

## SUPPLEMENTARY DATA

1  
2  
3  
4  
5  
6  
7  
8

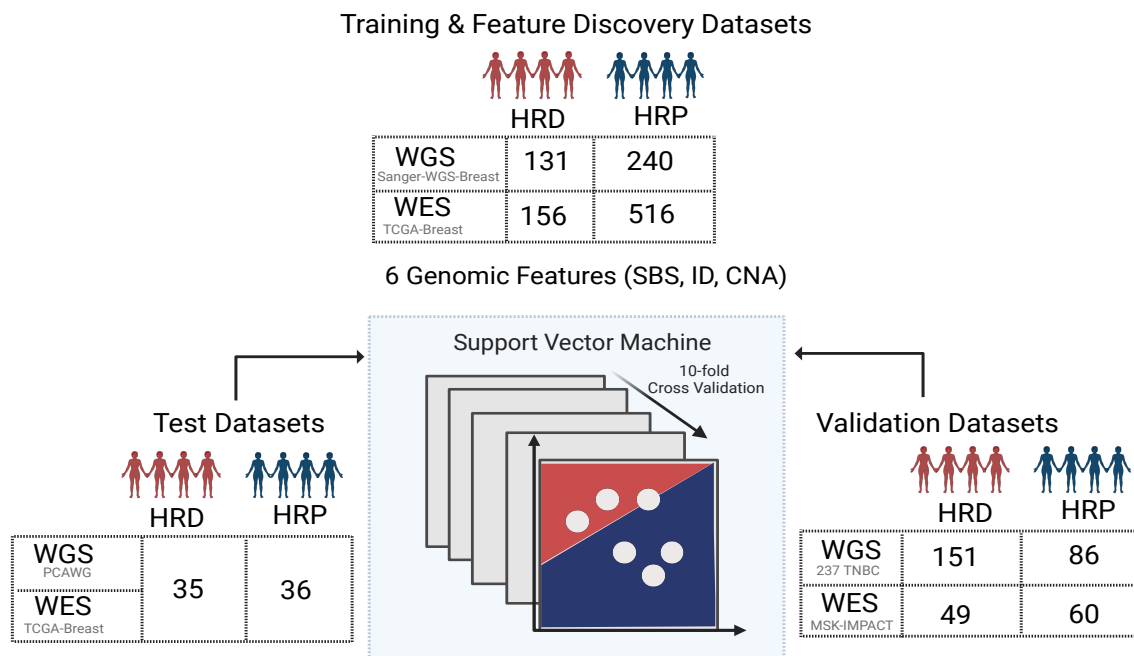
This supplementary data has been provided by the authors to give readers additional information about their work.

**Supplement to:** Abbasi A, Steele CD, Bergstrom EN, Khandekar A, Farswan A, *et al.* Detecting HRD in whole-genome and whole-exome sequenced breast and ovarian cancers.

