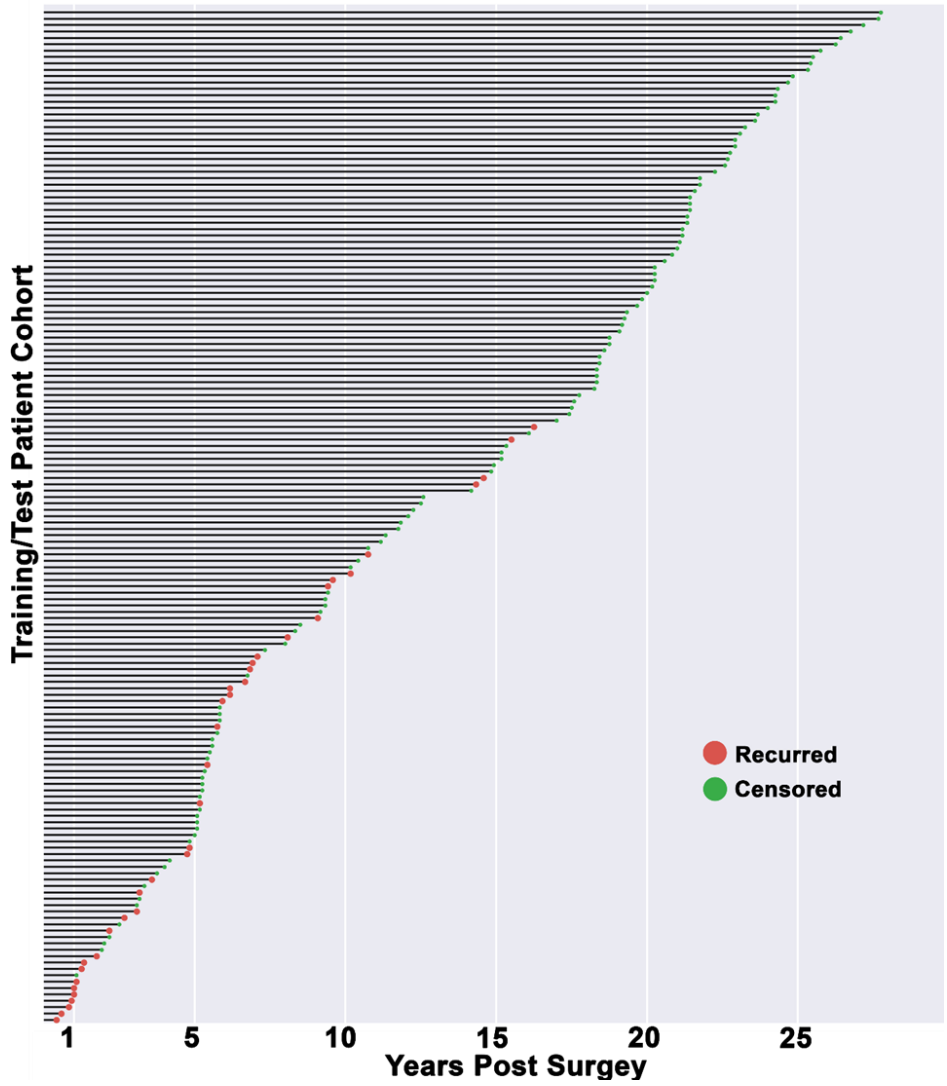


Figure 3.8: Recurrence distributions of the 159 patients in the training/test cohort, ordered according to earliest censored time or time of recurrence to last follow-up. Red points indicate a recurrence at the last follow up date while green points specify censoring.

Patients were mostly high-grade (69%), post-menopausal (80.5%), older than 50 (83.7%), and did not receive radiotherapy (73.6%). Additionally, almost all patients had complete excision with wide (>2mm) negative margins (97.5%). Within this training cohort, aside from an increased prevalence of high grade, patients who developed recurrence did not have any significant differences in the proportions of standard clinicopathological variables compared to patients who remained recurrence-free (**Table 3.5**).

Table 3.5: The distribution of baseline characteristics between patients who experienced ipsilateral recurrences versus those that did not in the training cohort. The χ2 p value signifies significant difference in proportions

Training Cohort Clinical Characteristics by Recurrence Status			
Baseline characteristic	**Recurred (N = 37)**	**Rec. Free (N = 122)**	**p value**
Patient age			
Median Age (range), years	55 (41 - 73)	57 (30 - 83)	0.3225
Age <50, n (%)	8 (21.6)	18 (14.7)	
Age>=50, n (%)	29 (78.4)	104 (85.3)	
Menopausal Status, n (%)			
Pre	9 (24.3)	22 (18.0)	0.3975
Post	28 (75.7)	100 (82.0)	
Presentation, n (%)			
Screening	16 (43.2)	69 (56.6)	0.155
Symptomatic	21 (56.8)	53 (43.4)	
Comedo Necrosis, n (%)			
No	15 (40.5)	45 (36.9)	0.6878
Yes	22 (59.5)	77 (63.1)	
Radiation, n (%)			
No	28 (75.7)	89 (72.9)	0.7419
Yes	9 (24.3)	33 (27.1)	
Grade, n (%)			
1	1 (2.7)	24 (19.8)	0.0425
2	7 (18.9)	17 (14.1)	
3	29 (78.4)	80 (66.1)	
Margins, n (%)			
Negative	35 (94.6)	119 (98.3)	0.2035
Positive	2 (5.4)	2 (1.7)	
Tumor Size			
Median Tumor Size	1.5 (0.3 - 5.0)	1.8 (0.1 - 14.5)	0.3022
Size <2.0, n (%)	23 (63.9)	65 (54.2)	
Size >=2.0, n (%)	13 (36.1)	55 (45.8)	

The validation cohort consisted of only high-grade (3) patients, but otherwise differed from the training cohort with higher rates of comedo necrosis (81.6%, p<0.0001), and slightly higher proportion of patients presenting at screening (64.9%, p=0.0316) **(Table 3.1 and Table 3.6).** Within this validation cohort only radiation has a significant proportional difference between patients who developed recurrence versus those who did not **(Table 3.7)**.

Table 3.6: Proportional differences in variable distributions between the training/testing cohort and the external validation cohort. P-values are for the chi-square test for proportions

Baseline characteristic	p value
Recurrence	0.1294
Radiation	0.2982
Margins	0.3071
Comedo Necrosis	<.0001
Grade	<.0001
Presentation	0.0316
Menopausal Status	0.3518
Size	0.7368
Age	0.2997

Table 3.7: Distribution of baseline characteristics between patients who experienced recurrence versus those that did not in the validation cohort. The χ2 p-value signifies significant difference in proportions

Validation Cohort Clinical Characteristics by Recurrence Status			
Baseline characteristic	**Recurred (N = 32)**	**Rec. Free (N = 153)**	**p value**
Patient age			
Median Age (range), years	60 (44 - 73)	59 (36 - 77)	0.5643
Age <50, n (%)	3 (9.4)	20 (13.1)	
Age>=50, n (%)	29 (90.6)	133 (86.9)	
Menopausal Status, n (%)			
Pre	5 (15.6)	24 (15.7)	0.9931
Post	27 (84.4)	129 (84.3)	
Presentation, n (%)			
Screening	20 (62.5)	100 (65.4)	0.758
Symptomatic	12(37.5)	53 (34.6)	
Comedo Necrosis, n (%)			
No	8 (25.0)	26 (17.0)	0.2876
Yes	24 (75.0)	127 (83.0)	
Radiation, n (%)			
No	30 (93.8)	115 (75.2)	0.0202
Yes	2 (6.2)	38 (24.8)	
Grade, n (%)			
1	0 (0)	0 (0)	-
2	0 (0)	0 (0)	
3	32(100)	153 (100)	
Margins, n (%)			
Negative	32 (100)	151 (98.7)	0.5155
Positive	0 (0)	2 (1.3)	
Tumor Size			
Median Tumor Size	1.7 (0.2 - 12.0)	1.8 (0.3 - 11.0)	0.3233
Size <2.0, n (%)	20 (62.5)	81 (52.9)	
Size >=2.0, n (%)	12 (37.5)	72 (47.1)	

3.4.2 Texture features differentiate significantly between annotated tissue regions

To develop a pipeline for automated annotation of various clinically relevant regions within DCIS tumor tissue sections, we found that overall accuracy leveled off at 10,359 50x50 pixel ground truth image tiles **(Fig 3.9)** from 32 training cohort slides. For developing the final annotation classifier, these ground truth areas were augmented (using rotation/transposition) to a total of 41,436 **(Fig 3.10A)**.

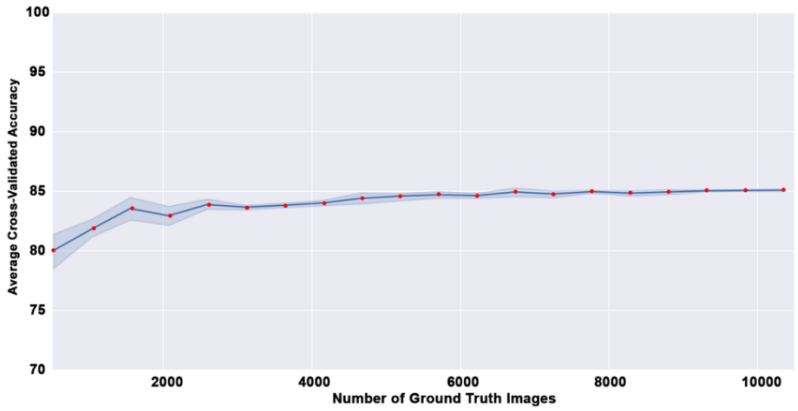


Figure 3.9: Effect of sample size used for ground truth annotation on cross-validated accuracy. Average k-fold accuracy of annotation prediction versus number of ground truth regions. Shaded bands represent 95% confidence intervals

Using the original (non-augmented) collection of ground truth regions, we observed that majority of our texture features possessed significant discriminatory ability between all annotated class combinations (**Fig 3.10B**). The classes with the most discriminatory texture features between them were cancer versus stroma (96% of features had a p-value <0.05). By contrast, texture features had the least discriminating power when it came to distinguishing stroma from blood vessels (only 80% of features were significant). Cross validation of the unaugmented ground truth collection resulted in an accuracy of 84.59%, with individual class distinction accuracies, not counting background, ranging from 75.8%-90.5% **(Fig 3.10C)** (with additional performance metrics shown in **Table 3.8**).

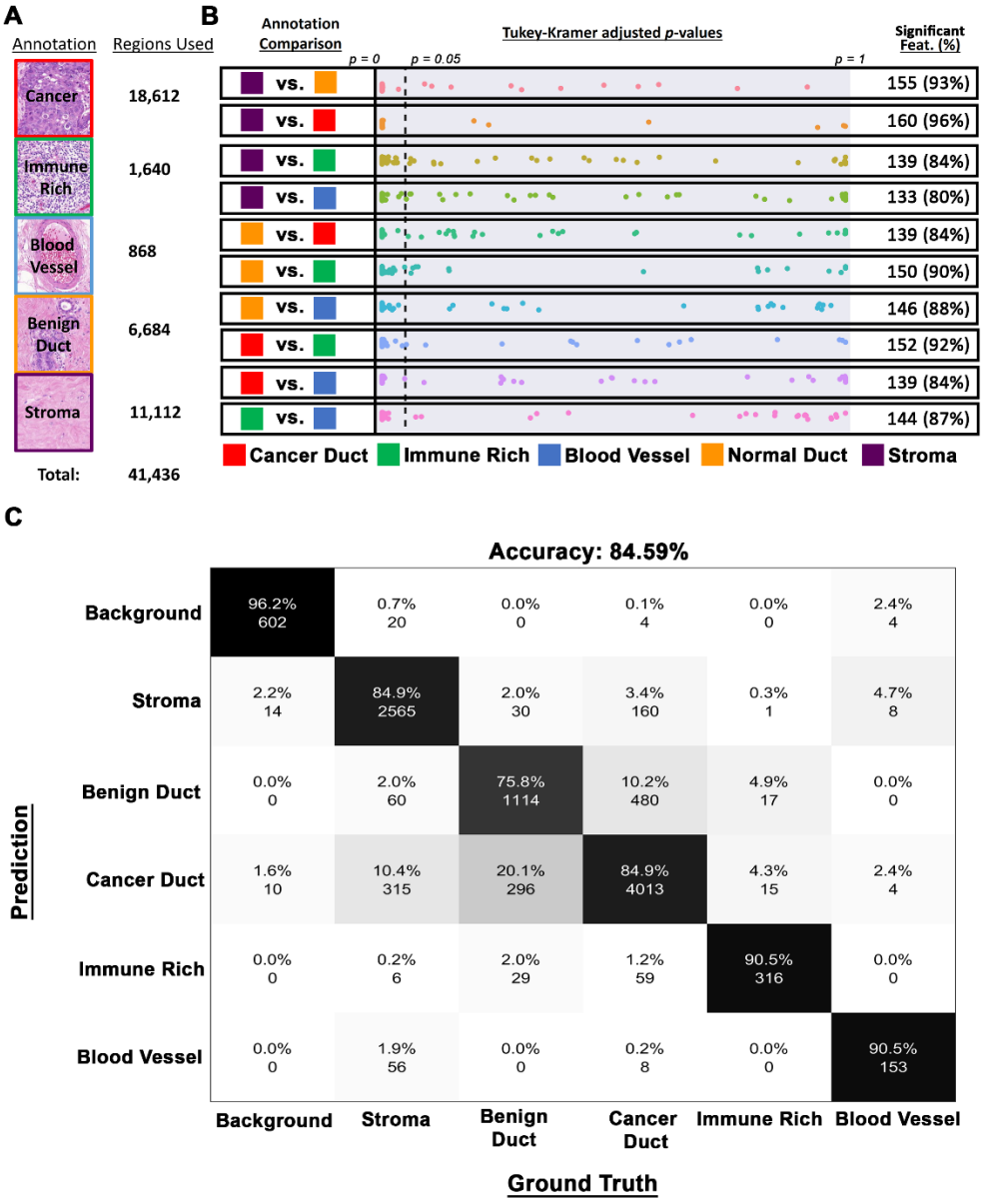


Figure 3.10: Full Slide Annotation. (A) List of annotation classes used, and representative examples, alongside the number of ground truth regions available to develop the texture-based annotation classifier. (B) Multivariate adjusted p-value (Tukey-Kramer) distributions for all 166 features (as points) between all annotated class comparisons. Reference dotted line indicates an adjusted p-value of 0.05, with features possessing significant discriminatory ability (p-values<0.05) situated on the left of it and summarized alongside. (C) Confusion matrix (which quantifies the performance of the class annotation model) comparing training ground truth data to the cross validated annotation classifier test set outputs. Analysis was performed on the original regions before 4-fold augmentation.

Table 3.8: Additional confusion matrix performance metrics for the annotation classifier.

Annotation	Sensitivity	Specificity	Precision	Recall	F-Score
Background	0.96	1.00	0.96	0.96	0.96
Stroma	0.85	0.97	0.92	0.85	0.88
Benign Ducts	0.76	0.94	0.67	0.76	0.71
Cancer Duct	0.85	0.89	0.86	0.85	0.86
Immune Rich	0.91	0.99	0.77	0.91	0.83
Blood Vessel	0.91	0.99	0.71	0.91	0.79

3.4.3 An 8-feature recurrence classifier significantly predicts recurrence risk

Thresholding at a 10-year follow-up maximized the number of significant whole slide features different between slides from patients who recurred versus those that did not progress **(Fig 3.11A)**. This follow up time is also consistent with many follow-up times in clinical studies [193] and with the fact that most DCIS patients recur within 10 years. Overall, around 1,238 (37%) whole slide features differed significantly ($p<0.05$) with a 10-year follow-up as compared to at most 25% for 5, 15, and 20-year follow-up time points.

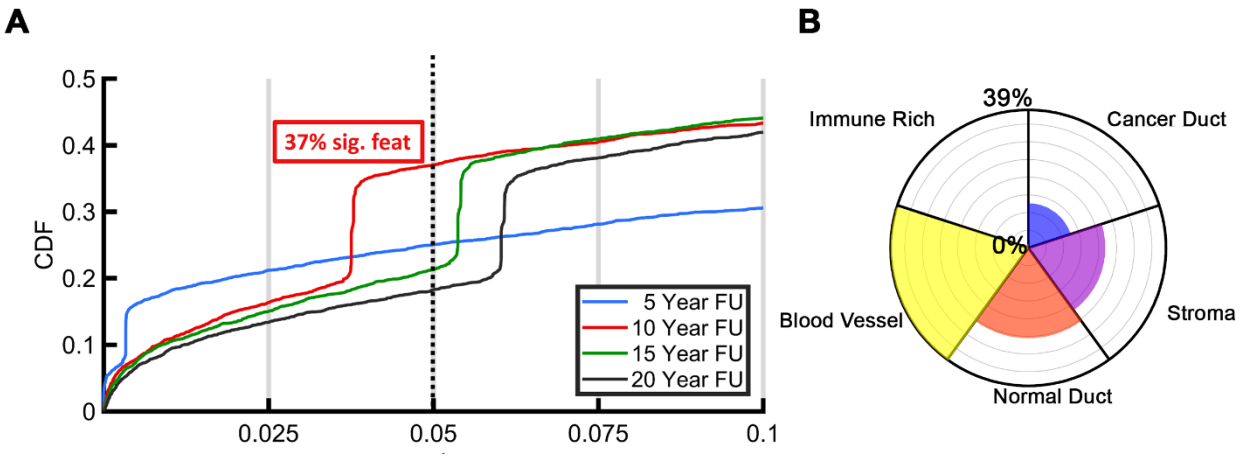


Figure 3.11: (A) The cumulative density function (CDF) of feature significance, noted by the t-test p-values, versus maximum follow-up (FU) time explored. Using 10-year recurrence, 37% of whole slide features were significantly (0.05) different between patients who developed recurrence by 10 years versus those that remained recurrence-free. (B) Within this 10-year follow-up recurrence distinction, the significant feature distribution by class difference is shown

in a radar plot, with the max fill (blood vessel features) indicating 39% of the filtered total significant features.

Testing 10-year recurrence risk model built with these filtered features (i.e. using all significant features prior to the sequential removal step in **Fig. 3.1J**), resulted in an average 5-fold cross validated accuracy around 80% ,regardless of the ML model **(Table 3.9)**, and a random forest high-risk group possessing a hazard ratio of 3.19 **(Fig 3A)**, almost equivalent to the performance of using the full feature set (accuracy: 80.8%; HR: 3.13). Interestingly, among the filtered whole slide features, the majority (88%) stemmed from non-cancer annotations and only 1% came from differences in lymphocyte dense properties between patients **(Fig 3.11B)**.

Table 3.9: Comparison of multiple machine learning algorithms to select the best model (and its associated features) for the recurrence classifier. 'No annotation' indicates the performance of a random forest model built without considering classes obtained from the first annotation step. Optimized models reflect performance after selection of optimal set of features. For each ML model, the model accuracy and high-risk group hazard ratio upon using either the full feature set or the optimized feature set, are shown.

Model	Average Accuracy (std.)	Average Hazard Ratio (std.)
Optimized Random Forest (RF)	**0.86 (0.010)**	**8.55 (1.272)**
Full Feature Non Annotated RF	0.78 (0.012)	3.08 (0.496)
Optimized Non Annotated RF	0.79 (0.008)	2.82 (0.283)
Full Feature SVM Model	0.80 (0.002)	1.21 (0.174)
Optimized SVM Model	0.80 (0.037)	7.27 (4.537)
Full Feature KNN Model	0.79 (0.004)	1.59 (0.347)
Optimized KNN Model	0.80 (0.009)	4.31 (0.537)

Choosing the most prognostic variables through the sequential forward selection though, resulted in half of the features being derived from cancer areas (**Fig 3.12B** with additional feature details in **Table 3.10**). The final 8-feature model lowered the misclassification rate to 0.101, achieved an average (of 100 iterations) cross validated accuracy above 86%, and yielded a model that robustly stratified the DCIS patients in our training cohort and identified a high-risk group with 8.5x higher recurrence risk by 10 years **(Fig 3.12A).** Fig 3C illustrates a typical Kaplan Meier survival curve from one of the model training iterations (out of the total 100) of the combined cross-validated test

sets. The slides classified into the high-risk group carry a recurrence-free survival (RFS) of only 24% compared to the 90% seen in the low-risk group. To show the importance of the initial machine learning annotation step (Fig 1A-G), a 'non-annotated' RF model built (with feature selection) without utilizing annotation classification (simply using the overall texture statistical moments of all the areas of the slides) resulted in a significantly lower accuracy (79%) and HR (2.82) **(Table 3.9)**.

