

Prognostic Significance of Nucleolar Assessment in Invasive Breast Cancer

Khlood A. Elsharawy^{1,2}, ¹ Michael S. Toss¹, Sara Raaft¹, Graham Ball³, Andrew R. Green¹, Mohammed A. Aleskandarany¹, Leslie W. Dalton⁴ and Emad A. Rakha¹

¹ Academic Pathology, Division of Cancer and Stem Cells, School of Medicine, University of Nottingham, Nottingham, UK

² Department of Zoology, Faculty of Science, Damietta University, Damietta, Egypt

³ School of Science and Technology, Nottingham Trent University, Clifton Campus, Nottingham, UK

⁴ Department of Histopathology, South Austin Hospital, Austin, TX USA

Correspondence:

Professor Emad Rakha

Department of Histopathology,
Nottingham University Hospital NHS Trust,
City Hospital Campus, Hucknall Road, Nottingham,
NG5 1PB, UK

Tel: (44) 0115-9691169

Fax: (44) 0115- 9627768

Email: Emad.rakha@nuh.nhs.uk
emadrakha@yahoo.com

Running Title: Nucleolar assessment in breast cancer

Abstract

Aims: Nucleolar morphometric features have a potential role in the assessment of aggressiveness of many cancers. However, the role of nucleoli in invasive breast cancer (IBC) is still unclear. This study aimed to investigate the optimal scoring method of nucleoli in IBC and their prognostic significance, and refine the grading of BC by incorporating the nucleolar score.

Methods and results: Digital images acquired from hematoxylin and eosin (H&E) stained sections from a large IBC cohort were divided into training (n=400) and validation (n=1200) sets were used in this study. Four different assessment methods including 1) modified Helpap's method, and counting prominent nucleoli (size $\geq 2.5\mu\text{m}$) in 2) 10 field views (10 FVs), 3) 5 FVs and 4) 1 FV were evaluated in the training set to identify the optimal method associated with the best performance and significant prognostic value. The optimal method was applied to the validation set and to an external validation set the Cancer Genome Atlas (TCGA) data (n=743). Scoring prominent nucleoli in 5 FVs, showed the highest inter-observer concordance rate (intraclass correlation coefficient = 0.8) and significant association with breast cancer specific survival (BCSS) ($p < 0.0001$). High nucleolar score was associated with younger age, larger tumor size and higher grade. Incorporating of nucleolar score in the Nottingham grade system showed higher significant association with survival than the conventional grade.

Conclusions: Quantification of nucleolar prominence in 5 FVs is a cost-efficient and reproducible morphological feature that can predict IBC behaviour and can provide an alternative to pleomorphism to improve BC grading performance.

Keywords: Nucleoli, invasive breast cancer, methods, prognosis

INTRODUCTION

Histological grade of invasive breast cancer (IBC), using the Nottingham grading system, is one of the strongest prognostic indices to guide management ¹⁻³. Nuclear pleomorphism, together with tubule formation and mitotic count, are used to determine histological grade using light microscopy. However, the subjective nature of the grading system results in subsequent variation among pathologists and hence variation in patients' management decisions ⁴. Amongst the three parameters of the Nottingham grading system, assessment of nuclear pleomorphism is considered the least reproducible component due to its subjective ill-defined scoring criteria, thus it shows the lowest level of inter-observer agreement ⁵. The agreement of assessing nuclear pleomorphism in IBC among expert breast pathologists ranges from 0.35 ⁶ to 0.59 ⁷ by Kappa statistics.

Morphological changes in the nucleoli are considered as one of the first histopathological characteristics of malignant transformation, along with abnormal mitotic figures, thickened and irregular nuclear membrane and coarse chromatin ⁸. Malignant cells represent a variety of changes in nucleolar structure including nucleolar composition, size, number and chromatin texture. The nucleolar size and number correlate with the increase in the grade of malignancy ⁹. MacCarty *et al* described that the nucleolus could show much larger size in relative proportion to the size of the nucleus in malignant cells regardless of the type or origin of the neoplasm ¹⁰. Nevertheless, nucleolar size plays a central proportional role in the rate of cell proliferation, and its morphology is closely related to cancer growth ¹¹⁻¹³. The major role of the nucleolus is synthesis of rRNA, processing and assembly of ribosomal subunits ¹⁴⁻¹⁶. Increased number and size of nucleoli indicates higher rate of ribosome biogenesis representing a major metabolic requisite for cell growth and proliferation ^{17, 18}. These changes in the nucleolus is linked to growth factors and oncogene proteins which induce cell proliferation such as epidermal growth factor ¹⁹, insulin-like growth factor ²⁰ and c-Myc ²¹.

Prominent nucleoli were shown to be associated with poor outcome in various tumors ⁸, including BC ²². Quantitative nucleolar morphometry predicts metastasis and biochemical recurrence in prostate cancer ²³. Furthermore, nucleolar size assessed by silver staining of argyrophilic nucleolar organiser regions (AgNOR) was reported to be associated with poor clinical outcome of BC in several studies ^{8, 24, 25}.

However, nucleolar morphometric features are not widely assessed in IBC, due to lack of reliable assays that can be routinely implemented in a clinical setting. With the progressive deployment of digital platforms in reporting pathology laboratories, the assessment of subcellular details including nucleoli might become a feasible practice to improve BC grading performance and to capture useful prognostic value using artificial intelligence techniques or visual eyeballing method. Therefore, in this study we sought to investigate the optimal method of nucleoli evaluation using digitised images, in terms of

reliability, reproducibility and prognostic significance, in large cohorts of BC. Also, the added value of nucleolar score to refine the performance of Nottingham grading system was assessed.

MATERIALS AND METHODS

Study Cohorts

Nottingham Cohort

This study was conducted on a large series (n=1600) of primary operable BC presented to Nottingham City Hospital, Nottingham, United Kingdom between 1999-2006. This is a well-characterised cohort with long term clinical follow-up (median 138 months, range 0-216 months) and detailed clinico-pathological data including patient's age at diagnosis, primary tumor histological grade, tumor size and histological tumor type, Nottingham Prognostic Index (NPI), molecular subtypes and outcome data. The latter includes breast cancer specific survival (BCSS), defined as time (in months) from the date of the primary surgery to the time of death from BC and distant metastasis free survival (DMFS), and defined as the time (in months) from the primary surgery until the first event of distant metastasis.

For the purpose of this study, this cohort was divided into two-groups: the first sequential 400 cases (25%) of the cohort were used as a training set and the subsequent 1200 cases were used as a validation set. Supplementary table 1 summarizes the Clinicopathological parameters of both sets.

Freshly cut, 4 μm thick, full face sections for the Nottingham cohort were stained with Haematoxylin and Eosin (H&E), one section per case from the tumor block that showed the largest tumor burden defined by $> 50\%$ of the section area. H&E slides were scanned into high definition digital images through a high-resolution (0.19 $\mu\text{m}/\text{pixel}$) scanning using high throughput scanner (Pannoramic 250 Flash III, 3D-Histech, Budapest, Hungary), followed by viewing the slides using Case Viewer Software program (version 2.2.0.85, 3D-Histech, Budapest, Hungary) on full screen panel (size = 21inch, resolution = 1366 \times 768). These tumors were graded using the digital images as previously described ²⁶.

External Validation Cohort

In addition, digital H&E IBC images from The Cancer Genome Atlas (TCGA) ²⁷ (n=743), have been used as further external validation set. These images were digitized at 40X objective magnification contains upward of 10^{10} pixels and has long term clinical follow-up (10 years) .All the images were directly downloaded from cBioPortal website and viewed on case viewer for nucleolar score ²⁸.

Scoring of Nucleoli

In this study, nucleoli were assessed using visual scoring of the digitised whole slide images according to the following four scoring methods:

(1) Modified Helpap's method ²⁹, that was based on results related to prior work ³⁰. This method stratified nucleoli into three scores based on their prominence. Nucleoli were assigned score 1 if no prominent nucleoli (i.e. inconspicuous), or single less prominent nucleoli that are difficult to see at 20x were observed. A nucleolar score 3 was assigned if prominent nucleoli, which were easily seen at 10x and were identified in at least 20% of the tumor ³¹ or dysmorphic/multiple nucleoli were present. Score 2 was assigned to those tumor with nucleoli not scored 1 or 3.

For objectivity enhancement, nucleoli were also scored at 40x, and were stratified using the number of prominent nucleoli (defined as $\geq 2.5\mu\text{m}$ in size or easily seen at 10x) per defined number of field views (FVs) on the screen;

(2) Counting in 10 FVs (equivalent to 1mm² area)

(3) 5 FVs (0.5mm²) and

(4) 1 FV (0.1mm²).

Counting was carried out within the nucleoli hotspot defined as area of conspicuous nucleoli observed by scanning the images at 10x. Examples of nucleolar score in 1FV are shown in Figure 1. The cases were scored based on the absolute number of prominent nucleoli and then cutoffs were applied to categorise the cases into 3 scores; 1, 2 and 3.

