Supplementary Figures and Tables

A

Tumor	TCGA n = 5155										
	BRCA	CRC	GBM	LIHC	LUAD	LUSC	OV	PAAD	PRAD	UCEC	SUM
Total n=	1058	580	420	364	536	497	520	177	488	515	5155
HRD-H	292	18	8	28	184	257	361	13	15	81	1247
HRD-L	766	562	412	336	352	240	159	164	473	434	3908
Missing data	43	84	188	16	76	46	430	4	97	48	1042
n=	1005	496	232	348	460	451	90	173	391	467	4113
HRD-H	281	15	6	26	162	232	57	13	10	68	870
HRD-L	724	481	226	322	298	219	33	160	381	399	3243

B

Tumor	CPTAC n = 573					
	GBM	LUAD	LSCC	PDA	UCEC	SUM
Total n=	99	111	109	153	101	573
HRD-H	0	15	33	4	3	55
HRD-L	99	96	76	149	98	518
Missing data	0	5	1	14	2	22
n=	-	106	108	139	99	452
HRD-H	-	14	33	4	3	54
HRD-L	-	92	75	135	96	497

C
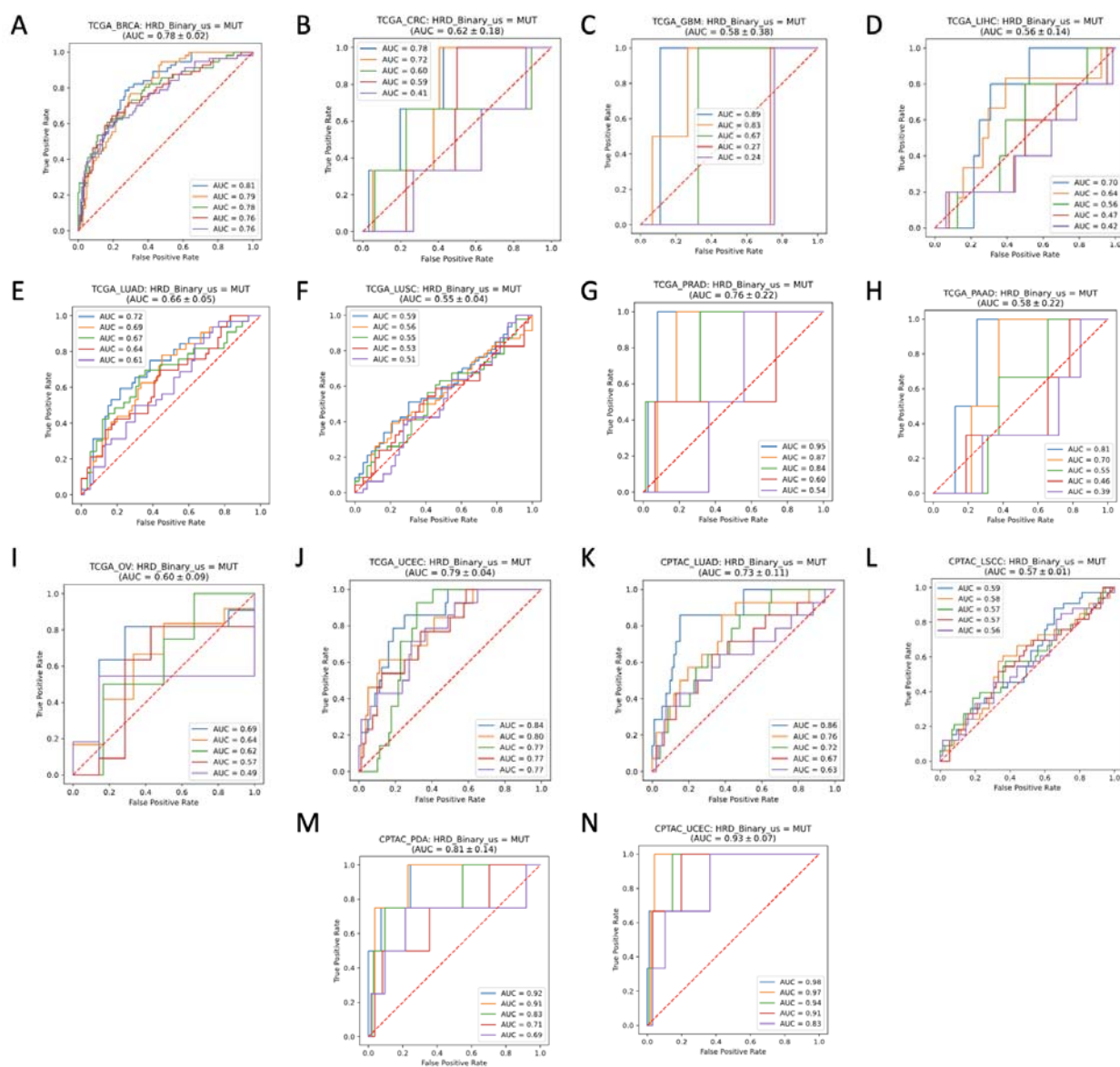

Supplementary Figure 1: Homologous recombination deficiency prevalences across the cohorts. (A) Overview of the total patient count (n=573) in the CPTAC cohort before merging the image data with the molecular data and afterward. (B) Overview of the total patient count (n=5,155) in the TCGA cohort before merging the image data with the molecular data and afterward. (C) Distribution of the homologous recombination deficiency high (HRD-H) and low (HRD-L) patient number among the different tumor types of The Cancer Genome Atlas (TCGA) and Clinical Proteomic Tumor Analysis