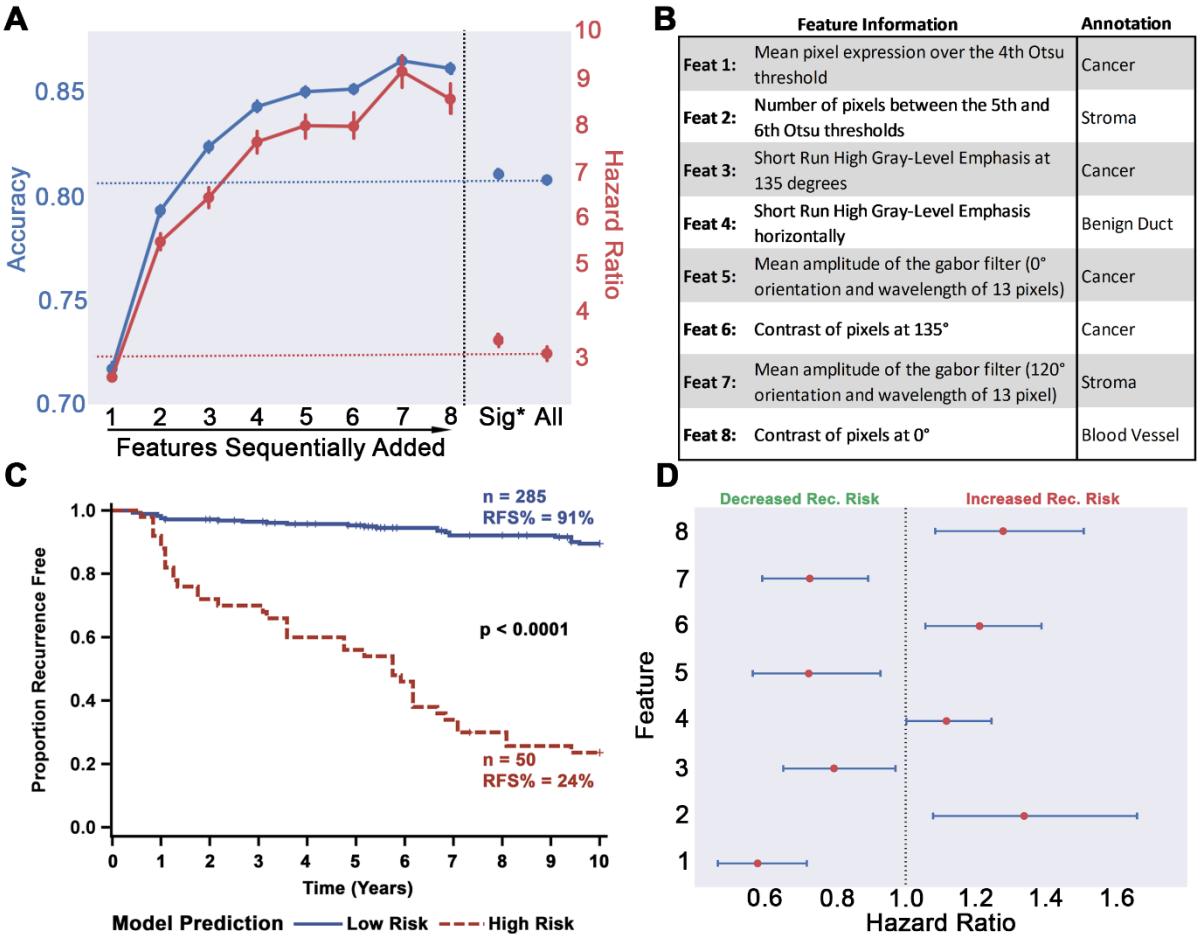


Figure 3.12: Full Slide Feature Selection for development of recurrence classifier. (A) The change in model accuracy and high-risk group hazard ratio with the sequential addition of features. The reference hazard ratio and accuracies, based on the model with all features, is shown in red and blue horizontal dashed lines respectively. The model which included all filtered features (Sig: p<0.05) is also shown for comparison. Bars on markers indicate 95% confidence intervals. (B) General feature descriptions, and the annotations from which they stem from, of the final 8-feature recurrence classification model. (C) Kaplan-Meier curves showing stratification of patient slides by the final recurrence classifier model. Data shown is*

based on slides used for the training cohort, wherein the test sets for each selected cross validated iteration were combined. Significance was measured using the log-rank test. (D) Univariate HR of the selected features, z-score transformed for illustrative purposes. All variables are significant and blue horizontal lines depict 95% confidence intervals. The fact that none of the confidence intervals cross the HR=1.0 reference line shows that these features are highly and unequivocally significant.

The 8 features selected for the final model, when evaluated as continuous variables in univariate analysis, all provided significant prognostic value, with half being associated with higher risk of recurrence and the other half providing a protective effect **(Fig 3.12D)**.

Table 3.10: Feature characteristics of the final 8-feature recurrence classification model. The significance shown is based on the t-test for each feature between patients who experienced recurrence within 10 years and those that did not. The misclassification cost is computed sequentially (for e.g., the misclassification cost for feature 3 is the cost for a model which includes features 1, 2 and 3). SFTA: Segmentation-based Fractal Texture Analysis, GLRL: Grey Level Run Length, GLCO: Grey Level Co-Occurrence

	Features Added							
Feature Information	Feat. 1	Feat. 2	Feat. 3	Feat. 4	Feat. 5	Feat. 6	Feat. 7	Feat. 8
Significance (p value)	0	0.0146	0.0151	0.0376	0.008	0.0041	0.0014	0.0146
Misclassification Cost After Add	0.256	0.194	0.167	0.14	0.128	0.125	0.113	0.101
General Category	STFA	STFA	GLRL	GLRL	Gabor	GLCO	Gabor	GLCO
Statistical Moment	Mean	SD	Mean	Mean	Skew	Kurt.	Skew	Skew
Tissue Annotation	Cancer	Stroma	Cancer	Normal	Cancer	Cancer	Stroma	BV
Overexpressed In	No Rec.	Rec.	No Rec.	Rec.	No Rec.	Rec.	No Rec.	Rec.

Dichotomizing patients into groups using the 2 mean cancer features (consisting of feature #1 and #3, as the mean moment and cancer annotations are the most intelligible combination for texture-based analyses), for interpretive purposes, showed conflicting effects. Alone, feature #1, calculates the hematoxylin staining, or blue color intensity, per pixel (or point) within the malignant ductal profile areas (above a certain Otsu method autogenerated threshold [194]) **(Fig 3.13A-D)**, very significantly stratified patients into two distinct risk groups **(Fig 3.13E)**, while feature #3 was unable to do so **(Fig 3.14)**. However, if patients were first split into high- and low-risk groups

through feature #1 **(Fig 3.14B)** followed by another stratification using feature #3, a significant difference in survival between the two subgroups was increased when compared to the stratification by feature #1 alone **(Fig 3.14C)**, showing the dependency of variables for maximizing prognostic relevance (High Risk Group HR for feature #1 alone=3.017, High Risk Group HR for features #1+#3=7.308).

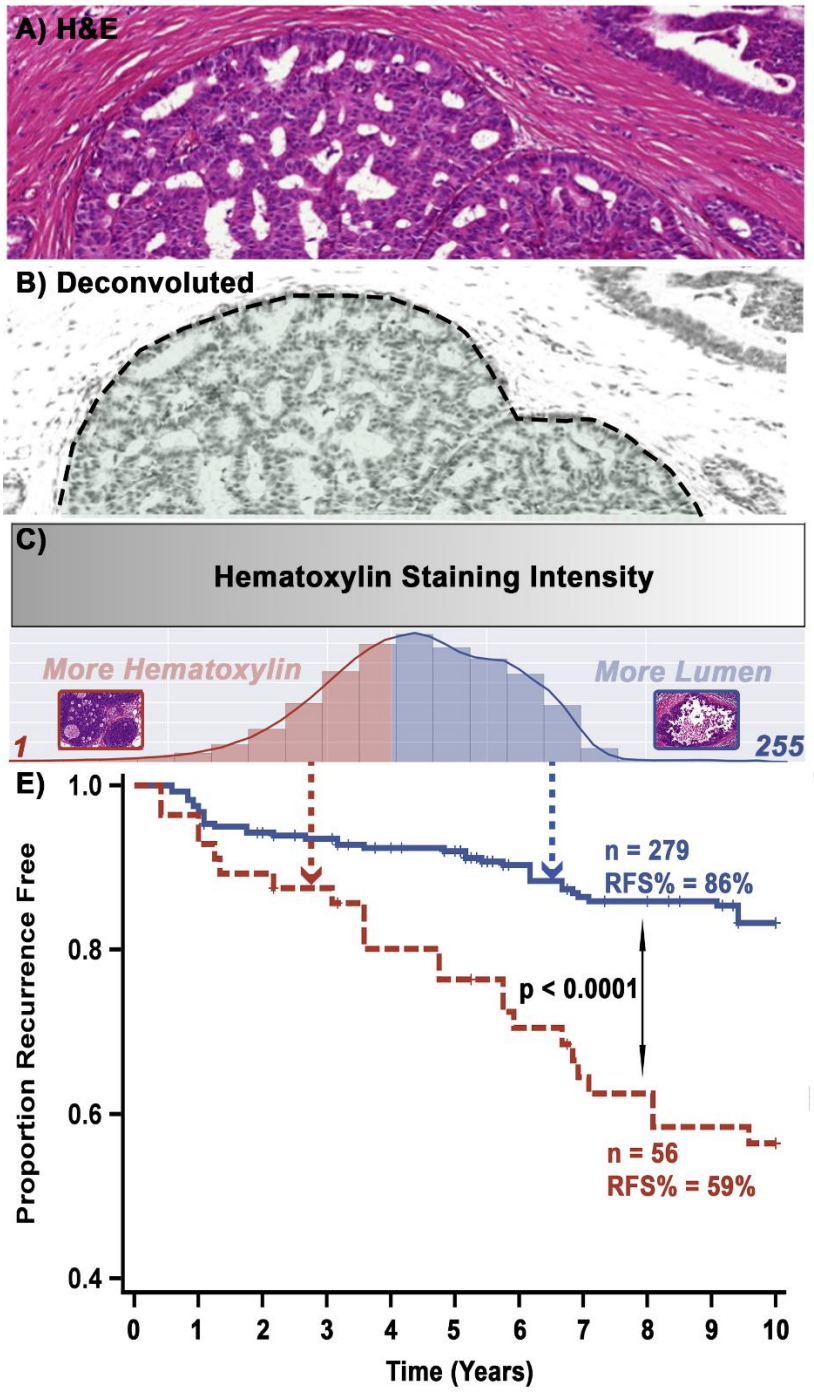


Figure 3.13: Interpretation and prognostic relevance of the most prognostic feature in our 8-feature DCIS recurrence risk prediction model. (A) An example "cancer" region with a cribriform architecture in an H&E-stained slide (prior to deconvolution). (B) The region shown in (A) after hematoxylin deconvolution. (C) Intense hematoxylin staining (relative to the image tile section) is represented by a grey level intensity of 1, while no staining is depicted by a grey level value of 255. The adaptive Otsu thresholds by progressively using a higher threshold. Therefore, if the cancer region has lumens, it would yield a higher average intensity (more white pixels) as compared to a solid pattern (no white pixels). Using an optimized threshold of 208 (D), it is

observed that full slides whose cancer regions have an average feature #1 above that cutoff recur significantly less than patients below that threshold (E).

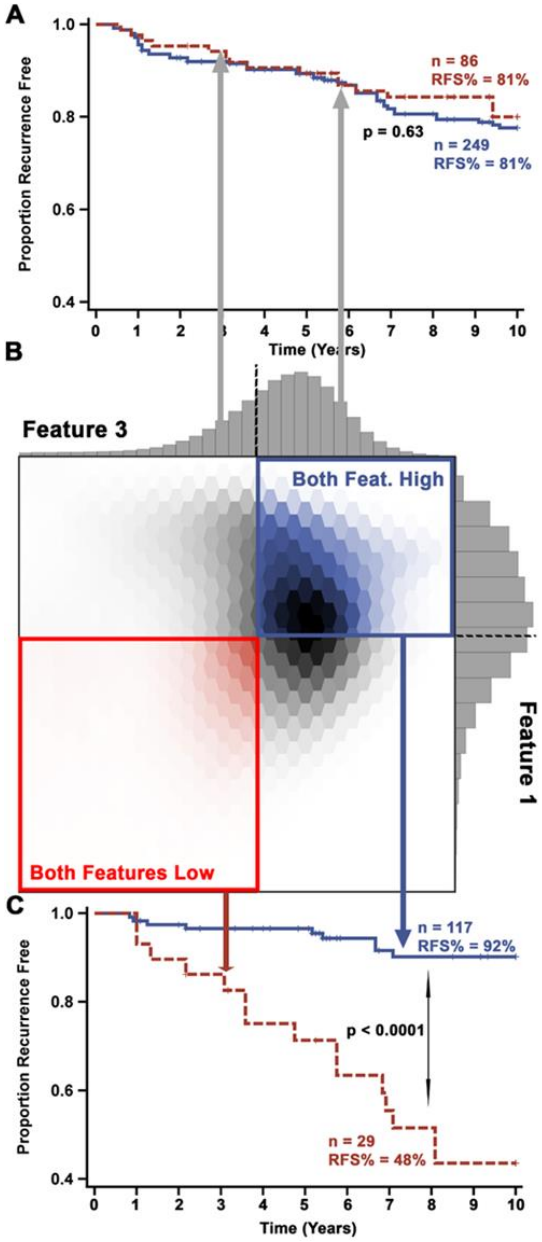
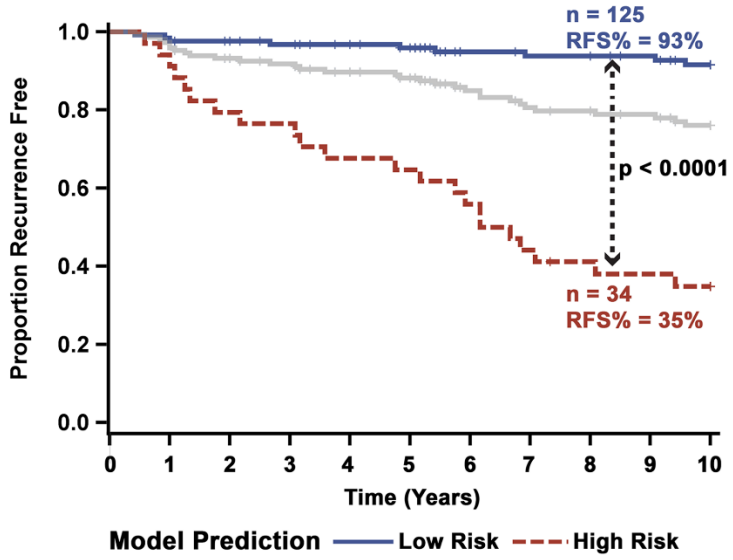


Figure 3.14: Combination of features produces optimal stratification. (A) Optimally stratifying patients by feature #3 provides little individual prognostic benefit. However, if patients are first split by feature #1, followed by feature #3 (B), a very significant survival difference can be observed between the high- and low-risk groups (C).

Applying the recurrence classifier based on the final 8 features at the patient level showed that the classifier significantly stratified the patients in the training cohort ($p<0.0001$). Patients

classified to the high-risk group (N=34) had an RFS of only 35%, compared to the 93% seen in patients in the low-risk group (N=125) **(Fig 3.15A)**. This significant stratification remained even if the analysis was performed after omitting patients with discordant slide classifications. This iteration had a univariate high-risk hazard ratio of 11.6 and retained its very high significance when controlling for necrosis, size, grade, margins, radiation therapy, and patient age **(Fig 3.15B)**.

A



B

Training Cohort Cox Regression							
Variables		Univariate Analysis			Multivariate Analysis		
		Hazard Ratio	95% Confidence interval	*P*-value	Hazard Ratio	95% Confidence interval	*P*-value
Recurrence Free Survival							
Predictive Model	High Risk vs. Low	11.617	5.334 - 25.303	<0.0001	12.542	5.435 - 28.946	<0.0001
Comedo Necrosis	Present vs. Absent	0.839	0.411 - 1.713	0.6302	0.651	0.255 - 1.661	0.3689
Size	per mm	0.979	0.957 - 1.003	0.0812	0.982	0.958 - 1.007	0.1533
Grade	3 vs. 1 and 2	1.252	0.560 - 2.801	0.5836	1.305	0.452 - 3.771	0.6223
Margin	Positive vs. Negative	1.226	0.167 - 8.994	0.8412	0.617	0.075 - 5.071	0.6536
Age	Per year	0.981	0.942 - 1.022	0.3669	1.007	0.965 - 1.052	0.7385
Radiotherapy	Yes vs. No	1.073	0.479 - 2.403	0.8636	1.113	0.439 - 2.818	0.8219

Figure 3.15: A) 5-fold Cross validated Kaplan-Meier curves of the training cohort. Significance is measured using the log-rank test and the grey line represents the un-stratified full cohort. B) Univariate and multivariate Cox regression analysis comparing the influence of common clinicopathological variables alongside the 8-feature recurrence risk prediction model for recurrence-free survival, on the training set (after 5-fold cross validation)

None of the clinical variables in the original cohort showed significant risk stratification ability in multivariate analysis, although grade was significant univariately **(Fig 3.15B and Fig 3.16)**. Moreover, the model provided a superior c-index (0.77) and model fit (AIC = 239.8) to the clinical

variables **(Fig 3.17)**. Additionally, select clinical variables neither improved the overall model nor add any prognostic relevance individually **(Fig 3.18)**.

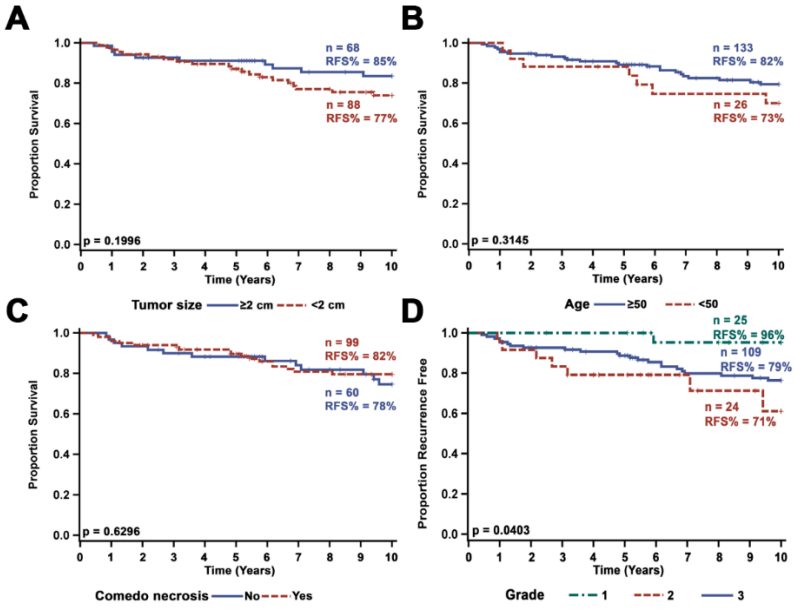


Figure 3.16: Stratification of patients in training cohort using standard clinical variables. Cross validated Kaplan-Meier curves of patient outcomes (Recurrence-free survival, RFS) stratified based on (A) tumor size, (B) patient age, (C) comedo necrosis status, and (D) Nottingham grade. Significance is measured through the log-rank test.