To check the reliability and reproducibility of the nucleolar evaluation methods, the training set was scored using all aforementioned methods by four observers, where the first observer scored all the cases, while the others scored 25% of the training set. The optimal method for nucleolar score was determined based on the level of reproducibility in terms of inter-observer concordance and the association with outcome.

The first validation set (n=1200 cases) was scored by two observers using the chosen method, where the first observer scored all the cases, while the other scored 50% of the cases to further assess the level of concordance of this method. Nucleolar score carried out by the first observer (KA) was considered in the final statistical analysis. For external validation of the optimal method chosen, digital images of scanned H&E full-face slides of the TCGA were scored by one observer.

In previous study we graded the Nottingham cohort twice using whole slide images by one observer ³², and data on the concordance of grade and pleomorphism was used in this study. Moreover, incorporation of the nucleolar score in the Nottingham Grading System was attempted to assess the performance of the Nottingham grade with the addition of this score to the three components and or replacement of nuclear pleomorphism.

Statistical analysis

The optimal cut-off point of nucleoli count against BCSS was defined using X-tile bioinformatics software version 3.6.1 (School of Medicine, Yale University, New Haven, USA)³³. Nucleoli were given three scores depending on these cut-off points (Supplementary Table 2). Also, new grade scores of Nottingham grading system were

obtained using cut-off points of total scores against BCSS using X-tile. IBM-SPSS statistical software 24.0 (SPSS, Chicago, IL, USA) was used in statistical analysis. The degree of inter-observer agreement was assessed through intra-class correlation coefficient (ICC) for continuous data. Fleiss' Kappa statistic was used to assess the concordance between more than two observers for categorical variables. Association between nucleoli count with different concordant and discordant cases was analyzed using Kruskal-Wallis test. Outcome analysis was assessed using Kaplan-Meier curves and the log-rank test. Cox proportional hazards multivariate regression modelling was used for the multivariate analysis. For all tests, p- values < 0.05 (two-tailed) were considered as statistically significant.

RESULTS

This study included two cohorts of IBC: 1) the Nottingham cohort (n=1600 cases) split into: a) a training set comprising 400 cases scored by four observers using four different scoring methods and b) a validation set (n=1200), which was scored by counting nucleoli in 5 FVs by two observers. 2) an external validation cohort; TCGA cohort (n=743) to assess the performance of the optimal scoring method.

A) *The training set:*

The highest degree of inter-observer agreement was observed in counting nucleoli in 5 FVs (ICC = 0.782), whereas the least concordance was seen with the modified Helpap's method (Fleiss' Kappa value = 0.417). The concordance rate of nucleolar score using different methods between the four observers is summarized in Supplementary Table 3.

The percentage of cases scored 3 using Modified Helpap's method, and at 10, 5 and 1 FV were 17%, 16%, 17%, and 18%, respectively. Their association with BCSS, DMFS are summarized in Table 1.

Higher nucleolar score showed significant association with BCSS ($p=0.011$, $p=0.013$, and $p=0.024$ for nucleolar score in 10, 5 and 1 FV, respectively) while modified Helpap's method showed no association ($p=0.195$) Figure 2. A similar observation was seen with DMFS Supplementary Figure 1.

B) *The validation set:*

The nucleoli were assessed in the validation set using the 5 FVs method, where the concordance rate between the two observers was (ICC = 0.981). Nucleolar score in the validation set showed that 534 cases (45%) had score 1, 40% of cases had score 2 and 15% of cases showed score 3.

Association of nucleoli with other clinicopathological parameters:

High nucleolar score was associated with parameters characteristic of aggressive tumor behavior including younger patient age, larger tumor size, higher tumor grade, stage 3, and poor Nottingham Prognostic Index (NPI). Only 10% of cases with nucleolar score 1 were high grade lesions in comparison with 40% and 63% of nucleolar score 2 and score 3, respectively ($p<0.0001$). Also, significant association was observed between nucleoli and number of positive axillary lymph nodes where 14% of cases with nucleolar score 3 showed more than 3 positive lymph nodes compared with 10% and 5% of cases with nucleolar score 2 and 1, respectively ($p<0.0001$).

Nucleolar score 3 tumors were more associated with oestrogen receptor negativity, progesterone receptor negativity and HER2 positivity (all $p<0.0001$). Significant associations were found between nucleolar score and histological subtypes of IBC where 97% of nucleolar score 3 tumors were of no specific histological type (NST) ($p<0.0001$). The association between nucleolar scores and various clinicopathological parameters are summarized in Table 2.

Association of nucleolar score with grade concordant and discordant cases:

To assess the value of using nucleolar score in cases associated with low grade concordance (cases with borderline features for grade using the Nottingham grading system), nucleolar score was applied to the subgroups of IBC based on grade concordance³². Significant association was observed between nucleolar score and grade concordant and discordant cases ($p < 0.0001$). 13% of grade concordant cases (G2/2) had nucleolar score 3 compared to 29% of discordant cases (G2/3) nucleoli was score 3. On other hand, when we compared between nucleolar score 2 in (G2/2) & (G2/3), we found that this percentage is decreased from 50% to 44%. Also, it was illustrated in the bar chart (supplementary figure 2) that (G2/3) has higher nucleolar score than (G2/2). Association between nucleoli counting in 5 FVs and concordant and discordant cases are shown in Supplementary Figure 2.

Outcome analysis:

In the whole cohort, significant associations were observed between high nucleolar score and shorter BCSS and DMFS (Log-rank=33.32, $p < 0.0001$ & Log-rank= 33.72, $p < 0.0001$ respectively) as shown in Figure 3.

Multivariate Cox regression model adjusting for the standard prognostic clinicopathological parameters including patient age, tumor size, and nodal stage showed nucleolar score as an independent predictor of survival (Table 4).

When the analysis was restricted to the most common type duct NST excluding special tumor types, incorporation of nucleolar score to the other grade components in Cox regression model showed that nucleoli was independent prognostic factor in survival prediction (hazard ratio=1.259, 95% confident interval= 1.03-1.53, p value=0.022) while pleomorphism was not significantly associated with outcome. Moreover, when pleomorphism was replaced with nucleolar score, the latter was independent predictor and showed more significant than nuclear pleomorphism (hazard ratio=0.004, 95% confident interval= 1.09-1.58, p value=0.004). Table 5 summarize the multivariate analysis results for various models used.

Nucleoli and the Nottingham grading system:

The incorporation of nucleolar score, as a replacement of nuclear pleomorphism, in the Nottingham grading system specially in NST cases Table 6 showed higher significant association with BCSS (Log-rank= 102, $p = 8.3 \times 10^{-23}$) than the normal grade (Log-rank=96, $p = 1.6 \times 10^{-21}$). In addition, when the nucleolar score was added to the existing grade components, it still showed higher significant association with survival than normal grade (Log-rank=104, $p = 2.8 \times 10^{-23}$), (Figure 4). Groups with new grade scores are illustrated in Supplementary Table 4.

A multivariate cox regression model including tumor size, nodal stage and tumor grade after replacing nuclear pleomorphism with nucleolar score or adding the nucleolar score

as an additional feature, showed that those grades are independently predictive of survival ($p= 8.2 \times 10^{-13}$ & $p= 6.1 \times 10^{-13}$, respectively) better than the normal grade ($p= 1.2 \times 10^{-13}$) (Supplementary Table 5).

Validation of nucleoli assessment in an external (TCGA) cohort:

Nucleolar score using the optimal method in the TCGA data set showed that 354 cases (48%) were of score 1, whereas scores 2 and 3 comprised 41% and 11% of cases, respectively. High nucleolar score was significantly associated with poor 10-year overall survival (OS) (Log-rank 12.81, $p = 0.002$).

Moreover, incorporation of nucleolar score in tumor grade using the Nottingham Scoring method for TCGA cases, which were assessed by one observer (L. Dalton), showed that incorporation of nucleolar score as an additional component to the grade had stronger association with OS (Log-rank= 13.5, $p=0.001$) than conventional grade (Log-rank = 11.9, $p=0.003$) and also an association was observed after replacing pleomorphism with nucleolar score in the grading system (Log-rank = 10.9, $p=0.004$) Figure 5.

DISCUSSION

The nucleolus is considered as a mirror of the cellular metabolic activity. Prominent nucleolus is associated with a high translational potential and is considered as an indicator of the cells' high demand for proteins (e.g. proto-oncogene proteins). Prominent nucleoli are an indication of cellular kinetics and cytochemical changes that occur in cancer cells³¹. In carcinoma, the association of nucleolar morphometric changes with poor patient prognosis is noteworthy and there is an increasing evidence from independent data that suggests an active role of the nucleolus in tumorigenesis^{15, 34, 35}. However, there are currently no consensus histopathological guidelines to evaluate nucleoli in BC clinical practice. Previous studies assessing the prognostic significance of nucleoli in BC lacked objective criteria for scoring nucleoli, and did not identify optimal cut-off points for prognostic stratification³¹. In the current study, nucleoli were assessed using a variety of scoring methods in a large cohort of invasive BC with long-term follow-up data to determine the most reproducible assessment method whilst also providing prognostic value. The applicability of incorporating nucleolar score into the routine grading system was also assessed.