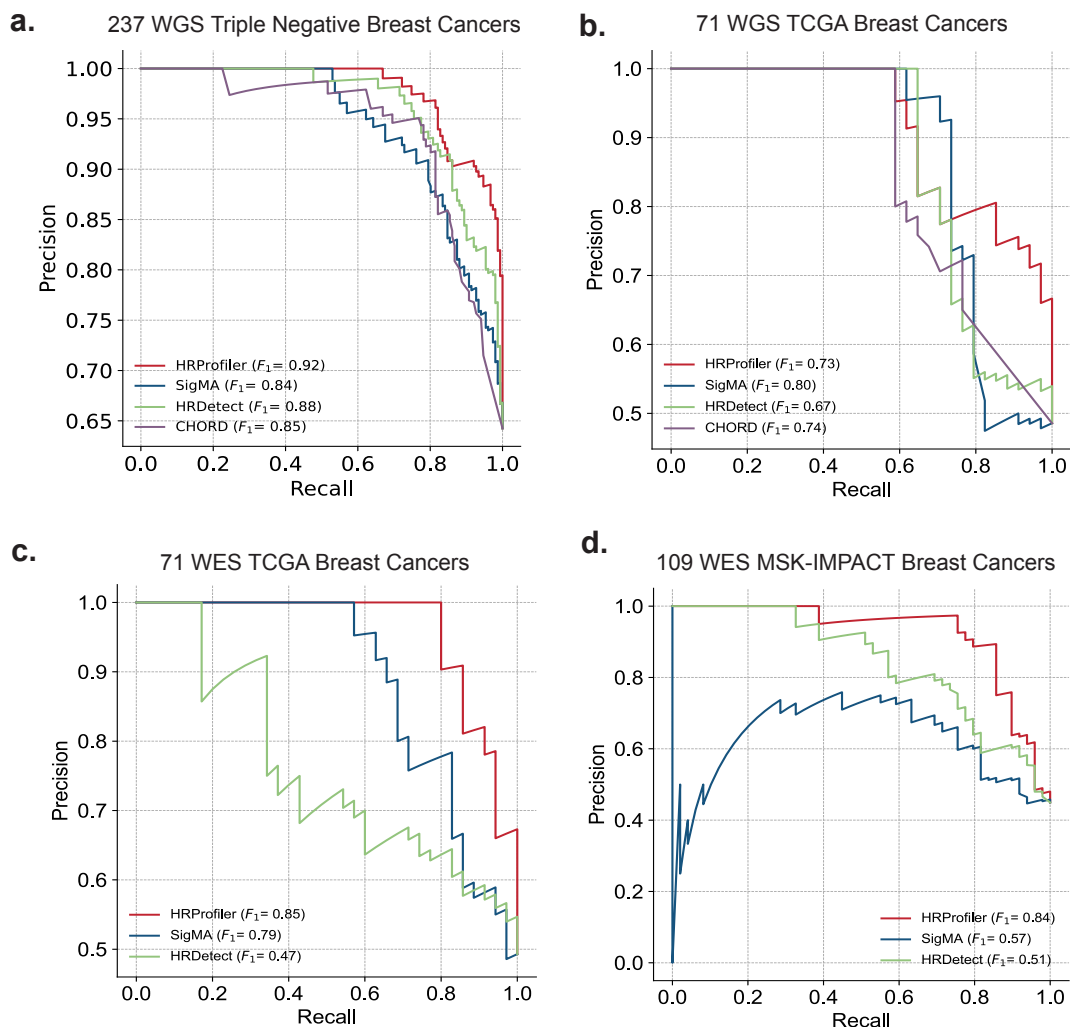
9	<b>Supplementary Data</b>	
10	<b>Section S1: Supplementary Figures S1-S7</b>	<b>3</b>
11	Figure S1: Datasets used for training, testing, and validating HRProfiler in breast cancers	
12	.....	3
13	Figure S2: Precision and recall of HRD tools on breast validation datasets using HRD genomic	
14	ground truth annotations .....	4
15	Figure S3: Evaluating CHORD for predicting survival to chemotherapy in whole-genome	
16	sequenced breast cancers.....	5
17	Figure S4: Evaluating the presence of defects in BRCA1/2 or HRD-associated signatures for	
18	predicting survival in chemotherapy-treated breast cancers.....	6
19	Figure S5: Datasets used for training, testing, and validating HRProfiler in ovarian cancers	
20	.....	7
21	Figure S6: Performance of HRD tools on external ovarian validation datasets using HRD genomic	
22	ground truth annotations.....	8
23	Figure S7: Evaluating the presence of defects in BRCA1/2 or HRD-associated signatures for	
24	predicting survival in PARP inhibitor treated ovarian cancer.....	9

## Section S1: Supplementary Figures

25 **Figure S1: Datasets used for training, testing, and validating HRProfiler in breast**  
 26 **cancer.** Schematic outline of the workflow for training, testing, and validating HRProfiler,  
 27 a support vector machine model for detecting homologous recombination deficient (HRD)  
 28 and homologous recombination proficient (HRP) breast cancers from whole-genome  
 29 sequenced (WGS) and whole-exome sequenced (WES) data. The model was trained  
 30 based on 6 genomic features, encompassing, single base substitutions (SBS), small  
 31 insertions and deletions (ID), and copy-number alternations (CN). Training and testing  
 32 data included samples from The Cancer Genome Atlas (TCGA), Sanger institute, and  
 33 Pan-Cancer Analysis of Whole Genomes (PCAWG) study. Validation datasets include  
 34 the independent Triple Negative Breast (TNBC) and the Memorial Sloan Kettering Cancer  
 35 Center's Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT) data.  
 36

