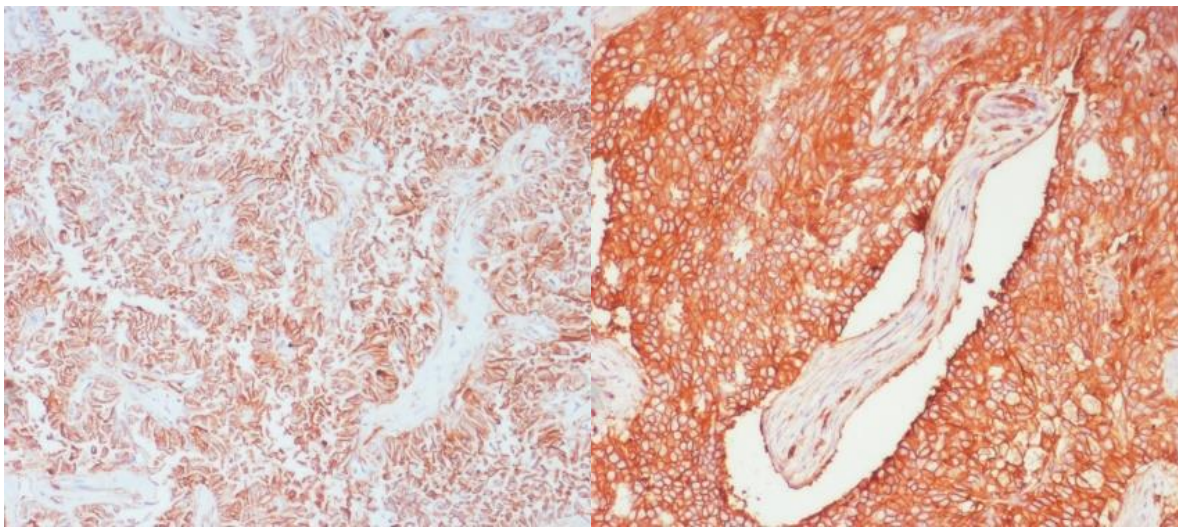


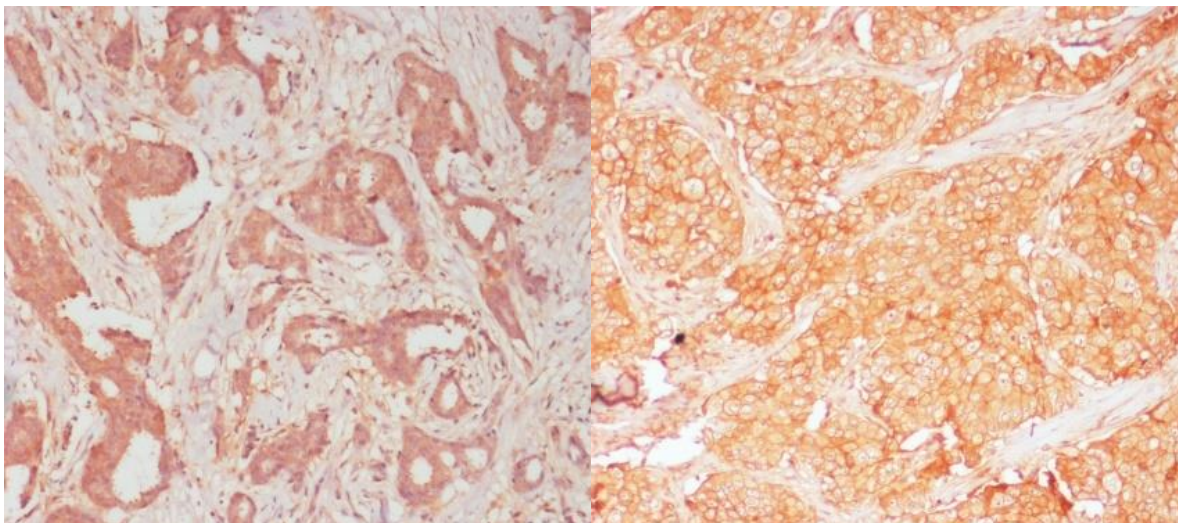
Supplementary Methods

Antibodies used for IHC-based testing of the biomarkers evaluated during classifier development are as follows: anti-CD44 (Thermo Scientific, Cat. # MA513890), anti-ABCC4 (Abnova, Cat. # H00010257-M03), anti-ABCC11 (Novus Biologicals, Cat. # NBP1-82623), anti-N-Cadherin (AbCam, Cat. # ab98952), anti-Cadherin-pan (Spring Biosciences, Cat. # E2364), anti-E-Cadherin (Abcam, Cat. # ab40772), anti-MAGE-A9 (Abcam, Cat. # ab49074), anti-MAGE-A11 (Abcam, Cat. # ab96236) anti-ABCG2 (Novus Biologicals, Cat. # NB110-93511), anti-APC (Labvision, Cat. # RB-9276-P1), anti-HIF2 α (Novus Biologicals, Cat. # NB100-132SS), anti-THY1 (Abnova, Cat. # PAB3877), anti-total β -catenin (Novus Biologicals, Cat. # NBP1-03615), anti-Active- β -catenin (Millipore, Cat. # 05-665), anti-FOXA1 (Abcam, Cat. # ab99892), anti-p-Nrf2 (Abcam, Cat. # ab76026), anti-CxCR4 (Abcam, Cat. # ab124824)

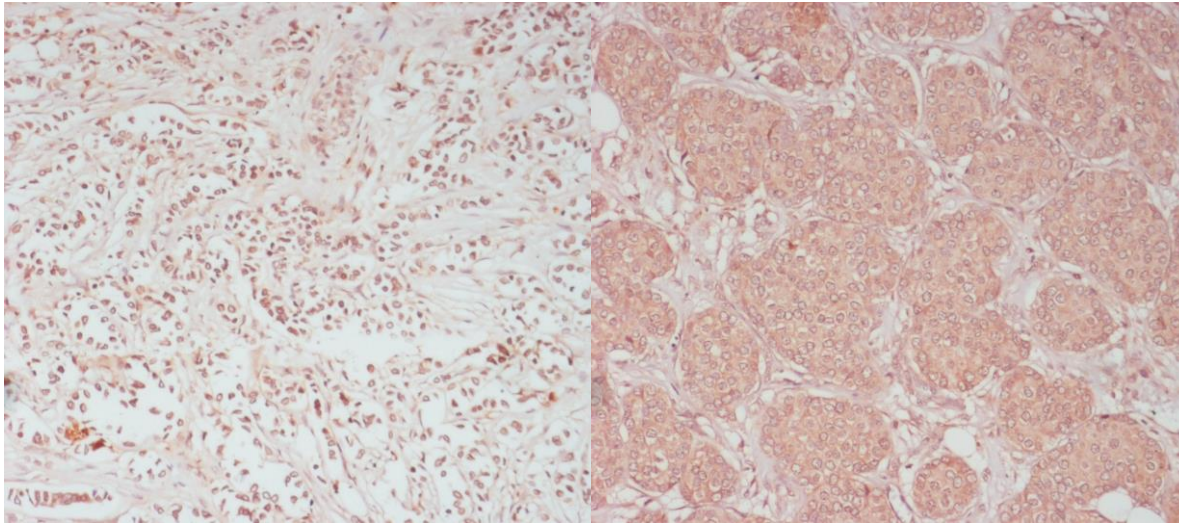
Supplementary Figure S1



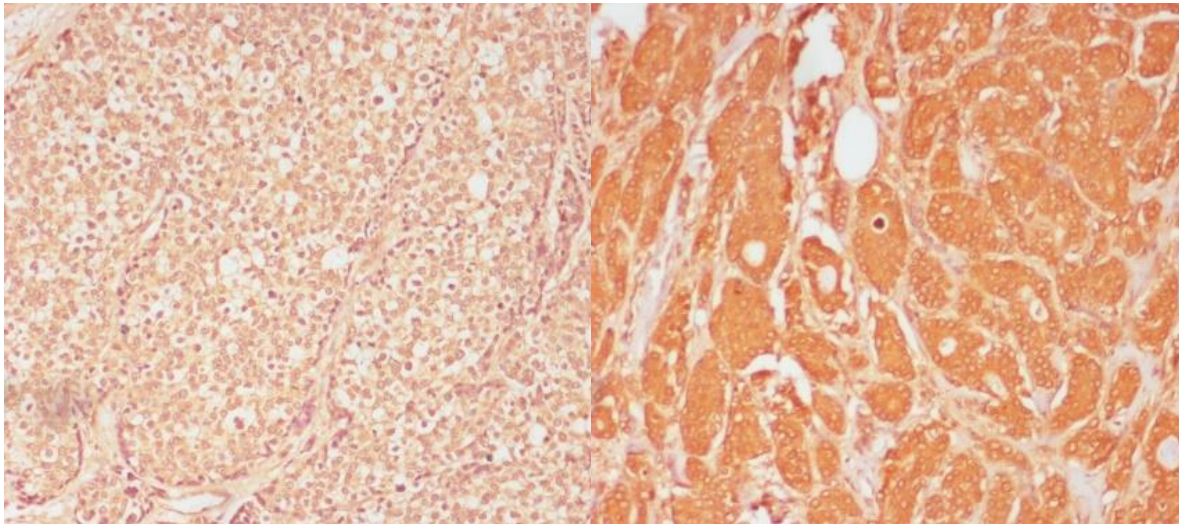
Part (a): CD44 low (left panel) and high (right panel) % staining



