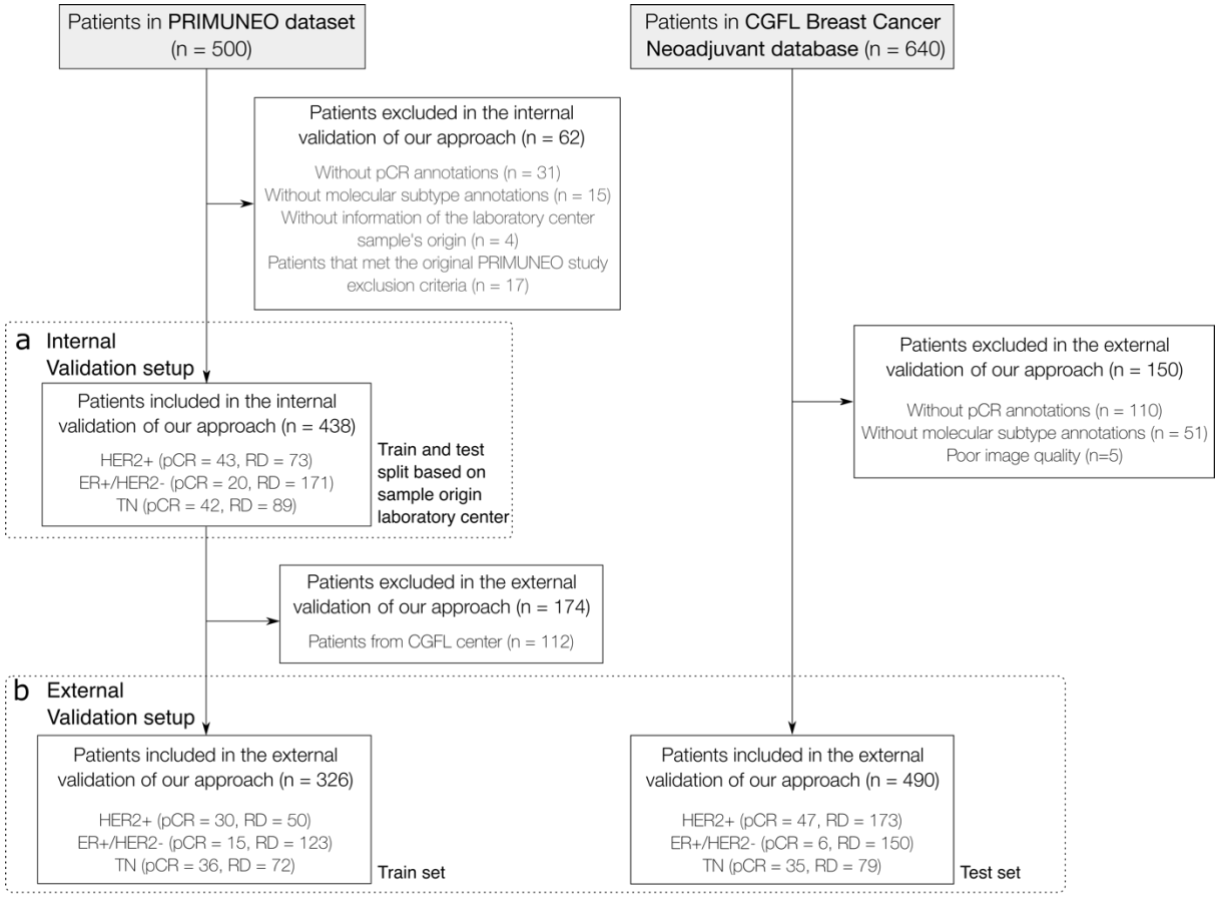
Breast-NEOprAIdict: a deep learning solution for predicting pathological complete response on biopsies of breast cancer patients treated with neoadjuvant chemotherapy.

Natalia Fernanda Valderrama[1,‡], Louis-Oscar Morel[1,‡], Daniel Tshokola Mweze[1], Valentin Derangère[2,3,4], Isabelle Desmoulins[5], Didier Mayeur[5], Courèche Kaderbhai[5], Silvia Ilie[5], Audrey Hennequin[5], Nicolas Roussot[5], Antony Bergeron[2], Françoise Beltjens[2], Carlo Pescia[6], Henri-Philippe Morel[7], Charles Coutant[8], Laurent Arnould[2], Nathan Vinçon[1,*], Sylvain Ladoire[2,3,4,5,9,*]

