Artificial Intelligence Image Analysis for Chromosomal Instability in Primary and Metastatic Breast Cancers

Joe Oakley[1]; Yikan Wang[1]; David S Klimstra[1]; Fresia Pareja[2]; Jorge Sergio Reis-Filho[2]

Paige.AI Inc., 11 Times Square, 37 Floor, New York, USA 2. Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY 10021, USA

Background:

Chromosomal Instability, the gain or loss of large portions of or entire chromosomes, has prognostic and therapeutic value in Breast Cancer. Current testing for CIN is molecular-based. We developed an AI-based method applied to digital whole slide images (WSI) of H&E-stained BC sections to identify BCs that may harbor CIN, utilizing molecular measures of CIN as 'ground truth'.

Design:

Whole slide images from 1841 primary and 1708 metastatic breast cancers sequenced using targeted sequencing (MSK-IMPACT assay) were retrieved. Tetraploidy, whole genome doubling (WGD), loss of heterozygosity (LOH), and fraction of genome altered (FGA) derived from MSK-IMPACT were used as ground truth for CIN. Cut-offs for LOH and FGA were determined by comparison to incidence of biallelic loss of *BRCA1/2* and *TP53* oncogenic alterations, respectively. A genome instability index (GI) based on FGA and LOH was derived and used as a separate ground truth for CIN in AI model development. Breast cancer tissue regions were identified within WSI using a pre-trained digital pathology foundation model, and embeddings for cancer tiles were used to train an aggregation network with 8-fold cross validation to detect CIN for each of the ground truths of tetraploidy, WGD, LOH, FGA, and GI index. The obtained models' performance were validated by selecting a representative model from the 8-fold cross validation and testing on an independent cohort of 1429 WSIs of H&E-stained sections from 793 primary and 604 metastatic breast cancer samples.

$$\text{Genomic Instability (GI) Index} = \frac{\text{FGA x LOH}}{((\text{FGA x LOH}) + (1-\text{FGA}) \times (1-\text{LOH}))}$$

Results:

Results were evaluated by receiver operator characteristic (ROC) area under the curve (AUC) for each of the developed AI models. In primary and metastatic breast cancer samples, the AI-based method predicted tetraploidy with AUC of 0.85 and 0.84, WGD with an AUC of 0.83 and 0.81, LOH with an AUC of 0.78 and 0.69, FGA with an AUC of 0.83 and 0.82, and GI index with an AUC of 0.85 and 0.81, respectively. Subtype stratification of results was performed for ER+, triple negative status, and HER2-Low or HER2-amplified status. AUCs in primary and metastatic settings were robust across these subtypes for models using tetraploidy, WGD, LOH, FGA or GI Index as their ground truth for CIN.

When tested on the independent validation cohort, the AI-based method showed best performance with representative models trained using metastatic examples, with AUCs of 0.77 for tetraploidy, 0.73 for WGD, 0.73 for LOH, 0.78 for FGA, and 0.77 for GI obtained for the primary and metastatic cases of the validation cohort.

Obtained AUC for Detection of Various Measures of CIN in All Primary and Metastatic Breast Cancers and Subtypes Tested

Breast Cancer Type	Tetraploidy	Whole Genome Doubling	Loss of Heterozygosity ≥20%	Fractional Genome Alteration ≥70%	Genomic Instability Index ≥0.2	Cases for Each Breast Cancer Type
	ROC AUC	ROC AUC	ROC AUC	ROC AUC	ROC AUC	Cases
Primary Breast Cancers, All Types	0.85	0.83	0.78	0.83	0.85	1841
Metastatic Breast Cancers, All Types	0.84	0.81	0.69	0.82	0.81	1708
Histologic Grade 3, Primary	0.85	0.81	0.78	0.81	0.84	927
ER Positive, Primary	0.84	0.83	0.73	0.81	0.82	1597
Triple Negative Breast Cancer Old*, Primary	0.84	0.78	0.82	0.76	0.87	285
Triple Negative Breast Cancer New**, Primary	0.80	0.78	0.79	0.70	0.89	180
Her2 Low Breast Cancer, Primary	0.86	0.82	0.78	0.84	0.85	992
Her2 Amplified Breast Cancer, Primary	0.72	0.86	0.71	0.80	0.72	118
Histologic Poorly Differentiated, Metastatic	0.83	0.79	0.70	0.82	0.84	291
ER Positive, Metastatic	0.83	0.81	0.67	0.81	0.79	1413
Triple Negative Breast Cancer Old*, Metastatic	0.81	0.79	0.75	0.81	0.87	346
Triple Negative Breast Cancer New**, Metastatic	0.83	0.82	0.74	0.85	0.85	205
Her2 Low Breast Cancer, Metastatic	0.84	0.83	0.69	0.83	0.81	865
Her2 Amplified Breast Cancer, Metastatic	0.78	0.68	0.63	0.76	0.77	193

