**A Computational Pathology Collagen Signature predictive of tamoxifen benefit in ductal carcinoma in situ: Results from a cohort within the UK/ANZ DCIS randomized trial**

**Authors**: Arpit Aggarwal, Sunil Badve, Jack Cuzick, Anant Madabhushi, Mangesh A. Thorat

**Background**

Disorder in collagen fiber architecture, characterized by alterations in collagen organization and distribution, plays a critical role in tumor progression. In this study, we developed a collagen-based computational pathology biomarker and evaluated its association with the benefits of adjuvant treatment in a clinical trial cohort.

**Methods**

The study utilized H&E-stained Whole Slide Images (WSIs) of DCIS from the UK/ANZ DCIS randomized controlled trial, which included 4 treatment allocations: no adjuvant treatment, radiotherapy, tamoxifen, or radiotherapy+tamoxifen. Two-thirds of ipsilateral breast events (IBE) from Tamoxifen treatment group were used for training, with each case (patient with IBE; n=96) matched to two controls by age (+/-7 years) and treatment (n=192). For the development of the collagen-tamoxifen score (CTS), 102 samples (IBE=35) out of 270 formed the training set (TrS), while 140 samples (IBE=20) formed the validation set (VaS). Machine learning models were employed to extract a series of features relating to collagen arrangement from routine H&E images. A logistic regression model was trained using these features and generated a continuous score (CTS). A 66th percentile risk score threshold was applied on CTS in TrS to stratify patients into low or high-risk groups in TrS and VaS.

**Results**

In TrS+VaS, 194 of 713 (27%) patients were classified as CTS-high. Among 481 patients with ER (clonal) status data, a weak negative correlation (spearman rho -0.19, p<0.0001) was seen between ER (clonal) and CTS. Specifically, 61% (109/178) of ER-negative DCIS were CTS-low, while 21% (63/303) of ER-positive DCIS were CTS-high. In CTS-high DCIS was associated with greater than 3-fold risk of IBE as compared with CTS-low DCIS in both TrS [HR = 4.54; 95% CI, 2.27-9.06, p<0.0001] and VaS [HR = 3.46; 95%CI, 1.41-8.48, p=0.0067]. In multivariate analyses in VaS (n=114, IBE=14), CTS (high vs. low) was the only independent predictor of IBE (HR=4.75; 95%CI, 1.31-17.26, p=0.018). In VaS subset with known ER-status (n=63, IBE=12), CTS was the only variable with borderline significant p value (0.08), ER (P=0.86), inclusion of CTS in the model made the model more informative, albeit with borderline significance [Δχ2 (1d.f.) 3.46; p=0.06].

**Conclusions**

Disorder in collagen architecture is associated with tamoxifen resistance in DCIS patients. Our computational pathology based collagen-tamoxifen score has a role, independent of ER-status, in predicting tamoxifen benefit in DCIS.