

RESEARCH ARTICLE

# The use of automated Ki67 analysis to predict Oncotype DX risk-of-recurrence categories in early-stage breast cancer

Satbir Singh Thakur<sup>1,2</sup>, Haocheng Li<sup>1,3</sup>, Angela M. Y. Chan<sup>2</sup>, Roxana Tudor<sup>1</sup>, Gilbert Bigras<sup>4</sup>, Don Morris<sup>1,2</sup>, Emeka K. Enwere<sup>2</sup>\*, Hua Yang<sup>5</sup>\*

**1** Department of Oncology, University of Calgary, Calgary, Alberta, Canada, **2** Translational Laboratories, Tom Baker Cancer Center, Calgary, Alberta, Canada, **3** Department of Community Health Sciences, University of Calgary, Calgary, Alberta, Canada, **4** Department of Pathology and Laboratory Medicine, University of Alberta, Edmonton, Alberta, Canada, **5** Department of Pathology and Laboratory Medicine, University of Calgary, Calgary, Alberta, Canada

\* These authors contributed equally to this work.

\* [emeka.enwere@ucalgary.ca](mailto:emeka.enwere@ucalgary.ca) (EKE); [hua.yang@cls.ab.ca](mailto:hua.yang@cls.ab.ca) (HY)



**OPEN ACCESS**

**Citation:** Thakur SS, Li H, Chan AMY, Tudor R, Bigras G, Morris D, et al. (2018) The use of automated Ki67 analysis to predict Oncotype DX risk-of-recurrence categories in early-stage breast cancer. PLoS ONE 13(1): e0188983. <https://doi.org/10.1371/journal.pone.0188983>

**Editor:** Aamir Ahmad, University of South Alabama Mitchell Cancer Institute, UNITED STATES

**Received:** June 6, 2017

**Accepted:** November 16, 2017

**Published:** January 5, 2018

**Copyright:** © 2018 Thakur et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** Research funding was obtained from Calgary Lab Services. Satbir Thakur is a recipient of a Post Doctoral Scholarship from Alberta Cancer Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** We would like to confirm that all authors approve of the submitted version of

## Abstract

Ki67 is a commonly used marker of cancer cell proliferation, and has significant prognostic value in breast cancer. In spite of its clinical importance, assessment of Ki67 remains a challenge, as current manual scoring methods have high inter- and intra-user variability. A major reason for this variability is selection bias, in that different observers will score different regions of the same tumor. Here, we developed an automated Ki67 scoring method that eliminates selection bias, by using whole-slide analysis to identify and score the tumor regions with the highest proliferative rates. The Ki67 indices calculated using this method were highly concordant with manual scoring by a pathologist (Pearson's  $r = 0.909$ ) and between users (Pearson's  $r = 0.984$ ). We assessed the clinical validity of this method by scoring Ki67 from 328 whole-slide sections of resected early-stage, hormone receptor-positive, human epidermal growth factor receptor 2-negative breast cancer. All patients had Oncotype DX testing performed (Genomic Health) and available Recurrence Scores. High Ki67 indices correlated significantly with several clinico-pathological correlates, including higher tumor grade (1 versus 3,  $P < 0.001$ ), higher mitotic score (1 versus 3,  $P < 0.001$ ), and lower Allred scores for estrogen and progesterone receptors ( $P = 0.002, 0.008$ ). High Ki67 indices were also significantly correlated with higher Oncotype DX risk-of-recurrence group (low versus high,  $P < 0.001$ ). Ki67 index was the major contributor to a machine learning model which, when trained solely on clinico-pathological data and Ki67 scores, identified Oncotype DX high- and low-risk patients with 97% accuracy, 98% sensitivity and 80% specificity. Automated scoring of Ki67 can thus successfully address issues of consistency, reproducibility and accuracy, in a manner that integrates readily into the workflow of a pathology laboratory. Furthermore, automated Ki67 scores contribute significantly to models that predict risk of recurrence in breast cancer.

this manuscript, and that this manuscript is not under consideration elsewhere. None of the authors have conflicts of interest to declare. The funding agencies (Calgary Lab Services and Alberta Cancer Foundation) had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

## Introduction

Breast cancer is the most common form of cancer among women, and is the second-leading cause of cancer-related death worldwide [1]. Treatment decisions for breast cancer are significantly influenced by subtype, which is determined from expression of estrogen receptor (ER), progesterone receptor (PgR), human epidermal growth factor receptor 2 (HER2), and the proliferation marker Ki67 [2]. With advancement of the field of genomics, various genetic tools have been developed which assist in the subtyping of breast cancers, in addition to the immunohistochemical markers listed above. Some of the multigene assays that are currently available for early-stage breast cancers include the Oncotype DX<sup>®</sup> (Genomic Health Inc.), MammaPrint<sup>®</sup> (Agendia BV), Prosigna<sup>®</sup> (PAM50) (NanoString Technologies Inc.) and EndoPredict<sup>®</sup> (Myriad Genetics Inc.) [3]. The percentage of Ki67-positive tumor cells, or the Ki67 index, is used clinically to distinguish between Luminal A and Luminal B subtypes [4]. Expression of Ki67 is also a good predictor of pathological complete response [5–8], response to chemotherapy [9–14], and likelihood of relapse [15–17].

In spite of the significant body of evidence supporting the clinical application of Ki67, its use is hampered by a few drawbacks. First, there are no universally-accepted guidelines for analysis and clinical interpretation of Ki67 staining. For instance, while the St. Gallen Expert Panel recommended a Ki67 cut-point of 14% for administration of neoadjuvant chemotherapy in 2011 [18], they subsequently withdrew their recommendation of a cut-point, given the variability across different centers [19]. Second, there is no consensus about whether the Ki67 index should be calculated as the percentage of Ki67-positive tumor cells in a tissue section, or from “hot spots” representing tumor regions in which cell proliferation is highest [20, 21]. Third, even with consensus amongst the cut-points and scoring methods, inter-observer variability in manual scoring still exists, even among pathologists [22–25]. While the hot spot method is ideal to represent the highest-proliferating region of a heterogeneous disease, individual biases in selecting these hot spots negatively affect inter-observer concordance in scoring. These issues have prompted the development of several multi-center teams and clinical studies to investigate the utility of automated scoring approaches for Ki67.

In the context of early-stage, ER/PgR-positive breast cancer, adjuvant chemotherapy is only effective in a small subgroup of patients with a high risk of relapse [26–29]. A number of multigene assays have emerged in recent years, which assist in clinical management of early-stage breast cancer patients by estimating the risk of recurrence [30–32]. These assays include Oncotype DX, which is used to support clinical decisions involving the administration of chemotherapy [29, 33–35]. The assay involves the use of gene expression analysis to generate a Recurrence Score; this is a continuous variable, ranging from 0 to 100, that indicates risk of cancer recurrence [29]. Based on the Recurrence Score, patients are classified into low (<18), intermediate (18–30), or high ( $\geq 31$ ) risk-of-recurrence groups. The genes assessed by Oncotype DX include those coding for ER, PgR, HER2 and Ki67, with the latter contributing the most to calculation of the Recurrence Score [29, 36, 37]. The high cost of multigene assays (~US\$4,000 per patient for Oncotype DX) has prompted multiple groups to create similar risk assessment models, by using biomarker expression data already collected as part of the standard of care [36, 38–41]. Nevertheless, the absence of reliable data on Ki67 remains a weak link with these efforts.

Here, we report the development of a robust automated method of Ki67 scoring, using the HALO<sup>®</sup> image analysis platform. Our approach combines the biological value of hot spot scoring with whole-slide analysis, to produce results that are unbiased and highly reproducible. We validated the clinical value of Ki67 scores obtained by this approach using standard clinico-pathological data and machine learning algorithms to predict the Oncotype DX risk

category of patients in our cohort. These results, upon validation, may provide a method of consistent and accurate Ki67 assessment that could be easily incorporated into current clinical practice.

## Materials and methods

### Patient selection

This study received ethical approval from the Conjoint Health Research Ethics Board at the University of Calgary. Patients included in this retrospective study were 328 women diagnosed with early stage breast cancer (ER/PgR-positive, HER2-negative, lymph node-negative, stages I to II) treated between 2014 and 2016 in Alberta, Canada. Samples from resected tissue specimens were submitted for Oncotype DX testing (Genomic Health). A database (The Alberta Provincial Oncotype DX Database) was developed by HY (Calgary, Canada) and GB (Edmonton, Canada) from the information obtained from the Oncotype DX testing and other clinico-pathological variables (Age, Tumor Size, Tumor Grade, Tumor Nuclear Grade, Tumor Architecture, Tumor Estrogen and Progesterone Receptor levels) available for these patients.

### Specimen selection and immunohistochemistry

For each case of resected breast cancer, the formalin-fixed, paraffin-embedded tissue block with the largest tumor cross-section was identified. One section (4  $\mu\text{m}$  thickness) was stained for hematoxylin & eosin (H&E). A consecutive section was stained with an antibody to Ki67 (clone MIB1, Dako [Santa Clara, California, USA], 1:200 dilution, EnVision FLEX low pH buffer). The slides were counterstained with FLEX hematoxylin (Dako) and permanently mounted per the manufacturer's protocol.