Table 1: The AUC for prediction of various measures of CIN in the tested primary and metastatic breast cancers and their subtypes was generally higher for primary tumors overall and in each subtype category and trended towards better prediction with larger amounts of WSI to train for that subtype. The "old" definition of triple negative breast cancer (*) indicates those breast cancers which were negative by IHC for ER and PR and HER2 2+ without amplification on in situ hybridization, HER2 1+ or HER2 0. The "new" definition of triple negative breast cancer (**) was defined as cases which were negative by IHC for ER and PR and on HER2 staining were IHC 0—thus including the definition of HER2 Low (HER2 IHC 1+ or 2+ without amplification on in situ hybridization) as HER2 positive.

Performance Metrics on the Independent Validation Cohort of Metastatic and Primary Breast Cancer WSIs Using the Representative Models Trained on Primary Breast Cancers (Top) or Metastatic Breast Cancers (Bottom) for the Measures of CIN Investigated.

Primary Breast Cancer Trained Model

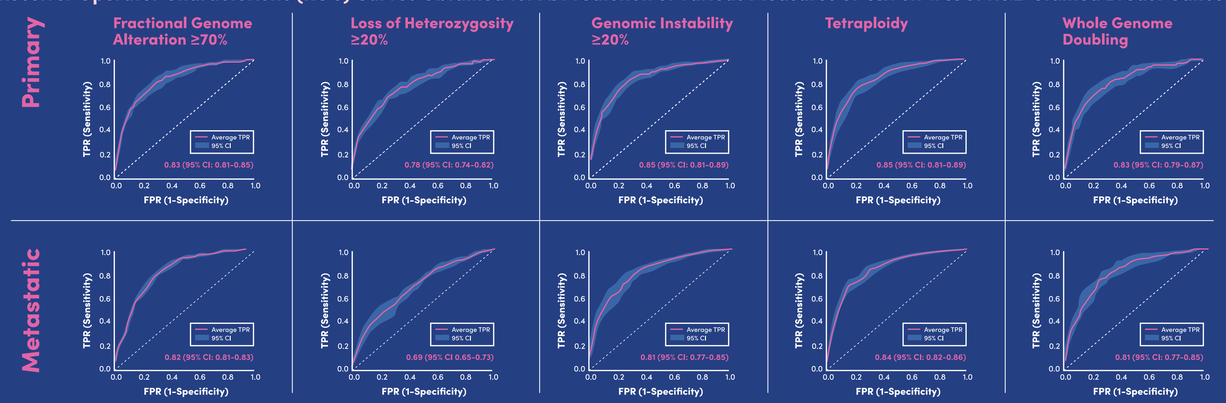
CIN Ground Truth	ROC AUC	Sensitivity	Specificity	PPV	NPV	Number of Cases Positive	Number of Cases Negative	Total Cases
Tetraploidy	0.73	0.72	0.64	0.67	0.69	652	629	1281
LOH > or = 20%	0.73	0.83	0.41	0.63	0.67	765	632	1397
FHA > or = 70%	0.73	0.96	0.17	0.45	0.86	575	822	1397
GI Index	0.73	0.91	0.33	0.67	0.71	832	565	1397
WGD	0.73	0.68	0.66	0.55	0.77	365	592	957