Assessment of nucleoli in the training set demonstrated that counting nucleoli in 5 FVs (0.5mm²) using digitised whole slide images was the optimal method. This inference was based on the following; firstly, nucleolar score in 5 FVs had the highest concordance rate between the observers. Secondly, this method highly represents nucleoli within the whole slide without the requirement to count nucleoli in wider areas which is time consuming. Counting in 10 FVs requires more than 100 prominent nucleoli to stratify the tumor. It was also more objective than the modified Helpap's method²⁹. Lastly, nucleolar score in 5 FVs showed significant association with the patient outcome.

In 5 FVs method, we defined the nucleoli as $\geq 2.5 \mu\text{m}$ or easily seen at 10x to provide a distinction of what was observed at the level of research versus what might be applied in clinical practice. We utilised an eyepiece reticle in which 2.5 microns was the breadth between two cross-hatches at 40x. It was apparent to the observers that 2.5 microns at 40x via the reticle correlated with what was easily seen at 10x and therefore the eyepiece reticle was made redundant. We think that in daily clinical practice the 10x rule could be used based on our experimental observations at 2.5 microns.

In addition, we followed the same method of counting mitotic figures in routine practice of grading by choosing field with enough degree of cellularity and we relied on hotspots. We chose the field for counting after scanning the whole tumour using low power at 10x to locate nucleoli hotspots. We used fixed number of areas (hotspots) to count the prominent nucleoli in all cases. Nucleoli were given three scores depending on cut-off point generated with X-tile against breast cancer specific survival (BCSS).

The main disadvantage of the modified Helpap's method is the lack of information on the number of field views or the number of cells with prominent nucleoli to be scored or

even the size of nucleoli to be considered. Although we followed the method recommended by Donizy et al. ³¹ in melanoma, which used 20% of the tumor cells with prominent nucleoli, it was still more subjective than counting a defined number of FVs. This might also explain the discrepancies in high nucleolar scores between different methods as the modified Helpap's method required >20% of the tumor cells with prominent nucleoli (i.e. thousands of cells), while in the other methods, fewer cells with prominent nucleoli were counted to consider a high score.

High nucleoli count showed significant association with poor BCSS and shorter DMFS in both training and validation cohorts and also overall survival in the external validation cohort. Association between prominent nucleoli and poor prognosis has also been reported in melanoma ³¹ and prostatic cancer ²³.

Based on the association between nucleolar score with concordant and discordant cases, we can rely on the nucleolar score if the tumor is high or low grade. When we compared between concordant cases with discordant cases we found that percentage of nucleolar score 3 in (G2/3) was higher than (G2/2). The statistical significance here is related to these variations in percentages between groups and not to the absolute percentage. So, we concluded that the nucleolar score could provide an additional feature to determine the grade of these borderline cases.

In our study, high nucleolar score showed higher association with NST than lobular and other special types which include invasive cribriform, invasive mucinous and tubular carcinoma. NST is the most common type of IBC that constitutes 40-75% of all mammary invasive carcinomas ³⁶ and is associated with poor prognosis in comparison to other types ³⁷. We therefore incorporated nucleolar score to the tumor grade in NST cases only.

Despite the objective improvements that have been made to BC grading methods, any assessment of morphological characteristics inevitably retains a subjective element and is heavily dependent on the pre-analytical parameters. Nuclear pleomorphism indicates shape, chromatin distribution and size of nucleolus. Lack of clear definitions of this grading criterion, their multiplicity, and subjectivity ^{5, 38} of measurement resulting in poor reproducibility, are strong motivations to replace it with other more objective components. Also, nuclear pleomorphism showed the least concordance among grade components between pathologists in several studies ^{6, 7, 38, 39}.

Multivariate Cox regression model incorporating nucleolar score in 5 FVs with the Nottingham grade components including pleomorphism ⁴⁰ in NST cases showed that nucleolar score was independent predictive factor of outcome. In addition, after removal of nuclear pleomorphism from the model's co-variates, nucleolar score showed a high significant association than pleomorphism. Therefore, nucleolar scores could be more predictive of outcome than pleomorphism. These findings support the hypothesis that

nucleoli could be incorporated into grade, following our approach of quantitation and scoring, to provide a more objective measurement than using pleomorphism.

When we replaced nuclear pleomorphism with nucleolar scores, it showed higher significant association with patient outcome than grade with pleomorphism. Also, when adding nucleolar score as an additional feature to the grade, it showed highly significant association with BCSS. These findings support our hypothesis that nucleolar score could be a promising parameter to be assessed to enhance the grading system.

In conclusion, assessment of nucleoli in H&E stained full face invasive breast carcinoma sections in 5 FVs using whole slide images is a reproducible and practical method to predict tumor behavior and progression and also provides more evidence for the reproducibility and reliability of using digitised images in clinical practice. The use of whole slide images technology in nucleolar assessment also opens up opportunities for computer-assisted classification for easier determination of 2.5 microns with improved standardisation and reproducibility of evaluation for refinement of methodology by training deep learning models. Application of this method in routine practice would aid in risk-stratification of invasive BC for more individualized patient management.

Acknowledgments

This research was supported and funded by the Egyptian Ministry of Higher Education and Scientific Research. Also, we thank the Innovate UK (PathLAKE consortium) for funding parts of this study.

REFERENCES

- 1.Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: Experience from a large study with long-term follow-up. *Histopathology* 1991;**19**;403-410.
- 2.Rakha EA, El-Sayed ME, Lee AH *et al.* Prognostic significance of nottingham histologic grade in invasive breast carcinoma. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2008;**26**;3153-3158.
- 3.Rakha EA, Reis-Filho JS, Baehner F *et al.* Breast cancer prognostic classification in the molecular era: The role of histological grade. *Breast cancer research : BCR* 2010;**12**;207.
- 4.Rakha EA, Ahmed MA, Aleskandarany MA *et al.* Diagnostic concordance of breast pathologists: Lessons from the national health service breast screening programme pathology external quality assurance scheme. *Histopathology* 2017;**70**;632-642.
- 5.Huang J, Bay BH, Tan PH. Nuclear morphology in breast cancer: Is pleomorphism an illusion? *Medical Hypotheses* 2000;**55**;26-28.
- 6.Sloane JP, Amendoeira I, Apostolikas N *et al.* Consistency achieved by 23 european pathologists from 12 countries in diagnosing breast disease and reporting prognostic features of carcinomas. European commission working group on breast screening pathology. *Virchows Archiv : an international journal of pathology* 1999;**434**;3-10.
- 7.Meyer JS, Alvarez C, Milikowski C *et al.* Breast carcinoma malignancy grading by bloom-richardson system vs proliferation index: Reproducibility of grade and advantages of proliferation index. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 2005;**18**;1067-1078.
- 8.Derenzini M, Montanaro L, Trere D. What the nucleolus says to a tumour pathologist. *Histopathology* 2009;**54**;753-762.
- 9.Zink D, Fischer AH, Nickerson JA. Nuclear structure in cancer cells. *Nature reviews. Cancer* 2004;**4**;677-687.
- 10.MacCarty WC. The value of the macronucleolus in the cancer problem. *The American Journal of Cancer* 1936;**26**;529.
- 11.Derenzini M, Trere D, Pession A, Montanaro L, Sirri V, Ochs RL. Nucleolar function and size in cancer cells. *The American journal of pathology* 1998;**152**;1291-1297.
- 12.Derenzini M, Trere D, Pession A, Govoni M, Sirri V, Chieco P. Nucleolar size indicates the rapidity of cell proliferation in cancer tissues. *The Journal of pathology* 2000;**191**;181-186.
- 13.Trere D, Ceccarelli C, Montanaro L, Tosti E, Derenzini M. Nucleolar size and activity are related to prb and p53 status in human breast cancer. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* 2004;**52**;1601-1607.
- 14.Lam YW, Trinkle-Mulcahy L, Lamond AI. The nucleolus. *Journal of Cell Science* 2005;**118**;1335.
- 15.Montanaro L, Trere D, Derenzini M. Nucleolus, ribosomes, and cancer. *The American journal of pathology* 2008;**173**;301-310.
- 16.Bywater MJ, Poortinga G, Sanij E *et al.* Inhibition of rna polymerase i as a therapeutic strategy to promote cancer-specific activation of p53. *Cancer cell* 2012;**22**;51-65.
- 17.Fischer AH, Bardarov S, Jr, Jiang Z. Molecular aspects of diagnostic nucleolar and nuclear envelope changes in prostate cancer. *Journal of cellular biochemistry* 2004;**91**;170-184.
- 18.Hernandez-Verdun D. The nucleolus: A model for the organization of nuclear functions. *Histochemistry and cell biology* 2006;**126**;135-148.
- 19.Zhong S, Zhang C, Johnson DL. Epidermal growth factor enhances cellular tata binding protein levels and induces rna polymerase i- and iii-dependent gene activity. *Molecular and cellular biology* 2004;**24**;5119-5129.
- 20.James MJ, Zomerdijk JC. Phosphatidylinositol 3-kinase and mtor signaling pathways regulate rna polymerase i transcription in response to igf-1 and nutrients. *The Journal of biological chemistry* 2004;**279**;8911-8918.
- 21.Grandori C, Gomez-Roman N, Felton-Edkins ZA *et al.* C-myc binds to human ribosomal DNA and stimulates transcription of rna genes by rna polymerase i. *Nature cell biology* 2005;**7**;311-318.