### Assessment of Ki67 index

Manual scores of Ki67 were generated by a pathologist (HY) from independent counting of at least three high-power fields (40X objective) of the most mitotically active area in a section, with a minimum of 500 invasive tumor cells counted per area. For automated Ki67 assessment, an image analysis algorithm was designed in the HALO<sup>®</sup> image analysis software platform (version 2.0.1145.14, Indica Labs, Corrales, New Mexico, USA) that identified both Ki67-positive nuclei (brown) and Ki67-negative nuclei (hematoxylin-positive, blue) from slides stained using chromogenic immunohistochemistry. Invasive breast cancers were marked on diagnostic H&E slides by a pathologist (HY). Both the H&E slides and serial tissue sections stained for Ki67 were digitized using an Aperio Scanscope XT slide scanner. The area to be analyzed on the Ki67 image was identified and annotated from the matched H&E using the Image Registration tool in the HALO<sup>®</sup> software. Once annotated, the entire tumor-bearing region was segmented into 500  $\times$  500  $\mu\text{m}$  square tiles and submitted for image analysis. The results included the number of nuclei detected, as well as the percentage of Ki67-positive nuclei from each tile. The hot-spot Ki67 index for each case was calculated as the mean percentage of Ki67-positive nuclei in the top five tiles having at least 500 cells per tile. The whole-slide Ki67 index was calculated by determining the percentage of Ki67-positive nuclei present in each tile in the marked area, and then calculating the mean per-tile percentage for all tiles on a slide. In this approach, no criterion of having minimum 500 cells per tile was followed, and scores from all tiles in the marked tumor-containing areas were used to calculate the final Ki67 index.

## Statistical analyses

Correlations involving the Ki67 index were statistically tested by Pearson's correlation coefficient (Pearson's  $r$  and two-tailed  $P$  values) and Lin's concordance correlation coefficient where appropriate using GraphPad PRISM (version 6.0). Correlations between Ki67 indices and various clinico-pathological variables were assessed by nonparametric tests (Kruskal-Wallis to compare among three or more groups; Wilcoxon rank sum for comparison across two groups) for continuous variables and Fisher's exact test for categorical factors in R (version 3.2.4).

## Random Forest modeling

A Random Forest machine learning paradigm was used to evaluate our ability to predict Oncotype DX risk-of-recurrence categories from available data. All modeling was performed in R (version 3.2.4), using the "randomForest" package (version 4.6–12). The data set was derived from the 199 patients from our cohort for whom we had complete data in all variables. The modelling was designed to predict the Oncotype DX-derived Recurrence Score for all patients in the test group. This was done using, in one instance, the first 13 clinico-pathological variables listed in [S2 Table](#) that were obtained from the patients' charts. The Recurrence Score obtained from this modelling was referred to as the predicted Recurrence Score, pRS. The Oncotype DX reports for each patient included gene expression scores for ER, PgR and HER2, as determined by quantitative RT-PCR [42]. No other Oncotype DX gene expression scores are available in the reports. To determine whether these data would improve the accuracy with which a Random Forest model could predict Recurrence Scores, we created a second modelling instance. In this instance, the contributing variables consisted of the first 13 clinico-pathological variables listed in [S2 Table](#), as well as the Oncotype DX gene expression scores for ER, PgR and HER2, obtained from reports. This Random Forest instance thus consisted of 16 variables that were again used to predict the actual Oncotype DX Recurrence Scores. The predicted Recurrence Scores in this instance were termed pRS<sub>odx</sub>, to differentiate them from the pRS that were derived in the absence of any Oncotype DX data. The predicted Recurrence Scores were then assigned to low, intermediate or high risk-of-recurrence groups, using the standard cut points applied to this test (low < 18, intermediate 18–30, high  $\geq$  31, range 0–100). Random Forest parameters for  $n_{tree}$ ,  $m_{try}$ , and  $samplesize$  were optimized using the method of Huang and Boutros [43], and set at  $n_{tree} = 1000$ ;  $m_{try} = 15$  or 12 for analysis including or not including Oncotype DX ER/PgR/HER2 data, respectively; and  $samplesize = 40$ . For each round of analysis, data from a randomly-selected 50% of patients were used to train the model, which was then applied to predict Recurrence Scores of the remaining patients. The predicted Recurrence Scores were assigned to risk-of-recurrence groups, as described above. Performance in each round was evaluated by constructing a confusion matrix (2x2 table) that matched each patient's predicted risk group against the Oncotype DX-determined risk group assignment. For the purposes of assessing predictive performance of the analyses, correct prediction that a patient was of low risk-of-recurrence constituted a true positive test result, and correct prediction that a patient was of high risk constituted a true negative result. Incorrect prediction that a high-risk patient was of low risk constituted a false positive, and incorrect prediction that a low-risk patient was of high risk was a false negative. Patients with intermediate risk (from either actual or predicted Oncotype DX scores) were not included in evaluation of performance. Each round of model-building and prediction also generated data on percentage Increased Mean Squared Error (%IncMSE). This is the difference in mean squared error of the predicted Recurrence Score, between when a specific variable is used in the Random Forest model, and when that variable is randomly permuted [44]. A higher %IncMSE thus indicates

greater predictive value of a variable. Results presented are from 1,000 independent rounds of cross-validation. Performance statistics were reported as mean ± standard deviation.

## Results

### Cohort description

The present study involved 328 patients who were treated for hormone receptor-positive, HER2-negative, early-stage breast cancer in Alberta between 2014 and 2016. The clinico-pathological characteristics of these patients are listed in Table 1. All patients in the cohort underwent Oncotype DX genomic testing and had available Recurrence Scores. Using the standard Genomic Health cut-points, there were 185 low-risk (62%), 110 intermediate-risk (28%) and 33 (10%) high-risk patients. Clinico-pathological characteristics of these patients were also grouped by Oncotype DX risk group (S1 Table).

### Performance evaluation of automated Ki67 analysis

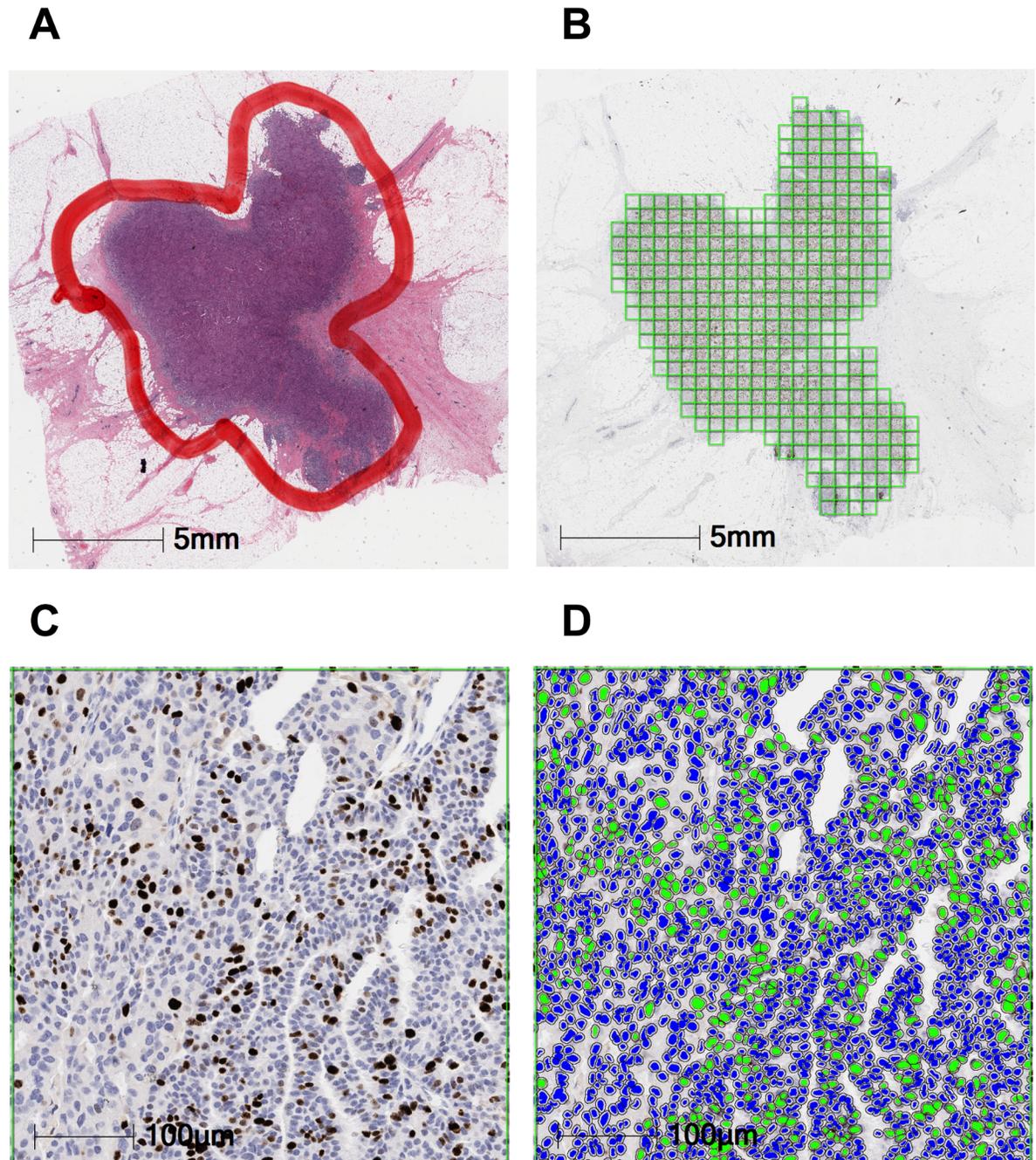
An overview of the analysis workflow is shown in Fig 1. Tumor areas were segmented into square tiles as the first step in the automated analysis. The median number of tiles per slide