[1] Ummon HealthTech SAS, Dijon, 21000, France
[2] Georges Francois Leclerc Cancer Center, Department of Biology and Pathology of tumors, Dijon, 21000, France
[3] Georges François Leclerc Cancer Center, Platform of Transfer in Biological Oncology, Dijon, 21000, France
[4] University of Burgundy-Franche Comté, Dijon, 21000, France
[5] Georges Francois Leclerc Cancer Centre, Department of Medical Oncology, Dijon, 21000, France
[6] University of Milan, Department of Molecular Medicine, Milan, 20122, Italy
[7] Technipath Lyon, Dommartin, 69380, France
[8] Georges Francois Leclerc Cancer Centre, Department of Surgical Oncology, Dijon, 21000, France
[9] INSERM U1231, Dijon, 21000, France
[*] nathanvincon@ummon.health, SLadoire@cgfl.fr
[*] Nathan Vinçon and Sylvain Ladoire contributed equally to this work.
[‡] Natalia Fernanda Valderrama and Louis-Oscar Morel contributed equally to this work.



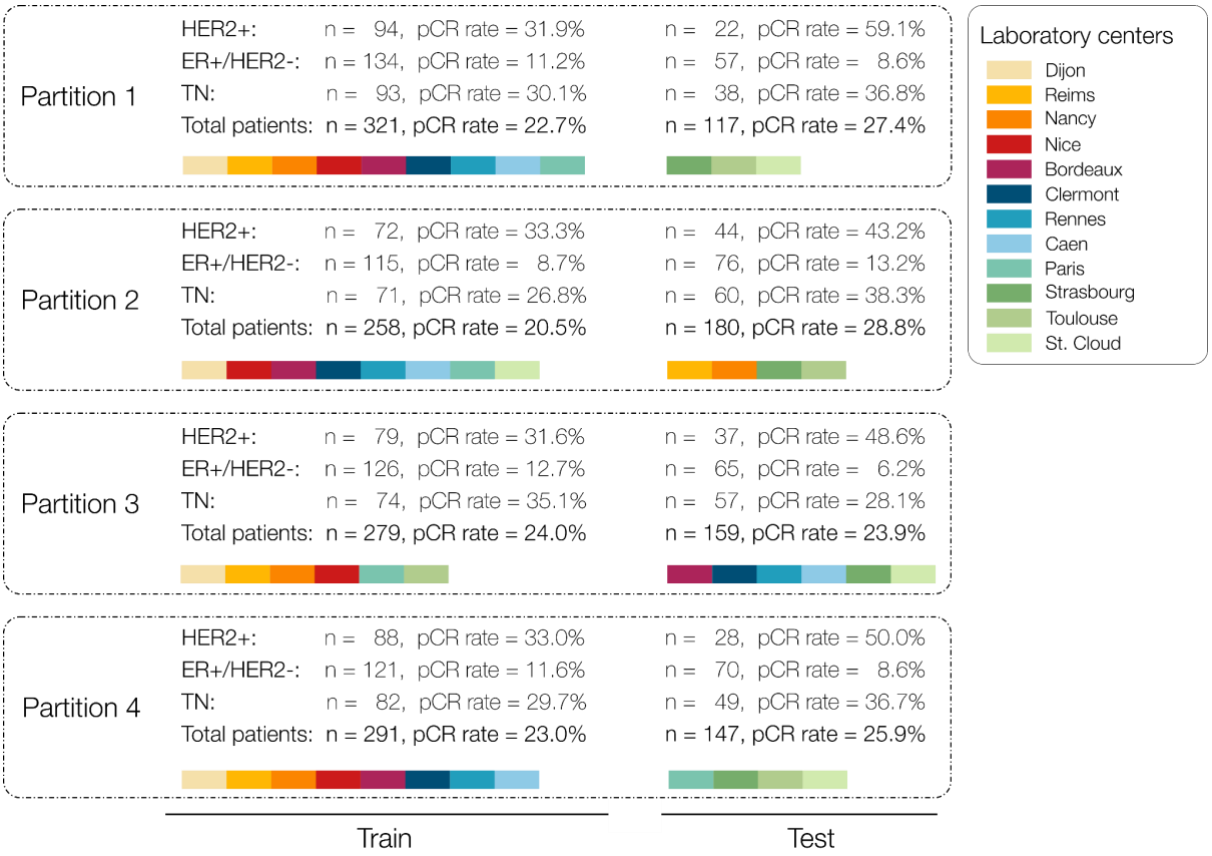
Supplementary Figure S1. Workflow for patient selection in the PRIMUNEO and CGFL Breast Cancer Neoadjuvant databases. The diagram illustrates the patient selection process in the PRIMUNEO and CGFL Breast Cancer Neoadjuvant databases. a) The chart specifies the included patients in PRIMUNEO to compose the internal validation setup. b) The training set for the external validation setup is then derived by excluding patients from the CGFL center in Dijon in the PRIMUNEO cohort. The test set is composed by the CGFL Breast Cancer Neoadjuvant database. The chart considers the number of patients with pCR or RD categorized by the tumor's molecular subtype.