- 22.Lorenzato M, Abboud P, Lechki C *et al.* Proliferation assessment in breast cancer: A double-staining technique for agnor quantification in mib-1 positive cells especially adapted for image cytometry. *Micron (Oxford, England : 1993)* 2000;**31**;151-159.
- 23.Diaconescu S, Diaconescu, D., & Toma, S. Nucleolar morphometry in prostate cancer. *Bulletin of the Transilvania University of Brasov.Medical Sciences* 2010; **3**;23-26.
- 24.Pich A, Chiusa L, Margaria E. Prognostic relevance of agnors in tumor pathology. *Micron (Oxford, England : 1993)* 2000;**31**;133-141.
25. Derenzini M, Trere D, O'Donohue MF, Ploton D. Interphase nucleolar organiser regions in tumour pathology. *Molecular biology in cellular pathology*: Wiley, 2003;137-152.
- 26.Rakha EA, Aleskandarani M, Toss MS *et al.* Breast cancer histologic grading using digital microscopy: Concordance and outcome association. *Journal of clinical pathology* 2018;**71**;680-686.
- 27.Ciriello G, Gatza ML, Beck AH *et al.* Comprehensive molecular portraits of invasive lobular breast cancer. *Cell* 2015;**163**;506-519.
- 28.Gao J, Aksoy BA, Dogrusoz U *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cbiportal. *Science signaling* 2013;**6**;pl1.
- 29.Helpap B, Knupffer J, Essmann S. Nucleolar grading of renal cancer. Correlation of frequency and localization of nucleoli to histologic and cytologic grading and stage of renal cell carcinomas. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 1990;**3**;671-678.
- 30.Dalton LW, Gerds TA. The advantage of discordance: An example using the highly subjective nuclear grading of breast cancer. *The American journal of surgical pathology* 2017;**41**;1105-1111.
- 31.Donizy P, Biecek P, Halon A, Maciejczyk A, Matkowski R. Nucleoli cytomorphology in cutaneous melanoma cells - a new prognostic approach to an old concept. *Diagnostic pathology* 2017;**12**;88.
- 32.Rakha EA, Aleskandarany MA, Toss MS *et al.* Impact of breast cancer grade discordance on prediction of outcome. *Histopathology* 2018;**73**;904-915.
- 33.Camp RL, Dolled-Filhart M, Rimm DL. X-tile. *A New Bio-Informatics Tool for Biomarker Assessment and Outcome-Based Cut-Point Optimization* 2004;**10**;7252-7259.
- 34.Montanaro L, Trere D, Derenzini M. Changes in ribosome biogenesis may induce cancer by down-regulating the cell tumor suppressor potential. *Biochimica et biophysica acta* 2012;**1825**;101-110.
- 35.Bacalini MG, Pacilli A, Giuliani C *et al.* The nucleolar size is associated to the methylation status of ribosomal DNA in breast carcinomas. *BMC cancer* 2014;**14**;361.
- 36.Moinfar F. *Essentials of diagnostic breast pathology: A practical approach*: Springer Science & Business Media, 2007.
- 37.Makki J. Diversity of breast carcinoma: Histological subtypes and clinical relevance. *Clinical medicine insights. Pathology* 2015;**8**;23-31.
- 38.Frierson HF, Jr., Wolber RA, Berean KW *et al.* Interobserver reproducibility of the nottingham modification of the bloom and richardson histologic grading scheme for infiltrating ductal carcinoma. *American journal of clinical pathology* 1995;**103**;195-198.
- 39.Rakha EA, Bennett RL, Coleman D, Pinder SE, Ellis IO. Review of the national external quality assessment (eqa) scheme for breast pathology in the uk. *Journal of clinical pathology* 2017;**70**;51-57.
- 40.Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: Experience from a large study with long-term follow-up. *Histopathology* 2002;**41**;154-161.

Table (1): Association of nucleolar score as defined by the four methods used and patient outcome in terms of breast cancer specific survival (BCSS) and distant metastasis free survival (DMFS) in the training set.

Methods	Nucleolar score Number of cases (%)	BCSS Log-rank	DMFS Log-rank
---------	--	------------------	------------------

	Score 1	Score 2	Score 3	(<i>p</i> -value)	(<i>p</i> -value)
Modified Helpap's method	146 (36)	187 (47)	67 (17)	3.3 (0.195)	3.4 (0.185)
Nucleoli scoring in 10 field views (1mm²) (at 40x)	130 (33)	204 (51)	66 (16)	9.1 (0.011)	6.7 (0.035)
Nucleoli scoring in 5 field views (0.5mm²) (at 40x)	130 (32)	203 (51)	67 (17)	8.8 (0.013)	6.6 (0.038)
Nucleoli count in 1 field view (0.1mm²) (at 40x)	131 (33)	197 (49)	72 (18)	7.4 (0.024)	6.2 (0.044)

Significant *p* values are in bold

Table (2): Association between nucleolar scores in 5 field views (0.5mm²) and clinicopathological parameters in Nottingham cohort (n=1600). Nucleolar scores cut-off was determined using X-tile. Bioinformatics software.

Parameter	Nucleolar score			Chi- square (χ²) (p-value)
	Score 1 N (%)	Score 2 N (%)	Score 3 N (%)	
Patient Age ≤ 50 years > 50 years	165 (25) 485 (75)	252 (39) 399 (61)	108 (40) 163 (60)	32.1 (<0.0001)
Tumour size ≤ 2cm > 2cm	470 (72) 180 (28)	398(61) 252 (39)	134 (49) 137 (51)	46.4 (<0.0001)

Tumour grade				
1	333 (50)	103 (16)	11 (4)	412.6 (<0.0001)
2	262 (40)	296 (44)	89 (33)	
3	69 (10)	264 (40)	173 (63)	
Stage				
1	434 (67)	389 (60)	151 (56)	23.6 (<0.0001)
2	174 (27)	189 (29)	76 (28)	
3	42 (6)	73 (11)	43 (16)	
Oestrogen receptor status				
Positive	611 (94)	511 (79)	155 (57)	175.4 (<0.0001)
Negative	39 (6)	140 (21)	116 (43)	
Progesterone receptor status				
Positive	405 (62)	323 (50)	98 (36)	56.3 (<0.0001)
Negative	245 (38)	328 (50)	173 (64)	
HER2 receptor status				
Positive	31 (5)	70 (11)	37 (14)	28.5 (<0.0001)
Negative	607 (95)	559 (89)	229 (86)	
Nottingham prognostic index (NPI)				
Good	344 (53)	157 (24)	29 (11)	216.3 (<0.0001)
Moderate	258 (40)	368 (57)	164 (61)	
Poor	48 (7)	125 (19)	76 (28)	
Lymph node positive (LN)				
Negative	473 (73)	416 (64)	168 (62)	28.4 (<0.0001)
Positive 1-3	144 (22)	170 (26)	46 (24)	
Positive >3	33 (5)	65 (10)	39 (14)	
Histological subtypes				
No specific type (NST)	526 (81)	596 (92)	261 (97)	63.2 (<0.0001)
Lobular	60 (9)	31 (5)	5 (2)	
Other special type*	64 (10)	22 (4)	3 (1)	

* Other special types include tubular, mucinous, cribriform, papillary, micropapillary
Significant *p* values are in bold

Table (3): Association between nucleolar scoring in 5 field views (0.5mm²) and concordant and discordant cases in the whole cohort (n=1600). Nucleolar score cut-offs were determined by X-tile.