A

Variable	Harrell's c-statistic (95% CI)
Predictive Model	**0.77 (0.69 - 0.85)**
Grade	0.52 (0.44 - 0.60)
Margin Status	0.50 (0.48 - 0.52)
Necrosis	0.52 (0.43 - 0.60)
Radiation	0.51 (0.43 - 0.58)
Age	0.56 (0.46 - 0.67)
Size	0.57 (0.48 - 0.67)

B

Model	Akaike Information Criterion
Null Model	278.33
Predictive Model	**239.80**
Clinopathological Variables	276.93

Figure 3.17: (A) The Harrell's c-statistic and 95% confidence interval for the 8-feature model and common clinopathological variables in the training cohort. (B) The Akaike Information Criterion

(AIC) comparing the fit of a null model (no variables), the 8-feature model, and a model composed of the common clinopathological variables (Grade, margins status, necrosis, radiation, age, and size). The lower the AIC value the better the model fits the recurrence data

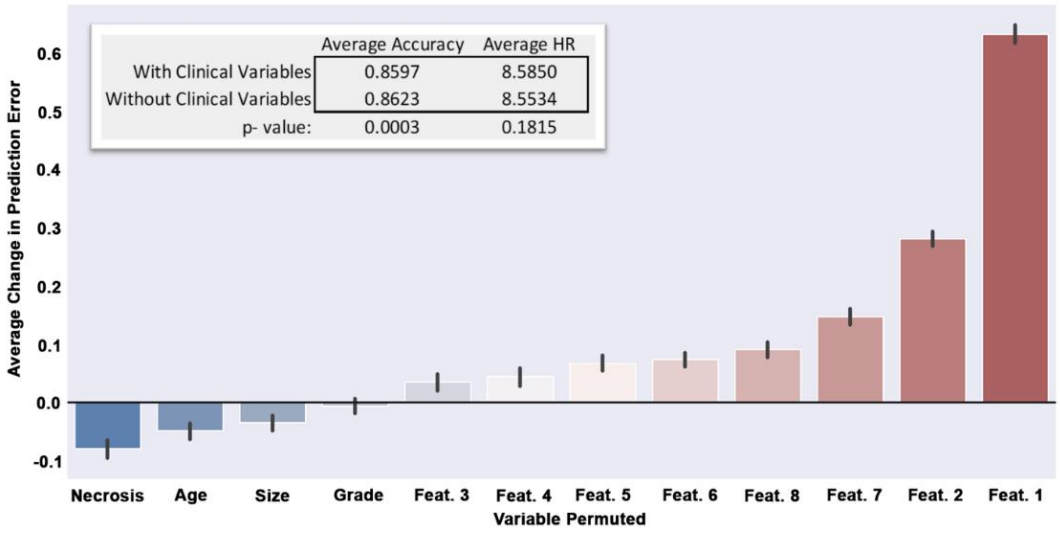


Figure 3.18: Impact of clinical features on model performance when clinical variables are concatenated with the 8 features of the recurrence classifier, within a random forest model. Averaged out-of-bag feature importance (and 95% confidence intervals) from 100 models shows that clinical features do not contribute positively to the overall performance of the model. Feature importance (i.e., how heavily the model relies on each given feature for the output prediction) is defined as the change in prediction error when the values of those variables are permuted (to, in effect, break the relationship between the feature and the model outcome) across out-of-bag observations. Hence larger error changes correspond to more vital variables. Insert: Average cross-validated accuracy and hazard ratios of models built with and without clinical variables show (yes/no) significant differences

Notably, the same model was able to significantly stratify high grade DCIS patients **(Fig 3.19A)**, low/intermediate grade DCIS patients **(Fig 3.19B)**, the subset of all patients who received adjuvant radiation therapy, and all patients treated with BCS alone **(Fig 3.19C-D)** into subgroups with high and low recurrence risks. Additionally, the model was able to identify patients at high-risk for both invasive **(Fig 3.20)** and DCIS recurrence (**Fig 3.21)**, even when controlling for clinicopathological variables. Converting the 8 selected features into a continuous risk score,

through an RSF, resulted in a significant (p<0.0001) prognostic model, with each unit increase providing incremental 5% higher 10-year recurrence risk **(Table 3.11).**

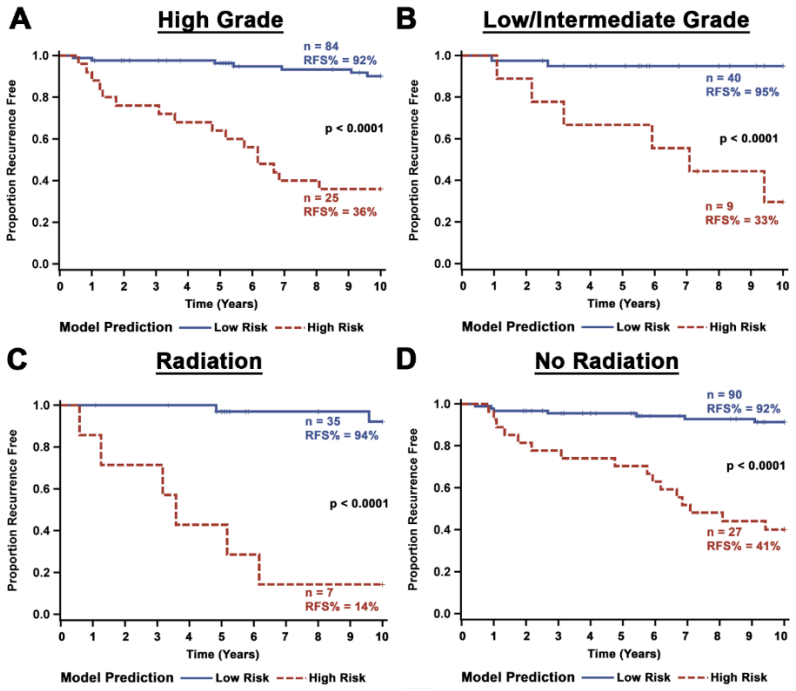
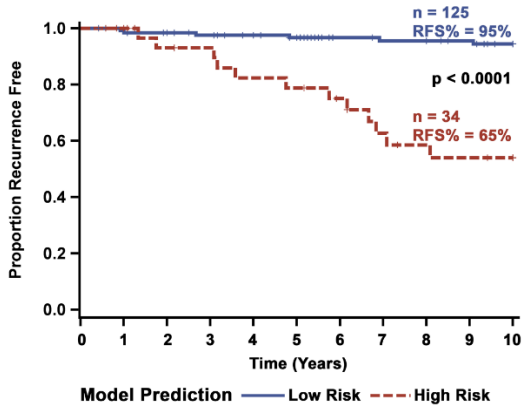


Figure 3.19: Cross validated Kaplan-Meier curves of patients within the training cohort, developed by combining the testing sets for a cross validated iteration. (A) The recurrence classifier model used with Grade 3 patients' slides only. (B) The recurrence classifier model used with Grade 1 and 2 patients' slides only. (C) Recurrence classifier used on slides from patients who received adjuvant radiation and (D) Recurrence classifier used on slides taken from patients treated with BCS alone.

A



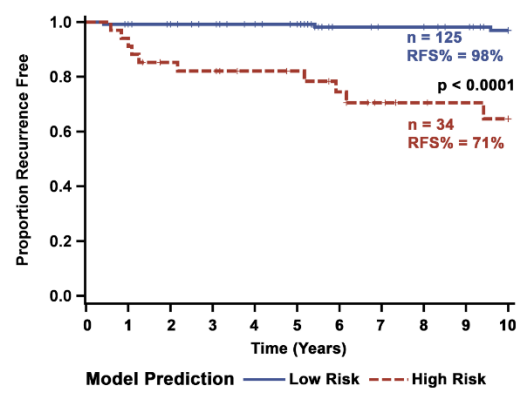
B

Training Cohort Cox Regression for Invasive Recurrence

Variables		Univariate Analysis			Multivariate Analysis		
		Hazard Ratio	95% Confidence interval	*P*-value	Hazard Ratio	95% Confidence interval	*P*-value
Recurrence Free Survival							
Predictive Model	High Risk vs. Low	9.837	3.681 - 26.288	<.0001	8.615	3.145 - 23.597	<0.0001
Comedo Necrosis	Present vs. Absent	0.954	0.370 - 2.462	0.9227	0.54	0.152 - 1.918	0.3406
Size	per mm	0.985	0.959 - 1.012	0.2696	0.3495	0.957 - 1.016	0.8754
Grade	1 vs. 2	-	-	0.9927	-	-	0.9919
Grade	1 vs. 3	-	-	0.9925	-	-	0.9919
Margin	Positive vs. Negative	-	-	0.9929	-	-	0.9946
Age	Per year	1.015	0.958 - 1.075	0.6138	1.038	0.974 - 1.106	0.2469
Radiotherapy	Yes vs. No	0.617	0.178 - 2.136	0.4460	0.745	0.191 - 2.908	0.6719

Figure 3.20: (A) Cross validated Kaplan-Meier curves of patients within the training cohort stratified by the trained recurrence classifier and using only invasive recurrence as an event. Significance is measured through the log-rank test. (B) Univariate and multivariate Cox regression analysis comparing the influence of common clinicopathological variables alongside the 8-feature recurrence risk prediction model for invasive recurrence-free survival, on the training set

A



B

Training Cohort Cox Regression for DCIS Recurrence

Variables		Univariate Analysis			Multivariate Analysis		
		Hazard Ratio	95% Confidence interval	*P*-value	Hazard Ratio	95% Confidence interval	*P*-value
Recurrence Free Survival							
Predictive Model	High Risk vs. Low	15.123	4.139 - 55.252	<.0001	18.181	3.563 - 92.780	0.0005
Comedo Necrosis	Present vs. Absent	0.705	0.237 - 2.099	0.5300	0.916	0.208 - 4.025	0.9074
Size	per mm	0.967	0.922 - 1.013	0.1543	0.971	0.922 - 1.023	0.2721
Grade	1 vs. 2	4.201	0.434 - 40.652	0.2152	1.596	0.139 - 18.299	0.7071
Grade	1 vs. 3	2.283	0.289 - 18.026	0.4336	0.862	0.081 - 9.159	0.9019
Margin	Positive vs. Negative	3.221	0.418 - 24.830	0.2618	1.537	0.133 - 17.772	0.7306
Age	Per year	0.944	0.892 - 1.000	0.0491	0.963	0.902 - 1.029	0.2688
Radiotherapy	Yes vs. No	1.927	0.629 - 5.901	0.2509	2.304	0.625 - 8.497	0.2102

Figure 3.21: (A) Cross validated Kaplan-Meier curves of patients within the training cohort stratified by the trained recurrence classifier and using only DCIS recurrence as an event. Significance is measured through the log-rank test. (B) Univariate and multivariate Cox regression analysis comparing the influence of common clinicopathological variables alongside

the 8-feature recurrence risk prediction model for DCIS recurrence-free survival, on the training set

Table 3.11: Univariate cox regression analysis of the impact that a random survival forest, trained with the 8 selected features, has on both the training (through combining the cross-validation test sets) and validation cohorts. Each unit risk is produced through the RSFs output 10-year recurrence survival function

Variables		Univariate Analysis		
		Hazard Ratio	**95% Confidence interval**	***P*-value**
Recurrence Free Survival				
Predictive Model (Training)	Per unit risk	1.051	1.037 - 1.065	<0.0001
Predictive Model (Validation)	Per unit risk	1.051	1.003 - 1.102	0.0358

3.4.4 Validation study confirms prognostic value of the 8-feature recurrence risk classifier

We proceeded to validate our 8-feature DCIS recurrence risk prediction model in an independent validation cohort of DCIS cases (n=185 from Nottingham University Hospital). Analyzing individual slides (treating each slide as an individual patient) using our previously-trained 8-feature classifier resulted in highly significant stratification of the validation cohort into high- and low-risk groups with regard to their RFS **(Fig 3.22)**. Patient-wise analysis led to further improvement in recurrence risk prediction. Ninety-two percent of patients classified into the low-risk stayed recurrence-free for 10 years, compared to only 54% of patients who are classified as high-risk **(Fig 3.23A)**. Removing patients with discordant cases did not adjust the model stratification.

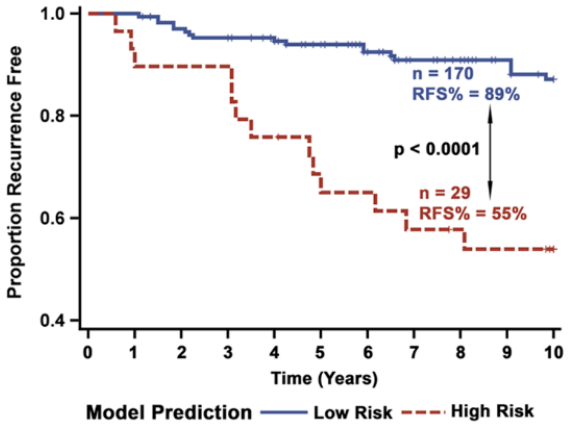
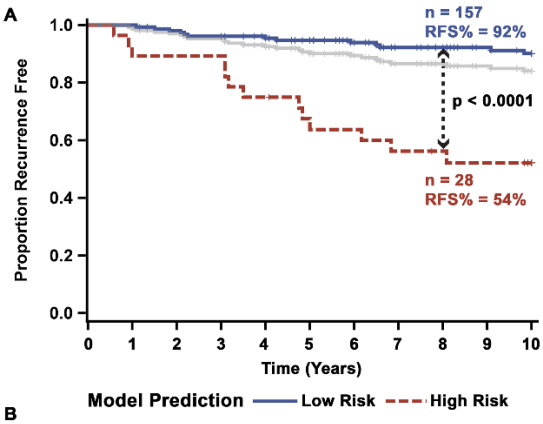


Figure 3.22: Kaplan-Meier curves of slides within the validation cohort stratified by the trained recurrence classifier model. Significance is measured through the log-rank test

While lower than the training/test cohort, the univariate hazard ratio of this classifier on the validation cohort patients is 6.4 (p<0.0001) and over 6.8 (p<0.0001) when controlling for necrosis, size, margin status, and age **(Fig 3.23B)**.

A



B

Validation Cohort Cox Regression

variables		Univariate Analysis			Multivariate Analysis		
		Hazard Ratio	95% Confidence interval	*P*-value	Hazard Ratio	95% Confidence interval	*P*-value
Recurrence Free Survival							
Predictive Model	Low Risk vs. High	6.39	2.961 - 13.789	<0.0001	6.752	3.066 - 14.865	<0.0001
Comedo Necrosis	Absent vs. Present	0.805	0.323 - 2.006	0.6417	1.087	0.423 - 2.798	0.862
Size	per mm	1.003	0.986 - 1.021	0.7131	1.008	0.990 - 1.026	0.3941
Margin	Negative vs. Positive	-	-	0.9964	-	-	0.9918
Age	Per year	1.031	0.979 - 1.085	0.2472	1.034	0.981 - 1.089	0.2106
Radiotherapy	Yes vs. No	0.372	0.088 - 1.579	0.18	0.397	0.091 - 1.733	0.2192

Figure 3.23: Validation of 8-feature DCIS recurrence risk prediction model in an independent validation cohort. (A) Kaplan-Meier curves showing robust stratification of patients in the validation cohort into high-risk of recurrence and low-risk of recurrence subgroups. Significance was measured using the log-rank test and the grey line represents the un-stratified full validation cohort. (B) Univariate and multivariate Cox regression analysis of the validation cohort comparing the influence of common clinicopathological variables on the recurrence risk predictive 8-feature model, for 10-year recurrence-free survival.

Once again, the model provided superior discrimination (c-index=0.69) and model fit (AIC=243) as compared to the clinicopathological variables **(Fig 3.24)**. Even though this validation cohort had very few patients recurring after radiotherapy, the 8-feature recurrence risk-predictive model was able to significantly predict long-term outcomes after radiotherapy **(Fig 3.25A)**.

A

Variable	Harrell's c-statistic (95% CI)
Predictive Model	**0.69 (0.59 - 0.78)**
Margin Status	0.50 (0.50 - 0.51)
Necrosis	0.53 (0.44 - 0.61)
Radiation	0.55 (0.49 - 0.61)
Age	0.55 (0.44 - 0.67)
Size	0.48 (0.37 - 0.60)

B

Model	Akaike Information Criterion
Null Model	260.53
Predictive Model	**243.00**
Clinopathological Variables	264.64

Figure 3.24: (A) The Harrell's c-statistic and 95% confidence interval for the 8-feature model and common clinopathological variables in the validation cohort. (B) The Akaike Information Criterion (AIC) comparing the fit of a null model (no variables), the 8-feature model, and a model composed of the common clinopathological variables (Margins status, necrosis, radiation, age, and size). The lower the AIC value the better the model fits the recurrence data.

Additionally, a clear high-risk subgroup was identified among patients treated with only BCS **(Fig 3.25B)**. Censoring the 8 patients whose recurrence was DCIS (rather than invasive disease) resulted in robust identification of patients at high-risk of recurrence as invasive disease, regardless of other clinicopathological variables **(Fig 3.26)**. Furthermore, although the number of events was limited, the model significantly identified a group at high risk of DCIS recurrence **(Fig 3.27)**.

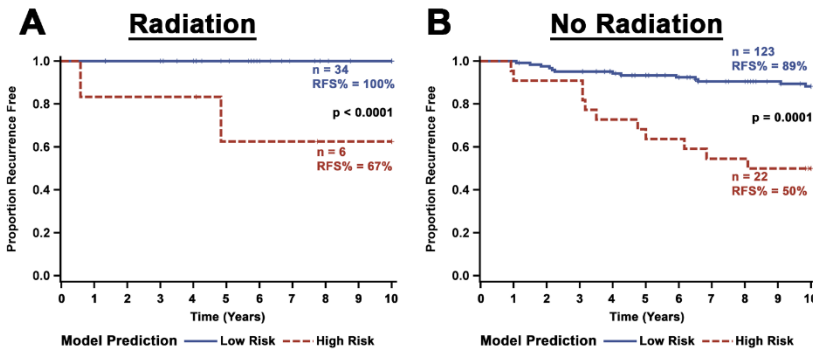
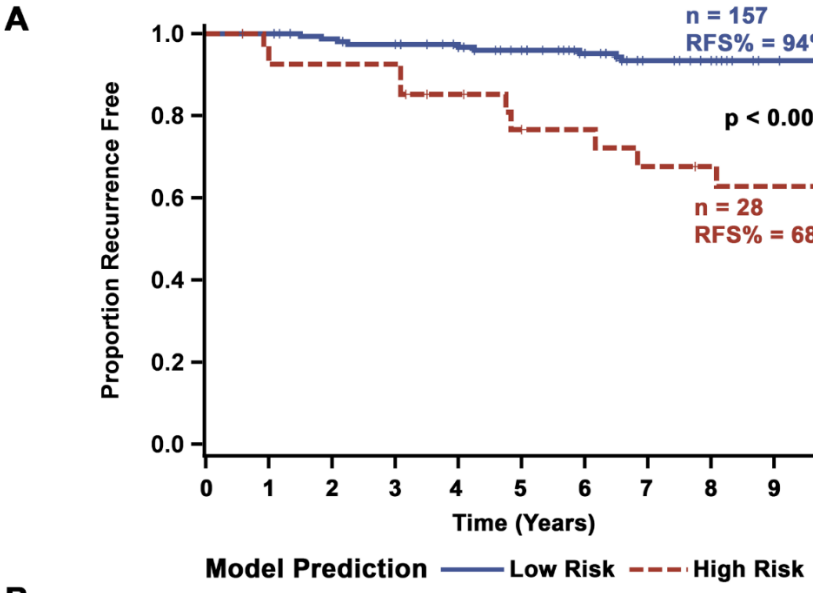


Figure 3.25: Kaplan-Meier curves of patients within the validation cohort, developed by combining the testing sets for a cross validated iteration. (A) Recurrence classifier model used on slides from patients who received adjuvant radiation, (B) Patients who were treated with BCS alone, and (C) using invasive recurrence as the event. Significance is measured through the log-rank test.

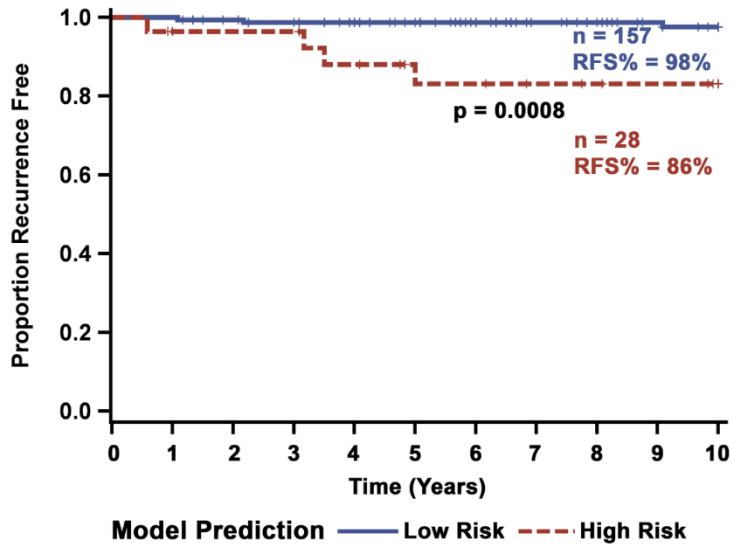


B

Validation Cohort Cox Regression for Invasive Recurrence							
Variables		Univariate Analysis			Multivariate Analysis		
		Hazard Ratio	95% Confidence interval	*P*-value	Hazard Ratio	95% Confidence interval	*P*-value
Recurrence Free Survival							
Predictive Model	High Risk vs. Low	5.759	2.339 - 14.178	0.0001	6.287	2.469 - 16.007	0.0001
Comedo Necrosis	Present vs. Absent	0.681	0.245 - 1.893	0.4618	0.948	0.328 - 2.739	0.9219
Size	per mm	1.008	0.990 - 1.027	0.3738	1.012	0.993 - 1.032	0.2035
Margin	Positive vs. Negative	-	-	0.9937	-	-	0.9942
Age	Per year	1.03	0.969 - 1.094	0.3423	1.036	0.975 - 1.101	0.2536
Radiotherapy	Yes vs. No	0.253	0.034 - 1.902	0.1819	0.289	0.037 - 2.246	0.2355

Figure 3.26: (A) Kaplan-Meier curves showing robust stratification of patients in the validation cohort into high-risk of recurrence and low-risk of recurrence subgroups and using only invasive recurrence as an event. (B) Univariate and multivariate Cox regression analysis comparing the influence of common clinicopathological variables alongside the 8-feature recurrence risk prediction model for invasive recurrence-free survival, on the validation set

A



B

Validation Cohort Cox Regression for DCIS Recurrence							
Variables		Univariate Analysis			Multivariate Analysis		
		Hazard Ratio	95% Confidence interval	*P*-value	Hazard Ratio	95% Confidence interval	*P*-value
Recurrence Free Survival							
Predictive Model	High Risk vs. Low	8.488	1.898 - 37.962	0.0051	8.289	1.840 - 37.342	0.0059
Comedo Necrosis	Present vs. Absent	1.418	0.171 - 11.791	0.7466	1.688	0.193 - 14.720	0.6358
Size	per mm	0.980	0.931 - 1.033	0.4555	0.985	0.932 - 1.041	0.5883
Margin	Positive vs. Negative	-	-	0.9961	-	-	0.9949
Age	Per year	1.034	0.937 - 1.142	0.5059	1.024	0.922 - 1.136	0.6619
Radiotherapy	Yes vs. No	0.704	0.084 - 5.894	0.7462	0.640	0.073 - 5.621	0.6876

Figure 3.27: (A) Kaplan-Meier curves showing robust stratification of patients in the validation cohort into high-risk of recurrence and low-risk of recurrence subgroups and using only DCIS recurrence as an event. (B) Univariate and multivariate Cox regression analysis comparing the influence of common clinicopathological variables alongside the 8-feature recurrence risk prediction model for DCIS recurrence-free survival, on the validation set

Additionally, using an RFS model for continuous risk resulted in a similar, significant (p=0.0358), hazard ratio as was seen in the training cohort (HR = 1.05 per unit increase) **(Table 3.11).**

3.5 Discussion

Limited understanding of progression of pre-invasive ductal lesions to invasive ones [144] and lack of clinicopathological [22] and molecular markers [23], which can predict recurrence, lead to uncertainty in therapeutic decision-making. Without a confident measure of recurrence risk, patients are often at risk for over- and under-treatment [32]. The aim of this study was to develop

a novel image analysis pipeline which could predict the 10-year ipsilateral recurrence risk in DCIS patients treated with BCS. We also show that our approach of class-annotating slide regions prior to feature extraction for recurrence prediction enhances our model's prognostic ability. Additionally, our two-tiered approach enables better interpretation of the features that our model uses for recurrence prediction; this is particularly important given that with machine learning approaches, it is often difficult to understand why the trained model responds in a particular way to a set of input data.