Supplementary Table S1. Description of the PRIMUNEO Database, including the number of patients exhibiting either pathological complete response (pCR) or residual disease (RD), categorized by the tumor's molecular subtype within each center.

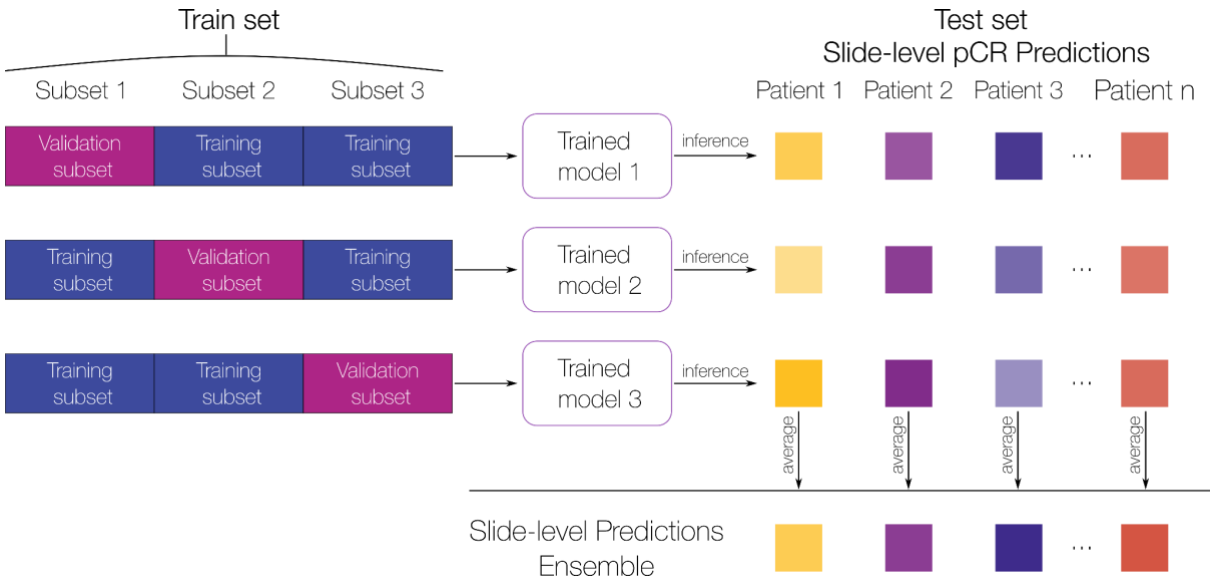
Center	Total patients	HER2+		ER+/HER2-		TN	
		pCR	RD	pCR	RD	pCR	RD
DIJON	112	13	23	5	48	6	17
STRASBOURG	69	10	3	4	24	9	19
REIMS	48	5	11	3	17	4	8
RENNES	23	2	4	0	6	2	9
NANCY	38	3	8	2	10	7	8
CAEN	34	3	6	0	14	3	8
NICE	26	2	4	4	8	2	6
PARIS	30	1	5	1	12	4	7
TOULOUSE	25	1	3	1	15	3	2
ST CLOUD	23	2	3	0	13	2	3
BORDEAUX	9	1	2	0	4	0	2
CLERMONT	1	0	1	0	0	0	0
TOTAL	**438**	**43**	**73**	**20**	**171**	**42**	**89**

Supplementary Table S2. CGFL breast cancer neoadjuvant database description according to tumor molecular subtype.