**Table 1. Clinico-pathological characteristics of patients in the study cohort.**

Age at Dx (Years)		Patients	%	n
	Mean	55	100	327*
	Range	29–85		
<b>Tumor Size (mm)</b>				
	Mean	18.1	100	316
	Range	2.5–57		
<b>Tumor Grade</b>				
	1	64	20.7	309
	2	196	63.4	
	3	49	15.8	
<b>Tumor Nuclear Grade</b>				
	1	32	10.3	309
	2	172	55.6	
	3	105	33.9	
<b>Tumor Differentiation</b>				
	1	31	10	309
	2	88	28.4	
	3	190	61.4	
<b>Tumor ER Intensity</b>				
	Weak	8	3.2	244
	Moderate	13	5.3	
	Strong	223	91.3	
<b>Tumor PgR Intensity</b>				
	Weak	14	6.1	228
	Moderate	65	28.5	
	Strong	149	65.3	
<b>Recurrence Score</b>				
	Low (<18)	185	56.4	328
	Intermediate (18–30)	110	33.6	
	High (>30)	33	10.0	

\*: Age at Dx not available for one patient. ER, estrogen receptor; PgR, progesterone receptor

<https://doi.org/10.1371/journal.pone.0188983.t001>



**Fig 1. Overview of the automated image analysis workflow.** An H&E of the tumor specimen was manually annotated by a pathologist (A). Annotations were transferred to the matching Ki67-stained slide, and segmented into tiles (B). Ki67-positive and -negative nuclei in each tile (C) were identified and counted by the analysis algorithm, which colored them green and blue, respectively (D).

<https://doi.org/10.1371/journal.pone.0188983.g001>

was 328 (range 11–1600, interquartile range 273). Only tiles with  $\geq 500$  cells were selected for analysis to determine Ki67 index. To evaluate the performance of the automated analysis, we compared the Ki67 indices from 45 randomly selected cases to manual scoring of the same. The manual scoring was performed by a breast pathologist (HY) who was blinded to the automated

assay's results. There was a high correlation between both analysis methods (Fig 2A; Pearson's  $r = 0.909$ ,  $P < 0.001$ ) with high levels of concordance (Lin's concordance = 0.881).

To test the ease of use and reproducibility of the method developed, 50 cases were selected at random, and subjected to automated analysis separately by one novice and one expert user. As expected, scores from both users were concordant and reproducible with minimal variance (Fig 2B; Pearson's  $r = 0.984$ ;  $P < 0.001$ ).

### Clinical validity of whole-slide hot spot analysis

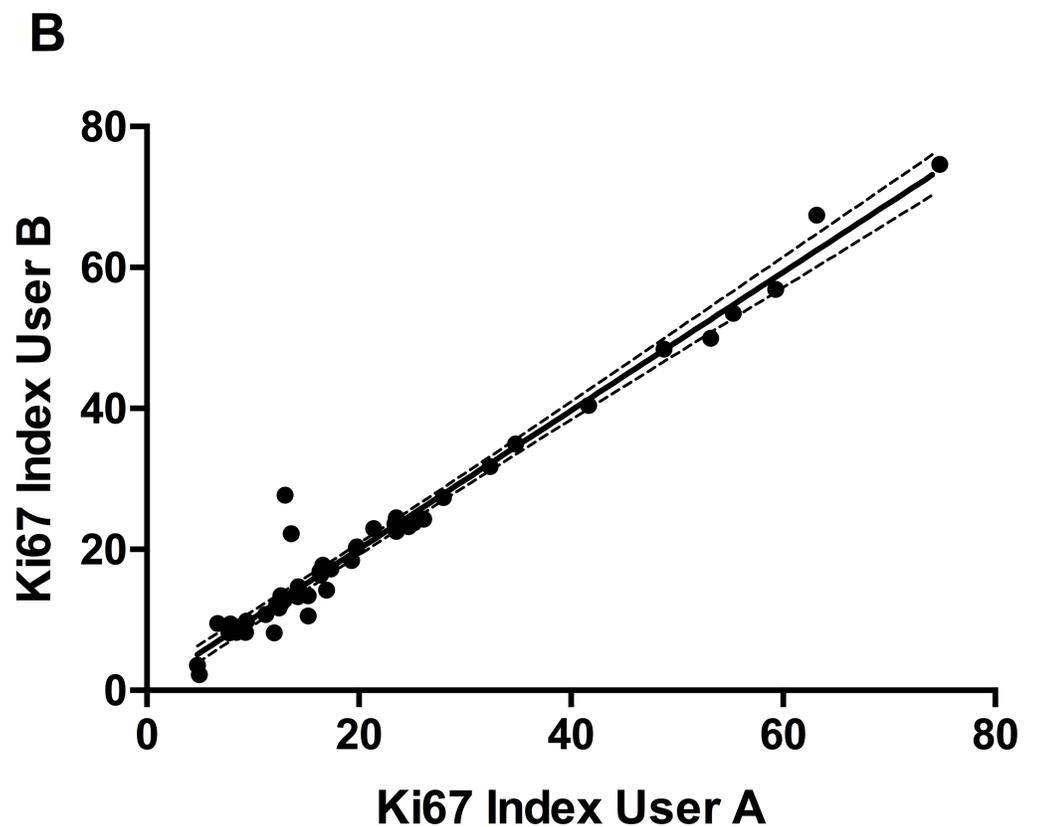
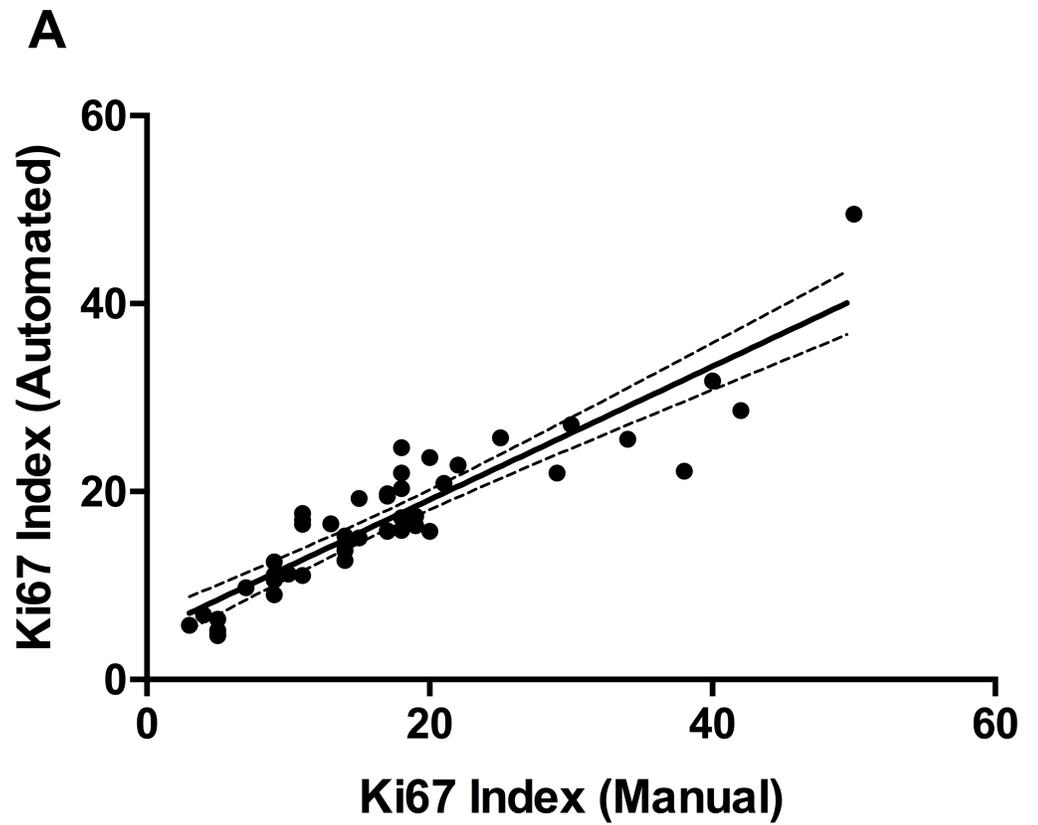
There is no consensus about the relative validity of Ki67 scoring of hot spots, as compared to scoring of the entire slide. The analysis of hot spots allows representation of the most aggressive tumor regions from otherwise heterogeneously-proliferative tumor specimens [20, 45, 46]; however, the utility of this approach is normally marred by inter-user bias. We hypothesized that the elimination of bias, as is the case with our assay, would allow the generation of Ki67 indices that were more strongly correlated with common prognostic factors. Correlations were assessed between whole-slide or hot-spot Ki67 indices and multiple clinico-pathological variables (Table 2). There were interactions between Ki67 indices generated from both approaches (whole-slide and hot-spot) and all three Oncotype DX risk groups. Ki67 indices also correlated positively with the Recurrence Scores, tumor grade and mitotic score, and negatively with ER and PgR Allred scores. There were no notable differences in the extent to which either whole-slide or hot-spot indices interacted with the clinico-pathological variables.