Predictably, most of the features selected for the final recurrence classifier model originate from tumor regions, whose cells show both gross morphological changes and nuclear alterations, such as deviations in heterochromatin [195]. The patterns and distribution of hematoxylin within cancer could reflect changes in both ductal architecture and cellular cytological features, both long mainstays of DCIS grading [196-203], and can be continuously quantified [153]. The surrounding stroma is composed of a collection of many varied cell types that also produce diverse hematoxylin staining patterns. Fibroblasts [157] and myofibroblasts [204], for example, have both been implicated in DCIS invasion and recurrence, and provide distinct hematoxylin distributions. As fibroblasts are rich in rough endoplasmic reticulum, they would be much more basophilic [55] and demonstrate different hematoxylin staining patterns compared to myofibroblasts. It should be noted, as a limitation, that the stroma is the principal area where the addition of eosin deconvolution into our pipeline would perhaps improve model performance due to stromal collagen diffusion and densities. Thickening of the ECM, through fibrous deposits such as collagen, promotes cancer progression [205], and since collagen is eosinophilic, its distribution and texture features would be best quantified with the eosin stain.

Benign epithelial ducts and blood vessels both provide a single feature towards the final recurrence classifier model. These classes' relative deficiency of selected features can perhaps be due to limitations for these annotation within the pipeline and/or these regions not being as

prognostically informative as compared to cancer or the surrounding stroma. Vascular heterogeneity has varied impact on breast tumor progression [206]. It is possible that this prognostic value is being harnessed through our recurrence classifier. However, our choice of H&E slides limits us to only studying the texture of vessels containing visible red blood cells within a relatively large section (image tile); a smaller sliding window, would perhaps uncover smaller, but relevant, vascularization. It is interesting that a feature of benign epithelial ducts was included in our final recurrence classifier. As our use of the 'benign epithelial duct' annotation is inclusive of everything but DCIS, it is possible that potentially prognostic information inherent in regions containing abnormal malignancy precursor cells is being captured by our feature. Proliferative, non-cancerous, alterations such as columnar cell lesions often co-occur with DCIS, suggesting their potential for malignant transformations and can be used as a marker for BC risk [207]. Importantly, these premalignant regions could also possess variation in hematoxylin staining patterns. For example, usual ductal hyperplasia [160], characteristically shows nuclear pseudo-inclusions [208], which would show a unique hematoxylin texture pattern. As the distinction between some benign areas and low grade DCIS is not clear [209], with potentially similar histological and nuclear features, it comes as no surprise that benign epithelial ducts and cancer duct annotations had a level of uncertainty. Further testing to differentiate annotations between non-benign and benign regions might be advisable to see if this distinction can glean additional prognostic and interpretable value. Immune-rich regions were notably absent in both filtered features and the final model, likely due to immune dense areas of lymphocyte infiltration not possessing significant variability in cell and nuclear morphology [210].

Based on the hematoxylin texture distribution of these annotated regions, our model consists of some features that are perhaps amenable to logical interpretation in terms of disease biology, and some that elude obvious explanation; yet, both types are useful prognostically. Interpretable texture features can correlate with accepted pathological principles, such as histology, and allow

for a continuous, quantifiable, and non-biased measure which is beyond the capacity of the human eye. Additionally, they instill more confidence in machine learning approaches, which often can be considered black boxes. On the other hand, texture features and patterns which may lack discriminatory ability per se, can still provide discriminatory information when their higher order spatial statistics (e.g., statistical moments) are considered [211]. These non-visually extractable features can supplement a pathologist's visual inspection to provide additional unbiased prognostic value [212]. Our final full slide recurrence classifier model includes both types of features, with a clear example demonstrated through the two mean cancer slide annotated textures (the more interpretable feature #1, and a less intuitively interpretable feature #3). The most significant feature in the model (i.e., feature #1) quantifies the average hematoxylin intensity at a high end threshold, which broadly represents the underlying average tissue architecture (by enabling luminal versus more solid areas to be distinguished), long shown to have some value predicting DCIS recurrence [213]. Furthermore, as this feature is a continuous measurement, it also presents a relative scale that a more broadly defined architectural pattern (such as a classification of cribriform architecture) cannot. This can be especially useful for comparing between mixed pattern cases, which are often present in DCIS [214] and underlie inter-observer variability among pathologists [215]. Our univariate analysis indicated that a lower value of feature #1 correlated strongly with a higher rate of recurrence, consistent with the empirical observation that more solid DCIS cases have poorer outcomes [213] and are often of higher grade [214]. Feature #3 on the other hand, does not grant such discernable interpretation for our data. The SRHGE (Short Runs High Gray-level Emphasis) is a second order texture feature that explains the joint distribution of spatial arrangement and grey level, which, notwithstanding, has had previous success in machine learning algorithms for cancer classification [216-218]. Interestingly, this feature also presents a prime example of the dependency of some of these features within our data and why a tree-based classifier can exploit such a relationship. On its own, feature #3 did not show significant stratification ability; however, if used on patients directly after splitting

them into high and low feature #1 groups, we observed a marked increase in stratification ability. This type of association is conserved in a tree-based algorithm as they allow for branching results which depend on upstream features.

In this study, we used a combination of 8 features to create a machine learning-based model to predict risk of DCIS recurrence. Our model demonstrated outstanding prognostic ability in two independent patient cohorts, commandingly outperforming traditional histopathological variables. Additionally, this model was able to create prognostic groups with over double the hazard ratio of risk groups created through the commercially available Oncotype DCIS score [219] and improved concordance to the DCIS nomogram [147]. In our validation cohort, the model was able to identify a high-risk group of patients that had almost a 50% chance of recurring within 10 years (versus <10% chance within the low-risk group).

Within subsets of patients treated with BCS alone or those receiving additional adjuvant radiation, the recurrence classifier model also identified patients likely to recur. Thus, our model can serve as a clinical tool to help with treatment decisions. For example, high-risk patients who may have undergone BCS alone might require more aggressive treatments (such as radiotherapy) to avert recurrence. While there is debate if adjuvant radiation even provides a significant reduction in breast cancer-specific mortality for DCIS [220], or if any observed survival benefit should be attributed to radiotherapy's potential systemic effects (as opposed to local disease control) [221], the impact of radiotherapy on reducing recurrence is significant. Additionally, our model identifies a low-risk group that has only an 8% 10-year risk of recurrence even without radiation. This result compares favorably to the low-risk group identified by the OncoType DX DCIS score (10.6% 10 year rec risk) [222], and can suggest de-escalation/elimination of radiation therapy for this patient subgroup. Thus, our model offers distinct clinical utility for high-grade patients (who have a high recurrence risk) as well as preliminary results for low-/intermediate-grade patients. Finally, our data has shown some potential in identifying patients who have a high risk of recurrence even

after adjuvant radiotherapy. Although the sample size is very limited for this cohort, our findings provide impetus to pursue a larger study exploring this aspect.

Our study has a few limitations. The first caveat is that both the training and validation cohorts originate from the same institution. Although the recurrence classifier model is ‘seeing’ samples from patients in the validation cohort for the first time, the cohorts are likely to have significant similarities in tissue processing, staining, and imaging protocols, and likely, patient demography. Thus, the generalizability of this model must be tested in additional external cohorts from diverse institutions. Additionally, our validation cohort consists entirely of high-grade patients. Although It is important to note that finding a reliable cost-efficient prognostic variable in high grade DCIS remains of utmost importance, as radiotherapy currently appears to be overused in high grade DCIS compared with the reported lower recurrence rates, the value of the model in lower grade lesions, and the view of safe radiation omission from these lower grade patients is a valid question that has to be validated in a subsequent study.

Although our model significantly stratified patients who received radiation, in both the training and validation cohort, the sample size is notably small and requires additional testing. Technical avenues for improvement include combining multiple image resolutions and sliding window sizes, as we had to balance the slide processing speed (20× would not be feasible to run a similar analysis on our current computers) while still preserving structural differences that would allow pathologists to distinguish all annotated classes. An intrinsic limitation of traditional ‘human-crafted feature-based’ ML is that feature engineering is limited to human knowledge. Alternatively, a deep learning approach, such as one involving convolutional neural networks, may be able to outperform this system and identify novel morphological signatures even more informative for patient recurrence risk prediction.

3.6 Conclusion

The model presented in this study robustly predicts DCIS recurrence risk and significantly outperforms traditional clinicopathologic variables. Simply inputting a scan of an H&E-stained DCIS tumor slide into this tool would allow the identification of patients who are at low-risk and likely do not even require adjuvant radiation, and those patients at such high-risk that an even more aggressive therapy may be advisable (such as systemic radiation[221]). Although this methodology is promising, it requires additional testing with more diverse samples and treatments before any clinical utility of this pipeline can be unequivocally established. Ultimately, our study provides proof-of-principle that such a pipeline can predict DCIS recurrence risk; in future studies, we hope to train this pipeline on images from core biopsies, as a treatment aware model, to predict patients' recurrence risk so that their entire treatment plan (including type of surgery and recommendations regarding radiotherapy) can be tailored based on their risk profile.

Copyright by
Sergey Klimov
2019

4 Predicting Pancreatic Neuroendocrine Tumor (PanNET) Metastasis Risk through a Multi-Label Deep Learning Approach

4.1 Abstract

Pancreatic neuroendocrine tumors (PanNET) are the second most common type of pancreatic cancer, and its incidence is steadily rising. Unlike its extremely aggressive cousin, pancreatic neuroendocrine carcinomas (PanNECs), PanNET patients have variable outcomes. Unfortunately, clinical prognostic markers are inadequate in identifying patient metastasis risk. To provide an accurate biomarker for this unmet clinical need, we develop a novel deep learning pipeline that first annotates whole slide images, identifies regions that are metastasis associated, and aggregates these regions into a probability distribution through which overall metastasis risk can be predicted.

H&E stained surgical resections of 90 patients (18 who develop metastasis) utilized for training and testing the sequential models. First a CNN was trained (through pathologist annotation) to identify tiles of cancer (separately as stromal poor/clearly delineated and stromal rich), stroma without cancer, normal parenchymal, and fat. Next, additional CNNs were trained, using patient metastasis outcome as the label, to predict if cancer and stromal annotated tiles came from slides from patients who ultimately experienced metastasis. Finally, the probability of every metastasis associable tile from a whole slide was aggregated into a distribution. Distribution statistics were extracted and fit into 18 machine learning models, with basic feature selection, to develop a model that outputs overall slide risk.

Annotation classification had a validated accuracy of 92.8%, with a sensitivity and specificity of over 90% for each label. Cross-validated tiles with high output probabilities (>99.9%) from the metastasis classifiers produced a sensitivity/specificity of 62%/89% for stromal tiles and 76%/69% for cancer tiles. Finally, a quadratic SVM was able to utilize a filter set of these features to provide a, leave-one-out cross validated, hazard ratio of over 8.00 regardless of cofounding variables.

Our 3 layered model fill's an unmet clinical need for accurately predicting metastasis risk for PanNET patients.

4.2 Introduction

Pancreatic neuroendocrine tumors (PanNETs) represent a rare subset (2%) of pancreatic neoplasms that affect up to 2 per every 1,000,000 new patients a year. PanNETS are epithelial cell cancers with neuroendocrine differentiation, often resembling cells making up the islets of Langorn (the endocrine producing cells within the pancreas). Like the aforementioned islets, PanNETS can also secrete hormones (notably insulin) into the bloodstream (noted as a 'functional' clinical type). Based on the most recent WHO classification system PanNETs are divided into well (WDNETs or 'ordinary') and poorly differentiated (PDNEC) categories though recent evidence suggests that these tumors are not in a continuum and should be regarded separately. Our study focus is on WDNETS (which we will refer to simply as PanNETS hereinafter), as the poorly differentiated versions risk stratification is much less ambiguous due to overall extremely pessimistic outcomes (with median survival typically under 2 years) [38].

Perhaps the biggest clinical hurdle in accurate patient prognosis for PANnet patients is the lack of viable biomarkers. The only markers routinely accepted in the clinic are those used to determine phases of the cell cycle: the mitotic count and the Ki-67 index. Unfortunately, these two grading mainstays are prone to quantification errors, with mitotic counts faulted by cells expressing mitotic mimics (such as cells undergoing pyknosis) and Ki-67 often showing poor visual concordance [223]. Aside from the technical issues with these markers, papers often argue for percent (1-2%) differences in grading to improve patient stratification. But these minute adjustments are unlikely to lead to prognostic shifts of any real magnitude. While the overall survival for ordinary PanNETS is quite high (60-70% 10-year survival), even the smallest lesions have shown a metastatic risk up to 15% in long term follow up. A model able to consistently provide accurate measure of metastatic risk would allow the oncology team confidence in recommending increased

surveillance or potentially using adjuvant targeted treatments. Some progress in developing models to determine recurrence free survival for PANnet patients has been done through IHC [40] and linear pathological combinative approaches [41]. But these models suffer from a lack of cross validation, and thus likely over-fit, and their results are not overtly superior to classic staging.

While there is a wide variation of morphological characteristics within PanNETS and morphological descriptors of high grade in NET's significantly impacts patient outcome [39], they are, surprisingly, not used for grading. Histological changes such as patterns of necrosis, variations in nuclear shape/atypia, chromatin clumping, and a reduction in tumor stroma, can represent a high-risk component even within the current well-differentiated grading of PanNETS[224]. To include a whole slide morphological analysis, we develop a novel multiclassification pipeline, which utilizes Convolutional Neural Networks (CNNs). CNNs have shown tremendous promise in identifying morphology distinct areas on digitized slides [225] and in directly identifying image patterns correlating to prognostic risk [226]. Taking advantage of these characteristics, our first CNN first annotates a full slide into relevant tissue regions. Next, we develop tissue specific CNNs to determine metastasis association for both cancer and adjacent stromal areas. Finally, we aggregate all of the results into a full slide metastasis association probability distribution, extract descriptive features of it, and use an ultimate layer of machine learning to determine an overall risk of metastasis for the patient.

4.3 Methodology

4.3.1 Study Population

This retrospective study cohort was obtained from surgical resections of patients, diagnosed as having PanNET, treated at Emory University hospital between 2002 and 2017 (**Table 4.1).** Patients presenting with metastasis during surgery, or those with a censored last-follow-up before a year were omitted from the analysis. In total, 90 patients were utilized for the study, with 20% developing metastasis at follow-up. Patient records were mined to obtain details of follow-up,

demography, and clinicopathological variables. Metastasis free survival was measured from surgical time to metastasis or last follow-up.

Table 4.1: Patient Characteristics. Descriptive data detailing the clinicopathological variables of the Emory PanNET cohort.

Patient Clinicopathological Characteristics	
Baseline characteristic	**Total (N = 90)**
Patient age	
Median Age (range), years	56 (19 - 82)
Age <50, n (%)	26 (28.9)
Age>=50, n (%)	53 (58.9)
Missing	11 (12.2)
Tumor Size	
Median Size (range), cm	3 (0.6 - 11)
Size <2.0, n (%)	34 (37.8)
Size>=2.0, n (%)	43 (47.8)
Missing	13 (14.4)
Sex, n (%)	
Male	35 (38.9)
Female	44 (48.9)
Missing	13 (14.4)
Metastasis status, n (%)	
Recurrence free	72 (80.0)
Recurred	18 (20.0)
Missing	0 (0.00)

4.3.2 Tumor Slide Selection

Full-face sections, from surgical resection tumor blocks, were pathologist analyzed and the most representative slides (based on tumor tissue and morphological variance) selected for analysis. Tissues were digitized using a 40x scanner (0.24 µm/pixel). Image quality was reviewed for focus and artifacts with any issues leading to a re-scan or omission respectively.

4.3.3 Automated Full Slide Annotation

For training a deep network towards predicting region annotation, the GoogLeNet architecture [227] was chosen and its terminal SoftMax layer was modified to classify 5 tissue classes: Cancer (separately as stromal poor/clearly delineated and stromal rich **(Fig 4.1)**, stroma without cancer, normal parenchymal, and fat. Pathologist annotated ground truth regions (non-overlapping 150x150x3 pixel tiles), for each class, were extracted using the MATLAB Image labeler app. Aside from the cancer/stroma class, which was determined by any cancer cluster within a stromal rich

region (fibrious stroma, fibrious septa, or hyalinized), training tiles almost exclusively came from non-mixed ground truth regions. The occasional intersection of annotations (edges) were labeled by the predominant class in the image tile.

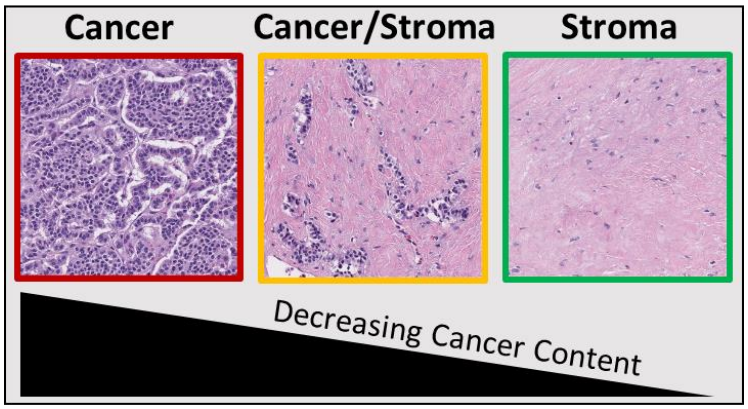


Figure 4.1: An example of the differences in tiles labeled as 'cancer (stroma poor)', 'cancer/stroma (hyalinized)', and stroma without cancer.

As single cell analysis/segmentation was outside the scope of this analysis, the WSI's were 4x down-sampled (**Fig 4.2** shows an example of the visual difference). This allowed for computational practicality while still providing clear visual discrimination of the gross scaled annotation regions. To take into account staining variability, and improve machine learning performance[166] , each tile, for both training and later classification, was color normalized [165] to a consistent H&E stain. To improve the generalizability and robustness of the CNN classifier, and reduce overfitting, without reducing the image quality [178], a thorough augmentation [225] of training tiles was performed.

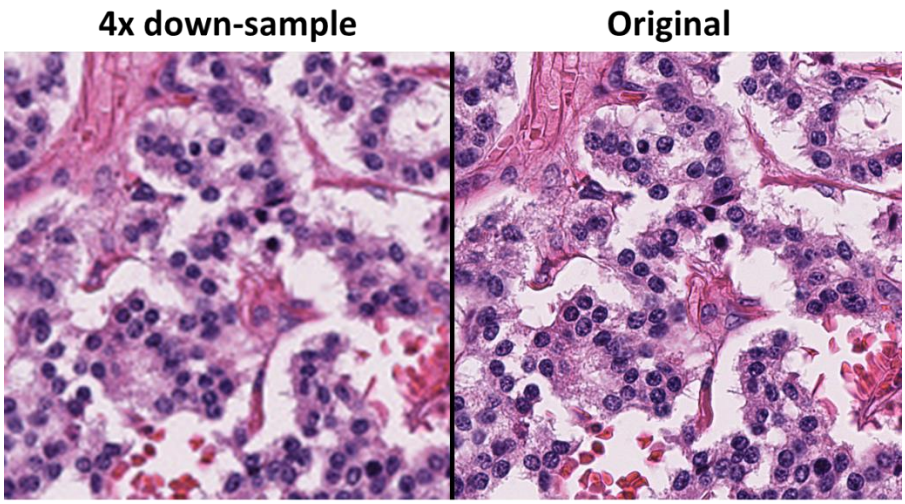


Figure 4.2: An example of the visual appearance of the digitized slide after 4x down sampling.

Tiles were altered through image orientation, hue adjustment, and blur/noise/contrast perturbations[225] to expand the final training set by a factor of 45 (**Fig 4.3 has an example of some augmentations**). The CNN was trained using the stochastic gradient descent with momentum, a batch size of 35 tiles, and a learning rate of 1e-4. The training tiles were re-shuffled at the start of each epoch and the training accuracy was measured at the end of each epoch. Training was performed until the, end of epoch, multiclass accuracy for each label was over 99%. The classifier was validated using external slides. Due to the scarcity of histology slides within the cohort, tiles from the validation slides were extracted with overlap (50%).