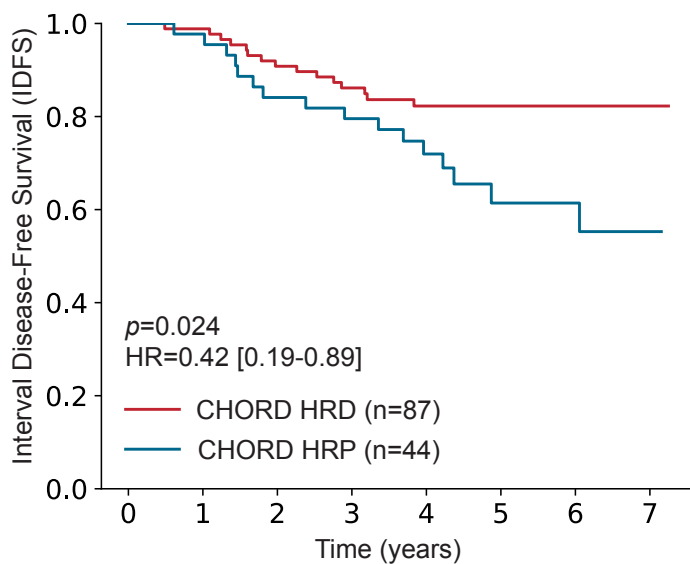


37 **Figure S2: Precision and recall of HRD tools on breast validation datasets using**  
 38 **HRD genomic ground truth annotations.** Precision and recall (PR) curves were  
 39 calculated for HRProfiler, SigMA, HRDetect, and CHORD. **(a)** PR curves for 237 whole-  
 40 genome sequenced (WGS) triple negative breast cancers. **(b)** PR curves for 71 WGS  
 41 breast cancers from The Cancer Genome Atlas (TCGA) project. **(c)** PR curves for 71  
 42 whole-exome sequenced (WES) breast cancers. **(d)** PR curves for 109 MSK-IMPACT  
 43 WES breast cancers. No PR curves are shown for CHORD in panels (c) and (d) as the  
 44 tool cannot be applied to WES data. The  $F_1$  scores, *i.e.*, the harmonic mean of precision  
 45 and recall, are shown for each tool within the respective panel's legend.

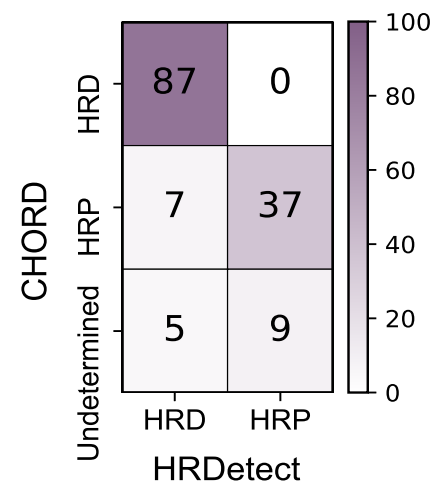


46 **Figure S3: Evaluating CHORD for predicting survival to chemotherapy in whole-**  
 47 **genome sequenced breast cancers.** All presented results are for 145 chemotherapy-  
 48 treated whole-genome sequenced (WGS) triple negative breast cancers. **(a)** Kaplan-  
 49 Meier curves for 131 breast cancers annotated as HRD and HRP by CHORD. Note that  
 50 14/145 samples were classified as undetermined by CHORD and these samples were  
 51 excluded from the survival analysis. The y-axis of the Kaplan-Meier curves reflects  
 52 Interval Disease Free Survival (IDFS), and the x-axis corresponds to time measured in  
 53 years. The p-value and hazard ratio (HR) are based on a Cox proportional hazards model  
 54 after adjusting for age and tumor grade. An 95% confidence interval is provided for the  
 55 HR within the Kaplan-Meier plot. **(b)** Confusion matrix comparing the HRD and HRP  
 56 annotations between CHORD and HRDetect.

a.