It is worth mentioning here that Ki67 scores from both approaches (whole-slide and hot-spot) were strongly correlated (Pearson  $r = 0.938$ ); however, the hot-spot indices were significantly higher than the whole-slide indices across the entire cohort (S1 Fig) and when patients were stratified into Oncotype DX risk-of-recurrence groups (S2 Fig).

Since Ki67 is the greatest contributor to the Oncotype DX assay's results, we evaluated the association between hot-spot Ki67 indices and Oncotype DX Recurrence Scores. The correlation between these, while modest, was significant (Fig 3A; Pearson's  $r = 0.553$ ,  $P < 0.001$ ). While the distribution of Ki67 indices for low-risk patients overlapped with those for intermediate-risk patients, the median Ki67 indices were significantly higher in the latter (Table 2). In contrast, Ki67 indices of high- and low-risk patients assorted into distinct clusters (Fig 3B). A clear difference in the Ki67 indices between high- and low-risk patients was observed (Pearson's  $r = 0.684$ ,  $P < 0.001$ ), with 97% patients in the high-risk category having Ki67 indices over 20%.

### The use of machine learning to infer Oncotype DX risk groups

Ki67 is the main contributor to the Oncotype DX recurrence score [29, 36, 37], and markers of proliferation similarly contribute to the risk-of-recurrence assessments of other multigene assays [30]. We asked if an integrated machine learning analysis of clinico-pathological data, including the Ki67 scores generated by our automated assay, could provide useful information about recurrence risk in these patients. We created and evaluated a Random Forest machine learning model, using a subset of 199 patients for whom we had complete data sets. While we used all available Oncotype DX data for model training, we focused our assessment of model performance on high- and low-risk patients. This is because the intermediate risk patients constitute an ambiguous group from both predictive and treatment standpoints [36, 41]. The outline for model training, evaluation and cross-validation is shown in Fig 4, the variables used are listed in S2 Table, and the summary results from 1,000 rounds of cross-validation are in Table 3. Predicted Recurrence Scores were obtained using 13 clinico-pathological variables alone (pRS) or in addition to the gene expression scores for ER, PgR and HER2 as obtained



**Fig 2. Performance evaluation of automated Ki67 analysis.** (A) Comparison of Ki67 indices (n = 45) as assessed by manual and automated scoring. Pearson's r = 0.909; Lin's concordance correlation coefficient = 0.881; P<0.001. (B) Assessment of inter-user concordance using the automated analysis (n = 50). Pearson's r = 0.984; P<0.001. Dashed lines indicate 95% confidence intervals.

<https://doi.org/10.1371/journal.pone.0188983.g002>

from the Oncotype DX reports (pRS<sub>odx</sub>). Random Forest models trained with Oncotype DX-derived expression data for ER, PgR and HER2 performed better than models trained without these data; however, the differences were modest, particularly for accuracy (1.0%) and negative predictive value (6.3%). Similar results were obtained when Ki67 indices from whole-slide analyses were used in place of the hot-spot data (Table 3). To determine which variables contributed the most to the accuracy of the model, we calculated the increase in mean squared error, which indicates the degree to which the model's accuracy would decrease if a specific variable were omitted. In the pRS<sub>odx</sub> models, PgR (determined by RT-PCR in the Oncotype DX assay) and Ki67 were the greatest contributors, together accounting for approximately 70% of the model's accuracy (Fig 5A). In the absence of Oncotype DX data, Ki67 was the highest contributor towards model accuracy, with its loss creating a mean squared error increase of 45.3% (Fig 5B).

### Discussion

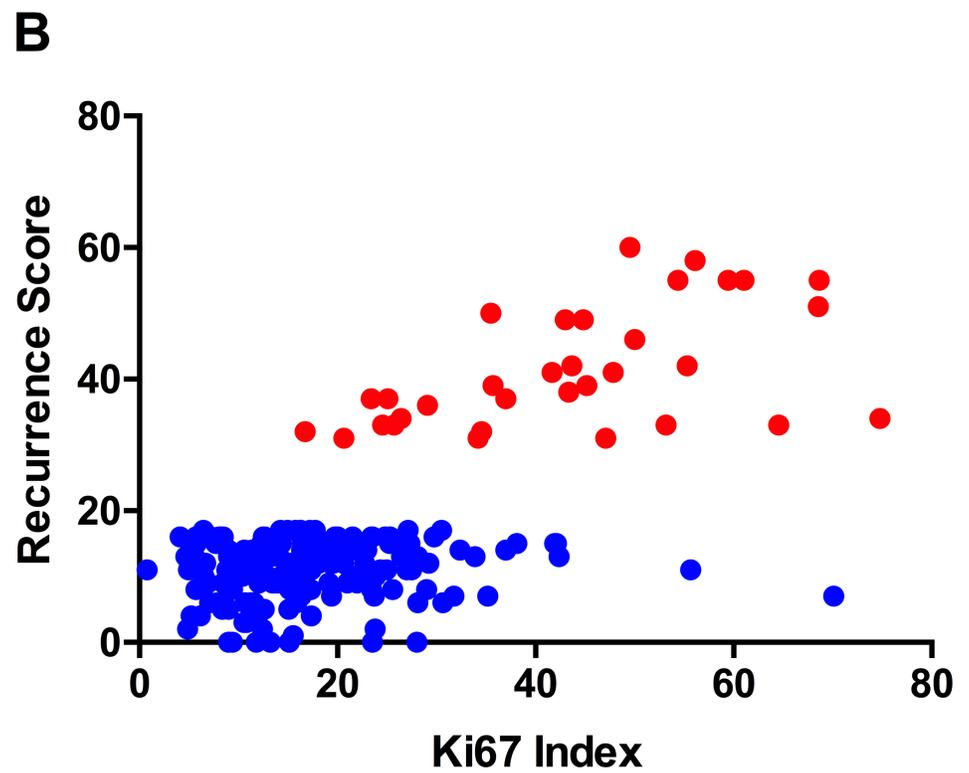
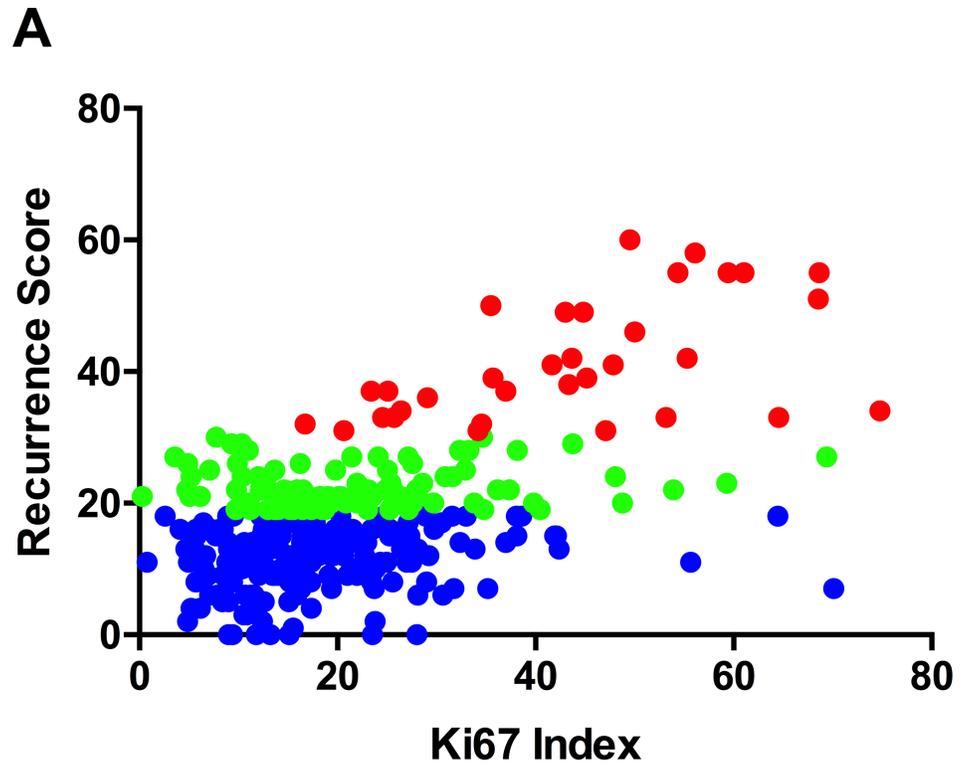
Ki67 is commonly used as a marker for proliferation, and has significant value as a prognostic biomarker in breast cancer [10, 20, 23, 47]. Nevertheless, it has proven challenging to establish standards for the quantification and interpretation of Ki67 in clinical practice. In routine