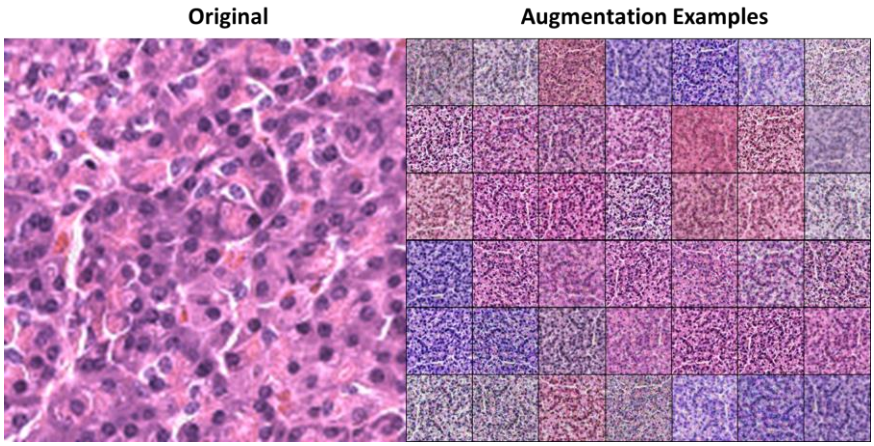


Figure 4.3: An example of the types of alterations utilized in training augmentation.

WSI annotation was performed by fully partitioning slides into non overlapping 150x150 tile segments, with background 'non-tissue' regions omitted from classification. These segmented tiles were independently processed with the trained CNN to produce a 5-dimensional output layer with the probability of the tiles belonging to one of the 5 classes. To assess the whole slide classification capability, CNN annotated cancer areas were overlaid with pathologist annotations and the Jaccard index calculated.

4.3.4 Metastasis Association Classifier

After annotating each WSI within our cohort, subsequent GoogLeNet classifiers were trained to predict metastasis association for tiles classified as either cancer or stroma (separately). For computational feasibility, and to reduce using weakly likelihood areas, only tiles predicted at a high enough confidence (95%) for either cancer or stromal were used (80% of total tiles). The WSI cohort was split through 5-fold cross validation. WSIs coming from different blocks of the same patient (n=14) were kept together in a 'fold'. This reduced potential bias by forcing each testing set to contain tiles from slides only from patients which the classifier was not trained with. The patient's distant metastasis status was used as the ground truth label for all tiles annotated from their slides. Each training set, for both the cancer and stromal CNN, was trained for 6 epochs (using stochastic gradient descent with momentum and dropping the initial learning rate of 0.001

by a factor of 0.1 at 4 epochs) before being applied towards the respective test set. To decrease overfitting to the training data, an L2 regularization was applied (0.0001). The predicted metastasis association labels (and scores) for tiles within each cross-validated testing set were concatenated to rebuild the full-sized cohort and used for further analysis.

4.3.5 Full Slide Feature Extraction and Metastasis Prediction

Metastasis associated probabilities, for both cancer and stromal tiles, were stitched together to form a WSI mask. Tiles without enough, directly adjacent, similarly annotated (tissue) neighbors (4 for cancer and 2 for stroma) were considered noise and filtered out. The distribution of the metastasis association score (0-100%) for the remaining tiles within each slide were the basis for the extraction of 150 ‘full slide’ features. These features were derived from histogram metrics of both individual tiles within the WSI (**Fig 4.4**) and after aggregation within a 10x10 tile area (**Fig 4.5**).Histogram metrics used included the statistical moments (mean, standard deviation, skewness, and kurtosis) and the tile counts/proportions, for both the overall metastasis associated probability tile distributions and within high probability (>90%) regions of each WSI.. Spatial features were bin counts of ‘spatially clustered’ metastasis associated groups with the assumption that ‘clusters’ of high-risk areas possess potential prognostic value beyond single regions.

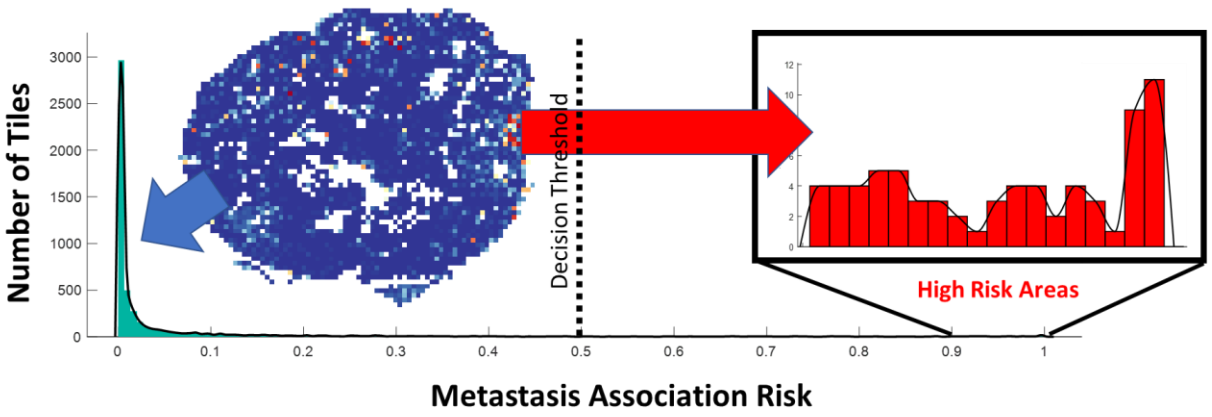


Figure 4.4: An example of a histogram of the metastasis associated probabilities within a full slide. Blue regions on the mask represent tiles associated with a non-metastatic patient and are overwhelmingly prevalent with most tiles having a 90% probability of not being associated with a metastatic patient (teal histogram). The few red regions, indicating high risk areas, are also highlighted in a separate histogram (showing the count of tiles with a high probability of metastasis association). Features extracted represent statistical moments (mean, standard deviation, skewness, and kurtosis) for the full histogram and on both extreme edges. Additionally, the overall frequency of high-risk regions and their relative proportions are also used.

Eighteen different machine learning models (**Table 4.2**) were trained using the full slide features as the input variables and the patient's metastasis information as the labels. Models were trained through a patient level, leave-one-out cross validation, wherein each left out set composed of only all the slides (if multiple) from a single patient. Patients with multiple slides (n = 14) were given a 'high-risk' prediction if any of their slides were predicted to metastasize. To improve accuracy [182], and reduce data dimensionality, simple feature filtering was performed. For each training fold, within the leave one out cross validation, a two-sample t-test was performed for all features between patients who experienced later metastasis versus those which did not. Consecutively increasing thresholds of the resulting t-score were tested for each ML model, with the final model selected based on the highest resulting accuracy. This model was further analyzed univariately using Kaplan-Meier survival analysis and multivariate analysis (alongside tumor size, patient age, and sex) using Cox Regression.

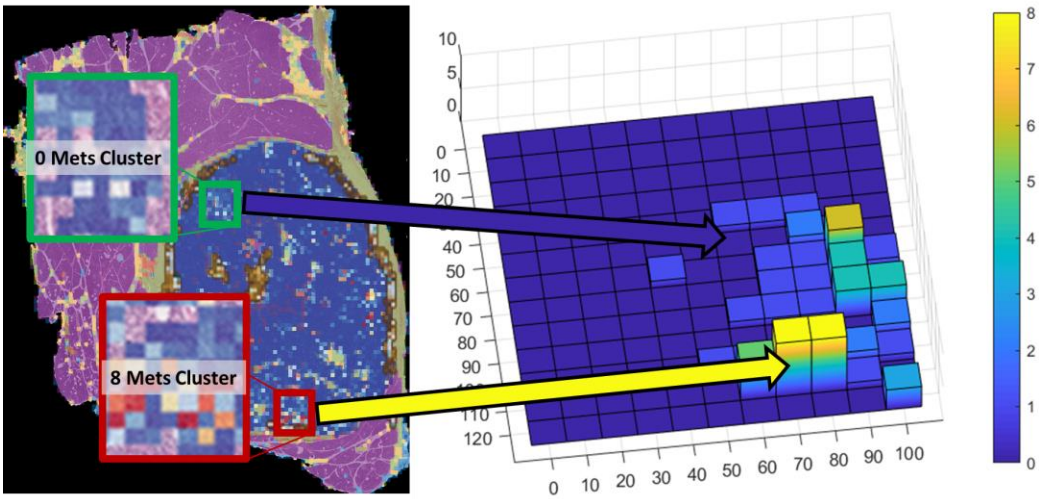


Figure 4.5: Example of the spatial cluster features used in the model. 10x10 tile areas are analyzed for frequency of various metastasis association probabilities (>50% and >90%). Features are the frequency of various 'clusters' (10x10 tile regions). An example of a frequency measure is seen above, where there are 2 clusters with at least 8 metastasis associated tiles

Table 4.2: List of machine learning models used for metastasis prediction

Machine Learning Models
1 - Fine Tree
2 - Medium Tree
3 - Coarse Tree
4 - Fine KNN
5 - Medium KNN
6 - Coarse KNN
7 - Cosine KNN
8 - Cubic KNN
9 - Weighted KNN
10 - Linear SVM
11 - Quadratic SVM
12 - Cubic SVM
13 - Fine Guassian SVM
14 - Medium Guassian SVM
15 - Coarse Guassian SVM
16 - Ensemble boosted trees
17 - Ensemble bagged trees
18 - RUSBoost trees

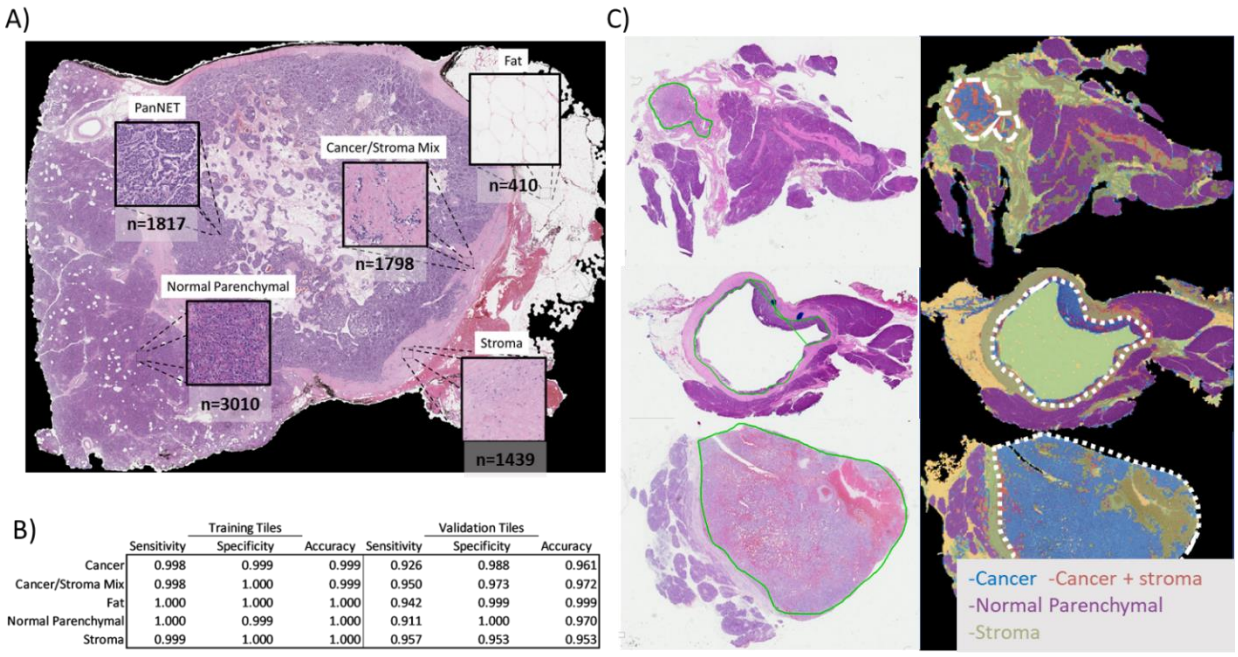
4.3.6 Statistical Analysis

For confusion matrix metrics accuracy was measured as the sum of true positives (TP) and true negatives (TN) divided by the total sum, sensitivity was the number of TP divided by total positives, and specificity was the number of TN divided by total negatives. Variable significance within survival analysis was measured though the log-rank test for Kaplan Meier curves and the Wald chi-square test for cox regression analysis. A p value of < 0.05 was considered significant for all results. Statistical analysis was performed using the SAS 9.4 software (Cary, NC, USA) and MATLAB R2018b (Natick, MA, USA).

4.4 Results

4.4.1 Deep Learning Discriminates between PanNET Tissue Annotations

A total of 8,474 non-overlapping 150x150 pixel, pathologist annotated, ground truth regions were extracted (**Fig 4.6A**). Augmenting this data resulted in 381,330 tiles utilized to train the annotation CNN. After 9 epochs the training data was almost perfectly classified (**Fig 4.6B**) and for the validation cohort produced an overall accuracy of 92.8% with greater than 90% sensitivity and specificity for every annotated class. The most accurate classification, in the validation cohort (n=42,976), was for stromal regions, which were properly classified over 95% of the times (**Table 4.3**) while the most discordance was found in cancer tiles, wherein 5% were improperly classified as stromal. Whole slide image classification provided strong concordance to (**Fig 4.6C**) pathologist annotations with a median Jaccard Index of 0.79. Cancer regions were always highlighted, with false positive areas associated to sparse edge/interface cases (generally not seen in training). These false positive areas have low probability (<95%), and thus were not used for downstream analysis.



B)

	Training Tiles			Validation Tiles		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
Cancer	0.998	0.999	0.999	0.926	0.988	0.961
Cancer/Stroma Mix	0.998	1.000	0.999	0.950	0.973	0.972
Fat	1.000	1.000	1.000	0.942	0.999	0.999
Normal Parenchymal	1.000	0.999	1.000	0.911	1.000	0.970
Stroma	0.999	1.000	1.000	0.957	0.953	0.953

Figure 4.6: A) Examples of the tissue annotation classes and the (non-augmented) ground truth count used for training. B) The multi-class sensitivity, specificity, and accuracy for the training and validation tiles. C) Examples of pathologist annotations (green solid line) for cancer regions versus automated whole slide annotation (major cancer regions outlined with a white dashed line)

Table 4.3: Row normalized confusion matrix for tissue annotation within the validation cohort

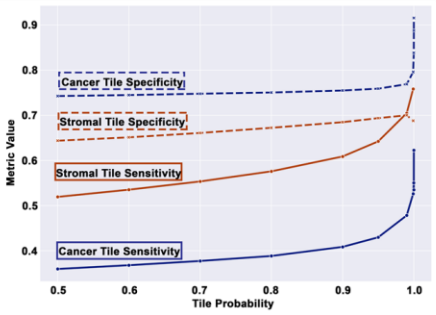
		Prediction				
		Cancer	Cancer/Stroma	Fat	Normal	Stroma
Ground Truth	Cancer	0.926	0.014	0.000	0.001	0.059
	Cancer/Stroma	0.027	0.950	0.000	0.000	0.023
	Fat	0.000	0.000	0.942	0.000	0.058
	Normal	0.013	0.041	0.001	0.911	0.034
	Stroma	0.008	0.035	0.000	0.000	0.957

4.4.2 Metastasis Association Provides Significant Prognostic Value

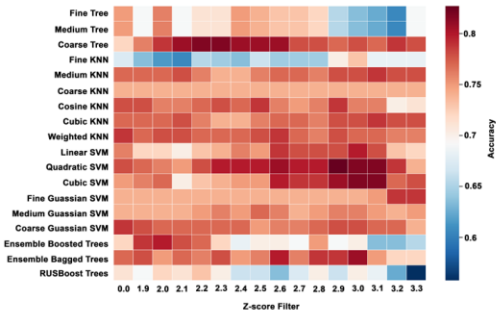
Within the full cohort, 430,318 cancer (both stroma rich/poor) and 211,361 stromal tile annotations were given with a greater than 95% CNN probability. Training the metastasis association CNN with these tiles (using the patients metastasis status as their labels) provided an overall test-set sensitivity/specificity of 36%/74% and 52%/65% for cancer and stromal tiles, respectively (**Fig 4.7A**). Analyzing tiles above a certain probability score resulted in an overall increase of both

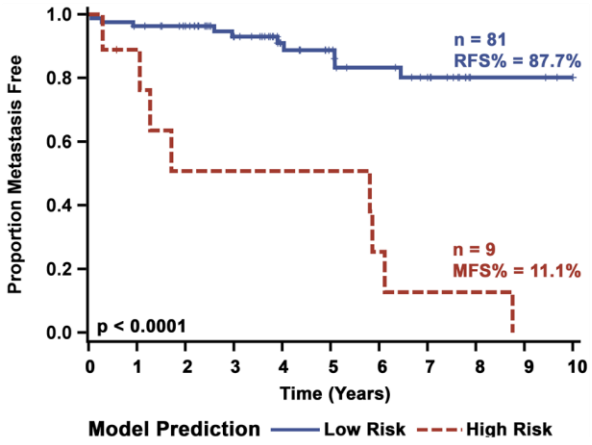
cancer and stromal performance measures. This resulted in a maximum of a 62% sensitivity and 89% specificity for cancer tiles classified with over a 99.999% probability (n=12,585) and 76% sensitivity and 69% specificity for stromal tiles classified with over a 99.9% probability (n=4,524). Predicting patient metastasis risk, using full slide features stemming from their metastasis associated tiles, resulted in a best leave-one-out cross validated accuracy of 78% using either a course gaussian SVM or a weighted KNN **(Fig 4.7B)**. An optimal model (accuracy 83%) was found using a quadratic SVM and filtering out features that were not significantly different (between patients in the training set who experienced metastasis versus those that did not) over a t-score value of +/- 2.9. Almost 90% of patients selected as high risk by this model metastasized within 10 years, as compared to only 13% of patients put in the low risk group **(Fig 4.7C).** The model stratification was highly significant, controlling for clinical variables **(Fig 4.7D)**.

A)



B)





D)

Cox Regression

Variables		Univariate Analysis			Multivariate Analysis		
		Hazard Ratio	95% Confidence interval	*P*-value	Hazard Ratio	95% Confidence interval	*P*-value
Recurrence Free Survival							
Predictive Model	High Risk vs. Low	8.891	3.474 - 22.758	<0.0001	8.056	2.673 - 24.276	0.0002
Sex	Male vs. Female	0.717	0.267 - 1.928	0.5099	1.343	0.452 - 3.992	0.5954
Age	per cm	1.034	0.992 - 1.079	0.115	1.015	0.973 - 1.058	0.4861
Size	per year	1.264	1.063 - 1.502	0.0081	1.287	1.054 - 1.572	0.0132

Figure 4.7: A) The cross validated sensitivity (or tiles which actual came from a patient who experienced metastasis/all tile predicted as metastasis associated) and specificity (or tiles which actual came from a patient who did not experienced metastasis/all tile predicted as non-metastasis associated) for various metastasis association probability cut points for the neural network built for cancer and stromal tiles. The cut points represent the CNN probability threshold at which a slide is analyzed. For example, at the tile probability cutoff of 0.7, the sensitivity/specificity was determined by only using tiles which had a metastasis/non-metastasis association >= 0.7. B) The leave-one-out cross validated accuracies of each model and different levels of feature filtering. A z-score filter of 0 represented models trained with all features while a z-score of 3.3 used models which only utilized the most significantly different features between the training metastasis and non-metastasis groups. C) Leave-one-out cross validated Kaplan-Meier curves of selected model for metastasis free survival. D) Univariate and multivariate Cox analysis comparing the selected model alongside common clinicopathological variables.

4.5 Discussion

Well differentiated pancreatic neuroendocrine tumors are a logical disease to study through computational pathology; it has a variability in outcomes, diverse morphology, and currently lacks sufficient prognostic biomarkers. Unlike neuroendocrine carcinomas, which are usually lethal, PanNET patients have an unpredictable prognosis (stemming mostly from metastasis risk), which almost entirely relies on Ki67 [224]. Neuroendocrine morphology, observed with H&E histology, has been shown to be a significant indicator of both prognosis and therapeutic response and is

linked to various important oncogenetic alterations [228]. Being able to identify and study these tumors through deep learning could reveal latent features important to disease progression. Additionally, the surrounding stroma, shown to have significant prognostic value in other cancers [52], is not well studied in PanNETs and could potentially provide supplementary prognostic value. Therefore, our study attempted to both identify these prognostic tissue regions within a WSI, employ separate deep learning networks to determine metastatic association risk within these regions, and analyze the probability distribution of metastatic association to predict an overall patient risk of metastasis.

At each step, our novel 3-step pipeline showed significant promise. Deep learning has consistently presented superhuman image classification, which remained consistent with the tissue labels for our PanNET cohort. In over 40,000 validation images, the CNN provided an overall accuracy of better than 92% for 5 separate annotations. Adding an additional label of metastasis association to tiles annotated as cancerous or adjacent stromal gave optimistic results. Unsurprisingly, low confidence metastasis associated areas, ones which the CNN gave 50-80% probability (for either mets or non-mets association), did not show meaningful discriminatory power. However, whenever this classifier output a prediction with a high probability (>99%), the discrimination significantly increased. This indicates that these stromal and cancer tiles possess morphological features, identifiable by a deep learning network, which can represent risk (or lack of) of these patients experiencing metastasis. Finally, when put together, a cross validated model was able to find a high-risk group of patients which had an over 8x higher risk of distant metastasis, higher than any biomarker for PanNET reported in literature. Although preliminary, these results show the potential of a multi-structured machine learning based model to provide enough risk stratification to help clinical decision planning.