Parameter	Nucleolar score			Chi- square (χ^2) (<i>p</i> -value)
	Score 1	Score 2	Score 3	
Concordant cases N (%)				
Grade 1-1	268 (79)	68 (20)	4 (1)	414.1 (<0.0001)
Grade 2-2	181 (37)	242 (50)	61 (13)	
Grade 3-3	46 (11)	210 (52)	151 (37)	

Discordant cases N (%)				
Grade 1-1	268 (79)	68 (20)	4 (1)	149.8 (<0.0001)
Grade 2-2	181 (37)	242 (50)	61 (13)	
Grade 1-2, 2-1	124 (61)	68 (34)	10 (5)	
Discordant cases N (%)				
Grade 2-2	181 (37)	242 (50)	61 (13)	115.4 (<0.0001)
Grade 3-3	46 (11)	210 (52)	151 (37)	
Grade 2-3, 3-2	43 (27)	72 (44)	47 (29)	

Significant *p* values are in bold

Table (4): Multivariate Cox regression analysis results for predictors of Breast Cancer Specific Survival in Nottingham cohort.

Parameters	Hazard ratio (HR)	95% confident interval (CI)		<i>p</i>-value
		Lower	Upper	
Nucleoli scoring	1.357	1.161	1.584	0.0001
Tumour size	1.561	1.222	1.994	0.0004
Stage	2.151	1.845	2.508	<0.0001
Patient age	1.217	0.956	1.548	0.11

Significant *p* values are in bold

Table (5): Multivariate Cox regression analysis of the components of grade as predictors of Breast Cancer Specific Survival, with and without inclusion of nucleolar scores.

Models	Parameters	Hazard ratio (HR)	95% confident interval (CI)		<i>p</i> -value
			Lower	Upper	
(A)	Nuclear pleomorphism	1.294	1.016	1.648	0.037
	Mitosis	1.473	1.239	1.751	<0.0001
	Tubule formation	2.178	1.446	3.281	<0.001
(B)	Nuclear pleomorphism	1.166	0.902	1.508	0.241
	Mitosis	1.454	1.223	1.729	<0.0001
	Tubule formation	2.125	1.411	3.201	<0.001
	Nucleoli scoring	1.259	1.034	1.533	0.022
(C)	Nucleoli scoring	1.314	1.093	1.580	0.004
	Mitosis	1.525	1.305	1.781	<0.0001
	Tubule formation	2.249	1.510	3.349	<0.0001

Significant *p* values are in bold

Table (6): Frequency of tumour grade and their association with outcome in terms of BCSS after adding nucleolar scoring to the grading system in no specific type (NST) cases (n=1198) (log-rank test).

Parameter	Tumour grade Number of cases (%)			Survival Log-rank (<i>p</i> -value)
	Grade 1	Grade 2	Grade 3	
Tumour Grade	337 (28)	523 (44)	338 (28)	96 (1.6 × 10⁻²¹)
Grade after replacing nuclear pleomorphism score with nucleoli score	199 (17)	602 (50)	397 (33)	102 (8.3 × 10⁻²³)
Grade after adding nucleoli score to the other three components of the grade	262 (22)	574 (48)	362 (30)	104 (2.8 × 10⁻²³)

Significant *p* values are in bold

FIGURE LEGENDS

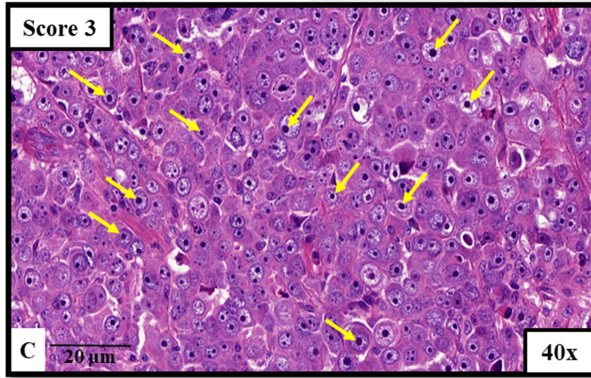
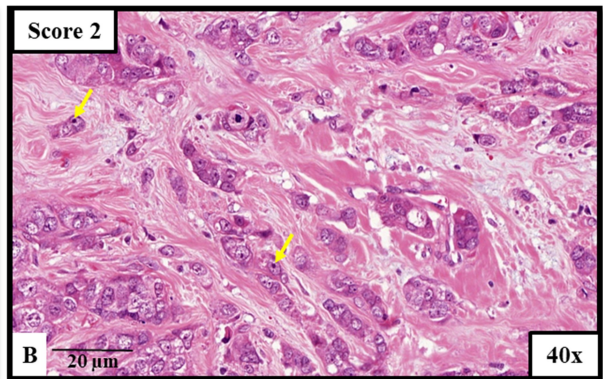
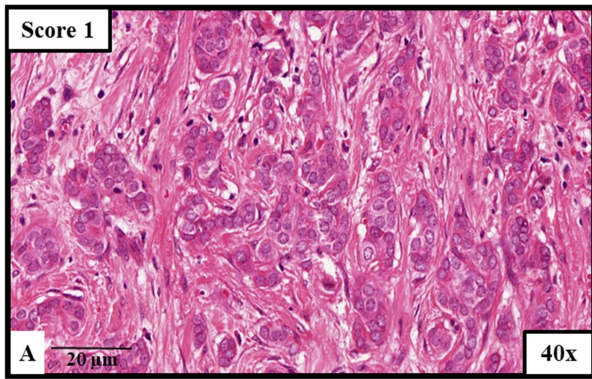
Figure (1): Examples of nucleolar scores 1, 2 and 3 in 1 field view (1FV) (yellow arrows indicate prominent nucleoli considered in counting), haematoxylin & eosin at x40.

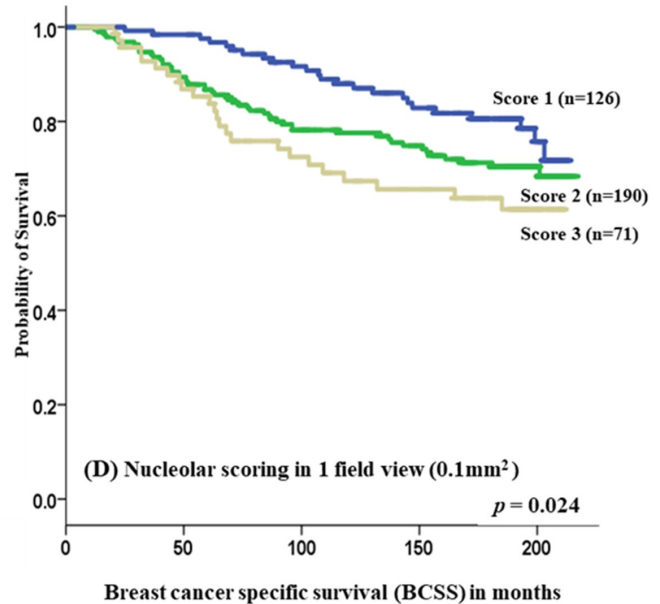
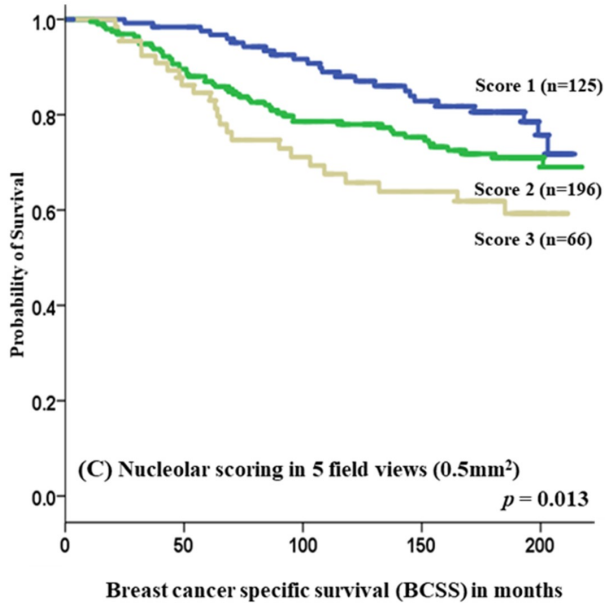
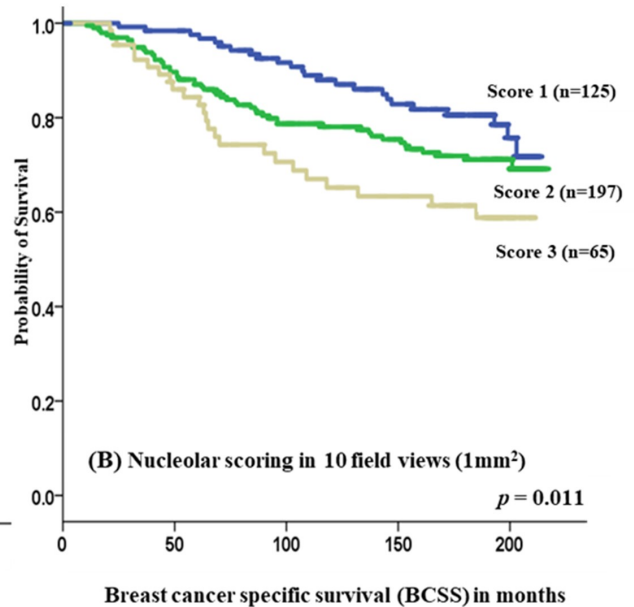
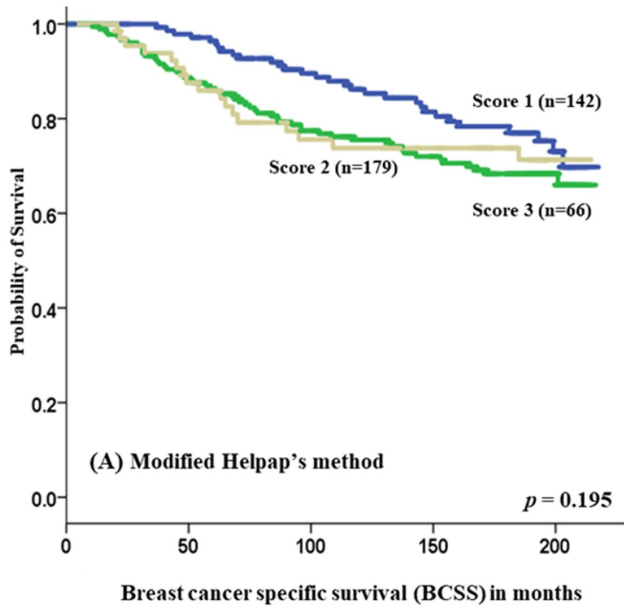
Figure (2): Kaplan-Meier survival plots showing association of nucleolar scores using the four different assessment methods with breast cancer specific survival (BCSS) in training set. A) Nucleolar scoring using Modified Helpap's method, B) Nucleolar scoring in 10 field views (1mm²), C) Nucleolar scoring in 5 field views (0.5mm²), and D) Nucleolar scoring in one field view (0.1mm²).