676 Consortium (CPTAC). Abbreviations: BRCA=breast invasive carcinoma; CRC=colorectal cancer;

677 GBM=glioblastoma; LIHC=liver cancer; LUAD=lung adenocarcinoma; LUSC/LSCC=lung squamous cell

678 carcinoma; OV=ovarian cancer; PAAD/PDA=pancreatic adenocarcinoma; PRAD=prostate

679 adenocarcinoma; UCEC=endometrial cancer

680

681

682 **Supplementary Figure 2: Receiving operating curve for the Internal Validation and tumor wise**

683 **external validation.** The Receiving operating curve (ROC) is shown for the five-fold internal cross-

684 validation experiment for each of the models in The Cancer Genome Atlas (TCGA) for the Homologous

685 recombination deficiency (HRD) binary score for (A) TCGA-BRCA, (B) TCGA-CRC, (C) TCGA-GBM,

686   (D) TCGA-LIHC, (E) TCGA-LUAD, (F) TCGA-LUSC, (G) TCGA-PAAD, (H) TCGA-PRAD, (I) TCGA-OV,

687   (J) TCGA-UCEC; Roc curves for the external validation on the Clinical Proteomic Tumor Analysis

688   Consortium (CPTAC) for each previously trained model for (K) CPTAC-LUAD, (L) CPTAC-LSCC, (M)

689   CPTAC-PDA, (N) CPTAC-UCEC. Abbreviations: BRCA=breast invasive carcinoma; CRC=colorectal

690   cancer; GBM=glioblastoma; LIHC=liver cancer; LUAD=lung adenocarcinoma; LUSC/LSCC=lung

691   squamous cell carcinoma; OV=ovarian cancer; PAAD/PDA=pancreatic adenocarcinoma;