Despite these encouraging results, this study has clear limitations. As our cohort is from a single hospital, and due to the relatively low incidence rate of PanNETs, our sample size is quite limited

in power and requires thorough external validation. Although our results stay significant when controlling for clinopathological variables, our retrospective cohorts lack Ki67 results, arguably the most important variable to control for, as it's the only real standard of care marker. Importantly, the deep learning networks likely would benefit from longer training or starting with more complex networks. But these concerns are outside the scope of this study due to data limitation and computational hardware constraints. Our study provides proof-of-principle that this unique, multi labeling, deep learning pipeline can predict PanNET metastasis risk using only surgical resection H&E tissue. Additionally, these results suggest that prognostic morphological patterns exist, for PanNET tissue, for both cancer and adjacent stromal regions, and provide evidence towards including them in clinical decision making. In future studies we hope to significantly increase our sample size, improve our network, and add in treatment options so that our model can hopefully provide predictive benefit.

5 Conclusion

Improperly identifying a patient's risk is an underlying, unacceptable, issue in healthcare. Approximately 10% of patient deaths, and up to 17% of adverse hospital events, result from an inability to form an actionable and timely measure of risk [229]. This is especially true for cancer, wherein a 'one size fits all' treatment approach is not uncommon, despite tremendous heterogeneity between patients cancers [230]. Beyond leading to either over or undertreatment, this inability to accurately identify optimal treatment contributes to an incredible resource waste [231]. This lack of personalization for patient cancer therapy, in general, is due to the fact that very few cancers have any clinically viable prognostic biomarkers developed and even fewer have ones capable of altering therapy decision making or predicting response [232].

As cancer moves more and more towards personalized therapy, an increasing amount of prognostic and predictive biomarkers will be required to finely tune clinical decision guidelines. To be optimally suitable for inclusion in the clinic the biomarkers will likely have to meet certain overarching criteria. Although no biomarker is immune to the uncertainty of cancer, the most importantly characteristic of a model is the ability to maximally stratify the 'split' of patient risk groups. Models which provide a 'low risk group' with too many events risk omitting potentially lifesaving therapy. Alternatively, if a model puts many 'true' low risk patients in their high risk groups they will unnecessarily open patients up to, potentially life threatening, long term and acute side effects from adjuvant therapy [233]. Additionally, the model should ideally be treatment facing. That is, it should have predictive properties [5] which would allow it to recommend (or argue against) a treatment option. While knowing if a patient is at high risk is good (prognostic), identifying if a patient is at high risk AFTER certain therapies is better, and a necessity for personalized medicine. Practically, the biomarker must be both safe and economically viable [5].

Markers which are too expensive for their benefit, such as Oncotype DCIS [33], are unlikely to get insurance (and consequentially patient) buy in.

The advancement of AI within cancer has led to monumental progress within biomarker development by applying varied algorithms towards diverse, multidimensional, data[234]. Models have been derived to utilize genetic data [235], clinicopathological variables [236], and imaging[237] to classify survival, recurrence, and susceptibility for cancers severely lacking in accurate biomarkers. For analyzing image, these models are more increasingly taking advantage of deep learning algorithms to provide superhuman classifications[238]. Our seminal work looks to further push this field through developing novel end to end approaches for WSI analysis and applying it towards determining various metrics of prognosis. Generally, we show the importance of multi-scale multi-label (region annotation classification followed by full slide risk determination). While other important publications have tried to show the ability of utilizing tiles for whole slide risk analysis, they either require pathologist annotation for risk regions [226] or had a limited annotation step and relatively minor prognostic value [52]. Our models thoroughly, and accurately, annotate WSIs and each model provided better than state-of-the-art stratification, regardless of cofounding clinopathological variables. Our models generally provided a level of treatment dependency. Our identification of the specific metastasis sites allows for biologically justifiable therapy (focused on the organ). For DCIS we explicitly showed our model validation in patients who have both received/been omitted radiotherapy. PanNET patients are the only ones who lack a directly correlated treatment decision, but as they have only been treated with resection, identifying the high-risk group could justifiably put them into a cohort requiring systemic therapy. While promising, this model requires the most varied and exhaustive validation. Additionally, our models are relatively cheap and, as it uses slices of tissue obtained from surgery, adds no risk. The image analysis models simply require slide digitization of samples already required in the

clinical workflow, whereas the distant metastasis model would require only 8 IHC TMA's, still an order of magnitude cheaper than clinical genetic signatures (~4,000 USD).

Our models, though, do have limitations. The overarching one being external, ideally blinded, validation. As the data required is quite precious and needs large cohort specific cohorts (specific antibodies or high-resolution resections) with long follow-up times, this is ongoing question that we are pursuing. While we have listed the specific technical limitations to each project, there exist general improvements, following the state of the art, that we are looking to address. Specifically deep learning networks which utilize multi-label multi-instance are perfect for our pipelines and have started to show significance for histopathology [239]. Furthermore, attaching a continuous measure of risk that encompasses follow-up time (such as Cox regression) is an adjustment that should be considered (either through survival forests [240] or Cox CNNs[241]). To improve prognostic capabilities, we would also like to link our histological analysis towards a genetic signature[226].

Ultimately our frameworks are easily transferred to other cancer types which utilize pathology (either H&E or IHC) and can be adapted to the biopsy setting (and thus theoretically provide neoadjuvant and surgical recommendations). Our models provide outstanding prognostic value, which in turn leads to more confident treatment recommendations from the clinician and provides hope for the patient. With a more confident therapy plan patients will be less likely to be over/undertreated and thus face less unnecessary side effects or recurrence. Importantly, our markers can be processed in a matter of minutes, even for full, max resolution, slides. Thus, these models will provide a potential tremendous health economic benefit at almost no cost. Ideally, we will test the potential of integrating our molecular model with our histological analysis in a cohort with multiple therapy options, to truly move towards the goal of personalized medicine..

REFERENCES

1. Siegel RL, Miller KD, Jemal A: **Cancer statistics, 2016**. *CA: a cancer journal for clinicians* 2016, **66**(1):7-30.
2. Mehlen P, Puisieux A: **Metastasis: a question of life or death**. *Nature reviews cancer* 2006, **6**(6):449.
3. Gupta GP, Massagué J: **Cancer Metastasis: Building a Framework**. *Cell* 2006, **127**(4):679-695.
4. Goossens N, Nakagawa S, Sun X, Hoshida Y: **Cancer biomarker discovery and validation**. *Translational cancer research* 2015, **4**(3):256.
5. Ludwig JA, Weinstein JN: **Biomarkers in cancer staging, prognosis and treatment selection**. *Nature Reviews Cancer* 2005, **5**(11):845.
6. Eble JN, Tavassoli FA, Devilee P: **Pathology and genetics of tumours of the breast and female genital organs**: Iarc; 2003.
7. Weigelt B, Peterse JL, Van't Veer LJ: **Breast cancer metastasis: markers and models**. *Nature reviews cancer* 2005, **5**(8):591.
8. Onitilo AA, Engel JM, Greenlee RT, Mukesh BN: **Breast cancer subtypes based on ER/PR and Her2 expression: comparison of clinicopathologic features and survival**. *Clinical medicine & research* 2009, **7**(1-2):4-13.
9. Smid M, Wang Y, Zhang Y, Sieuwerts AM, Yu J, Klijn JGM, Foekens JA, Martens JWM: **Subtypes of Breast Cancer Show Preferential Site of Relapse**. *Cancer Research* 2008, **68**(9):3108-3114.
10. Anders CK, Carey LA: **Biology, Metastatic Patterns, and Treatment of Patients with Triple-Negative Breast Cancer**. *Clinical Breast Cancer* 2009, **9**:S73-S81.
11. Kennecke H, Yerushalmi R, Woods R, Cheang MCU, Voduc D, Speers CH, Nielsen TO, Gelmon K: **Metastatic Behavior of Breast Cancer Subtypes**. *Journal of Clinical Oncology* 2010, **28**(20):3271-3277.
12. Luck AA, Evans AJ, Green AR, Rakha EA, Paish C, Ellis IO: **The influence of basal phenotype on the metastatic pattern of breast cancer**. *Clinical oncology* 2008, **20**(1):40-45.
13. Powles T, Paterson A, McCloskey E, Schein P, Scheffler B, Tidy A, Ashley S, Smith I, Ottestad L, Kanis J: **Reduction in bone relapse and improved survival with oral clodronate for adjuvant treatment of operable breast cancer [ISRCTN83688026]**. *Breast Cancer Research* 2006, **8**(2):R13.
14. Broders AC: **Carcinoma in situ contrasted with benign penetrating epithelium**. *Journal of the American Medical Association* 1932, **99**(20):1670-1674.
15. Ma X-J, Salunga R, Tuggle JT, Gaudet J, Enright E, McQuary P, Payette T, Pistone M, Stecker K, Zhang BM: **Gene expression profiles of human breast cancer progression**. *Proceedings of the National Academy of Sciences* 2003, **100**(10):5974-5979.
16. Barclay J, Ernster V, Kerlikowske K, Grady D, Sickles EA: **Comparison of risk factors for ductal carcinoma in situ and invasive breast cancer**. *Journal of the National Cancer Institute* 1997, **89**(1):76-82.
17. Heselmeyer-Haddad K, Garcia LYB, Bradley A, Ortiz-Melendez C, Lee W-J, Christensen R, Prindiville SA, Calzone KA, Soballe PW, Hu Y: **Single-cell genetic analysis of ductal carcinoma in situ and invasive breast cancer reveals enormous tumor heterogeneity yet conserved genomic imbalances and gain of MYC during progression**. *The American journal of pathology* 2012, **181**(5):1807-1822.

18. Page DL, Dupont WD, Rogers LW, Jensen RA, Schuyler PA: **Continued local recurrence of carcinoma 15–25 years after a diagnosis of low grade ductal carcinoma in situ of the breast treated only by biopsy**. *Cancer* 1995, **76**(7):1197-1200.
19. Boughey JC, Gonzalez RJ, Bonner E, Kuerer HM: **Current treatment and clinical trial developments for ductal carcinoma in situ of the breast**. *The oncologist* 2007, **12**(11):1276-1287.
20. Sanders ME, Schuyler PA, Dupont WD, Page DL: **The natural history of low-grade ductal carcinoma in situ of the breast in women treated by biopsy only revealed over 30 years of long-term follow-up**. *Cancer* 2005, **103**(12):2481-2484.
21. Kuerer HM: **Rational individualised selection of adjuvant therapy for ductal carcinoma in situ**. *The Lancet Oncology* 2011, **12**(1):2-3.
22. Wang S-Y, Shamliyan T, Virnig BA, Kane R: **Tumor characteristics as predictors of local recurrence after treatment of ductal carcinoma in situ: a meta-analysis**. *Breast cancer research and treatment* 2011, **127**(1):1-14.
23. Lari SA, Kuerer HM: **Biological markers in DCIS and risk of breast recurrence: a systematic review**. *Journal of Cancer* 2011, **2**:232.
24. Ward EM, DeSantis CE, Lin CC, Kramer JL, Jemal A, Kohler B, Brawley OW, Gansler T: **Cancer statistics: breast cancer in situ**. *CA: a cancer journal for clinicians* 2015, **65**(6):481-495.
25. Ceilley E, Jagsi R, Goldberg S, Kachnic L, Powell S, Taghian A: **The management of ductal carcinoma in situ in North America and Europe**. *Cancer* 2004, **101**(9):1958-1967.
26. Darby S, McGale P, Peto R, Granath F, Hall P, Ekbom A: **Mortality from cardiovascular disease more than 10 years after radiotherapy for breast cancer: nationwide cohort study of 90 000 Swedish women**. *Bmj* 2003, **326**(7383):256-257.
27. Zablotska LB, Neugut AI: **Lung carcinoma after radiation therapy in women treated with lumpectomy or mastectomy for primary breast carcinoma**. *Cancer* 2003, **97**(6):1404-1411.
28. Darby SC, McGale P, Taylor CW, Peto R: **Long-term mortality from heart disease and lung cancer after radiotherapy for early breast cancer: prospective cohort study of about 300 000 women in US SEER cancer registries**. *The lancet oncology* 2005, **6**(8):557-565.
29. Giordano SH, Kuo Y-F, Freeman JL, Buchholz TA, Hortobagyi GN, Goodwin JS: **Risk of cardiac death after adjuvant radiotherapy for breast cancer**. *Journal of the National Cancer Institute* 2005, **97**(6):419-424.
30. Silverstein MJ, Lagios MD, Groshen S, Waisman JR, Lewinsky BS, Martino S, Gamagami P, Colburn WJ: **The influence of margin width on local control of ductal carcinoma in situ of the breast**. *New England Journal of Medicine* 1999, **340**(19):1455-1461.
31. MacDonald HR, Silverstein MJ, Mabry H, Moorthy B, Ye W, Epstein MS, Holmes D, Silberman H, Lagios M: **Local control in ductal carcinoma in situ treated by excision alone: incremental benefit of larger margins**. *The American journal of surgery* 2005, **190**(4):521-525.
32. Groen EJ, Elshof LE, Visser LL, Emiel JT, Winter-Warnars HA, Lips EH, Wesseling J: **Finding the balance between over-and under-treatment of ductal carcinoma in situ (DCIS)**. *The Breast* 2017, **31**:274-283.
33. Raldow AC, Sher D, Chen AB, Recht A, Punglia RS: **Cost effectiveness of the oncotype DX DCIS score for guiding treatment of patients with ductal carcinoma in situ**. *Journal of Clinical Oncology* 2016, **34**(33):3963-3968.
34. Nofech-Mozes S, Spayne J, Rakovitch E, Hanna W: **Prognostic and predictive molecular markers in DCIS: a review**. *Advances in anatomic pathology* 2005, **12**(5):256-264.
35. Schott M, Klöppel G, Raffel A, Saleh A, Knoefel WT, Scherbaum WA: **Neuroendocrine neoplasms of the gastrointestinal tract**. *Deutsches Arzteblatt international* 2011, **108**(18):305-312.

36. Halfdanarson TR, Rabe KG, Rubin J, Petersen GM: **Pancreatic neuroendocrine tumors (PNETs): incidence, prognosis and recent trend toward improved survival**. *Annals of oncology : official journal of the European Society for Medical Oncology* 2008, **19**(10):1727-1733.
37. Singhi AD, Klimstra DS: **Well-differentiated pancreatic neuroendocrine tumours (Pan NET s) and poorly differentiated pancreatic neuroendocrine carcinomas (Pan NEC s): concepts, issues and a practical diagnostic approach to high-grade (G3) cases**. *Histopathology* 2018, **72**(1):168-177.
38. Reid MD, Balci S, Saka B, Adsay NV: **Neuroendocrine tumors of the pancreas: current concepts and controversies**. *Endocrine pathology* 2014, **25**(1):65-79.
39. Tang LH, Untch BR, Reidy DL, O'Reilly E, Dhall D, Jih L, Basturk O, Allen PJ, Klimstra DS: **Well-differentiated neuroendocrine tumors with a morphologically apparent high-grade component: a pathway distinct from poorly differentiated neuroendocrine carcinomas**. *Clinical Cancer Research* 2016, **22**(4):1011-1017.
40. Viúdez A, Carvalho FL, Maleki Z, Zahurak M, Laheru D, Stark A, Azad NZ, Wolfgang CL, Baylin S, Herman JG: **A new immunohistochemistry prognostic score (IPS) for recurrence and survival in resected pancreatic neuroendocrine tumors (PanNET)**. *Oncotarget* 2016, **7**(18):24950.
41. Gao H, Liu L, Wang W, Xu H, Jin K, Wu C, Qi Z, Zhang S, Liu C, Xu J: **Novel recurrence risk stratification of resected pancreatic neuroendocrine tumor**. *Cancer letters* 2018, **412**:188-193.
42. Kelgiorgi D, Dervenis C: **Pancreatic neuroendocrine tumors: the basics, the gray zone, and the target**. *F1000Research* 2017, **6**:663-663.
43. Louis DN, Gerber GK, Baron JM, Bry L, Dighe AS, Getz G, Higgins JM, Kuo FC, Lane WJ, Michaelson JS: **Computational pathology: an emerging definition**. *Archives of pathology & laboratory medicine* 2014, **138**(9):1133-1138.
44. Nawaz S, Heindl A, Koelble K, Yuan Y: **Beyond immune density: critical role of spatial heterogeneity in estrogen receptor-negative breast cancer**. *Modern Pathology* 2015, **28**(6):766.
45. Naik S, Doyle S, Agner S, Madabhushi A, Feldman M, Tomaszewski J: **Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology**. In*: 2008*: IEEE: 284-287.
46. Gertych A, Ing N, Ma Z, Fuchs TJ, Salman S, Mohanty S, Bhele S, Velásquez-Vacca A, Amin MB, Knudsen BS: **Machine learning approaches to analyze histological images of tissues from radical prostatectomies**. *Computerized Medical Imaging and Graphics* 2015, **46**:197-208.
47. Xu Y, Zhu J-Y, Eric I, Chang C, Lai M, Tu Z: **Weakly supervised histopathology cancer image segmentation and classification**. *Medical image analysis* 2014, **18**(3):591-604.
48. Yu K-H, Zhang C, Berry GJ, Altman RB, Ré C, Rubin DL, Snyder M: **Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features**. *Nature communications* 2016, **7**.
49. Rana A, Yauney G, Lowe A, Shah P: **Computational Histological Staining and Destaining of Prostate Core Biopsy RGB Images with Generative Adversarial Neural Networks**. In: *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA): 2018*: IEEE; 2018: 828-834.
50. Moolayil J: **An Introduction to Deep Learning and Keras**. In: *Learn Keras for Deep Neural Networks.* edn.: Springer; 2019: 1-16.
51. Hamidinekoo A, Denton E, Rampun A, Honnor K, Zwiggelaar R: **Deep learning in mammography and breast histology, an overview and future trends**. *Medical image analysis* 2018, **47**:45-67.
52. Beck AH, Sangoi AR, Leung S, Marinelli RJ, Nielsen TO, Van De Vijver MJ, West RB, Van De Rijn M, Koller D: **Systematic analysis of breast cancer morphology uncovers stromal features associated with survival**. *Science translational medicine* 2011, **3**(108):108ra113-108ra113.

53. Ademuyiwa FO, Miller A, O'Connor T, Edge SB, Thorat MA, Sledge GW, Levine E, Badve S: **The effects of oncotype DX recurrence scores on chemotherapy utilization in a multi-institutional breast cancer cohort**. *Breast cancer research and treatment* 2011, **126**(3):797-802.
54. Sapino A, Roepman P, Linn SC, Snel MH, Delahaye LJ, Van Den Akker J, Glas AM, Simon IM, Barth N, De Snoo FA: **MammaPrint molecular diagnostics on formalin-fixed, paraffin-embedded tissue**. *The Journal of Molecular Diagnostics* 2014, **16**(2):190-197.
55. Chan JK: **The Wonderful Colors of the Hematoxylin–Eosin Stain in Diagnostic Surgical Pathology**. *International journal of surgical pathology* 2014, **22**(1):12-32.
56. Kenny PA, Lee GY, Myers CA, Neve RM, Semeiks JR, Spellman PT, Lorenz K, Lee EH, Barcellos-Hoff MH, Petersen OW: **The morphologies of breast cancer cell lines in three-dimensional assays correlate with their profiles of gene expression**. *Molecular oncology* 2007, **1**(1):84-96.
57. Galea MH, Blamey RW, Elston CE, Ellis IO: **The Nottingham Prognostic Index in primary breast cancer**. *Breast cancer research and treatment* 1992, **22**(3):207-219.
58. Group UCSW: **United States cancer statistics: 1999–2013 incidence and mortality web-based report**. *Atlanta, GA* 2016.
59. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D: **Global cancer statistics**. *CA: A Cancer Journal for Clinicians* 2011, **61**(2):69-90.
60. Weigelt B, Peterse JL, van't Veer LJ: **Breast cancer metastasis: markers and models**. *Nat Rev Cancer* 2005, **5**(8):591-602.
61. Nguyen DX, Bos PD, Massague J: **Metastasis: from dissemination to organ-specific colonization**. *Nat Rev Cancer* 2009, **9**(4):274-284.
62. Lee YTNM: **Breast carcinoma: pattern of metastasis at autopsy**. *Journal of surgical oncology* 1983, **23**(3):175-180.
63. Rakha EA, El-Sayed ME, Green AR, Lee AHS, Robertson JF, Ellis IO: **Prognostic markers in triple-negative breast cancer**. *Cancer* 2007, **109**(1):25-32.
64. Fidler IJ: **The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited**. *Nat Rev Cancer* 2003, **3**(6):453-458.
65. Luzzi KJ, MacDonald IC, Schmidt EE, Kerkvliet N, Morris VL, Chambers AF, Groom AC: **Multistep Nature of Metastatic Inefficiency : Dormancy of Solitary Cells after Successful Extravasation and Limited Survival of Early Micrometastases**. *The American Journal of Pathology* 1998, **153**(3):865-873.
66. Weigelt B, Glas AM, Wessels LFA, Witteveen AT, Peterse JL, van't Veer LJ: **Gene expression profiles of primary breast tumors maintained in distant metastases**. *Proceedings of the National Academy of Sciences* 2003, **100**(26):15901-15905.
67. Smith MR, Saad F, Coleman R, Shore N, Fizazi K, Tombal B, Miller K, Sieber P, Karsh L, Damião R *et al*: **Denosumab and Bone Metastasis-Free Survival in Men With Castration-Resistant Prostate Cancer: Results of a Global Phase 3, Randomised, Placebo-Controlled Trial**. *Lancet* 2012, **379**(9810):39-46.
68. Hillner BE, Ingle JN, Chlebowski RT, Gralow J, Yee GC, Janjan NA, Cauley JA, Blumenstein BA, Albain KS, Lipton A *et al*: **American Society of Clinical Oncology 2003 Update on the Role of Bisphosphonates and Bone Health Issues in Women With Breast Cancer**. *Journal of Clinical Oncology* 2003, **21**(21):4042-4057.
69. Coleman R, Body JJ, Aapro M, Hadji P, Herrstedt J: **Bone health in cancer patients: ESMO Clinical Practice Guidelines**. *Annals of Oncology* 2014, **25**(suppl 3):iii124-iii137.
70. Steeg PS, Camphausen KA, Smith QR: **Brain metastases as preventive and therapeutic targets**. *Nat Rev Cancer* 2011, **11**(5):352-363.
71. Palmieri D, Lockman PR, Thomas FC, Hua E, Herring J, Hargrave E, Johnson M, Flores N, Qian Y, Vega-Valle E *et al*: **Vorinostat Inhibits Brain Metastatic Colonization in a Model of Triple-**

Negative Breast Cancer and Induces DNA Double-Strand Breaks. *Clinical Cancer Research* 2009, **15**(19):6148-6157.

