An Artificial Intelligence-Based Predictor of *CDH1* Biallelic Mutations and Invasive Lobular Carcinoma

Jorge S. Reis-Filho[1], Fresia Pareja[1], Fatemeh Derakhshan[1], David N. Brown[2], Jillian Sue[3], Pier Selenica[1], Yi Kan Wang[3], Arnaud Da Cruz Paula[1], Sheena Eb[3], Monami Banerjee[3], Zahra Ebrahimzadeh[3], Manuel Isava[3], Matthew Lee[3], Ran Godrich[3], Adam Casson[3], Ruben Padron[3], George Shaikovski[3], Alexander van Eck[3], Antonio Marra[1], Higinio Dopeso[1], Hannah Y Wen[1], Edi Brogi[1], Matthew Hanna[1], Chris Kanan[3], Jeremy D. Kunz[3], Felipe C. Geyer[3], Carla Lebowitz[3], David S. Klimstra[1], Leo Grady[3], Thomas J. Fuchs[3]

San Antonio Breast Cancer Symposium - December 7-10, 2021

INTRODUCTION

Invasive lobular carcinoma (ILC), the most frequent special histologic subtype of breast cancer (BC), is identifiable by pathologic assessment given its distinctive discohesive growth pattern, largely caused by *CDH1* inactivation. Compared to common forms of BC, ILCs display lower responses to chemotherapy (1) and selective estrogen receptor modulators (2). The low interobserver agreement for the diagnosis of ILC, however, renders the inclusion of histologic subtyping in therapeutic decision-making challenging.

CDH1 encodes for E-cadherin, a protein that mediates homophylic-homotypic adhesion in epithelial cells. In breast cancer, *CDH1* bi-allelic genetic alterations are almost restricted to ILCs.

Artificial intelligence (AI)-based algorithms hold promise for improving pathologic diagnosis; their performance, however, depends on the ground truth labeling used.

AIM

Here, we seek to develop an AI-based methodology for detection of ILC using '*CDH1* biallelic mutations' (i.e., mutation + loss-of-heterozygosity of the wild-type allele or two pathogenic somatic mutations) as ground truth, reasoning that in BC, >95% of *CDH1* bi-allelic inactivation is found in ILCs.

MATERIALS AND METHODS

- We developed a convolutional neural network system to detect *CDH1* biallelic genetic inactivation (AI-CDH1) using whole slide images of 1,100 primary BCs with available MSK-IMPACT targeted sequencing data.
- The model was trained using a 10-fold cross-validation method to detect *CDH1* biallelic mutations.
- We evaluated the performance of the AI-CDH1 classifier to predict the lobular phenotype and *CDH1* status using original and revised labels, following a histopathologic re-review of the histologic type and *CDH1* status curation. The latter was conducted by incorporating information on biallelic *CDH1* inactivation beyond *CDH1* mutations (homozygous deletions, deleterious structural rearrangements, and loss-of-heterozygosity and gene promoter methylation).

Table 1. Training and evaluation sets used by the AI-based model.

Set	Positive samples/ fold	Negative samples/ fold
Training set	85.2 (SD=2.6)	562.8 (SD=10.5)
Evaluation set	14.2 (SD=2.0)	93.8 (SD=9.1)

RESULTS

Detection of CDH1 mutations based on AI analysis of H&E

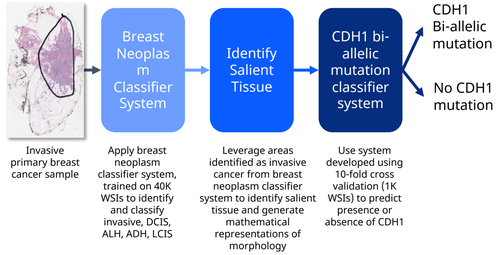
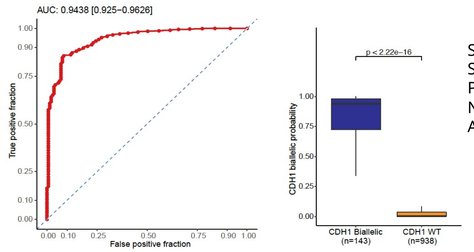