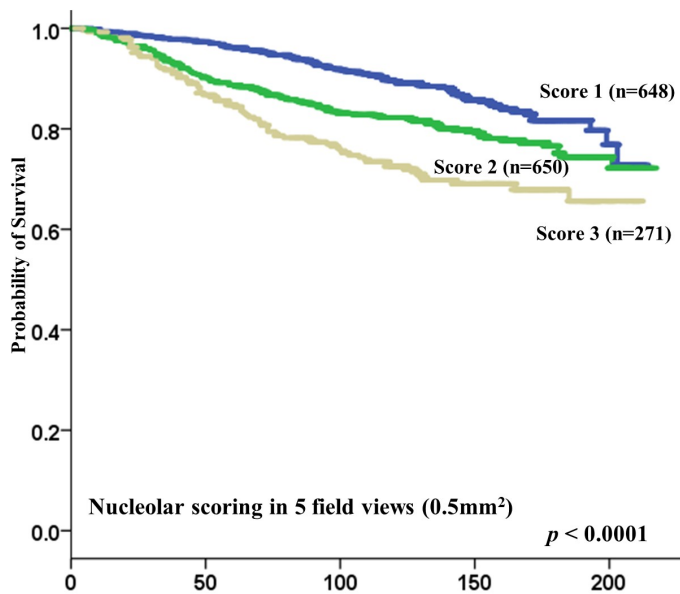
Figure (3): Kaplan-Meier plots showing association of nucleolar scoring in 5 field views (0.5mm²) with (A) Breast cancer specific survival (BCSS) and (B) Distant metastasis free survival (DMFS) in Nottingham cohort.

Figure (4): Kaplan-Meier survival plots showing breast cancer specific survival (BCSS) of (A) Nottingham tumor grade, (B) Nottingham grade after replacing nuclear pleomorphism with nucleolar score, (C) Nottingham grade after adding nucleolar score to its components in no-specific type (NST) cases.

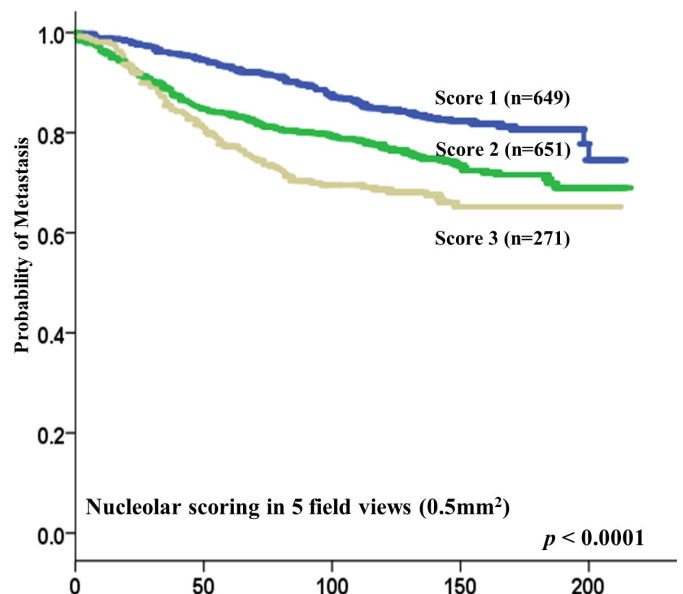
Figure (5): Kaplan-Meier survival plot showing overall survival (OS) of (A) Nucleolar scoring in 5 field views (B) Tumor grade (C) Tumor grade after replacing nuclear pleomorphism with nucleolar score, (D) Tumor grade after adding nucleolar score to its components in TCGA data.



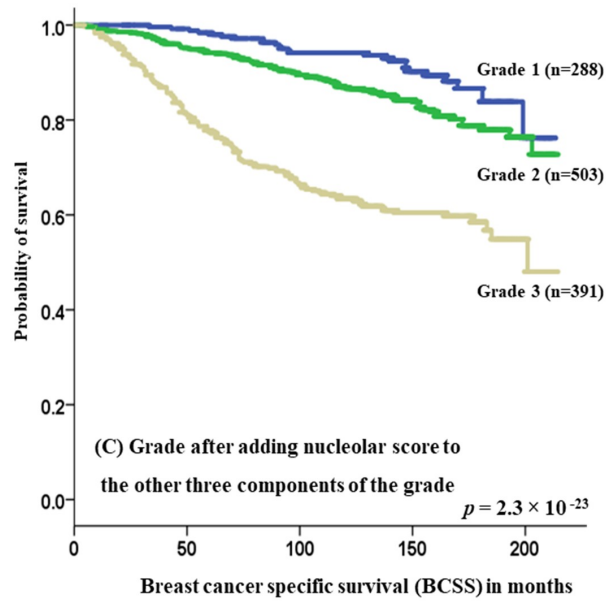
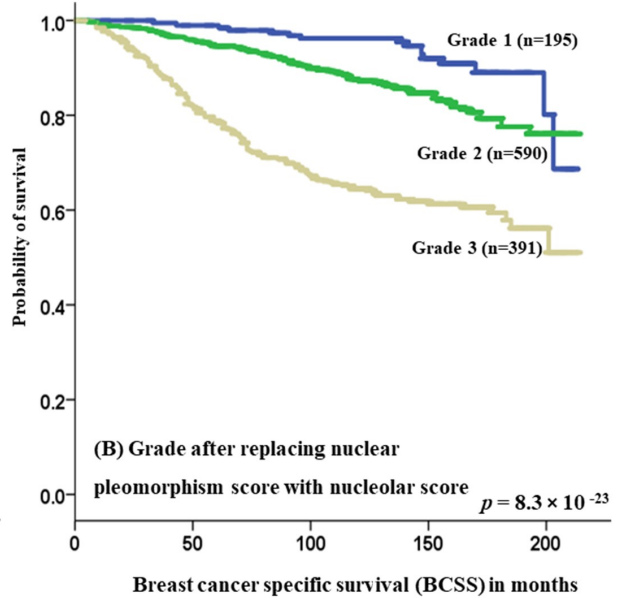
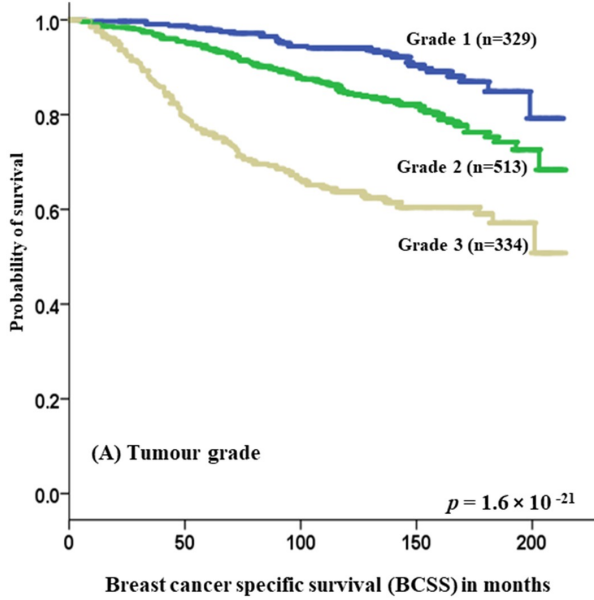