72. Massard C, Zonierek J, Gross-Goupil M, Fizazi K, Szczylik C, Escudier B: **Incidence of brain metastases in renal cell carcinoma treated with sorafenib**. *Annals of Oncology* 2010, **21**(5):1027-1031.
73. Hiratsuka S, Watanabe A, Aburatani H, Maru Y: **Tumour-mediated upregulation of chemoattractants and recruitment of myeloid cells predetermines lung metastasis**. *Nat Cell Biol* 2006, **8**(12):1369-1375.
74. Chen WS, Wei SJ, Liu JM, Hsiao M, Kou-Lin J, Yang WK: **Tumor invasiveness and liver metastasis of colon cancer cells correlated with cyclooxygenase-2 (COX-2) expression and inhibited by a COX-2–selective inhibitor, etodolac**. *International journal of cancer* 2001, **91**(6):894-899.
75. Minn AJ, Kang Y, Serganova I, Gupta GP, Giri DD, Doubrovin M, Ponomarev V, Gerald WL, Blasberg R, Massagué J: **Distinct organ-specific metastatic potential of individual breast cancer cells and primary tumors**. *The Journal of clinical investigation* 2005, **115**(1):44-55.
76. GUPTA GP, MINN AJ, KANG Y, SIEGEL PM, SERGANOVA I, CORDÓN-CARDO C, OLSHEN AB, GERALD WL, MASSAGUÉ J: **Identifying Site-specific Metastasis Genes and Functions**. *Cold Spring Harbor Symposia on Quantitative Biology* 2005, **70**:149-158.
77. Landemaine T, Jackson A, Bellahcène A, Rucci N, Sin S, Abad BM, Sierra A, Boudinet A, Guinebretière J-M, Ricevuto E *et al*: **A Six-Gene Signature Predicting Breast Cancer Lung Metastasis**. *Cancer Research* 2008, **68**(15):6092-6099.
78. Caulfield T, Evans J, McGuire A, McCabe C, Bubela T, Cook-Deegan R, Fishman J, Hogarth S, Miller FA, Ravitsky V *et al*: **Reflections on the Cost of "Low-Cost" Whole Genome Sequencing: Framing the Health Policy Debate**. *PLoS Biology* 2013, **11**(11).
79. Rakha EA, Elsheikh SE, Aleskandarany MA, Habashi HO, Green AR, Powe DG, El-Sayed ME, Benhasouna A, Brunet JS, Akslen LA *et al*: **Triple-negative breast cancer: distinguishing between basal and nonbasal subtypes**. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2009, **15**(7):2302-2310.
80. Rakha EA, El-Sayed ME, Green AR, Lee AH, Robertson JF, Ellis IO: **Prognostic markers in triple-negative breast cancer**. *Cancer* 2007, **109**(1):25-32.
81. Abd El-Rehim DM, Pinder SE, Paish CE, Bell J, Blamey RW, Robertson JF, Nicholson RI, Ellis IO: **Expression of luminal and basal cytokeratins in human breast carcinoma**. *J Pathol* 2004, **203**(2):661-671.
82. Rakha EA, Putti TC, Abd El-Rehim DM, Paish C, Green AR, Powe DG, Lee AH, Robertson JF, Ellis IO: **Morphological and immunophenotypic analysis of breast carcinomas with basal and myoepithelial differentiation**. *J Pathol* 2006, **208**(4):495-506.
83. Abd El-Rehim DM, Ball G, Pinder SE, Rakha E, Paish C, Robertson JF, Macmillan D, Blamey RW, Ellis IO: **High-throughput protein expression analysis using tissue microarray technology of a large well-characterised series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses**. *International journal of cancer Journal international du cancer* 2005, **116**(3):340-350.
84. Galea MH, Blamey RW, Elston CW, et al: **The Nottingham Prognostic Index in primary breast cancer.** *Br Cancer Res Treat* 1992, **22**: 207-219.
85. Abduljabbar R, Negm OH, Lai CF, Jerjees DA, Al-Kaabi M, Hamed MR, Tighe PJ, Buluwela L, Mukherjee A, Green AR *et al*: **Clinical and biological significance of glucocorticoid receptor (GR) expression in breast cancer**. *Breast cancer research and treatment* 2015, **150**(2):335-346.
86. Habashy HO, Rakha EA, Aleskandarany M, Ahmed MA, Green AR, Ellis IO, Powe DG: **FOXO3a nuclear localisation is associated with good prognosis in luminal-like breast cancer**. *Breast cancer research and treatment* 2011, **129**(1):11-21.

87. Mahmoud SM, Paish EC, Powe DG, Macmillan RD, Grainge MJ, Lee AH, Ellis IO, Green AR: **Tumor-infiltrating CD8+ lymphocytes predict clinical outcome in breast cancer**. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2011, **29**(15):1949-1955.
88. Jerjees DA, Negm OH, Alabdullah ML, Mirza S, Alkaabi M, Hameed MR, Abduljabbar R, Muftah A, Nolan CC, Green AR *et al*: **The mammalian target of rapamycin complex 1 (mTORC1) in breast cancer: the impact of oestrogen receptor and HER2 pathways**. *Breast cancer research and treatment* 2015, **150**(1):91-103.
89. Elsheikh SE, Green AR, Rakha EA, Samaka RM, Ammar AA, Powe D, Reis-Filho JS, Ellis IO: **Caveolin 1 and Caveolin 2 are associated with breast cancer basal-like and triple-negative immunophenotype**. *British journal of cancer* 2008, **99**(2):327-334.
90. Alshareeda AT, Negm OH, Green AR, Nolan CC, Tighe P, Albarakati N, Sultana R, Madhusudan S, Ellis IO, Rakha EA: **KPNA2 is a nuclear export protein that contributes to aberrant localisation of key proteins and poor prognosis of breast cancer**. *British journal of cancer* 2015, **112**(12):1929-1937.
91. Meyers J, Mandrekar J: **Cutpoint Determination Methods in Survival Analysis using SAS®: Updated% FINDCUT macro**.
92. Bull AD: **Convergence rates of efficient global optimization algorithms**. *Journal of Machine Learning Research* 2011, **12**(Oct):2879-2904.
93. Gelbart MA, Snoek J, Adams RP: **Bayesian optimization with unknown constraints**. *arXiv preprint arXiv:14035607* 2014.
94. Park S, Han W, Kim J, Kim MK, Lee E, Yoo TK, Lee HB, Kang YJ, Kim YG, Moon HG *et al*: **Risk Factors Associated with Distant Metastasis and Survival Outcomes in Breast Cancer Patients with Locoregional Recurrence**. *Journal of Breast Cancer* 2015, **18**(2):160-166.
95. Pogoda K, Niwińska A, Murawska M, Pieńkowski T: **Analysis of pattern, time and risk factors influencing recurrence in triple-negative breast cancer patients**. *Medical Oncology (Northwood, London, England)* 2013, **30**(1).
96. Gupta GP, Perk J, Acharyya S, de Candia P, Mittal V, Todorova-Manova K, Gerald WL, Brogi E, Benezra R, Massagué J: **ID genes mediate tumor reinitiation during breast cancer lung metastasis**. *Proceedings of the National Academy of Sciences* 2007, **104**(49):19506-19511.
97. Ruiz de Garibay G, Herranz C, Llorente A, Boni J, Serra-Musach J, Mateo F, Aguilar H, Gómez-Baldó L, Petit A, Vidal A *et al*: **Lymphangioleiomyomatosis Biomarkers Linked to Lung Metastatic Potential and Cell Stemness**. *PLoS ONE* 2015, **10**(7).
98. Barnholtz-Sloan JS, Sloan AE, Davis FG, Vigneau FD, Lai P, Sawaya RE: **Incidence Proportions of Brain Metastases in Patients Diagnosed (1973 to 2001) in the Metropolitan Detroit Cancer Surveillance System**. *Journal of Clinical Oncology* 2004, **22**(14):2865-2872.
99. Weil RJ, Palmieri DC, Bronder JL, Stark AM, Steeg PS: **Breast Cancer Metastasis to the Central Nervous System**. *The American Journal of Pathology* 2005, **167**(4):913-920.
100. Klos KJ, O'Neill BP: **Brain Metastases**. *The Neurologist* 2004, **10**(1):31-46.
101. Arnold SM, Young AB, Munn RK, Patchell RA, Nanayakkara N, Markesbery WR: **Expression of p53, bcl-2, E-Cadherin, Matrix Metalloproteinase-9, and Tissue Inhibitor of Metalloproteinases-1 in Paired Primary Tumors and Brain Metastasis**. *Clinical Cancer Research* 1999, **5**(12):4028-4033.
102. Deeken JF, Löscher W: **The Blood-Brain Barrier and Cancer: Transporters, Treatment, and Trojan Horses**. *Clinical Cancer Research* 2007, **13**(6):1663-1674.
103. Voduc KD, Nielsen TO, Perou CM, Harrell JC, Fan C, Kennecke H, Minn AJ, Cryns VL, Cheang MC: **[alpha] B-crystallin expression in breast cancer is associated with brain metastasis**. *npj Breast Cancer* 2015, **1**:15014.

104. Malin D, Strekalova E, Petrovic V, Deal AM, Ahmad AA, Adamo B, Miller CR, Ugolkov A, Livasy C, Fritchie K *et al*: **αB-crystallin: a Novel Regulator of Breast Cancer Metastasis to the Brain**. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2014, **20**(1):56-67.
105. Cardoso F, Costa A, Norton L, Cameron D, Cufer T, Fallowfield L, Francis P, Gligorov J, Kyriakides S, Lin N *et al*: **1st International consensus guidelines for advanced breast cancer (ABC 1)**. *The Breast* 2012, **21**(3):242-252.
106. Cleere DW: **Triple-negative breast cancer: a clinical update**. *Community Oncology* 2010, **7**(5):203-211.
107. Liedtke C, Mazouni C, Hess KR, André F, Tordai A, Mejia JA, Symmans WF, Gonzalez-Angulo AM, Hennessy B, Green M *et al*: **Response to Neoadjuvant Therapy and Long-Term Survival in Patients With Triple-Negative Breast Cancer**. *Journal of Clinical Oncology* 2008, **26**(8):1275-1281.
108. Largillier R, Ferrero JM, Doyen J, Barriere J, Namer M, Mari V, Courdi A, Hannoun-Levi JM, Ettore F, Birtwisle-Peyrottes I *et al*: **Prognostic factors in 1,038 women with metastatic breast cancer**. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* 2008, **19**(12):2012-2019.
109. Hu G, Kang Y, Wang XF: **From breast to the brain: unraveling the puzzle of metastasis organotropism**. *Journal of molecular cell biology* 2009, **1**(1):3-5.
110. Lorusso G, Ruegg C: **New insights into the mechanisms of organ-specific breast cancer metastasis**. *Seminars in cancer biology* 2012, **22**(3):226-233.
111. Roodman GD: **Mechanisms of Bone Metastasis**. *New England Journal of Medicine* 2004, **350**(16):1655-1664.
112. Mundy GR: **Metastasis: Metastasis to bone: causes, consequences and therapeutic opportunities**. *Nat Rev Cancer* 2002, **2**(8):584-593.
113. Winczura P, Sosińska-Mielcarek K, Duchnowska R, Badzio A, Lakomy J, Majewska H, Pęksa R, Pieczyńska B, Radecka B, Dębska-Szmich S *et al*: **Immunohistochemical Predictors of Bone Metastases in Breast Cancer Patients**. *Pathology Oncology Research* 2015, **21**(4):1229-1236.
114. Mihai R, Stevens J, McKinney C, Ibrahim NBN: **Expression of the calcium receptor in human breast cancer—a potential new marker predicting the risk of bone metastases**. *European Journal of Surgical Oncology (EJSO)* 2006, **32**(5):511-515.
115. Kumar R, Wang R-A, Bagheri-Yarmand R: **Emerging roles of MTA family members in human cancers**. *Seminars in Oncology*, **30**:30-37.
116. Kai L, Wang J, Ivanovic M, Chung YT, Laskin WB, Schulze-Hoepfner F, Mirochnik Y, Satcher RL, Jr., Levenson AS: **Targeting prostate cancer angiogenesis through metastasis-associated protein 1 (MTA1)**. *The Prostate* 2011, **71**(3):268-280.
117. Pakala SB, Rayala SK, Wang RA, Ohshiro K, Mudvari P, Reddy SDN, Zheng Y, Pires R, Casimiro S, Pillai MR *et al*: **MTA1 Promotes STAT3 Transcription and Pulmonary Metastasis in Breast Cancer**. *Cancer Res* 2013, **73**(12):3761-3770.
118. Dankof A, Fritzsche FR, Dahl E, Pahl S, Wild P, Dietel M, Hartmann A, Kristiansen G: **KPNA2 protein expression in invasive breast carcinoma and matched peritumoral ductal carcinoma in situ**. *Virchows Archiv* 2007, **451**(5):877-881.
119. Dahl E, Kristiansen G, Gottlob K, Klaman I, Ebner E, Hinzmann B, Hermann K, Pilarsky C, Dürst M, Klinkhammer-Schalke M *et al*: **Molecular Profiling of Laser-Microdissected Matched Tumor and Normal Breast Tissue Identifies Karyopherin α2 as a Potential Novel Prognostic Marker in Breast Cancer**. *Clinical Cancer Research* 2006, **12**(13):3950-3960.
120. Cavallaro U, Christofori G: **Cell adhesion and signalling by cadherins and Ig-CAMs in cancer**. *Nat Rev Cancer* 2004, **4**(2):118-132.

121. Aleskandarany MA, Negm OH, Green AR, Ahmed MA, Nolan CC, Tighe PJ, Ellis IO, Rakha EA: **Epithelial mesenchymal transition in early invasive breast cancer: an immunohistochemical and reverse phase protein array study**. *Breast cancer research and treatment* 2014, **145**(2):339-348.
122. Hazan RB, Phillips GR, Qiao RF, Norton L, Aaronson SA: **Exogenous expression of N-cadherin in breast cancer cells induces cell migration, invasion, and metastasis**. *J Cell Biol* 2000, **148**(4):779-790.
123. Bernard-Gallon D, Bosviel R, Delort L, Fontana L, Chamoux A, Rabiau N, Kwiatkowski F, Chalabi N, Satih S, Bignon YJ: **DNA repair gene ERCC2 polymorphisms and associations with breast and ovarian cancer risk**. *Molecular Cancer* 2008, **7**:36.
124. Smolarz B, Makowska M, Samulak D, Michalska MM, Mojs E, Wilczak M, Romanowicz H: **Single nucleotide polymorphisms (SNPs) of ERCC2, hOGG1, and XRCC1 DNA repair genes and the risk of triple-negative breast cancer in Polish women**. *Tumour Biology* 2014, **35**(4):3495-3502.
125. Smid M, Wang Y, Klijn JG, Sieuwerts AM, Zhang Y, Atkins D, Martens JW, Foekens JA: **Genes associated with breast cancer metastatic to bone**. *Journal of Clinical Oncology* 2006, **24**(15):2261-2267.
126. Bohn OL, Nasir I, Brufsky A, Tseng GC, Bhargava R, MacManus K, Chivukula M: **Biomarker profile in breast carcinomas presenting with bone metastasis**. *Int J Clin Exp Pathol* 2009, **3**(2):139-146.
127. Perry JK, Kannan N, Grandison PM, Mitchell MD, Lobie PE: **Are trefoil factors oncogenic?** *Trends in Endocrinology & Metabolism*, **19**(2):74-81.
128. Buache E, Etique N, Alpy F, Stoll I, Muckensturm M, Reina-San-Martin B, Chenard MP, Tomasetto C, Rio MC: **Deficiency in trefoil factor 1 (TFF1) increases tumorigenicity of human breast cancer cells and mammary tumor development in TFF1-knockout mice**. *Oncogene* 2011, **30**(29):3261-3273.
129. PREST SJ, MAY FEB, WESTLEY BR: **The estrogen-regulated protein, TFF1, stimulates migration of human breast cancer cells**. *The FASEB Journal* 2002, **16**(6):592-594.
130. Sheikh MS, Shao Z-M, Chen J-C, Hussain A, Jetten AM, Fontana JA: **Estrogen receptor-negative breast cancer cells transfected with the estrogen receptor exhibit increased RARα gene expression and sensitivity to growth inhibition by retinoic acid**. *Journal of Cellular Biochemistry* 1993, **53**(4):394-404.
131. Bos PD, Zhang XH-F, Nadal C, Shu W, Gomis RR, Nguyen DX, Minn AJ, van de Vijver MJ, Gerald WL, Foekens JA: **Genes that mediate breast cancer metastasis to the brain**. *Nature* 2009, **459**(7249):1005-1009.
132. Park SH, Noh SJ, Kim KM, Bae JS, Kwon KS, Jung SH, Kim JR, Lee H, Chung MJ, Moon WS *et al*: **Expression of DNA Damage Response Molecules PARP1, γH2AX, BRCA1, and BRCA2 Predicts Poor Survival of Breast Carcinoma Patients()**. *Translational Oncology* 2015, **8**(4):239-249.
133. Spain BH, Larson CJ, Shihabuddin LS, Gage FH, Verma IM: **Truncated BRCA2 is cytoplasmic: Implications for**. *Proceedings of the National Academy of Sciences of the United States of America* 1999, **96**(24):13920-13925.
134. Allegra CJ, Aberle DR, Ganschow P, Hahn SM, Lee CN, Millon-Underwood S, Pike MC, Reed SD, Saftlas AF, Scarvalone SA: **National Institutes of Health State-of-the-Science Conference statement: diagnosis and management of ductal carcinoma in situ September 22–24, 2009**. *Journal of the National Cancer Institute* 2010, **102**(3):161-169.
135. Ernster VL, Barclay J, Kerlikowske K, Wilkie H, Ballard-Barbash R: **Mortality among women with ductal carcinoma in situ of the breast in the population-based surveillance, epidemiology and end results program**. *Archives of internal medicine* 2000, **160**(7):953-958.

136. Worni M, Akushevich I, Greenup R, Sarma D, Ryser MD, Myers ER, Hwang ES: **Trends in treatment patterns and outcomes for ductal carcinoma in situ**. *Journal of the National Cancer Institute* 2015, **107**(12):djv263.
137. Groen EJ, Elshof LE, Visser LL, Rutgers EJT, Winter-Warnars HAO, Lips EH, Wesseling J: **Finding the balance between over- and under-treatment of ductal carcinoma in situ (DCIS)**. *The Breast* 2017, **31**:274-283.
138. Carraro DM, Elias EV, Andrade VP: **Ductal carcinoma in situ of the breast: morphological and molecular features implicated in progression**. *Bioscience reports* 2014, **34**(1):e00090.
139. Cowell CF, Weigelt B, Sakr RA, Ng CKY, Hicks J, King TA, Reis-Filho JS: **Progression from ductal carcinoma in situ to invasive breast cancer: revisited**. *Molecular oncology* 2013, **7**(5):859-869.
140. Provenzano E, Hopper JL, Giles GG, Marr G, Venter DJ, Armes JE: **Histological markers that predict clinical recurrence in ductal carcinoma in situ of the breast: an Australian population-based study**. *Pathology* 2004, **36**(3):221-229.
141. Gilleard O, Goodman A, Cooper M, Davies M, Dunn J: **The significance of the Van Nuys prognostic index in the management of ductal carcinoma in situ**. *World Journal of Surgical Oncology* 2008, **6**:61-61.
142. Rudloff U, Jacks LM, Goldberg JI, Wynveen CA, Brogi E, Patil S, Van Zee KJ: **Nomogram for predicting the risk of local recurrence after breast-conserving surgery for ductal carcinoma in situ**. *Journal of clinical oncology* 2010, **28**(23):3762-3769.
143. Lagios MD, Silverstein MJ: **Risk of recurrence of ductal carcinoma in situ by oncotype Dx technology: some concerns**. *Cancer* 2014, **120**(7):1085-1085.
144. Cowell CF, Weigelt B, Sakr RA, Ng CK, Hicks J, King TA, Reis-Filho JS: **Progression from ductal carcinoma in situ to invasive breast cancer: revisited**. *Molecular oncology* 2013, **7**(5):859-869.
145. Gorringe KL, Fox SB: **Ductal carcinoma in situ (DCIS) biology, biomarkers and diagnosis**. *Frontiers in oncology* 2017, **7**:248.
146. Fuchs TJ, Buhmann JM: **Computational pathology: Challenges and promises for tissue analysis**. *Computerized Medical Imaging and Graphics* 2011, **35**(7):515-530.
147. Gertych A, Swiderska-Chadaj Z, Ma Z, Ing N, Markiewicz T, Cierniak S, Salemi H, Guzman S, Walts AE, Knudsen BS: **Convolutional neural networks can accurately distinguish four histologic growth patterns of lung adenocarcinoma in digital slides**. *Scientific reports* 2019, **9**(1):1483-1483.
148. Baek HJ, Kim HS, Kim N, Choi YJ, Kim YJ: **Percent change of perfusion skewness and kurtosis: a potential imaging biomarker for early treatment response in patients with newly diagnosed glioblastomas**. *Radiology* 2012, **264**(3):834-843.
149. Uppaluri R, Hoffman EA, Sonka M, Hunninghake GW, McLennan G: **Interstitial lung disease: a quantitative study using the adaptive multiple feature method**. *American journal of respiratory and critical care medicine* 1999, **159**(2):519-525.
150. Ing N, Huang F, Conley A, You S, Ma Z, Klimov S, Ohe C, Yuan X, Amin MB, Figlin R: **A novel machine learning approach reveals latent vascular phenotypes predictive of renal cancer outcome**. *Scientific reports* 2017, **7**(1):13190.
151. Sertel O, Kong J, Catalyurek UV, Lozanski G, Saltz JH, Gurcan MN: **Histopathological image analysis using model-based intermediate representations and color texture: Follicular lymphoma grading**. *Journal of Signal Processing Systems* 2009, **55**(1-3):169.
152. Parker J, Dance DR, Davies DH, Yeoman LJ, Michell MJ, Humphreys S: **Classification of ductal cacinoma in situ by image analysis of calcifications from digital mammograms**. *The British journal of radiology* 1995, **68**(806):150-159.
153. Young IT, Verbeek P, Mayall BH: **Characterization of chromatin distribution in cell nuclei**. *Cytometry Part A* 1986, **7**(5):467-474.