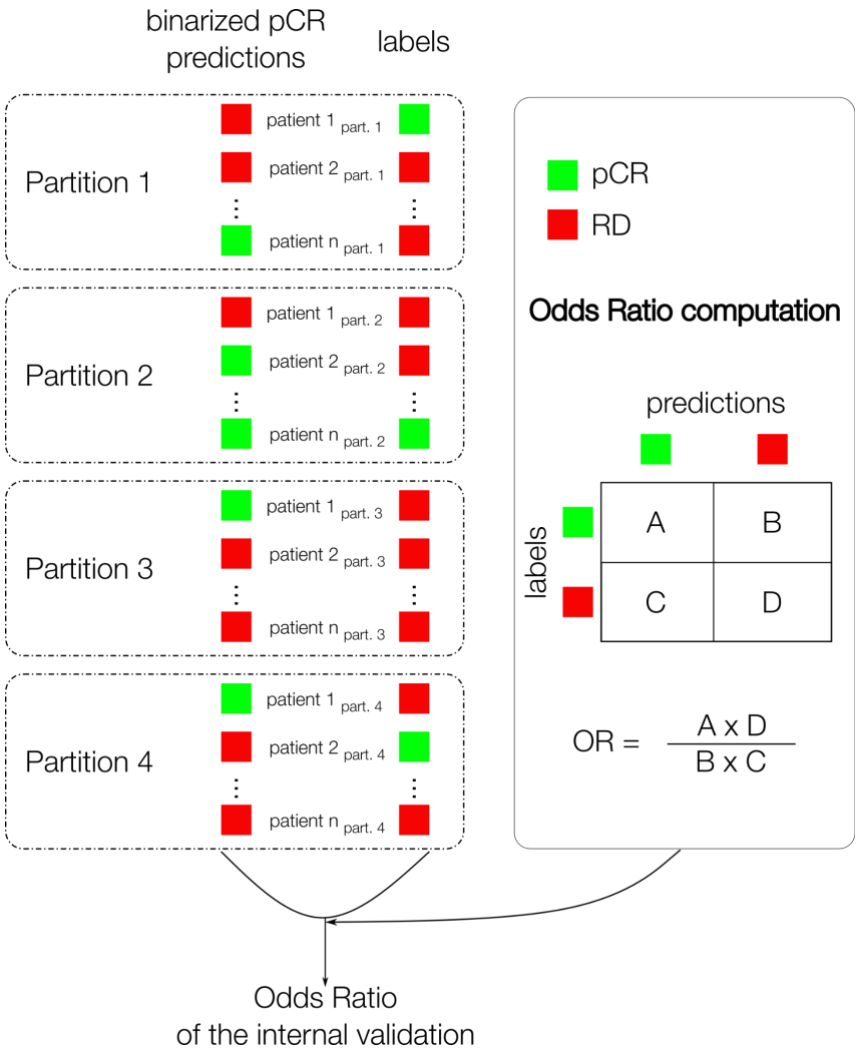
Cancer subtype	Number of patients	pCR	RD	Prevalence
HER2+	220	47	173	0.214
ER+/HER2-	156	6	150	0.038
TN	114	35	79	0.307
Total	**490**	**88**	**402**	**0.180**



Supplementary Figure S2. The chart shows the specifications of the laboratory, total number of patients and pCR rate (%) in each partition for the training and test subsets in the internal validation study. The distribution of the patients is also shown by molecular subtype. Best viewed in color.