692   PRAD=prostate adenocarcinoma; UCEC=endometrial cancer

693

Cross cancer external validation ROCs

694
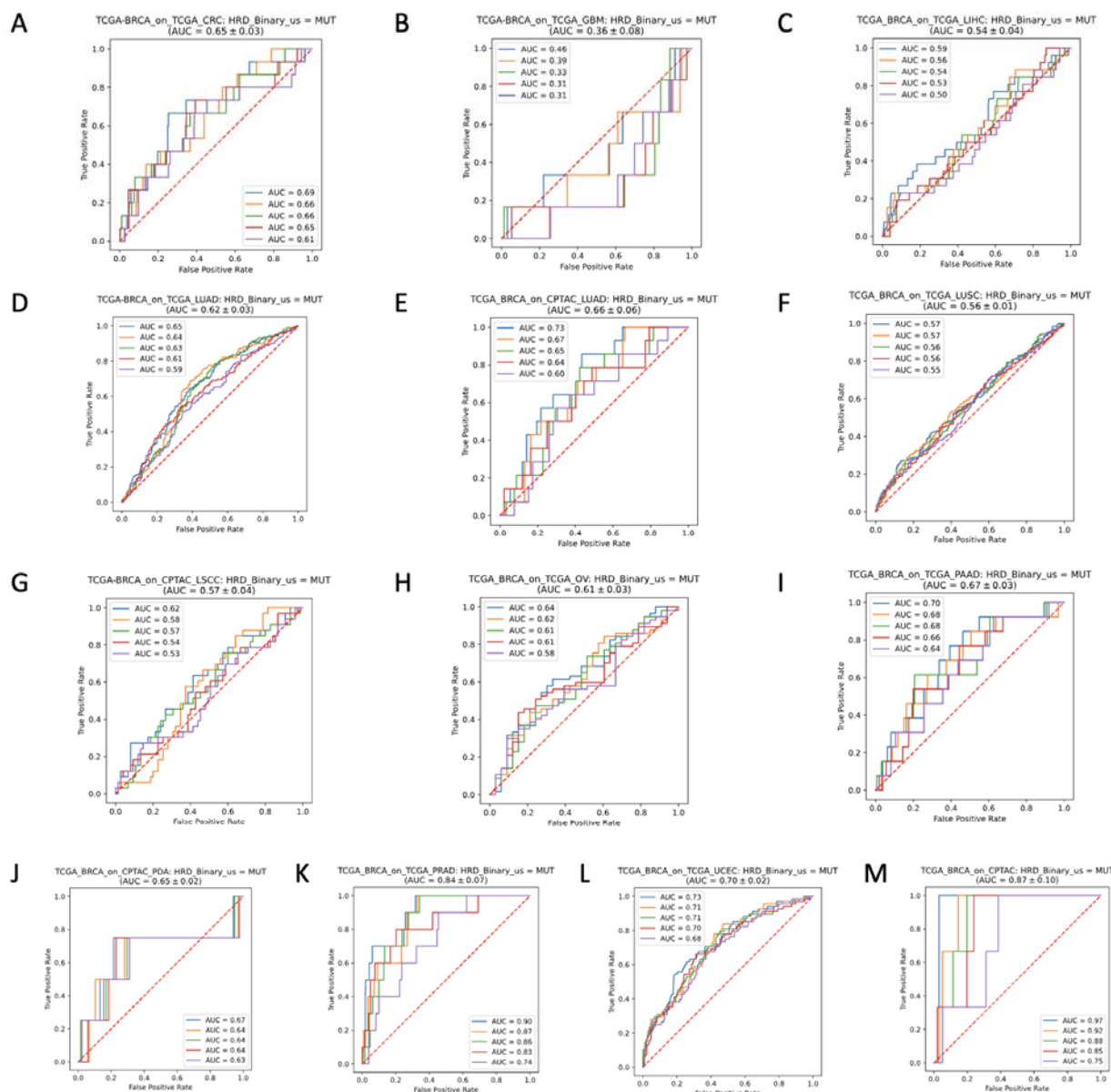
695

696 **Supplementary Figure 3: Receiving operating curve for the cross-cancer external validation.** The

697 Receiving operating curve (ROC) is shown for the cross-cancer external validation experiment for each

698 model trained on The Cancer Genome Atlas (TCGA) breast cancer (BRCA) cohort for the Homologous