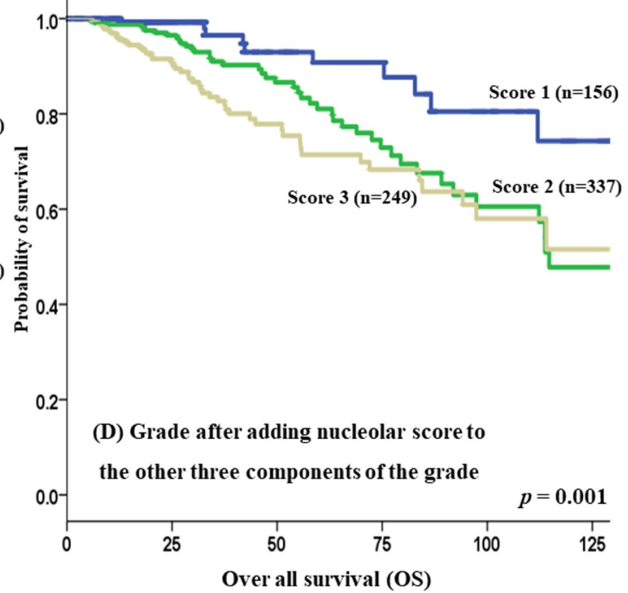
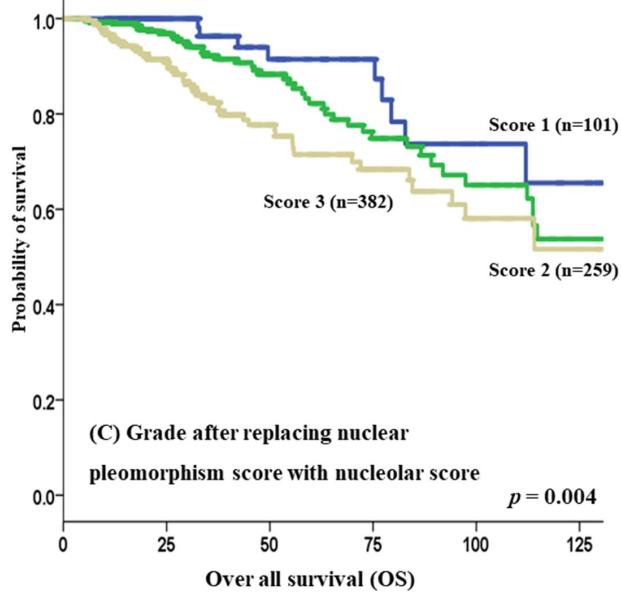
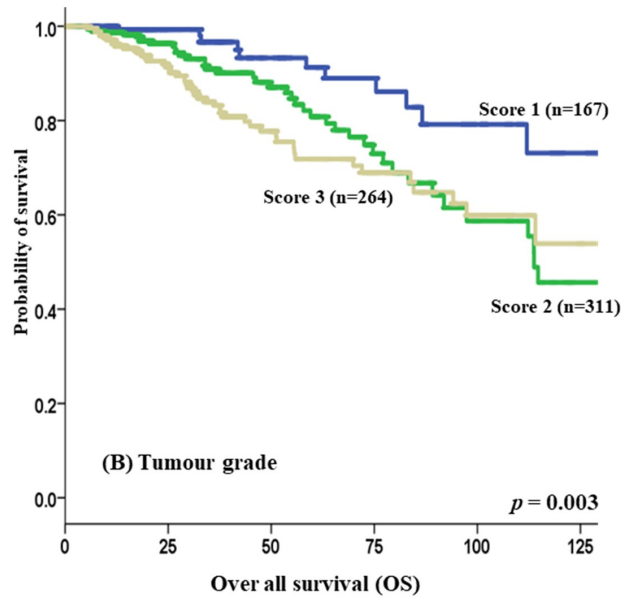
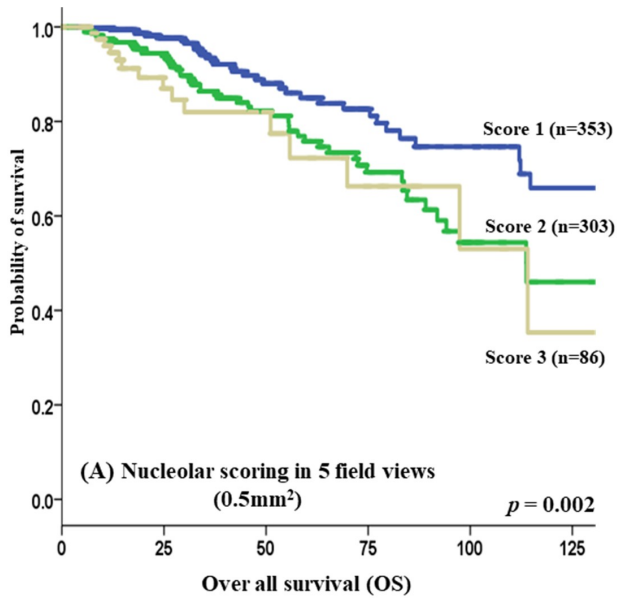


(A) Breast cancer specific survival (BCSS) in months



(B) Distant metastasis free survival (DMFS) in months





Supplementary Tables

Supplementary Table (1): Clinicopathological characteristics of the cases in both training (n=400) and validation sets (n=1200) of the study cohort.

Parameter	Training set Number of cases (%)	Validation set Number of cases (%)
Patient Age ≤ 50 years > 50 years	136 (35) 251 (65)	389 (33) 796 (67)
Tumour size ≤ 2cm > 2cm	225 (58) 162 (42)	777 (66) 407 (34)
Tumour grade 1 2 3	89 (22) 181 (45) 130 (33)	358 (30) 466 (39) 376 (31)
Stage 1 2 3	246 (64) 101 (26) 40 (10)	728 (62) 338 (28) 118 (10)
Lymph node status (LN) Negative Positive 1-3 Positive >3	236 (59) 115 (29) 46 (12)	812 (69) 277 (23) 96 (8)
Oestrogen receptor status Positive Negative	309 (80) 78 (20)	968 (82) 217 (18)
Progesterone receptor status Positive Negative	215 (56) 172 (44)	611 (52) 574 (48)
HER2 receptor status Positive Negative	42 (11) 340 (89)	96 (8) 1055 (92)
Nottingham prognostic index (NPI) Good Moderate Poor	106 (27) 217 (56) 64 (17)	424 (36) 573 (48) 185 (16)
Survival status Alive Dead	243 (71) 100 (29)	883 (82) 197 (18)
Distant metastasis Yes No	116 (30) 271 (70)	241 (20) 943 (80)

Supplementary Table (2): Cut-off points of nucleoli count as generated by the X-tile Bioinformatics software based on the association with breast cancer specific survival (BCSS).

Parameter	Definitions of Nucleoli Count		
	Score 1	Score 2	Score 3
Nucleoli count in 10 field views (1mm²) (at 40x)	0-4	5-101	> 101
Nucleoli count in 5 field views (0.5mm²)(at 40x)	0-2	3-50	> 50
Nucleoli count in 1 field view (0.1mm²) (at 40x)	0	1-9	> 9

Supplementary Table (3): Nucleolar scoring inter- observer concordance results in the training set (n=400).

Parameter	95% confident interval (CI)		Inter-observer reliability
	Lower	Upper	
Modified Helpap's method	0.359	0.474	0.417*
Nucleoli count in 10 field views (1mm²) (at 40x)	0.612	0.786	0.712 ^a
Nucleoli count in 5 field views (0.5mm²) (at 40x)	0.703	0.844	0.782 ^a
Nucleoli count in 1 field view (0.1mm²) (at 40x)	0.692	0.839	0.774 ^a

^a Performed using Intra-class correlation coefficient.

*Performed using Fleiss' Kappa test between four observers.

Supplementary Table (4): Incorporation of **nucleolar scores** into the Nottingham Grading System. Cut-off points of grade scores were generated by X-tile software based on association with breast cancer specific survival (BCSS) in no specific type (NST) cases.