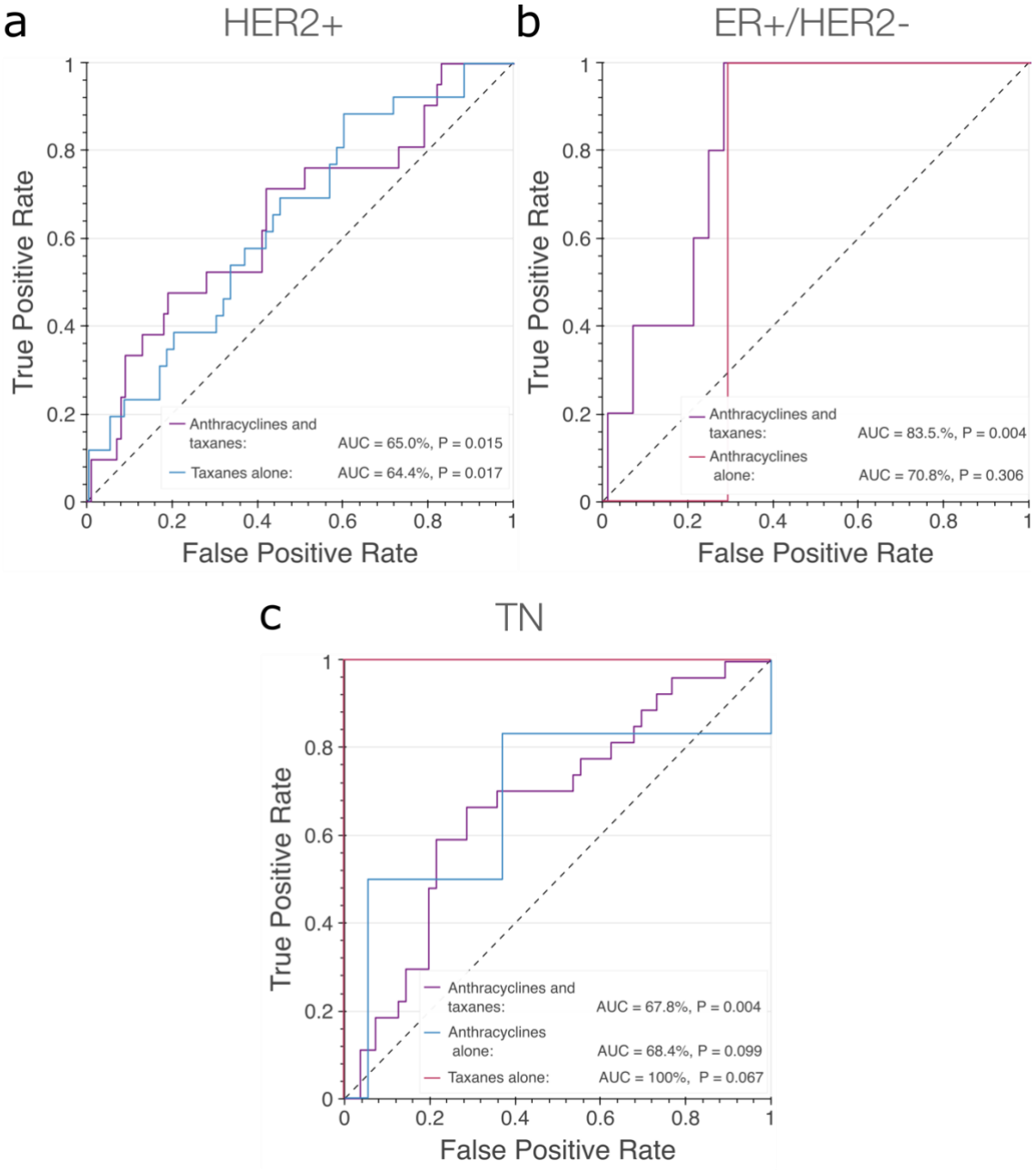
Supplementary Figure S3. Three-Fold Cross-Validation training strategy.



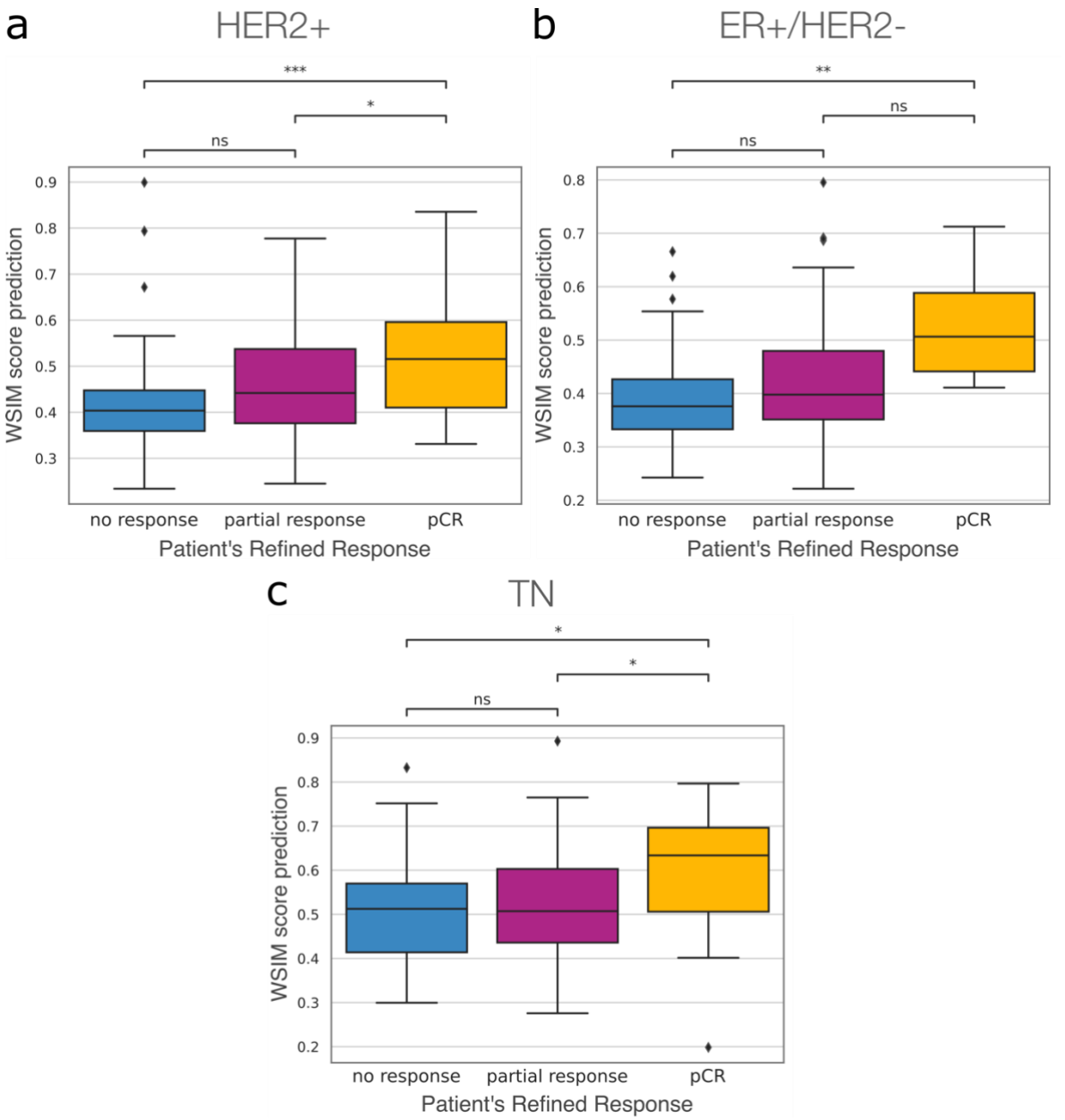
Supplementary Figure S4. Prediction ensemble for computation of OR metrics.