699 recombination deficiency (HRD) binary score on (A) TCGA-CRC, (B) TCGA-GBM, (C) TCGA-LIHC, (D)
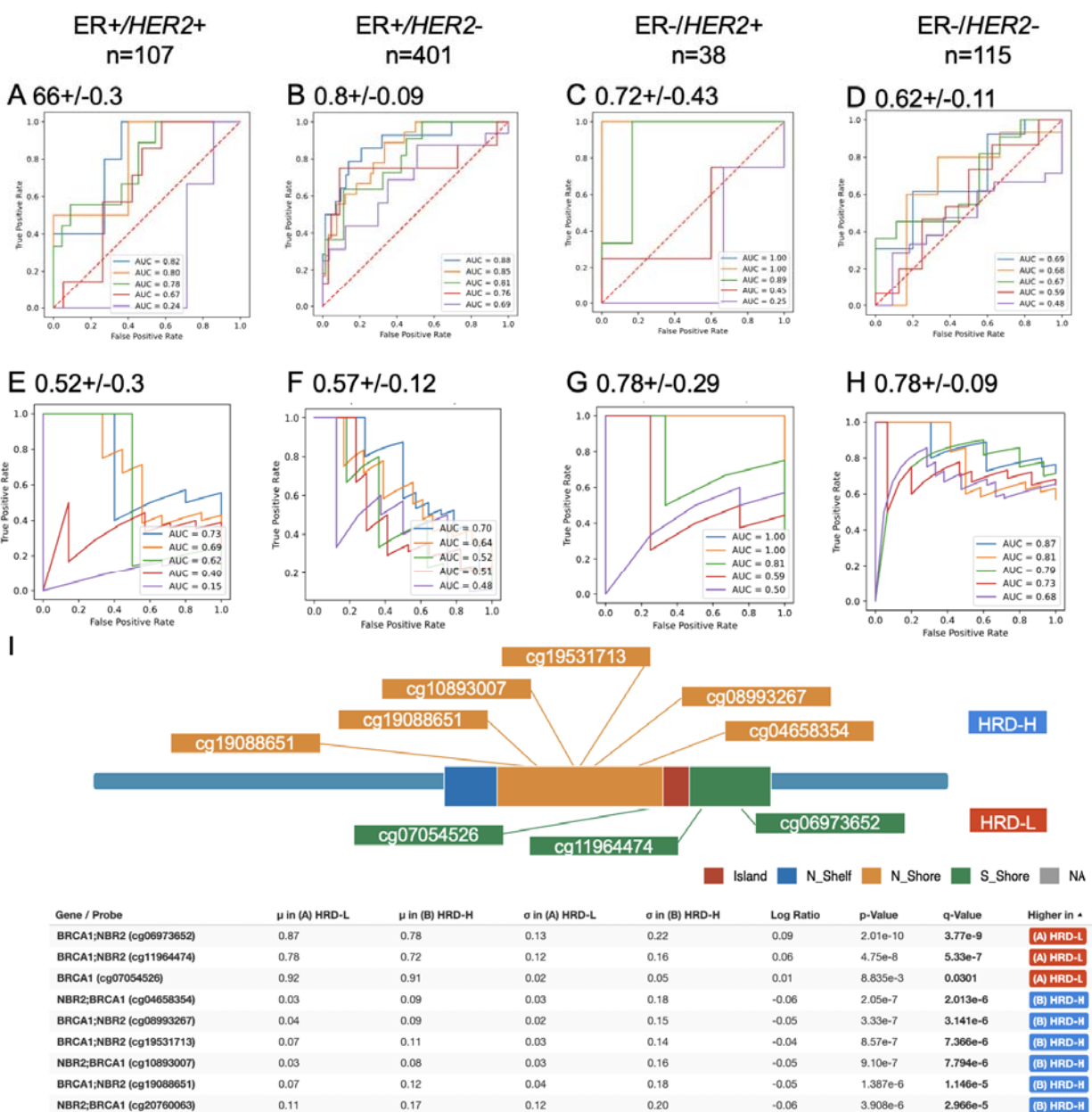
700 TCGA-LUAD, (E) CPTAC-LUAD, (F) TCGA-LUSC, (G) CPTAC-LSCC, (H) TCGA-OV, (I) TCGA-PAAD,

701 (J) CPTAC-PDA, (K) TCGA-PRAD, (L) TCGA-UCEC, (M) CPTAC-UCEC. Abbreviations: BRCA=breast

702 invasive carcinoma; CRC=colorectal cancer; GBM=glioblastoma; LIHC=liver cancer; LUAD=lung

703 adenocarcinoma; LUSC/LSCC=lung squamous cell carcinoma; OV=ovarian cancer;