**Table 2. Assessment of interaction between Ki67 indices and clinico-pathological variables.**

RS	Ki67 Index (Whole-slide)	P Value	Ki67 Index (Hot-spot)	P Value
Low	6.86	Int. vs Low = 0.012	15.84	Int. vs Low = 0.005
Intermediate	8.33	High vs Low < 0.001	19.21	High vs Low < 0.001
High	20.21	High vs Int. < 0.001	43.66	High vs Int. < 0.001
<b>Tumor Grade</b>				
1	8.15	2 vs 1 = 0.297	16.49	2 vs 1 = 0.704
2	6.53	3 vs 1 < 0.001	16.08	3 vs 1 < 0.001
3	7.46	3 vs 2 < 0.001	33.88	3 vs 2 < 0.001
<b>ER Score</b>				
ER<5	16.96	0.004	36.98	0.002
ER = 5	7.05		16.46	
<b>PgR Score</b>				
PgR<5	8.98	0.012	20.63	0.008
PgR = 5	6.66		15.27	
<b>Nuclear Grade</b>				
1	7.11	2 vs 1 = 0.555	17.10	2 vs 1 = 0.232
2	6.85	3 vs 1 = 0.046	16.06	3 vs 1 = 0.129
3	10.5	3 vs 2 < 0.001	23.83	3 vs 2 < 0.001
<b>Mitotic Score</b>				
1	6.63	2 vs 1 < 0.001	15.68	2 vs 1 < 0.001
2	11.95	3 vs 1 < 0.001	25.99	3 vs 1 < 0.001
3	17.54	3 vs 2 = 0.005	35.93	3 vs 2 = 0.008

RS, recurrence score; Int, intermediate-risk group; ER, estrogen receptor; PgR, progesterone receptor.

<https://doi.org/10.1371/journal.pone.0188983.t002>



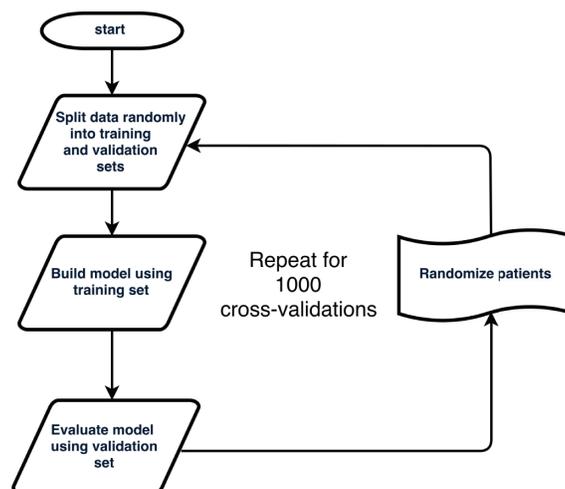
**Fig 3. Assessment of correlation between hot-spot Ki67 index and Oncotype DX scores.** (A) Plot showing association between Ki67 index and Oncotype DX low-risk (blue), intermediate-risk (green) and high-risk (red) groupings. Pearson's  $r = 0.5533$ ;  $P < 0.001$ . (B) Plot showing association between Ki67 and Oncotype DX low- and high-risk groupings. Pearson's  $r = 0.684$ ;  $P < 0.001$ .

<https://doi.org/10.1371/journal.pone.0188983.g003>

diagnostic procedures, Ki67 scoring, as performed manually by a pathologist, involves visual inspection of a limited number of tumor cells [23]. This method suffers from considerable inter-user discordance, creating difficulties in identification and validation of cut-points. In this report, we present a simple, automated scoring method for Ki67 assessment which addresses some of the key challenges currently associated with the analysis of this biomarker. Furthermore, this method combines the advantages of an unbiased whole-slide analysis with the clinical value of identifying the hot spots of highest tumor proliferation. The method features very high concordance against expert manual scoring, and between users. Finally, the Ki67 indices derived from this method correlate as expected with other clinico-pathological variables, and allow accurate inference of Oncotype DX risk-of-recurrence groups.

While several other groups have applied automated methods to the quantification of Ki67 [23, 48, 49], including some large multi-center studies [50, 51], our approach addresses a number of limitations evident in those studies. Many hormone receptor-positive breast cancers exhibit differences in proliferative rate across their spatial extents. In such tumors, the hot spots of higher-than-average proliferative indices are clinically meaningful [52, 53]. Consequently, automated analysis of tissue microarrays, or of manually-selected tumor regions, incur the same potential for selection bias as do manual scoring approaches [20, 54]. We addressed this problem by segmenting each slide image into tiles prior to analysis; since the entire image was subsequently analyzed, the hot spots emerged naturally as the tiles with the highest Ki67 scores. Furthermore, as the Ki67 index represents the mean of the top five tiles, we captured Ki67 data from multiple hot spots across each slide, as well as from non-hot spot regions. The result was a Ki67 index that accounted for proliferative heterogeneity without sacrificing either accuracy or robustness of performance. In contrast, for whole-slide Ki67 index, Ki67 scores from the tiles comprising the entire marked area on the slide were assessed.

The key advantage of computer-assisted analysis of Ki67 is consistency of scoring, particularly between users [23, 25]. The version of the analysis algorithm described here represented



**Fig 4. Outline of Random Forest training and evaluation workflow.**

<https://doi.org/10.1371/journal.pone.0188983.g004>

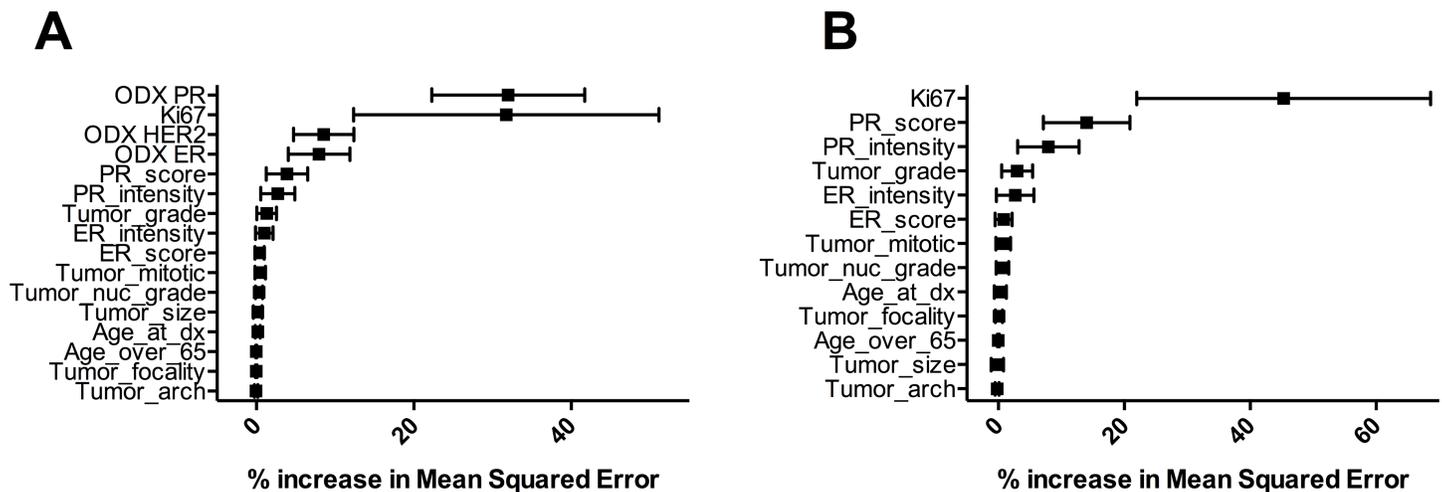
**Table 3. Summary performance of the Random Forest models predicting Oncotype DX risk groups.** Recurrence Scores were predicted using the clinico-pathological variables listed in S2 Table alone (pRS), or using the S2 Table variables in addition to gene expression scores for ER, PgR and HER2 that were included in the official Oncotype DX reports (pRS<sub>odx</sub>). Evaluation of performance of the Random Forest models was based on the extent to which the models correctly predicted, or failed to predict, each patient’s actual low- or high-risk Oncotype DX category. Values represent the mean outcomes ± standard deviations over 1,000 testing iterations.

Parameter	Using hot-spot Ki67 scores		Using whole-slide Ki67 scores	
	pRS <sub>odx</sub> vs. RS (%)	pRS vs. RS (%)	pRS <sub>odx</sub> vs. RS (%)	pRS vs. RS (%)
Accuracy	97.8 ± 1	96.8 ± 2	98.2 ± 2	97.2 ± 2
Sensitivity	99.4 ± 0.9	98.7 ± 1	99.9 ± 0.5	99.5 ± 1
Specificity	81.9 ± 16	80.1 ± 17	80.2 ± 17	75.5 ± 19
Positive predictive value	98.2 ± 1	97.8 ± 1	98.2 ± 2	97.5 ± 2
Negative predictive value	94.1 ± 10	87.8 ± 12	98.8 ± 6	94.4 ± 12

RS, Oncotype DX Recurrence Score.