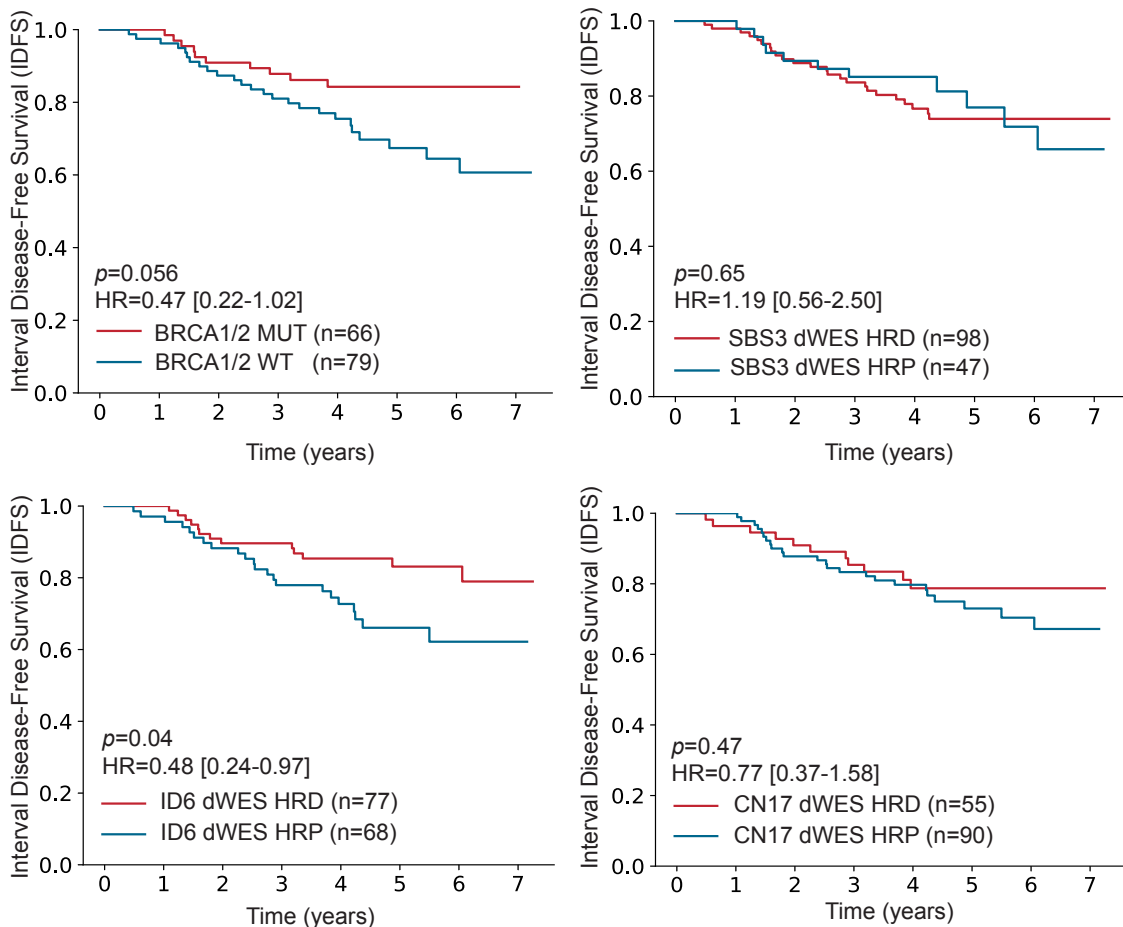


b.