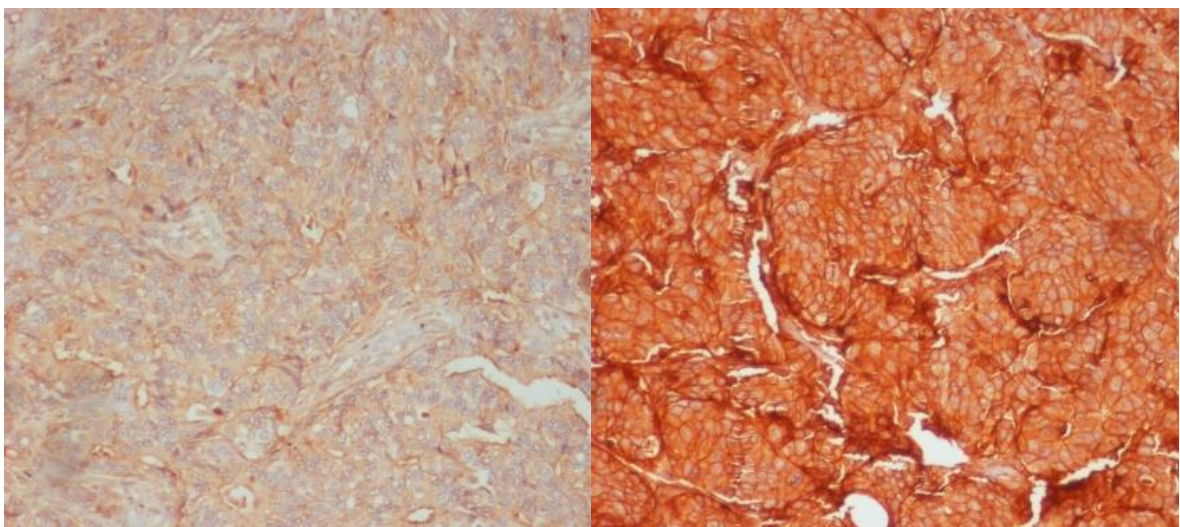
Part (b): ABCC4 low (left panel) and high (right panel) % staining