<https://doi.org/10.1371/journal.pone.0188983.t003>

the end-point of an optimization process performed on a subset of slides, which was used to analyze all 328 patient samples without any extra optimization. While the automated analysis demonstrated high concordance against a pathologist’s manual scoring, the key demonstration of utility was in the inter-user comparison. With less than 15 minutes of educational-instruction, a novice user of the software was successfully trained to perform an analysis of 50 slides, obtaining almost perfect concordance with an expert user. The only manual intervention required, and thus the only sources of inter-user variance, were in transferring the pathologist’s annotations to the digital images, and visual quality assessment of the top-scoring tiles. The automated analysis thus demonstrates a number of features—validity against the gold standard, reproducibility, and ease of use—that could facilitate its implementation in a clinical environment.



**Fig 5. Contribution of individual variables to the accuracy of the respective Random Forest models, as assessed by increases in mean squared error for models created without each variable.** Graphs represent models after 1,000 cycles of validation trained with both clinico-pathological data and Oncotype DX expression data for ER, PgR and HER2 (A) or with clinico-pathological data alone (B). Error bars represent standard deviation from the mean. ER intensity, estrogen receptor staining intensity; ER score, estrogen receptor expression score (immunohistochemistry); PR intensity, progesterone receptor staining intensity; PR score, progesterone receptor expression score (immunohistochemistry); ODX ER, Oncotype DX estrogen receptor gene expression score; ODX HER2, Oncotype DX HER2 expression score; ODX PR, Oncotype DX progesterone receptor gene expression score; Tumor\_arch, tumor differentiation score; Tumor\_nuc\_grade, tumor nuclear grade.

<https://doi.org/10.1371/journal.pone.0188983.g005>

While other studies have demonstrated the prognostic value of Ki67 as assessed by whole slide- and hot spot-focused analysis methods, the value of the hybrid analysis approach we applied in this study was not immediately apparent. We consequently used multiple methods to evaluate the validity of our Ki67 data. Previous studies highlighted associations between tumor proliferation and a variety of clinico-pathological endpoints, such as tumor grade, stage, and ER/PgR expression [55–57]. We observed in this study that Ki67 indices, derived from either whole-slide or hot-spot analysis approaches, correlated significantly with all of these. It is thus likely that, at least for the purposes highlighted in this study, whole-slide and hot-spot Ki67 analyses may be equally useful, so long as they are accurately determined. Ki67 is the greatest contributor to Oncotype DX recurrence score [31], and a strong correlation between Ki67 index and recurrence score could provide evidence supporting the clinical validity of our approach [37, 48, 58]. We observed a particularly strong correlation between Ki67 index and the high and low Oncotype DX risk-of-recurrence categories. Finally, we applied a multivariate machine learning approach to determine if Ki67 indices could contribute to prediction of these Oncotype DX categories. Remarkably, a model trained solely on clinico-pathological data, biomarker expression and Ki67 indices in both hot spot and whole-slide approaches predicted high- and low-risk of recurrence groups with 97% accuracy. In this context, Ki67 index was the most significant contributor to the accuracy of the model. Our results are similar to those published recently by Kim and colleagues [41], who used an identical machine learning approach to predict Oncotype DX risk-of-recurrence status from clinico-pathological and biomarker expression data. Given that Oncotype DX prediction methods, including the linear regression Magee equations [36, 40, 59], utilize potentially unreliable data from manual Ki67 scoring, it is likely that the accuracy of such methods would be significantly improved with robust Ki67 data from automated analyses.

Our study has a number of shortcomings that need to be addressed in the future. First, this study was conducted in a single academic center; consequently, it is not clear what logistical problems may occur in implementing this method in a multi-institutional setup. Secondly, the manual scores in this study were generated by a single pathologist so the exact measure of the inter-observer variability in calculating Ki67 index manually cannot be obtained and compared with the automatically generated Ki67 indices. Also, the manual scores were generated using a single tissue section from each patient with the largest tumor cross sectional area. It is possible that the single section selected may not represent the true mitotic nature of the tumor, due to intra-tumoral heterogeneity. The methodology as presented involves the use of expensive, proprietary software (HALO<sup>®</sup>). However, the fundamentals should be reproducible on a sufficiently robust open-source framework, such as CognitionMaster [41, 60]. This needs to be tested and could determine the extent of variability in results amongst different image analysis software. Furthermore, development and assessment of the machine learning model was only possible on a subset of patients (199) for whom we had complete data sets, since the modeling paradigm is not tolerant of missing data. In this, as in similar studies [36, 40, 59], there were few patients in the high risk-of-recurrence group; further testing of the model needs to be carried out in a larger patient population, with a more equitable distribution of patients in all three risk categories. The small cohort size also precluded our ability to evaluate a locked-down Random Forest model on a separate, independent group of patients. Such an approach would constitute a more thorough validation of the concepts introduced in the current proof-of-principle work.

In the present study, we have addressed the issue of Ki67 scoring which to date remains a major hurdle in clinical use of Ki67 as a biomarker. We present an automated, robust and reproducible method for quantification of Ki67 from whole-slide sections of breast cancer. The consistency and ease-of-use of this Ki67 analysis approach may facilitate the standardized

scoring of this biomarker across multiple clinical centers. We show that the Ki67 index derived with this method contributes significantly to prediction of Oncotype DX risk-of-recurrence status, when implemented in a multivariate machine learning model. The utility of predictive modeling, bolstered by the availability of accurate Ki67 data, may reduce the need for expensive multigene assays to assess risk of recurrence.

## Supporting information

**S1 Fig. Comparison of Ki67 indices from whole-slide and hot-spot analyses.**  
(DOCX)

**S2 Fig. Comparison of mean Ki67 indices from whole-slide and hot-spot analyses, across different Oncotype DX risk-of-recurrence groups.** Error bars represent standard error of mean; \* =  $P < 0.05$ .  
(DOCX)

**S1 Table. Clinico-pathological characteristics of patients, grouped by Oncotype DX risk-of-recurrence group.** P values were generated by Kruskal-Wallis, Wilcoxon or Fisher's exact tests as appropriate.  
(DOCX)

**S2 Table. Variables used as model inputs in the Random Forest analyses.** Oncotype DX Recurrence Score was the predicted variable in all cases.  
(DOCX)

## Author Contributions

**Conceptualization:** Satbir Singh Thakur, Gilbert Bigras, Emeka K. Enwere, Hua Yang.

**Data curation:** Roxana Tudor, Hua Yang.

**Formal analysis:** Satbir Singh Thakur, Haocheng Li, Angela M. Y. Chan, Emeka K. Enwere.

**Funding acquisition:** Satbir Singh Thakur, Don Morris, Hua Yang.

**Investigation:** Emeka K. Enwere, Hua Yang.

**Methodology:** Satbir Singh Thakur, Emeka K. Enwere, Hua Yang.

**Project administration:** Don Morris, Emeka K. Enwere, Hua Yang.

**Resources:** Gilbert Bigras, Emeka K. Enwere.

**Software:** Emeka K. Enwere.

**Supervision:** Don Morris, Emeka K. Enwere, Hua Yang.

**Validation:** Angela M. Y. Chan.

**Writing – original draft:** Satbir Singh Thakur, Emeka K. Enwere.

**Writing – review & editing:** Satbir Singh Thakur, Emeka K. Enwere.

## References

1. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer; [Internet]. 2013. Available from: <http://globocan.iarc.fr>.
2. Harris LN, Ismaila N, McShane LM, Andre F, Collyar DE, Gonzalez-Angulo AM, et al. Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast