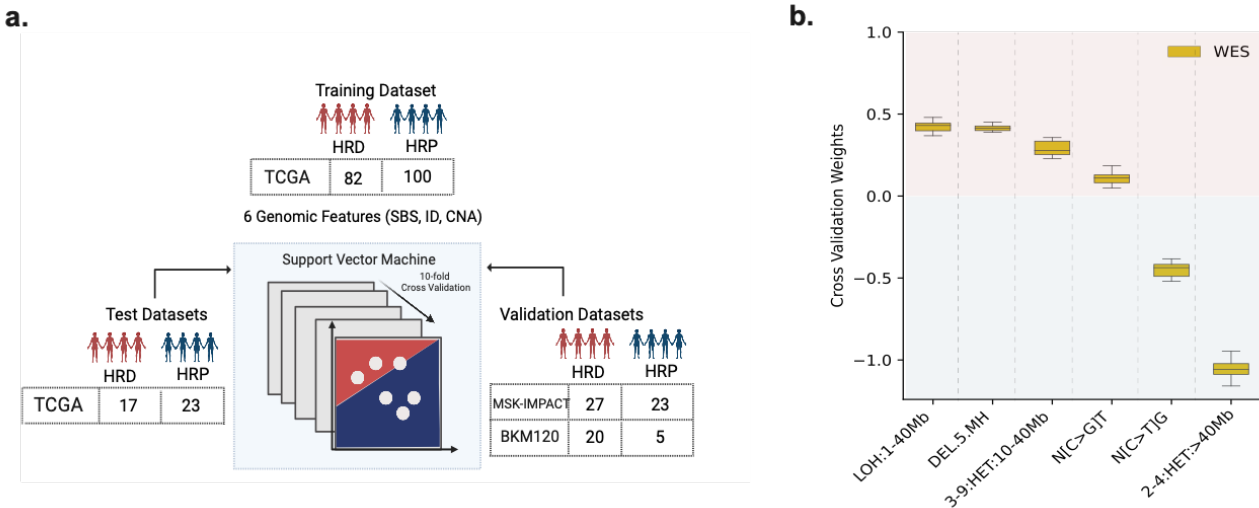


57 **Figure S4: Evaluating the presence of defects in *BRCA1/2* or HRD-associated**  
 58 **signatures for predicting survival in chemotherapy-treated breast cancers.** All  
 59 presented results are for 145 chemotherapy-treated triple negative breast cancers down-  
 60 sampled to whole-exomes (dWES). Kaplan-Meier curves and hazard ratios (HRs) for  
 61 samples annotated as HRD and HRP by either a defect in *BRCA1/2* or by the presence  
 62 of HRD-associated signatures SBS3, CN17, or ID6. The p-values and HRs are based on  
 63 a Cox proportional hazards model after adjusting for age and tumor grade. 95%  
 64 confidence intervals are provided for the HRs within the Kaplan-Meier plots. The y-axes  
 65 on all Kaplan-Meier curves reflect Interval Disease Free Survival (IDFS), and the x-axes  
 66 correspond to time measured in years.