Supplementary Table S3. Breast-NEOprAIdict performance is measured with sensitivity (recall), PPV positive predictive value (precision), specificity, and NPV negative predictive value on external validation for HER2+, ER+/HER2- and TN breast cancer molecular subtypes.

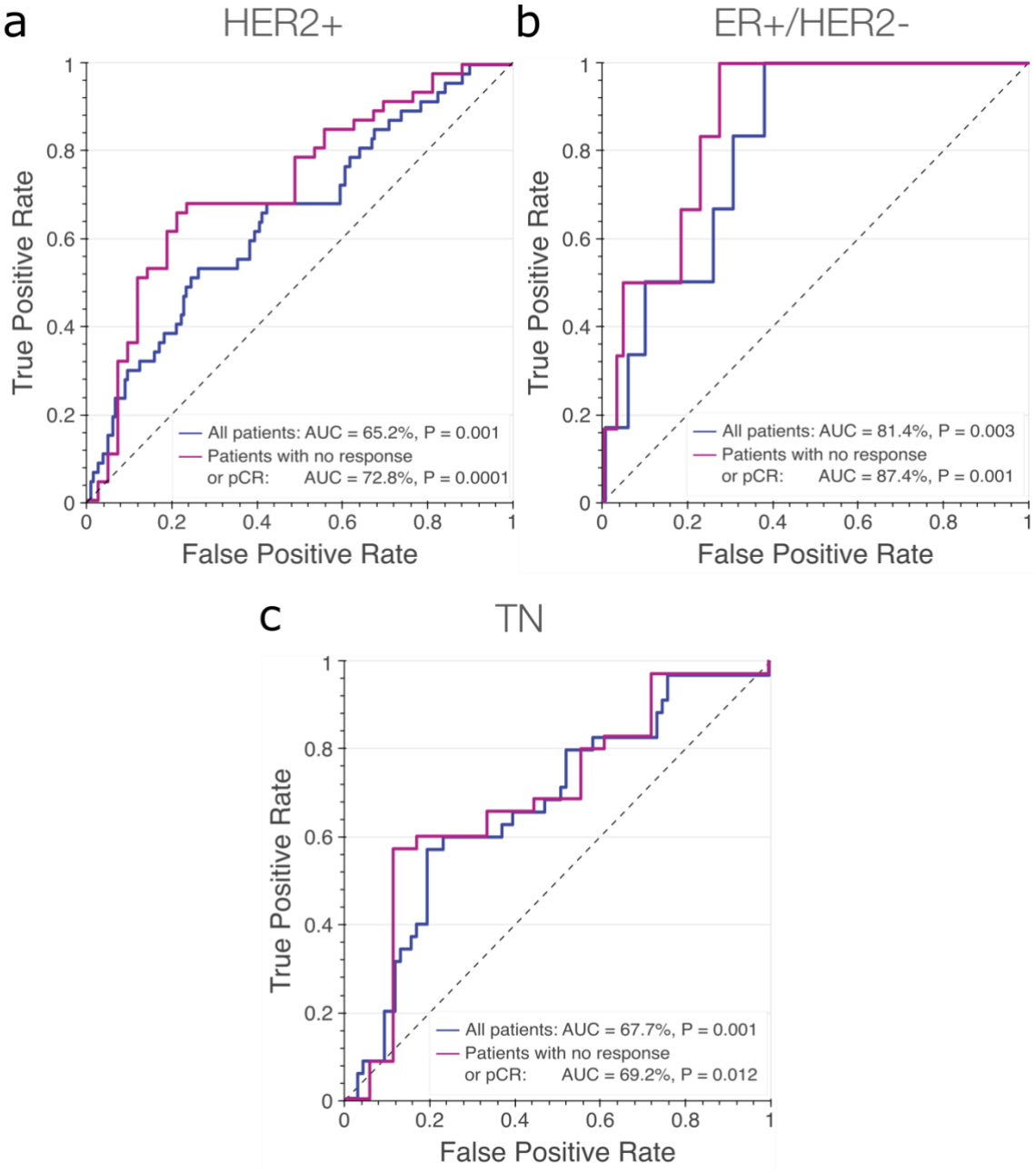
Metrics	Sensitivity	Specificity	PPV	NPV
HER2+	0.872	0.283	0.248	0.891
ER+/HER2-	1	0.567	0.084	1
TN	0.800	0.430	0.383	0.829