Part (c): ABCC11 low (left panel) and high (right panel) % staining



Part (d): N-Cadherin moderate (left panel) and high (right panel) % staining

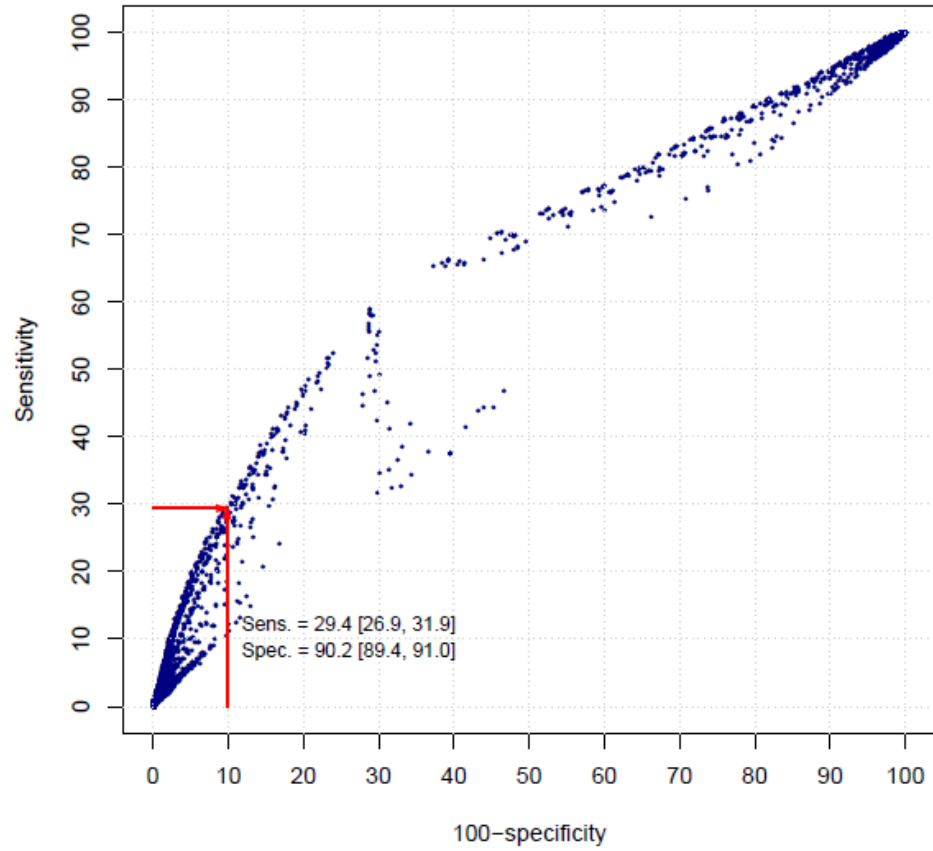


Part (e): Pan-Cadherin low (left panel) and high (right panel) % staining

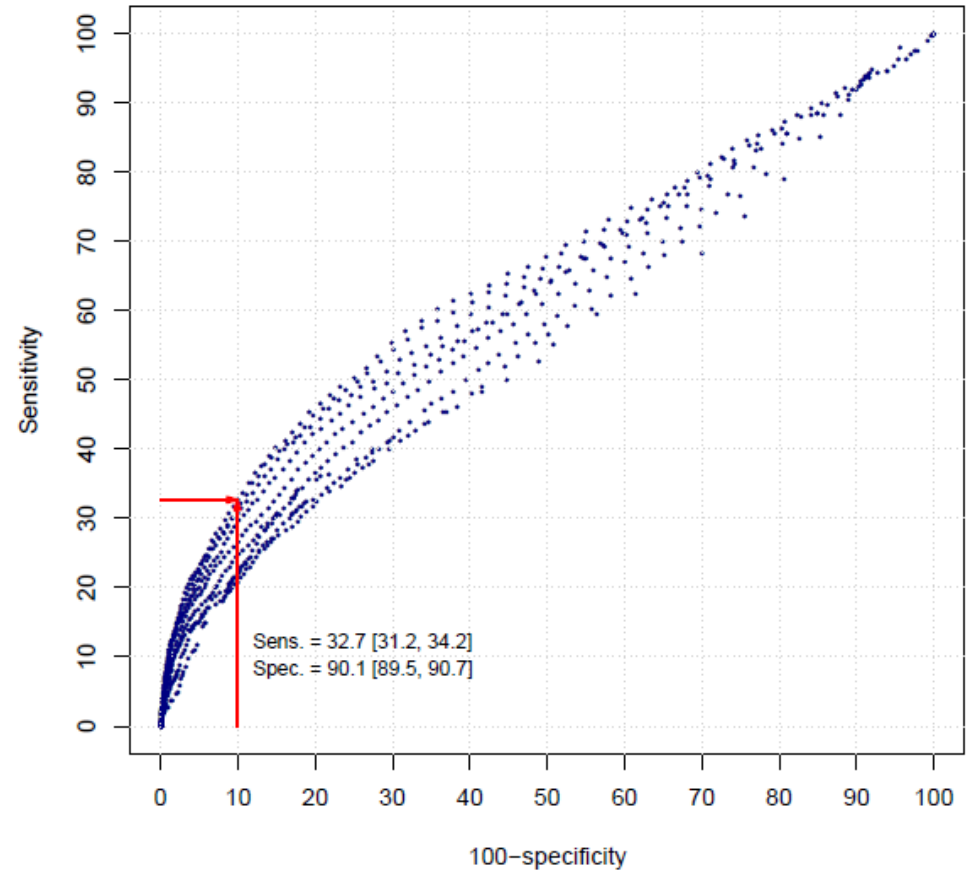
Figure S1: Representative images of IHC staining of CanAssist-Breast biomarkers