- Cancer: American Society of Clinical Oncology Clinical Practice Guideline. *J Clin Oncol*. 2016; 34(10):1134–50. <https://doi.org/10.1200/JCO.2015.65.2289> PMID: 26858339
3. Kwa M, Makris A, Esteva FJ. Clinical utility of gene-expression signatures in early stage breast cancer. *Nat Rev Clin Oncol*. 2017; 14(10):595–610. <https://doi.org/10.1038/nrclinonc.2017.74> PMID: 28561071
  4. Geyer FC, Rodrigues DN, Weigelt B, Reis-Filho JS. Molecular classification of estrogen receptor-positive/luminal breast cancers. *Adv Anat Pathol*. 2012; 19(1):39–53. <https://doi.org/10.1097/PAP.0b013e31823fafa0> PMID: 22156833
  5. Sinn HP, Schneeweiss A, Keller M, Schlombs K, Laible M, Seitz J, et al. Comparison of immunohistochemistry with PCR for assessment of ER, PR, and Ki-67 and prediction of pathological complete response in breast cancer. *BMC Cancer*. 2017; 17(1):124. <https://doi.org/10.1186/s12885-017-3111-1> PMID: 28193205
  6. Kurozumi S, Inoue K, Takei H, Matsumoto H, Kurozumi M, Horiguchi J, et al. ER, PgR, Ki67, p27(Kip1), and histological grade as predictors of pathological complete response in patients with HER2-positive breast cancer receiving neoadjuvant chemotherapy using taxanes followed by fluorouracil, epirubicin, and cyclophosphamide concomitant with trastuzumab. *BMC Cancer*. 2015; 15:622. <https://doi.org/10.1186/s12885-015-1641-y> PMID: 26345461
  7. Nishimura R, Osako T, Okumura Y, Hayashi M, Arima N. Clinical significance of Ki-67 in neoadjuvant chemotherapy for primary breast cancer as a predictor for chemosensitivity and for prognosis. *Breast Cancer*. 2010; 17(4):269–75. <https://doi.org/10.1007/s12282-009-0161-5> PMID: 19730975
  8. Yoshioka T, Hosoda M, Yamamoto M, Taguchi K, Hatanaka KC, Takakuwa E, et al. Prognostic significance of pathologic complete response and Ki67 expression after neoadjuvant chemotherapy in breast cancer. *Breast Cancer*. 2015; 22(2):185–91. <https://doi.org/10.1007/s12282-013-0474-2> PMID: 23645542
  9. Denkert C, Budczies J, von Minckwitz G, Wienert S, Loibl S, Klauschen F. Strategies for developing Ki67 as a useful biomarker in breast cancer. *Breast*. 2015; 24 Suppl 2:S67–72.
  10. Sanchez-Munoz A, Navarro-Perez V, Plata-Fernandez Y, Santonja A, Moreno I, Ribelles N, et al. Proliferation Determined by Ki-67 Defines Different Pathologic Response to Neoadjuvant Trastuzumab-Based Chemotherapy in HER2-Positive Breast Cancer. *Clin Breast Cancer*. 2015; 15(5):343–7. <https://doi.org/10.1016/j.clbc.2015.01.005> PMID: 25752727
  11. Fasching PA, Heusinger K, Haerberle L, Niklos M, Hein A, Bayer CM, et al. Ki67, chemotherapy response, and prognosis in breast cancer patients receiving neoadjuvant treatment. *BMC Cancer*. 2011; 11:486. <https://doi.org/10.1186/1471-2407-11-486> PMID: 22081974
  12. Kim KI, Lee KH, Kim TR, Chun YS, Lee TH, Park HK. Ki-67 as a predictor of response to neoadjuvant chemotherapy in breast cancer patients. *J Breast Cancer*. 2014; 17(1):40–6. <https://doi.org/10.4048/jbc.2014.17.1.40> PMID: 24744796
  13. Brown JR, DiGiovanna MP, Killelea B, Lannin DR, Rimm DL. Quantitative assessment Ki-67 score for prediction of response to neoadjuvant chemotherapy in breast cancer. *Lab Invest*. 2014; 94(1):98–106. <https://doi.org/10.1038/labinvest.2013.128> PMID: 24189270
  14. Sheri A, Dowsett M. Developments in Ki67 and other biomarkers for treatment decision making in breast cancer. *Ann Oncol*. 2012; 23 Suppl 10:x219–27.
  15. Pathmanathan N, Balleine RL. Ki67 and proliferation in breast cancer. *J Clin Pathol*. 2013; 66(6):512–6. <https://doi.org/10.1136/jclinpath-2012-201085> PMID: 23436927
  16. Joensuu K, Leidenius M, Kero M, Andersson LC, Horwitz KB, Heikkila P. ER, PR, HER2, Ki-67 and CK5 in Early and Late Relapsing Breast Cancer-Reduced CK5 Expression in Metastases. *Breast Cancer (Auckl)*. 2013; 7:23–34.
  17. Vincent-Salomon A, Hajage D, Rouquette A, Cedenot A, Gruel N, Alran S, et al. High Ki67 expression is a risk marker of invasive relapse for classical lobular carcinoma in situ patients. *Breast*. 2012; 21(3):380–3. <https://doi.org/10.1016/j.breast.2012.03.005> PMID: 22531230
  18. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thurlimann B, Senn HJ, et al. Strategies for subtypes—dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann Oncol*. 2011; 22(8):1736–47. <https://doi.org/10.1093/annonc/mdr304> PMID: 21709140
  19. Harbeck N, Thomssen C, Gnant M. St. Gallen 2013: brief preliminary summary of the consensus discussion. *Breast Care (Basel)*. 2013; 8(2):102–9.
  20. Dowsett M, Nielsen TO, A'Hern R, Bartlett J, Coombes RC, Cuzick J, et al. Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group. *J Natl Cancer Inst*. 2011; 103(22):1656–64. <https://doi.org/10.1093/jnci/djr393> PMID: 21960707

21. Jang MH, Kim HJ, Chung YR, Lee Y, Park SY. A comparison of Ki-67 counting methods in luminal Breast Cancer: The Average Method vs. the Hot Spot Method. *PLoS One*. 2017; 12(2):e0172031. <https://doi.org/10.1371/journal.pone.0172031> PMID: 28187177
22. Shui R, Yu B, Bi R, Yang F, Yang W. An interobserver reproducibility analysis of Ki67 visual assessment in breast cancer. *PLoS One*. 2015; 10(5):e0125131. <https://doi.org/10.1371/journal.pone.0125131> PMID: 25932921
23. Klauschen F, Wienert S, Schmitt WD, Loibl S, Gerber B, Blohmer JU, et al. Standardized Ki67 Diagnostics Using Automated Scoring—Clinical Validation in the GeparTrio Breast Cancer Study. *Clin Cancer Res*. 2015; 21(16):3651–7. <https://doi.org/10.1158/1078-0432.CCR-14-1283> PMID: 25501130
24. Polley MY, Leung SC, McShane LM, Gao D, Hugh JC, Mastropasqua MG, et al. An international Ki67 reproducibility study. *J Natl Cancer Inst*. 2013; 105(24):1897–906. <https://doi.org/10.1093/jnci/djt306> PMID: 24203987
25. Varga Z, Diebold J, Dommann-Scherrer C, Frick H, Kaup D, Noske A, et al. How reliable is Ki-67 immunohistochemistry in grade 2 breast carcinomas? A QA study of the Swiss Working Group of Breast- and Gynecopathologists. *PLoS One*. 2012; 7(5):e37379. <https://doi.org/10.1371/journal.pone.0037379> PMID: 22662150
26. Anampa J, Makower D, Sparano JA. Progress in adjuvant chemotherapy for breast cancer: an overview. *BMC Med*. 2015; 13:195. <https://doi.org/10.1186/s12916-015-0439-8> PMID: 26278220
27. Sueta A, Yamamoto Y, Hayashi M, Yamamoto S, Inao T, Ibusuki M, et al. Clinical significance of pretherapeutic Ki67 as a predictive parameter for response to neoadjuvant chemotherapy in breast cancer: is it equally useful across tumor subtypes? *Surgery*. 2014; 155(5):927–35. <https://doi.org/10.1016/j.surg.2014.01.009> PMID: 24582496
28. Darb-Esfahani S, Loibl S, Muller BM, Roller M, Denkert C, Komor M, et al. Identification of biology-based breast cancer types with distinct predictive and prognostic features: role of steroid hormone and HER2 receptor expression in patients treated with neoadjuvant anthracycline/taxane-based chemotherapy. *Breast Cancer Res*. 2009; 11(5):R69. <https://doi.org/10.1186/bcr2363> PMID: 19758440
29. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med*. 2004; 351(27):2817–26. <https://doi.org/10.1056/NEJMoa041588> PMID: 15591335
30. Braunstein LZ, Taghian AG. Molecular Phenotype, Multigene Assays, and the Locoregional Management of Breast Cancer. *Semin Radiat Oncol*. 2016; 26(1):9–16. <https://doi.org/10.1016/j.semradonc.2015.08.002> PMID: 26617205
31. Gyorffy B, Hatzis C, Sanft T, Hofstatter E, Aktas B, Pusztai L. Multigene prognostic tests in breast cancer: past, present, future. *Breast Cancer Res*. 2015; 17:11. <https://doi.org/10.1186/s13058-015-0514-2> PMID: 25848861
32. Sinn P, Aulmann S, Wirtz R, Schott S, Marme F, Varga Z, et al. Multigene Assays for Classification, Prognosis, and Prediction in Breast Cancer: a Critical Review on the Background and Clinical Utility. *Geburtshilfe Frauenheilkd*. 2013; 73(9):932–40. <https://doi.org/10.1055/s-0033-1350831> PMID: 24771945
33. Albain KS, Barlow WE, Shak S, Hortobagyi GN, Livingston RB, Yeh IT, et al. Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial. *Lancet Oncol*. 2010; 11(1):55–65. [https://doi.org/10.1016/S1470-2045\(09\)70314-6](https://doi.org/10.1016/S1470-2045(09)70314-6) PMID: 20005174
34. Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol*. 2006; 24(23):3726–34. <https://doi.org/10.1200/JCO.2005.04.7985> PMID: 16720680
35. Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF, et al. Prospective Validation of a 21-Gene Expression Assay in Breast Cancer. *N Engl J Med*. 2015; 373(21):2005–14. <https://doi.org/10.1056/NEJMoa1510764> PMID: 26412349
36. Klein ME, Dabbs DJ, Shuai Y, Brufsky AM, Jankowitz R, Puhalla SL, et al. Prediction of the Oncotype DX recurrence score: use of pathology-generated equations derived by linear regression analysis. *Mod Pathol*. 2013; 26(5):658–64. <https://doi.org/10.1038/modpathol.2013.36> PMID: 23503643
37. Sahebjam S, Aloyz R, Pilavdzic D, Brisson ML, Ferrario C, Bouganim N, et al. Ki 67 is a major, but not the sole determinant of Oncotype Dx recurrence score. *Br J Cancer*. 2011; 105(9):1342–5. <https://doi.org/10.1038/bjc.2011.402> PMID: 21970880
38. Cuzick J, Dowsett M, Pineda S, Wale C, Salter J, Quinn E, et al. Prognostic value of a combined estrogen receptor, progesterone receptor, Ki-67, and human epidermal growth factor receptor 2 immunohistochemical score and comparison with the Genomic Health recurrence score in early breast cancer. *J Clin Oncol*. 2011; 29(32):4273–8. <https://doi.org/10.1200/JCO.2010.31.2835> PMID: 21990413