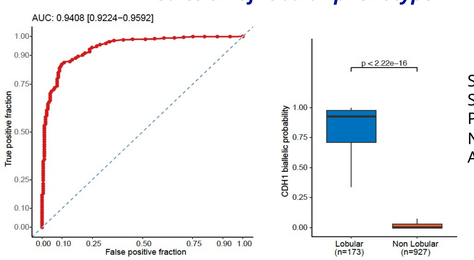
Figure 1. Schematic representation of AI-based algorithm for detection of *CDH1* biallelic mutations.

Prediction of biallelic CDH1 mutations



Sensitivity: 91.6%
Specificity: 85.9%
PPV: 49.8%
NPV: 98.5%
Accuracy: 86.7%

Prediction of lobular phenotype



Sensitivity: 89.0%
Specificity: 86.7%
PPV: 55.6%
NPV: 97.7%
Accuracy: 87.1%

Figure 2. Performance of the AI-based predictor for the detection of biallelic *CDH1* mutation and a lobular phenotype. Receiver operating characteristic plots (*left*) and box plots (*right*) of **A**) prediction of bi-allelic *CDH1* mutation status and **B**) of lobular phenotype.

Histologic features versus model prediction

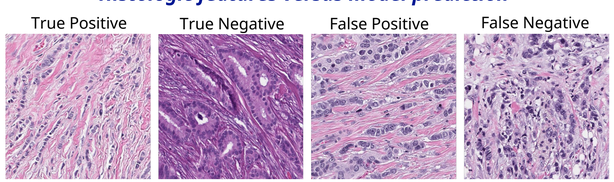


Figure 3. Representative micrographs of **A**) true positive classic ILC; **B**) true negative invasive ductal carcinoma (IDC) ; **C**) false positive IDC; **D**) false negative pleomorphic ILC. False positive tumors often displayed architectural lobular features but lacked the cytological features diagnostic of lobular neoplasms; whilst false negative tumors often displayed pleomorphic features.

Review of histologic and molecular features

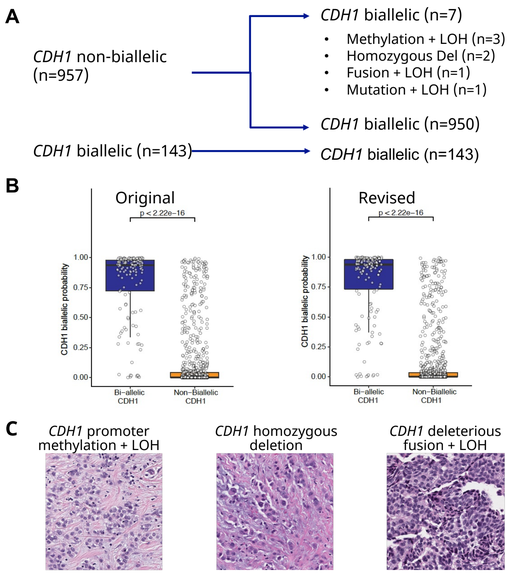


Figure 4. *CDH1* biallelic inactivation following incorporation of different molecular mechanisms A) Original and revised *CDH1* biallelic inactivation and **B**) prediction by AI-based algorithm **C**) Representative micrographs of cases with *CDH1* biallelic inactivation through different molecular mechanisms