154. Hoque A, Lippman SM, Boiko IV, Atkinson EN, Sneige N, Sahin A, Weber DM, Risin S, Lagios MD, Schwarting R: **Quantitative nuclear morphometry by image analysis for prediction of recurrence of ductal carcinoma in situ of the breast**. *Cancer Epidemiology and Prevention Biomarkers* 2001, **10**(3):249-259.
155. Axelrod DE, Miller NA, Lickley HL, Qian J, Christens-Barry WA, Yuan Y, Fu Y, Chapman J-AW: **Effect of quantitative nuclear image features on recurrence of ductal carcinoma in situ (DCIS) of the breast**. *Cancer informatics* 2008, **6**:99.
156. Chapman J-AW, Miller NA, Lickley HLA, Qian J, Christens-Barry WA, Fu Y, Yuan Y, Axelrod DE: **Ductal carcinoma in situ of the breast (DCIS) with heterogeneity of nuclear grade: prognostic effects of quantitative nuclear assessment**. *BMC cancer* 2007, **7**(1):174.
157. Sadlonova A, Novak Z, Johnson MR, Bowe DB, Gault SR, Page GP, Thottassery JV, Welch DR, Frost AR: **Breast fibroblasts modulate epithelial cell proliferation in three-dimensional in vitro co-culture**. *Breast Cancer Research* 2005, **7**(1):R46-R59.
158. Knopfelmacher A, Fox J, Lo Y, Shapiro N, Fineberg S: **Correlation of histopathologic features of ductal carcinoma in situ of the breast with the oncotype DX DCIS score**. *Modern Pathology* 2015, **28**(9):1167.
159. Teo NB, Shoker B, Jarvis C, Martin L, Sloane J, Holcombe C: **Angiogenesis and invasive recurrence in ductal carcinoma in situ of the breast**. *European Journal of Cancer* 2003, **39**(1):38-44.
160. Page DL, Dupont WD: **Anatomic markers of human premalignancy and risk of breast cancer**. *Cancer* 1990, **66**(S14):1326-1335.
161. Allred DC, Wu Y, Mao S, Nagtegaal ID, Lee S, Perou CM, Mohsin SK, O'Connell P, Tsimelzon A, Medina D: **Ductal carcinoma in situ and the emergence of diversity during breast cancer evolution**. *Clinical cancer research* 2008, **14**(2):370-378.
162. Pöllänen I, Braithwaite B, Haataja K, Ikonen T, Toivanen P: **Current analysis approaches and performance needs for whole slide image processing in breast cancer diagnostics**. In*: 2015*: IEEE: 319-325.
163. Lester SC, Bose S, Chen Y-Y, Connolly JL, de Baca ME, Fitzgibbons PL, Hayes DF, Kleer C, O'Malley FP, Page DL: **Protocol for the examination of specimens from patients with ductal carcinoma in situ of the breast**. *Archives of pathology & laboratory medicine* 2009, **133**(1):15-25.
164. Goode A, Gilbert B, Harkes J, Jukic D, Satyanarayanan M: **OpenSlide: A vendor-neutral software foundation for digital pathology**. *Journal of pathology informatics* 2013, **4**(1):27.
165. Reinhard E, Adhikhmin M, Gooch B, Shirley P: **Color transfer between images**. *IEEE Computer graphics and applications* 2001, **21**(5):34-41.
166. Janowczyk A, Basavanhally A, Madabhushi A: **Stain normalization using sparse autoencoders (StaNoSA): Application to digital pathology**. *Computerized Medical Imaging and Graphics* 2017, **57**:50-61.
167. Ruifrok AC, Johnston DA: **Quantification of histochemical staining by color deconvolution**. *Analytical and quantitative cytology and histology* 2001, **23**(4):291-299.
168. Gonzalez RC, R.E. Woods, S.L. Eddins: **Chapter 11**. In: *Digital Image Processing Using MATLAB.* edn. New Jersey: Prentice Hall; 2003.
169. Haralick RM, Shanmugam K: **Textural features for image classification**. *IEEE Transactions on systems, man, and cybernetics* 1973, **3**(6):610-621.
170. Galloway MM: **Texture analysis using gray level run lengths**. *Computer Graphics and Image Processing* 1975, **4**(2):172-179.
171. Wei X: **Gray Level Run Length Matrix Toolbox v1.0**. In.; 2007.
172. Costa A: **alceufc/sfta**. In. https://www.mathworks.com/matlabcentral/fileexchange/37933-alceufc-sfta: MATLAB Central File Exchange; 11/2/2016.

173. Costa AF, Humpire-Mamani G, Traina AJM: **An efficient algorithm for fractal analysis of textures**. In: *Graphics, Patterns and Images (SIBGRAPI), 2012 25th SIBGRAPI Conference on: 2012*: IEEE; 2012: 39-46.
174. Kuse M: **Gabor Image Features**. In. http://www.mathworks.com/matlabcentral/fileexchange/38844-gabor-image-features?focused=5249516&tab=function: MATLAB Central File Exchange; 10/30/2012.
175. Kovesi P: **Symmetry and asymmetry from local phase**. In: *Tenth Australian joint conference on artificial intelligence: 1997*: Citeseer; 1997: 2-4.
176. Kovesi P: **Image features from phase congruency**. *Videre: Journal of computer vision research* 1999, **1**(3):1-26.
177. Kuse M, Wang Y-F, Kalasannavar V, Khan M, Rajpoot N: **Local isotropic phase symmetry measure for detection of beta cells and lymphocytes**. *Journal of pathology informatics* 2011, **2**.
178. Araújo T, Aresta G, Castro E, Rouco J, Aguiar P, Eloy C, Polónia A, Campilho A: **Classification of breast cancer histology images using convolutional neural networks**. *PloS one* 2017, **12**(6):e0177544.
179. Breiman L: **Random forests**. *Machine learning* 2001, **45**(1):5-32.
180. Knopfelmacher A, Fox J, Lo Y, Shapiro N, Fineberg S: **Correlation of histopathologic features of ductal carcinoma in situ of the breast with the oncotype DX DCIS score**. *Modern Pathology* 2015, **28**(9):1167-1173.
181. Vaupel P, Kallinowski F, Okunieff P: **Blood flow, oxygen and nutrient supply, and metabolic microenvironment of human tumors: a review**. *Cancer research* 1989, **49**(23):6449-6465.
182. Guyon I, Elisseeff A: **An introduction to variable and feature selection**. *Journal of machine learning research* 2003, **3**(Mar):1157-1182.
183. Ishwaran H, Kogalur UB: **Random survival forests for R**. *R news* 2007, **7**(2):25-31.
184. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS: **Random survival forests**. *The annals of applied statistics* 2008, **2**(3):841-860.
185. Donker M, Litière S, Werutsky G, Julien J-P, Fentiman IS, Agresti R, Rouanet P, de Lara CT, Bartelink H, Duez N: **Breast-conserving treatment with or without radiotherapy in ductal carcinoma in situ: 15-year recurrence rates and outcome after a recurrence, from the EORTC 10853 randomized phase III trial**. *Journal of Clinical Oncology* 2013, **31**(32):4054-4059.
186. Wärnberg F, Garmo H, Emdin S, Hedberg V, Adwall L, Sandelin K, Ringberg A, Karlsson P, Arnesson L-G, Anderson H: **Effect of radiotherapy after breast-conserving surgery for ductal carcinoma in situ: 20 years follow-up in the randomized SweDCIS trial**. *Journal of Clinical Oncology* 2014, **32**(32):3613-3618.
187. Simon RM, Subramanian J, Li M-C, Menezes S: **Using cross-validation to evaluate predictive accuracy of survival risk classifiers based on high-dimensional data**. *Briefings in Bioinformatics* 2011, **12**(3):203-214.
188. Akaike H: **A new look at the statistical model identification**. In: *Selected Papers of Hirotugu Akaike.* edn.: Springer; 1974: 215-222.
189. Harrell FE, Lee KL, Mark DB: **Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors**. *Statistics in medicine* 1996, **15**(4):361-387.
190. Kremers WK: **Concordance for survival time data: fixed and time-dependent covariates and possible ties in predictor and time**. *Mayo Foundation* 2007.
191. Ishwaran H, Kogalur U: **Random Forests for Survival, Regression and Classification (RF-SRC), R package version 1.6**. *URL http://CRAN R-project org/package= randomForestSRC* 2014.
192. Contal C, O'Quigley J: **An application of changepoint methods in studying the effect of age on survival in breast cancer**. *Computational statistics & data analysis* 1999, **30**(3):253-270.

193. Silverstein MJ, Barth A, Poller DN, Gierson ED, Colburn WJ, Waisman JR, Gamagami P: **Ten-year results comparing mastectomy to excision and radiation therapy for ductal carcinoma in situ of the breast**. *European Journal of Cancer* 1995, **31**(9):1425-1427.
194. Liao P-S, Chen T-S, Chung P-C: **A fast algorithm for multilevel thresholding**. *J Inf Sci Eng* 2001, **17**(5):713-727.
195. Carone DM, Lawrence JB: **Heterochromatin instability in cancer: from the Barr body to satellites and the nuclear periphery**. In: *Seminars in cancer biology: 2013*: Elsevier; 2013: 99-108.
196. Mouriquand J, Pasquier D: **Fine needle aspiration of breast carcinoma: a preliminary cytoprognostic study**. *Acta cytologica* 1980, **24**(2):153-159.
197. Fisher ER, Redmond C, Fisher B: **Histologic grading of breast cancer**. *Pathology annual* 1980, **15**(Pt 1):239.
198. Hunt C, Ellis I, Elston C, Locker A, Pearson D, Blamey R: **Cytological grading of breast carcinoma—a feasible proposition?** *Cytopathology* 1990, **1**(5):287-295.
199. Robinson I, McKee G, Nicholson A, Jackson P, Cook M, D'Arcy J, Kissin M: **Prognostic value of cytological grading of fine-needle aspirates from breast carcinomas**. *The Lancet* 1994, **343**(8903):947-949.
200. Yu GH, Cajulis RS, De Frias DV: **Tumor cell (dys) cohesion as a prognostic factor in aspirate smears of breast carcinoma**. *American journal of clinical pathology* 1998, **109**(3):315-319.
201. Taniguchi E, Yang Q, Tang W, Nakamura Y, Shan L, Nakamura M, Sato M, Mori I, Sakurai T, Kakudo K: **Cytologic grading of invasive breast carcinoma**. *Acta cytologica* 2000, **44**(4):587-591.
202. Khan M, Haleem A, Al Hassani H, Kfoury H: **Cytopathological grading, as a predictor of histopathological grade, in ductal carcinoma (NOS) of breast, on air-dried Diff-Quik smears**. *Diagnostic cytopathology* 2003, **29**(4):185-193.
203. Fan F, Namiq AL, Tawfik OW, Thomas PA: **Proposed prognostic score for breast carcinoma on fine needle aspiration based on nuclear grade, cellular dyscohesion and bare atypical nuclei**. *Diagnostic cytopathology* 2006, **34**(8):542-546.
204. Dabiri S, Talebi A, Shahryari J, Safizadeh H: **Distribution of myofibroblast cells and microvessels around invasive ductal carcinoma of the breast and comparing with the adjacent range of their normal-to-DCIS zones**. *Archives of Iranian medicine* 2013, **16**(2):93.
205. Provenzano PP, Inman DR, Eliceiri KW, Knittel JG, Yan L, Rueden CT, White JG, Keely PJ: **Collagen density promotes mammary tumor initiation and progression**. *BMC medicine* 2008, **6**(1):11.
206. Senchukova MA, Nikitenko NV, Tomchuk ON, Zaitsev NV, Stadnikov AA: **Different types of tumor vessels in breast cancer: morphology and clinical value**. *Springerplus* 2015, **4**(1):512.
207. Aroner SA, Collins LC, Schnitt SJ, Connolly JL, Colditz GA, Tamimi RM: **Columnar cell lesions and subsequent breast cancer risk: a nested case-control study**. *Breast Cancer Research* 2010, **12**(4):R61.
208. Ip Y-T, Dias Filho MA, Chan JK: **Nuclear inclusions and pseudoinclusions: friends or foes of the surgical pathologist?** *International journal of surgical pathology* 2010, **18**(6):465-481.
209. Page DL, Dupont WD, Rogers LW, Rados MS: **Atypical hyperplastic lesions of the female breast. A long-term follow-up study**. *Cancer* 1985, **55**(11):2698-2708.
210. Stanton SE, Disis ML: **Clinical significance of tumor-infiltrating lymphocytes in breast cancer**. *Journal for immunotherapy of cancer* 2016, **4**(1):59.
211. Julesz B: **Visual pattern discrimination**. *IRE transactions on Information Theory* 1962, **8**(2):84-92.
212. Tourassi GD: **Journey toward computer-aided diagnosis: role of image texture analysis**. *Radiology* 1999, **213**(2):317-320.

213. Pinder S, Duggan C, Ellis I, Cuzick J, Forbes J, Bishop H, Fentiman I, George W: **A new pathological system for grading DCIS with improved prediction of local recurrence: results from the UKCCCR/ANZ DCIS trial**. *British Journal of Cancer* 2010, **103**(1):94-100.
214. Scripcaru G, Zardawi IM: **Mammary ductal carcinoma in situ: a fresh look at architectural patterns**. *International journal of surgical oncology* 2012, **2012**.
215. DOUGLAS-JONES A, Gupta S, Attanoos R, Morgan J, Mansel R: **A critical appraisal of six modern classifications of ductal carcinoma in situ of the breast (DCIS): correlation with grade of associated invasive carcinoma**. *Histopathology* 1996, **29**(5):397-409.
216. Filipczuk P, Fevens T, Krzyzak A, Monczak R: **Computer-aided breast cancer diagnosis based on the analysis of cytological images of fine needle biopsies**. *IEEE Transactions on Medical Imaging* 2013, **32**(12):2169-2178.
217. Sun X, Chuang S-H, Li J, McKenzie F: **Automatic diagnosis for prostate cancer using run-length matrix method**. In: *Medical Imaging 2009: Computer-Aided Diagnosis: 2009*: International Society for Optics and Photonics; 2009: 72603H.
218. Yang F, Thomas MA, Dehdashti F, Grigsby PW: **Temporal analysis of intratumoral metabolic heterogeneity characterized by textural features in cervical cancer**. *European journal of nuclear medicine and molecular imaging* 2013, **40**(5):716-727.
219. Solin LJ, Gray R, Baehner FL, Butler SM, Hughes LL, Yoshizawa C, Cherbavaz DB, Shak S, Page DL, Sledge Jr GW: **A multigene expression assay to predict local recurrence risk for ductal carcinoma in situ of the breast**. *Journal of the National Cancer Institute* 2013, **105**(10):701-710.
220. Narod SA, Iqbal J, Giannakeas V, Sopik V, Sun P: **Breast cancer mortality after a diagnosis of ductal carcinoma in situ**. *JAMA oncology* 2015, **1**(7):888-896.
221. Giannakeas V, Sopik V, Narod SA: **Association of radiotherapy with survival in women treated for ductal carcinoma in situ with lumpectomy or mastectomy**. *JAMA network open* 2018, **1**(4):e181100-e181100.
222. Rakovitch E, Nofech-Mozes S, Hanna W, Sutradhar R, Baehner FL, Miller DP, Fong C, Gu S, Tuck A, Sengupta S: **Multigene expression assay and benefit of radiotherapy after breast conservation in ductal carcinoma in situ**. *JNCI: Journal of the National Cancer Institute* 2017, **109**(4).
223. Tang LH, Gonen M, Hedvat C, Modlin IM, Klimstra DS: **Objective quantification of the Ki67 proliferative index in neuroendocrine tumors of the gastroenteropancreatic system: a comparison of digital image analysis with manual methods**. *The American journal of surgical pathology* 2012, **36**(12):1761-1770.
224. Salaria SN, Shi C: **Pancreatic neuroendocrine tumors**. *Surgical pathology clinics* 2016, **9**(4):595-617.
225. Gertych A, Swiderska-Chadaj Z, Ma Z, Ing N, Markiewicz T, Cierniak S, Salemi H, Guzman S, Walts AE, Knudsen BS: **Convolutional neural networks can accurately distinguish four histologic growth patterns of lung adenocarcinoma in digital slides**. *Scientific reports* 2019, **9**(1):1483.
226. Mobadersany P, Yousefi S, Amgad M, Gutman DA, Barnholtz-Sloan JS, Vega JEV, Brat DJ, Cooper LA: **Predicting cancer outcomes from histology and genomics using convolutional networks**. *Proceedings of the National Academy of Sciences* 2018, **115**(13):E2970-E2979.
227. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A: **Going deeper with convolutions**. In: *Proceedings of the IEEE conference on computer vision and pattern recognition: 2015*; 2015: 1-9.
228. Basturk O, Yang Z, Tang LH, Hruban RH, Adsay NV, McCall CM, Krasinskas AM, Jang K-T, Frankel WL, Balci S: **The high grade (WHO G3) pancreatic neuroendocrine tumor category is morphologically and biologically heterogeneous and includes both well differentiated and poorly differentiated neoplasms**. *The American journal of surgical pathology* 2015, **39**(5):683.

229. Ball J, Balogh E, Miller BT: **Improving diagnosis in health care**: National Academies Press; 2015.
230. Dalton WS, Friend SH: **Cancer biomarkers—an invitation to the table**. *Science* 2006, **312**(5777):1165-1168.
231. Newman-Toker DE, McDonald KM, Meltzer DO: **How much diagnostic safety can we afford, and how should we decide? A health economics perspective**. *BMJ Qual Saf* 2013, **22**(Suppl 2):ii11-ii20.
232. Sidransky D: **Emerging molecular markers of cancer**. *Nature Reviews Cancer* 2002, **2**(3):210.
233. Panel NIoHCD: **National Institutes of Health Consensus Development Conference statement: adjuvant therapy for breast cancer, November 1–3, 2000**. *JNCI Monographs* 2001, **2001**(30):5-15.
234. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI: **Machine learning applications in cancer prognosis and prediction**. *Computational and structural biotechnology journal* 2015, **13**:8-17.
235. Waddell M, Page D, Shaughnessy Jr J: **Predicting cancer susceptibility from single-nucleotide polymorphism data: a case study in multiple myeloma**. In: *Proceedings of the 5th international workshop on Bioinformatics: 2005*: ACM; 2005: 21-28.
236. Stojadinovic A, Nissan A, Eberhardt J, Chua TC, Pelz JO, Esquivel J: **Development of a Bayesian Belief Network Model for personalized prognostic risk assessment in colon carcinomatosis**. *The American Surgeon* 2011, **77**(2):221-230.
237. Exarchos KP, Goletsis Y, Fotiadis DI: **Multiparametric decision support system for the prediction of oral cancer reoccurrence**. *IEEE Transactions on Information Technology in Biomedicine* 2012, **16**(6):1127-1134.
238. Wang D, Khosla A, Gargeya R, Irshad H, Beck AH: **Deep learning for identifying metastatic breast cancer**. *arXiv preprint arXiv:160605718* 2016.
239. Mercan C, Mercan E, Aksoy S, Shapiro LG, Weaver DL, Elmore JG: **Multi-instance multi-label learning for whole slide breast histopathology**. In: *Medical Imaging 2016: Digital Pathology: 2016*: International Society for Optics and Photonics; 2016: 979108.
240. Ishwaran H, Lu M: **Random survival forests**. *Wiley StatsRef: Statistics Reference Online* 2008:1-13.
241. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y: **DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network**. *BMC medical research methodology* 2018, **18**(1):24.