704 PAAD/PDA=pancreatic adenocarcinoma; PRAD=prostate adenocarcinoma; UCEC=endometrial cancer

Gene / Probe	μ in (A) HRD-L	μ in (B) HRD-H	σ in (A) HRD-L	σ in (B) HRD-H	Log Ratio	p-Value	q-Value	Higher in ▲
BRCA1;NBR2 (cg06973652)	0.87	0.78	0.13	0.22	0.09	2.01e-10	3.77e-9	(A) HRD-L
BRCA1;NBR2 (cg11964474)	0.78	0.72	0.16	0.16	0.06	4.75e-8	5.33e-7	(A) HRD-L
BRCA1 (cg07054526)	0.92	0.91	0.02	0.05	0.01	8.835e-3	0.0301	(A) HRD-L
NBR2;BRCA1 (cg04658354)	0.03	0.09	0.03	0.18	-0.06	2.05e-7	2.013e-6	(B) HRD-H
BRCA1;NBR2 (cg08993267)	0.04	0.09	0.02	0.15	-0.05	3.33e-7	3.141e-6	(B) HRD-H
BRCA1;NBR2 (cg19531713)	0.07	0.11	0.03	0.14	-0.04	8.57e-7	7.366e-6	(B) HRD-H
NBR2;BRCA1 (cg10893007)	0.03	0.08	0.03	0.16	-0.05	9.10e-7	7.794e-6	(B) HRD-H
BRCA1;NBR2 (cg19088651)	0.07	0.12	0.04	0.18	-0.05	1.387e-6	1.146e-5	(B) HRD-H
NBR2;BRCA1 (cg20760063)	0.11	0.17	0.12	0.20	-0.06	3.908e-6	2.966e-5	(B) HRD-H

705

706 **Supplementary Figure 4: Subgroup analysis and overview the *BRCA1* promotor methylations in**

707 **TCGA-BRCA**. The Receiving operating curve (ROC) and Precision Recall curve (PRC) are shown for

708 the five-fold internal cross-validation experiment for each of the models in The Cancer Genome Atlas -

709 breast cancer (TCGA-BRCA) cohort for the Homologous recombination deficiency (HRD) score. ROC

710 curve is represented for the four different subgroups (A) estrogen receptor positive (ER+) and *HER2*+

711 (B) ER+ and *HER2*- (C) ER negative (ER-) and *HER2*+ (D) ER- and *HER2*-. The PRC curve is shown

712 for (E) ER+/HER2+, (F) ER+/HER2-, (G) ER-/HER2+, (H) ER-/HER2-. (I) Sketched representation of

713   the occurring promotor methylations (accessed with HM27 and HM450) in the *BRCA1* gene for the

714   ground truth Homologous recombination deficiency high (HRD-H) and low (HRD-L) subgroups.

715
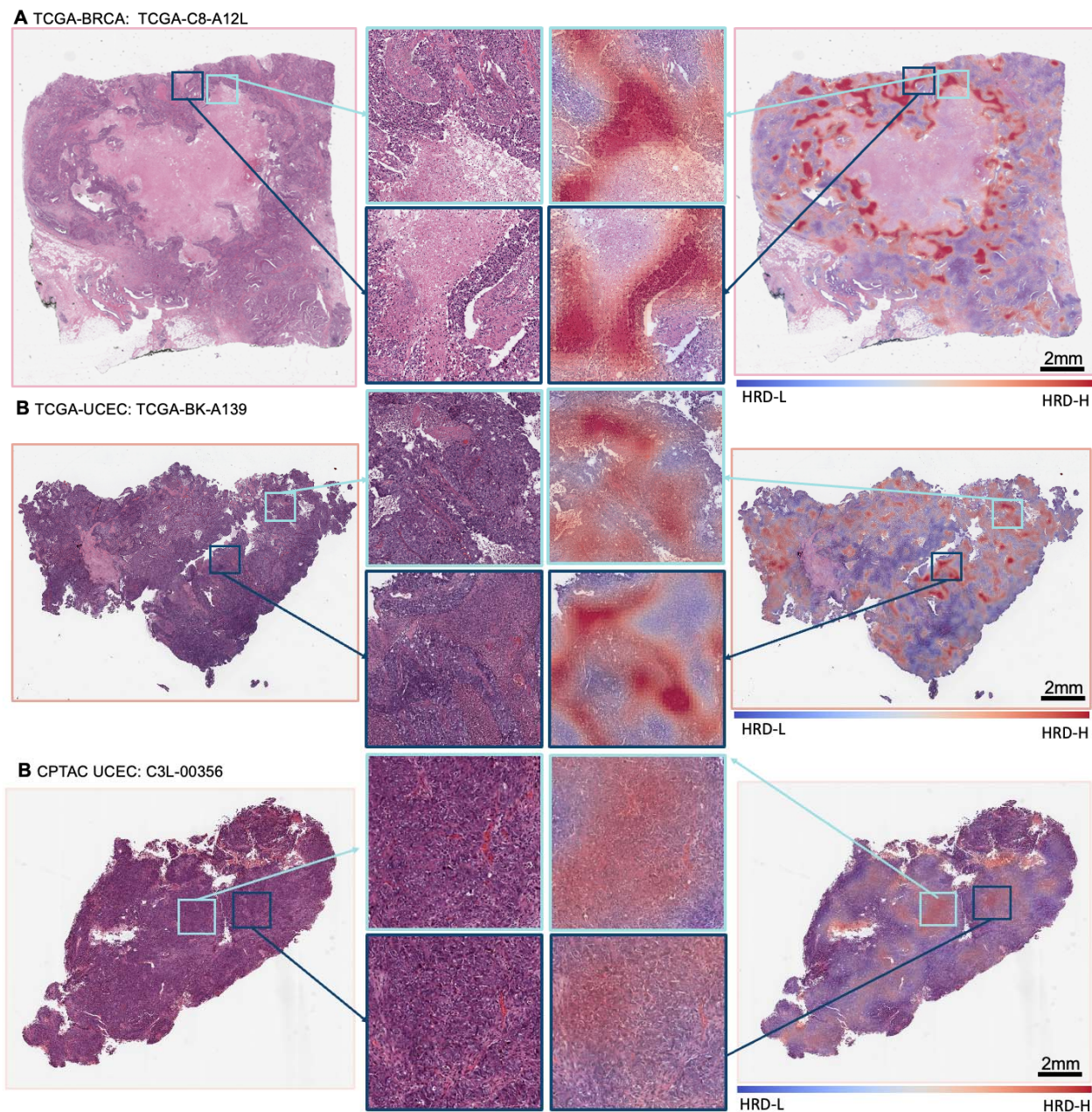
Supplementary Figure 5: Morphological features of Homologous recombination deficiency in