39. Harowicz MR, Robinson TJ, Dinan MA, Saha A, Marks JR, Marcom PK, et al. Algorithms for prediction of the Oncotype DX recurrence score using clinicopathologic data: a review and comparison using an independent dataset. *Breast Cancer Res Treat.* 2017; 162(1):1–10. <https://doi.org/10.1007/s10549-016-4093-4> PMID: 28064383
40. Turner BM, Skinner KA, Tang P, Jackson MC, Soukiazian N, Shayne M, et al. Use of modified Magee equations and histologic criteria to predict the Oncotype DX recurrence score. *Mod Pathol.* 2015; 28(7):921–31. <https://doi.org/10.1038/modpathol.2015.50> PMID: 25932962
41. Kim HS, Umbricht CB, Illei PB, Cimino-Mathews A, Cho S, Chowdhury N, et al. Optimizing the Use of Gene Expression Profiling in Early-Stage Breast Cancer. *J Clin Oncol.* 2016; 34(36):4390–7. <https://doi.org/10.1200/JCO.2016.67.7195> PMID: 27998227
42. Khoury T, Yan L, Liu S, Bshara W. Oncotype DX RT-qPCR assay for ER and PR correlation with IHC: a study of 3 different clones. *Appl Immunohistochem Mol Morphol.* 2015; 23(3):178–87. <https://doi.org/10.1097/PAI.000000000000078> PMID: 24992175
43. Huang BF, Boutros PC. The parameter sensitivity of random forests. *BMC Bioinformatics.* 2016; 17(1):331. <https://doi.org/10.1186/s12859-016-1228-x> PMID: 27586051
44. Chen X, Ishwaran H. Random forests for genomic data analysis. *Genomics.* 2012; 99(6):323–9. <https://doi.org/10.1016/j.ygeno.2012.04.003> PMID: 22546560
45. Arima N, Nishimura R, Osako T, Nishiyama Y, Fujisue M, Okumura Y, et al. A Comparison of the Hot Spot and the Average Cancer Cell Counting Methods and the Optimal Cutoff Point of the Ki-67 Index for Luminal Type Breast Cancer. *Oncology.* 2016; 90(1):43–50. <https://doi.org/10.1159/000441525> PMID: 26613521
46. Varga Z, Cassoly E, Li Q, Oehlschlegel C, Tapia C, Lehr HA, et al. Standardization for Ki-67 assessment in moderately differentiated breast cancer. A retrospective analysis of the SAKK 28/12 study. *PLoS One.* 2015; 10(4):e0123435. <https://doi.org/10.1371/journal.pone.0123435> PMID: 25885288
47. Ohno S, Chow LW, Sato N, Masuda N, Sasano H, Takahashi F, et al. Randomized trial of preoperative docetaxel with or without capecitabine after 4 cycles of 5-fluorouracil- epirubicin-cyclophosphamide (FEC) in early-stage breast cancer: exploratory analyses identify Ki67 as a predictive biomarker for response to neoadjuvant chemotherapy. *Breast Cancer Res Treat.* 2013; 142(1):69–80. <https://doi.org/10.1007/s10549-013-2691-y> PMID: 24122389
48. Abubakar M, Howat WJ, Daley F, Zabaglo L, McDuffus LA, Blows F, et al. High-throughput automated scoring of Ki67 in breast cancer tissue microarrays from the Breast Cancer Association Consortium. *J Pathol Clin Res.* 2016; 2(3):138–53. <https://doi.org/10.1002/cjp2.42> PMID: 27499923
49. Mohammed ZM, McMillan DC, Elsberger B, Going JJ, Orange C, Mallon E, et al. Comparison of visual and automated assessment of Ki-67 proliferative activity and their impact on outcome in primary operable invasive ductal breast cancer. *Br J Cancer.* 2012; 106(2):383–8. <https://doi.org/10.1038/bjc.2011.569> PMID: 22251968
50. Howat WJ, Blows FM, Provenzano E, Brook MN, Morris L, Gazinska P, et al. Performance of automated scoring of ER, PR, HER2, CK5/6 and EGFR in breast cancer tissue microarrays in the Breast Cancer Association Consortium. *J Pathol Clin Res.* 2015; 1(1):18–32. <https://doi.org/10.1002/cjp2.3> PMID: 27499890
51. Abubakar M, Orr N, Daley F, Coulson P, Ali HR, Blows F, et al. Prognostic value of automated Ki67 scoring in breast cancer: a centralised evaluation of 8088 patients from 10 study groups. *Breast Cancer Res.* 2016; 18(1):104. <https://doi.org/10.1186/s13058-016-0765-6> PMID: 27756439
52. Honma N, Horii R, Iwase T, Saji S, Younes M, Ito Y, et al. Ki-67 evaluation at the hottest spot predicts clinical outcome of patients with hormone receptor-positive/HER2-negative breast cancer treated with adjuvant tamoxifen monotherapy. *Breast Cancer.* 2015; 22(1):71–8. <https://doi.org/10.1007/s12282-013-0455-5> PMID: 23479208
53. Tashima R, Nishimura R, Osako T, Nishiyama Y, Okumura Y, Nakano M, et al. Evaluation of an Optimal Cut-Off Point for the Ki-67 Index as a Prognostic Factor in Primary Breast Cancer: A Retrospective Study. *PLoS One.* 2015; 10(7):e0119565. <https://doi.org/10.1371/journal.pone.0119565> PMID: 26177501
54. Besusparis J, Plancoulaine B, Rasmussen A, Augulis R, Green AR, Ellis IO, et al. Impact of tissue sampling on accuracy of Ki67 immunohistochemistry evaluation in breast cancer. *Diagn Pathol.* 2016; 11(1):82. <https://doi.org/10.1186/s13000-016-0525-z> PMID: 27576949
55. Inwald EC, Klinkhammer-Schalke M, Hofstadter F, Zeman F, Koller M, Gerstenhauer M, et al. Ki-67 is a prognostic parameter in breast cancer patients: results of a large population-based cohort of a cancer registry. *Breast Cancer Res Treat.* 2013; 139(2):539–52. <https://doi.org/10.1007/s10549-013-2560-8> PMID: 23674192
56. Cass JD, Varma S, Day AG, Sangrar W, Rajput AB, Raptis LH, et al. Automated Quantitative Analysis of p53, Cyclin D1, Ki67 and pERK Expression in Breast Carcinoma Does Not Differ from Expert

- Pathologist Scoring and Correlates with Clinico-Pathological Characteristics. *Cancers (Basel)*. 2012; 4(3):725–42.
57. Sun J, Chen C, Wei W, Zheng H, Yuan J, Tu YI, et al. Associations and indications of Ki67 expression with clinicopathological parameters and molecular subtypes in invasive breast cancer: A population-based study. *Oncol Lett*. 2015; 10(3):1741–8. <https://doi.org/10.3892/ol.2015.3461> PMID: 26622743
  58. Baxter E, Gondara L, Lohrisch C, Chia S, Gelmon K, Hayes M, et al. Using proliferative markers and Oncotype DX in therapeutic decision-making for breast cancer: the B.C. experience. *Curr Oncol*. 2015; 22(3):192–8. <https://doi.org/10.3747/co.22.2284> PMID: 26089718
  59. Flanagan MB, Dabbs DJ, Brufsky AM, Beriwal S, Bhargava R. Histopathologic variables predict Oncotype DX recurrence score. *Mod Pathol*. 2008; 21(10):1255–61. <https://doi.org/10.1038/modpathol.2008.54> PMID: 18360352
  60. Wienert S, Heim D, Kotani M, Lindequist B, Stenzinger A, Ishii M, et al. CognitionMaster: an object-based image analysis framework. *Diagn Pathol*. 2013; 8:34. <https://doi.org/10.1186/1746-1596-8-34> PMID: 23445542