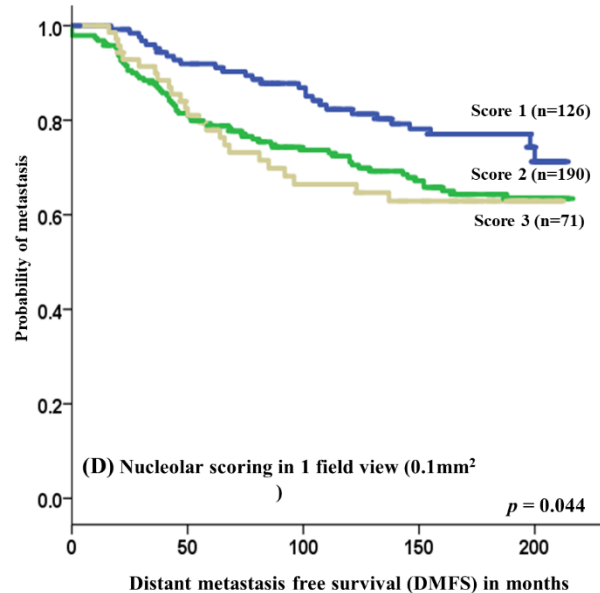
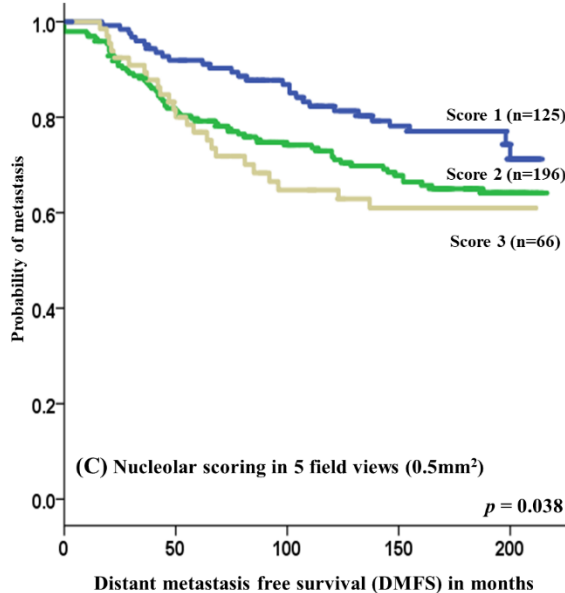
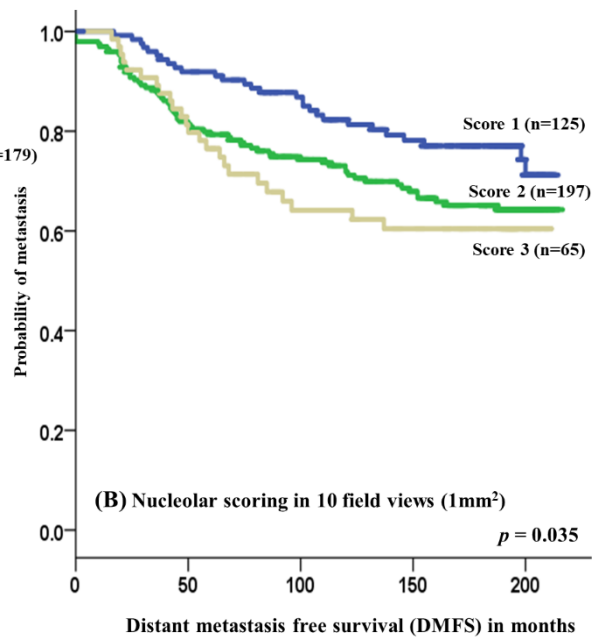
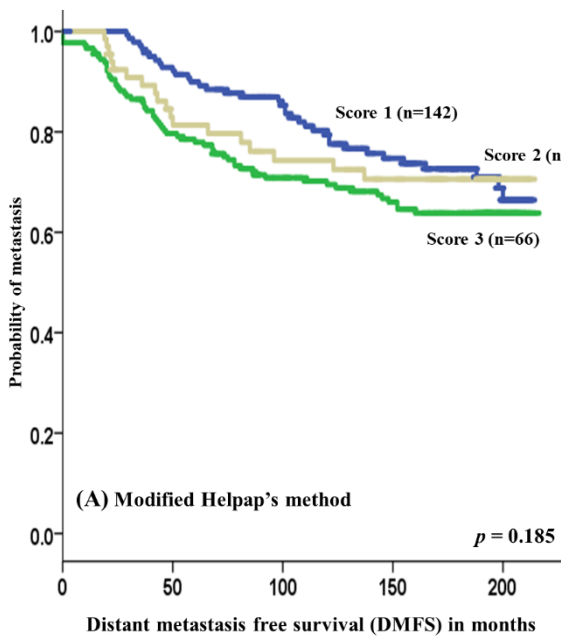
Groups	Score	Equivalent Grade
Group (1) Grade after replacing nuclear pleomorphism score with nucleolar score	Total score 3, 4	Grade 1
	Total score 5,6	Grade 2
	Total score 7,8,9	Grade 3
Group (2) Grade after adding nucleolar score to the other three components of the grade	Total score 4, 5 ,6	Grade 1
	Total score 7, 8, 9	Grade 2
	Total score 10,11,12	Grade 3

Supplementary Table (5): Results of multivariate Cox regression analysis in no-specific type (NST) cases (n=1198).

Models	Parameters	Hazard ratio (HR)	95% confident interval (CI)		Significance <i>p</i> -value
			Lower	Upper	
(A)	Grade	2.019	1.677	2.432	1.2 × 10⁻¹³
	Tumour size	1.591	1.222	2.070	0.001
	Stage	2.037	1.722	2.410	1.1 × 10⁻¹⁶
(B)	Grade after replacing nuclear pleomorphism score with nucleolar score	2.182	1.762	2.702	8.2 × 10⁻¹³
	Tumour Size	1.493	1.141	1.952	0.003
	Stage	2.010	1.697	2.381	6.1 × 10⁻¹⁶

(C)	Grade after adding nucleolar score to the other three components of the grade	2.083	1.706	2.544	6.1×10^{-13}
	Tumour size	1.518	1.161	1.983	0.002
	Stage	2.027	1.712	2.400	2.6×10^{-16}

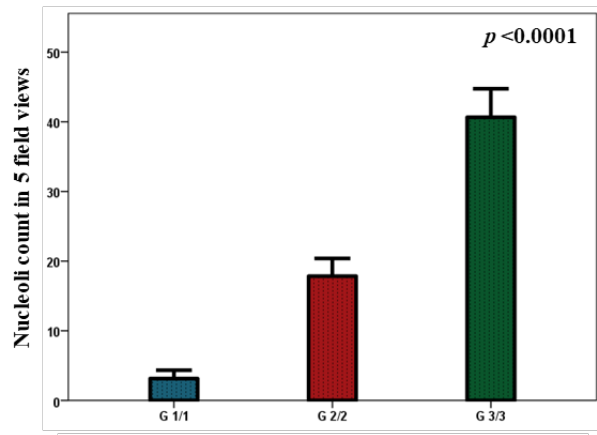
Significant p values are in bold



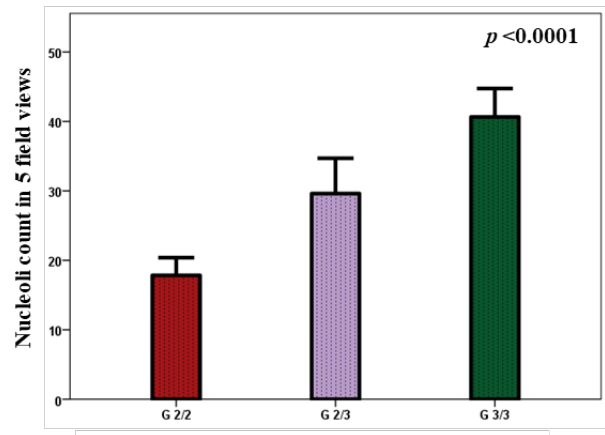
Supplementary figures

Supplementary Figure (1): Kaplan–Meier survival plot showing association of the four different scoring methods of nucleoli with distant metastasis free survival (DMFS) in training set. These show **nucleolar scoring** A) using Modified Helpap' method, B) in 10 field views (1mm²), C) in 5 field views (0.5mm²), and D) in one field view method (0.1mm²).

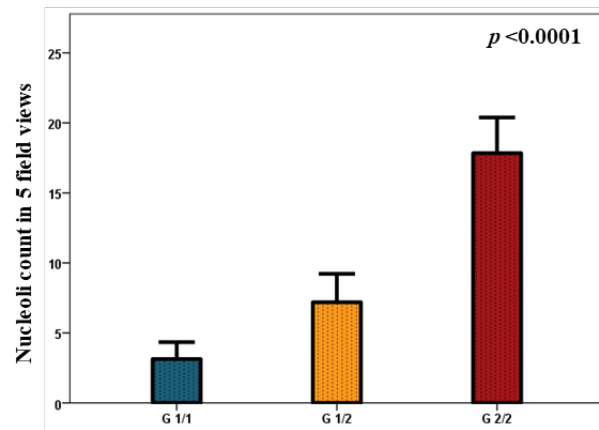
Supplementary Figure (2): Bar charts showing association of nucleoli counting in 5 field views (0.5mm²) and (A) Grade concordant cases (B) Grade high discordant cases (C) Grade low discordant cases in Nottingham cohort.



(A) Grade concordant cases



(B) Grade high discordant cases



(C) Grade low discordant cases