breast and endometrial cancer. Whole Slide Image (WSI) and classification heatmap (ground truth:

Homologous recombination deficiency high (HRD-H) and prediction: HRD-H) with magnifications of two

different regions. The model was trained on The cancer genome atlas (TCGA) breast cancer (BRCA)

cohort and deployed cross cancer wise. Top true positive predicted patients are shown for (A) TCGA-

721   BRCA, (B) Clinical Proteomic Tumor Analysis Consortium (CPTAC) endometrial cancer (UCEC) and

722   (C) TCGA-UCEC.
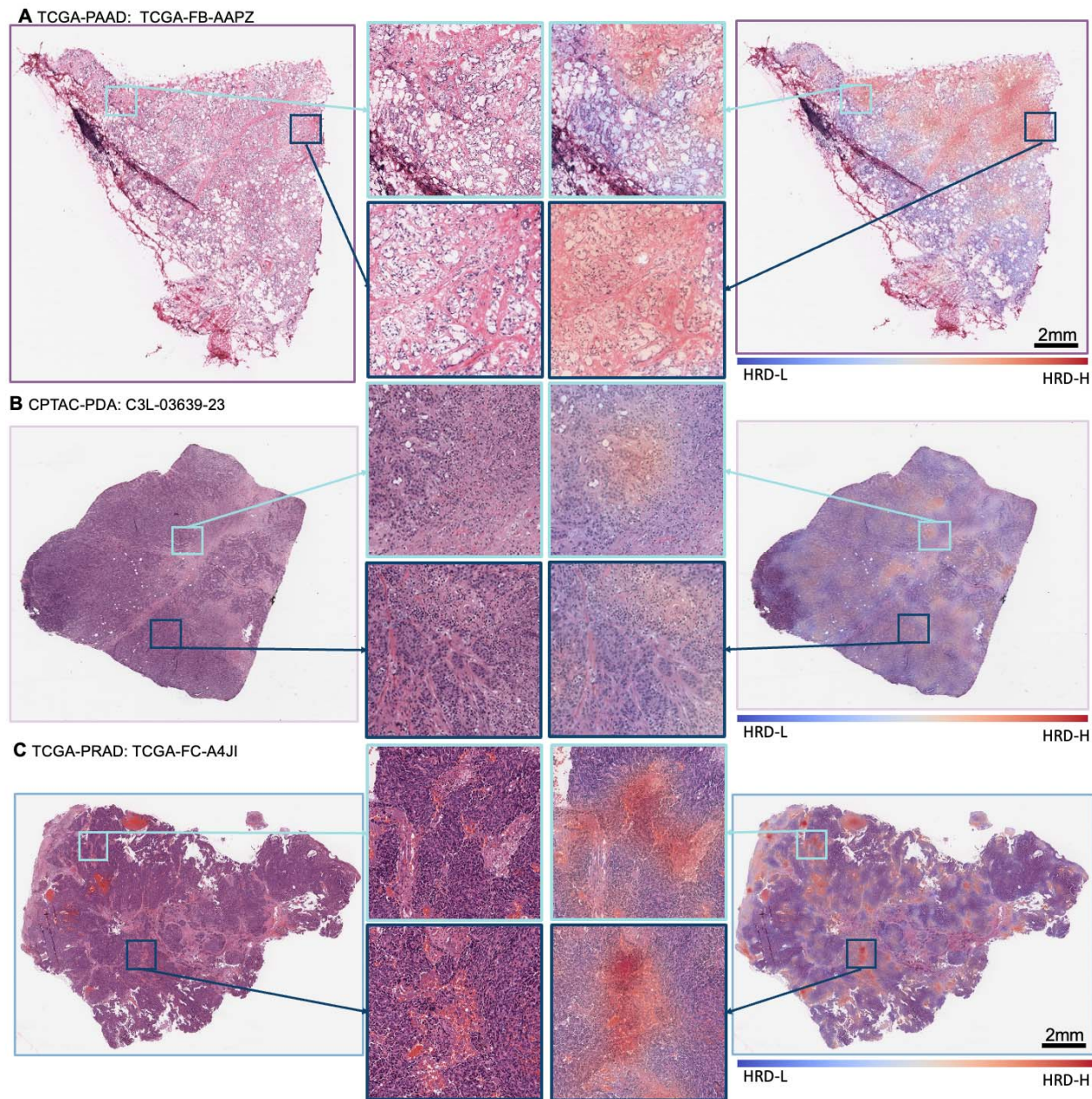
723

724   **Supplementary Figure 6: Morphological features of Homologous recombination deficiency in**

725   **pancreatic and prostate adenocarcinoma.** Whole Slide Image (WSI) and classification heatmap

726   (ground truth: Homologous recombination deficiency high (HRD-H) and prediction: HRD-H) with

727   magnifications of two different regions. The model was trained on The cancer genome atlas (TCGA)

728   breast cancer (BRCA) cohort and deployed cross cancer wise. Top true positive predicted patients are

729   shown for (A) TCGA pancreatic adenocarcinoma (PAAD), (B) Clinical Proteomic Tumor Analysis

730   Consortium (CPTAC) pancreatic adenocarcinoma (PDA) and (C) TCGA prostate adenocarcinoma

731   (PRAD).

732

733   **Supplementary Table 1: All raw statistical results.** All raw experimental results related to Figure 2,

734   including receiving operating curve (ROC) with 95% confidence interval (CI), Precision-Recall Curve