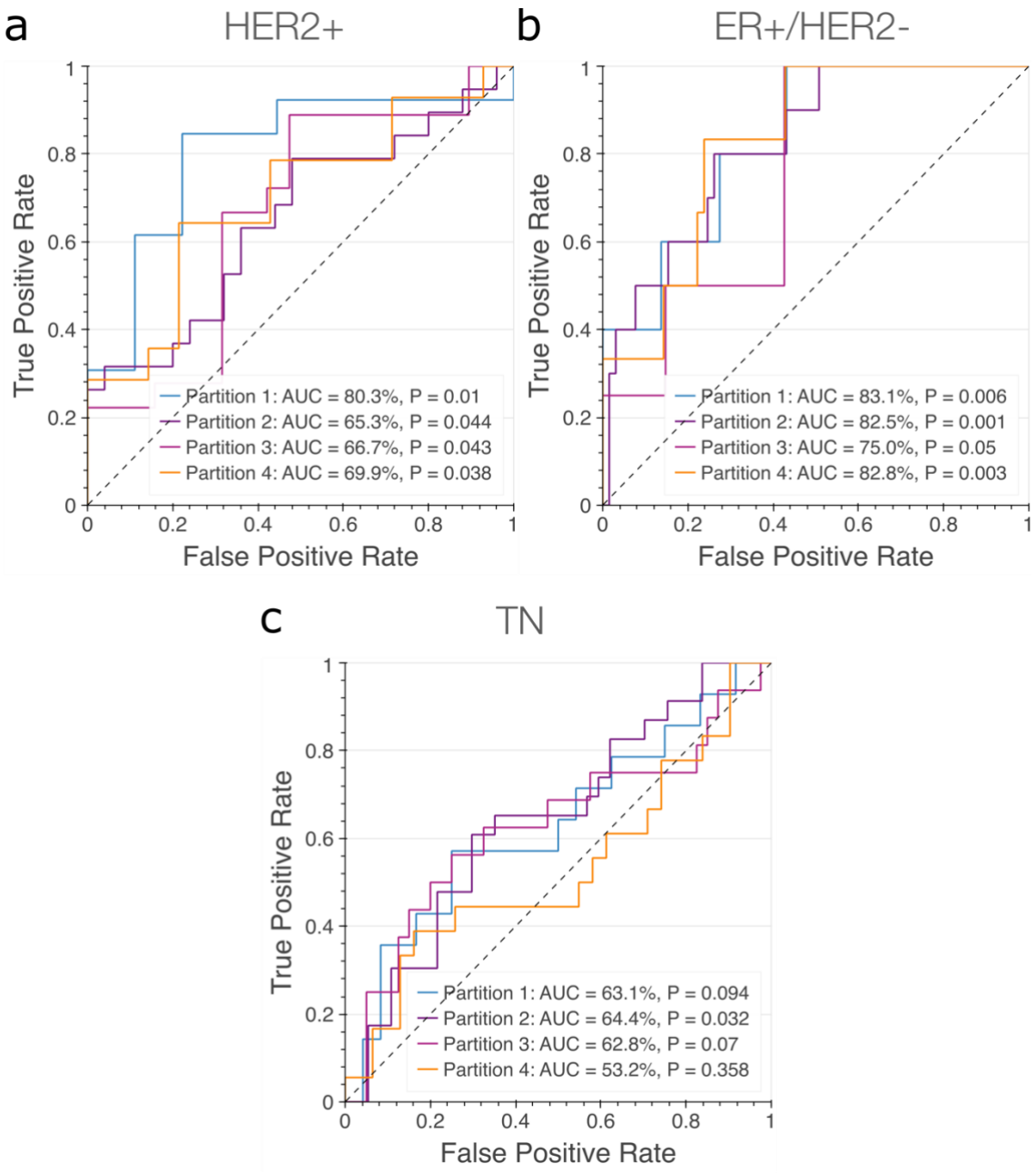
Supplementary Figure S5. AUC in patients with A) HER2+, B) ER+/HER2- and C) TN tumors in the external validation study, stratifying patients by the received treatment (Taxanes, Anthracyclines, or both).