Metastatic Breast Cancer Trained Model

CIN Ground Truth	ROC AUC	Sensitivity	Specificity	PPV	NPV	Number of Cases Positive	Number of Cases Negative	Total Cases
Tetraploidy	0.77	0.84	0.53	0.65	0.77	652	629	1281
LOH > or = 20%	0.73	0.70	0.63	0.70	0.64	765	632	1397
FHA > or = 70%	0.78	0.75	0.67	0.62	0.80	575	822	1397
GI Index	0.77	0.77	0.62	0.75	0.64	832	565	1397
WGD	0.73	0.81	0.51	0.50	0.81	365	592	957

Table 2: The ROC AUC, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) are shown for the models trained on primary or metastic breast cancers (top and bottom, respectively) to identify the various measures of ground truth. Each model presented was tested against the entire validation cohort of both primary and metastatic breast cancers. The number of cases positive and negative for each ground truth of CIN in the cohort are shown, including the total number of cases with a ground truth molecular result for that CIN metric. Performance was slightly better for the models trained on metastatic breast cancers for all ground truths of CIN tested. Performance overall on the validation cohorts for each ground truth of CIN was comparable to the training set ROC AUC for each ground truth of CIN (Table 1), validating the models and confirming generalizability of the underlying AI digital pathology analysis algorithms for these tasks.

Receiver Operator Characteristic (ROC) Curves Obtained for AI Prediction of Various Measures of CIN on WSI of H&E-Stained Breast Cancers



Normalized Distribution of Cases with Fractional Genome Alteration (FGA) and Loss of Heterozygosity (LOH) Versus TP53 Mutation Status And Biallelic Loss of BRCA1 and/or BRCA2

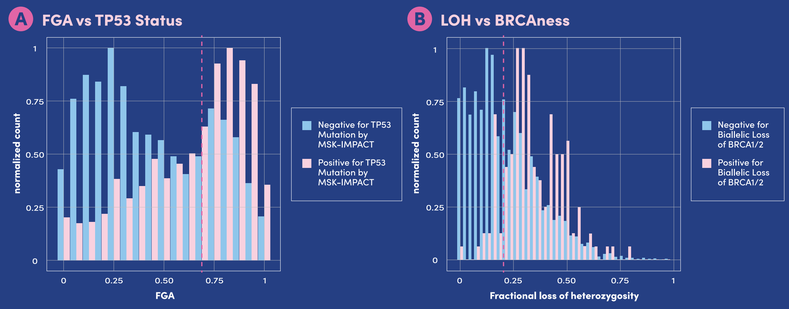


Figure 1: Panel A shows the normalized distribution of breast cancers tested based on the percentage of FGA demonstrated. Blue bars are wild type for TP53 mutations while pink bars are cases with oncogenic mutations in TP53. The dotted bright pink line shows a cut-off of 70% FGA, above which (to the right in the graph) cases become enriched for TP53 oncogenic mutations, consistent with high measurements of FGA. This 70% FGA cut-off was used as the basis for ground truth for training and evaluating an AI model to recognize CIN defined by 70% or greater FGA on WSI of H&E-stained breast cancers. Panel B shows the normalized distribution of breast cancers tested based on the percentage of LOH in those cases. Blue bars represent cases with no biallelic loss of BRCA1 or BRCA2, while pink bars are cases with biallelic loss of BRCA1 and/or BRCA2. The dotted bright pink line shows a cut-off of 20% LOH, above which (to the right in the graph) cases become enriched for biallelic loss of the homologous recombination repair associated BRCA genes, consistent with homologous repair deficiency presenting as LOH in association with biallelic losses of these genes. Thus, this 20% LOH cut-off was used as the basis for ground truth for training and evaluating an AI model to recognize CIN defined by LOH on WSI of H&E-stained breast cancers.

Conclusions:

- Measures of CIN yield phenotypic features that can be robustly identified by AI analysis of H&E WSI in primary and metastatic BCs, with performance confirmed on an independent validation set.
- Subtype stratification of breast cancers found preservation of AUC results across clinically relevant breast cancer subtypes.
- Metastatic sample trained models performed better on the validation cohort across all measures of CIN.
- This study provides the basis for development of AI-based tools to detect CIN not only in breast cancer but across cancer types, and a means to test CIN broadly within clinical trials or the clinic with future development and study.

Paige®