Supplementary Figure S2

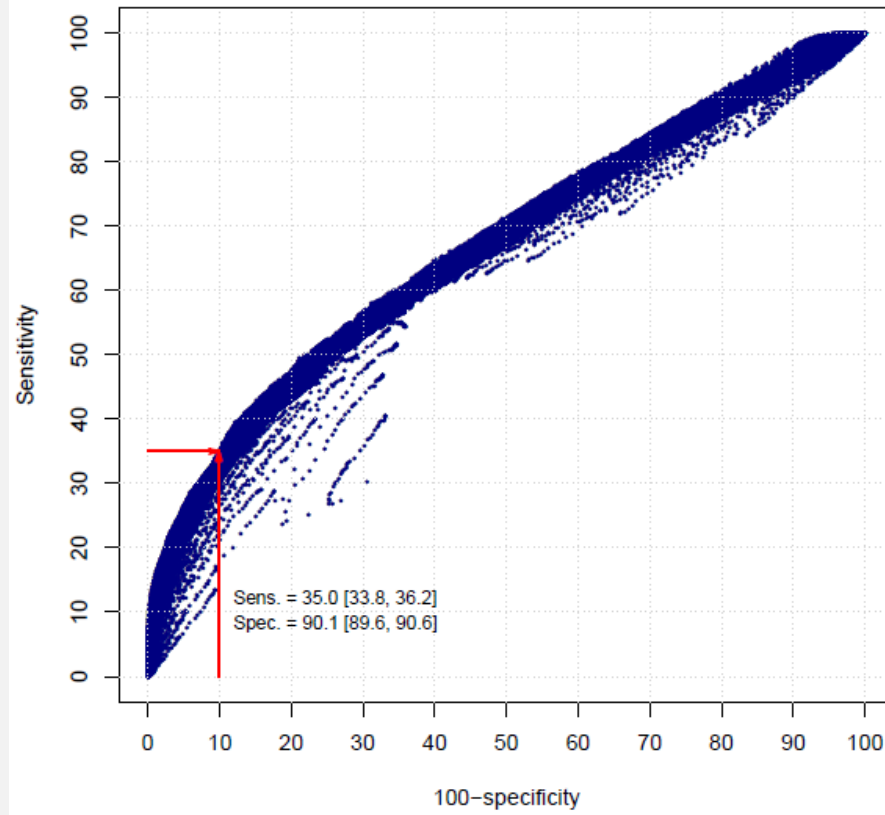
(a)



(b)



(c)



(d)

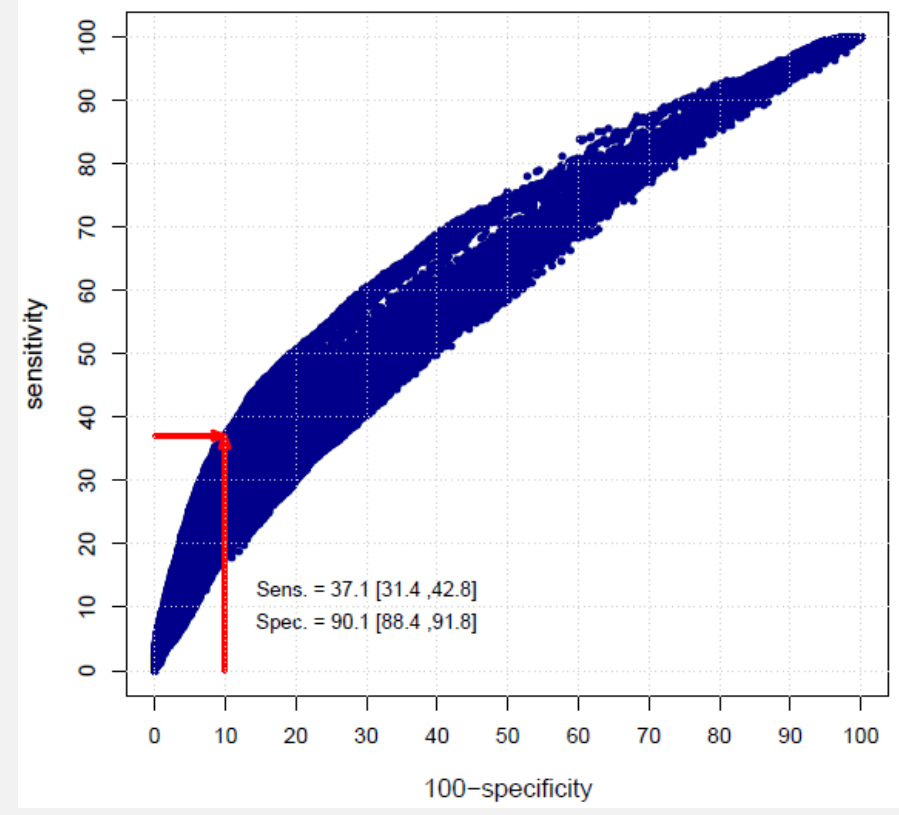


Figure S2: ROG analysis of different machine learning-based techniques: (a) linear SVM ROG (b) RF ROG (c) ESL ROG (c) MLP ROG and corresponding sensitivities (Sens.) and specificities (Spec.)