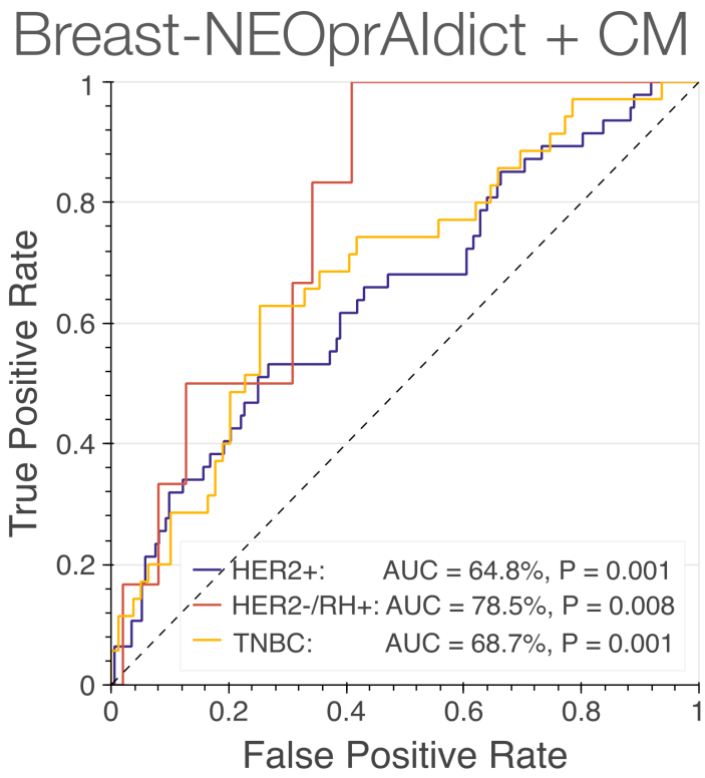
Supplementary Figure S6. Comparison of Breast-NEOprAIdict score prediction distribution within non-responder, partial responder, and complete responder patients with a) HER2+, b) ER+/HER2- and c) TN tumors. The one-sided Mann-Whitney-Wilcoxon test was used to compare the response categories. ns: 0.05 < p-value <= 1; *: 0.01 < p-value <= 0.05; **: 0.001 < p-value <= 0.01, ***: 0.0001 < p <= 0.001.



Supplementary Figure S7. Evaluation of Breast-NEOprAIdict pCR prediction performance in all patients compared to those who did not respond or exhibited complete response to Neoadjuvant Chemotherapy (excluding patients with partial response defined in the refined labels). We show the results for patients with a) HER2+, b) ER+/HER2- and c) TN tumors in external validation. Legends in the figures provide details on the Area Under the Curve (AUC) and associated p-values (P).



Supplementary Figure S8. AUC in patients with a) HER2+, b) ER+/HER2- and c) TN tumors in the internal validation study. The patients' pCR predictions are obtained by averaging Breast-NEOprAIdict and CM predictions.



Supplementary Figure S9. AUC in patients with HER2+, ER+/HER2-, and TN tumors in the external validation study. The patients' pCR predictions are obtained by averaging Breast-NEOprAIdict and CM predictions.

Supplementary Table S4. Ablation study on the ensemble of the EfficientNet B7-based model and the ViT-S/16-based model in the external validation. AUC in patients with HER2+, ER+/HER2-, and TN tumors is shown.

Methods	**EfficientNet B7-based model**		**ViT-S/16-based model**		**Breast-NEOprAIdict**	
	AUC	***P***	**AUC**	***P***	**AUC**	***P***
HER2+	0.632	0.003	0.599	0.019	0.652	0.001
ER+/HER2-	0.776	0.01	0.746	0.02	0.814	0.003
TN	0.661	0.003	0.647	0.006	0.677	0.001