735   (PRC) with 95% confidence interval (CI), p-values and Homologous recombination deficiency (HRD)

736   high (HRD-H) and HRD-low (HRD-L) patient numbers based on the ground truth, for internal 5-fold

737   cross-validation on The Cancer Genome Atlas (TCGA) external validation on Clinical Proteomic Tumor

738   Analysis Consortium (CPTAC). [Supplementary_Table_1_All_statistical_results.xlsx] in separate file

739

740   **Supplementary Table 2: Homologous recombination deficiency score Tables.** Training data and

741   calculated homologous recombination deficiency score (HRD) out of the three subscores loss of

742   heterozygosity (LOH), telomeric allelic imbalance (TAI) and large-scale state transitions (LST) available

743   as continuous (HRDsum) and binary (HRD_Binary) target with a chosen cut off of HRD-L<42 HRD-

744   H>=42 for patients of The Cancer Genome Atlas (TCGA, Sheet1) and Clinical Proteomic Tumor

745   Analysis Consortium (CPTAC, Sheet2).

746

747   **Supplementary Table 3: Weblink for customized Homologous recombination deficiency (HRD)**

748   **subgroups.** Weblink for accessing the clinical and molecular characteristics for both ground truth and

749   prediction Homologous recombination Deficiency (HRD) subgroups at www.cbioportal.org for The

750   Cancer Genome Atlas breast cancer (TCGA-BRCA) Pan Cancer Atlas 2018 study and the TCGA-

751   BRCA Firehose Legacy cohort.