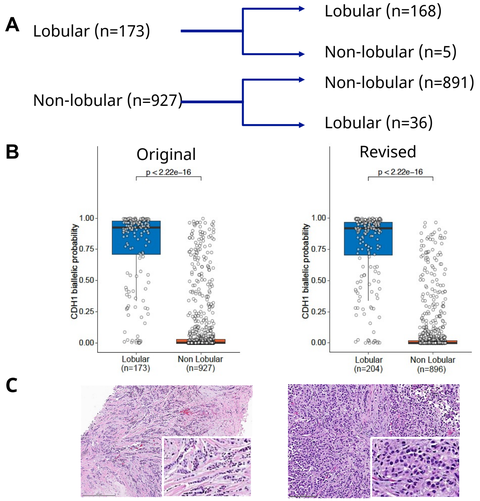
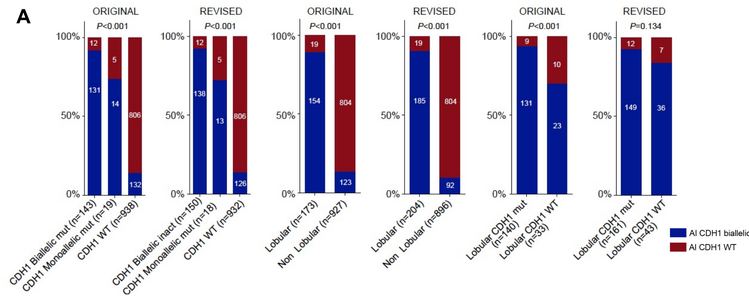
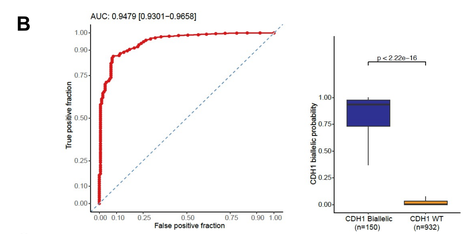


Figure 5. Reclassification of histologic subtype upon pathology re-review A) Reclassification of lobular and non-lobular phenotype. **B**) Prediction of original (left) and revised (right) lobular phenotype. **C**) Representative micrographs of cases originally labeled 'non-lobular' and reclassified as 'lobular'.

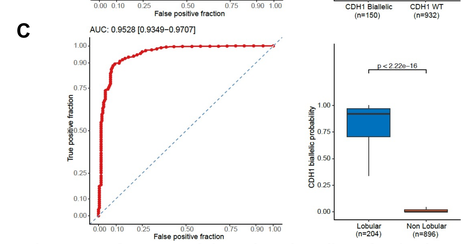
Performance of the AI-based CDH1 mutation predictor upon CDH1 status and diagnosis curation





Prediction of CDH1 biallelic inactivation

Sensitivity 92.0%
Specificity 86.5%
PPV 52.3%
NPV 98.5%
Accuracy 87.2%



Prediction of lobular phenotype upon re-review

Sensitivity 90.7%
Specificity 89.7%
PPV 66.8%
NPV 97.7%
Accuracy 89.9%

Figure 6. Performance of the AI-based predictor for the detection of a *CDH1* biallelic inactivation and lobular phenotype upon *CDH1* status and diagnosis curation. (B-C) ROC plots (*left*) and box plots (*right*) of **B**) prediction of bi-allelic *CDH1* biallelic inactivation and **C**) of lobular phenotype.

CONCLUSIONS

By training a machine learning system to detect '*CDH1* biallelic mutations', as ground truth rather than histologic diagnosis of lobular carcinoma, which might be confounded by human subjectivity, we developed an AI-based system that can detect ILCs accurately, providing a new paradigm for the use of orthogonal 'ground truth' in the development of AI-based cancer classification systems.

REFERENCES

1. Marmor et al. Cancer 2017
2. Metzger Filho et al. JCO 2015

ACKNOWLEDGMENTS

J.S.R.-F. is funded in part by the Breast Cancer Research Foundation. JSRF and FP are funded in part by a National Institutes of Health (NIH)/National Cancer Institute (NCI) P50CA24774901 grant and. FP is funded in part by an NIH/NCI K12 CA184746 grant. Research reported here was partly funded by a Cancer Center Support Grant of the NIH/NCI (grant P30CA008748)