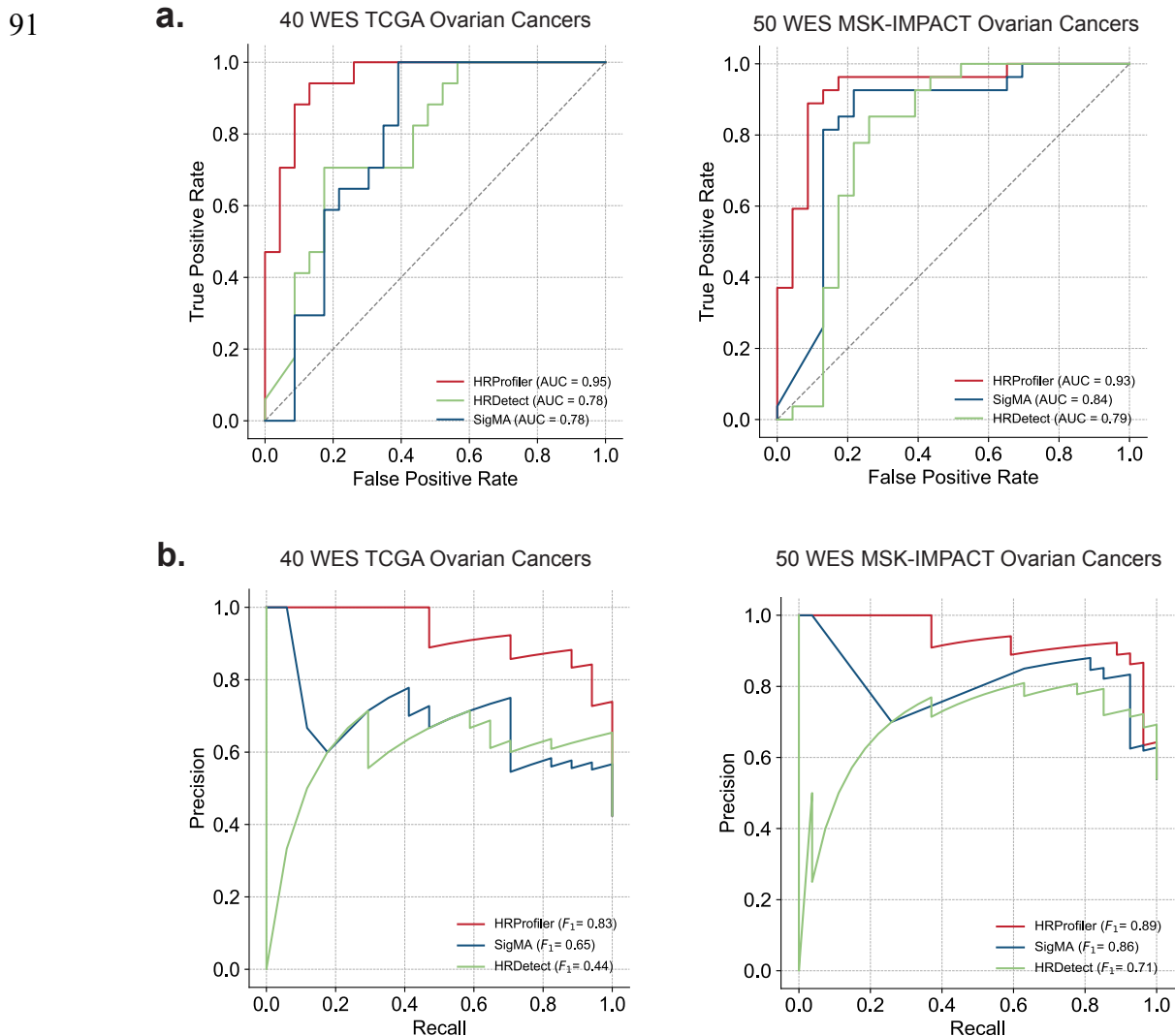
67



68 **Figure S5: Datasets and features used for training, testing, and validating**  
 69 **HRProfiler in ovarian cancer (a)** Schematic outline of the workflow for training, testing,  
 70 and validating HRProfiler, a support vector machine model for detecting homologous  
 71 recombination deficient (HRD) and homologous recombination proficient (HRP) ovarian  
 72 cancers from whole-exome sequenced data. The model was trained based on 6 genomic  
 73 features, encompassing, single base substitutions (SBS), small insertions and deletions  
 74 (ID), and copy-number alternations (CN). Training and testing data included samples from  
 75 The Cancer Genome Atlas (TCGA) project. Validation datasets include the independent  
 76 Memorial Sloan Kettering Cancer Center’s Integrated Mutation Profiling of Actionable  
 77 Cancer Targets (MSK-IMPACT) dataset and samples from a phase Ib trial of the PARP  
 78 inhibitor olaparib in combination with the PI3K inhibitor (BKM120 cohort). **(b)** The average  
 79 10-fold cross validation weights of the six features derived from WES ovarian training  
 80 dataset using a linear-kernel support vector machine. Positive weights reflect features  
 81 predictive for HRD samples, while negative weights correspond to features predictive for  
 82 HRP samples.



83 **Figure S6: Performance of HRD tools on external ovarian validation datasets using**  
 84 **HRD genomic ground truth annotations. (a)** Receiver operating characteristic (ROC)  
 85 curves and **(b)** precision and recall curves were calculated for HRProfiler, SigMA, and  
 86 HRDetect on a held-out test dataset of 40 whole-exome sequenced (WES) ovarian  
 87 samples from The Cancer Genome Atlas (TCGA) project and on an external validation  
 88 dataset of 50 WES MSK-IMPACT ovarian cancer. The areas under the ROC (AUCs) as  
 89 well as the  $F_1$  scores, *i.e.*, the harmonic mean of precision and recall, are shown for each  
 90 tool within the respective legend of each panel.





92 **Figure S7: Evaluating the presence of defects in *BRCA1/2* or HRD-associated**  
 93 **signatures for predicting survival in PARP inhibitor treated ovarian cancers.**  
 94 Progression free survival (PFS) across 25 PARPi treated ovarian cancers stratified based  
 95 on presence of **(a)** *BRCA1/2* mutations, **(b)** SBS3, **(c)** CN17, or **(d)** ID6. Listed p-values  
 96 and hazard ratios (HRs) are based on a Cox proportional hazards model after adjusting  
 97 for age at diagnosis and tumor stage. 95% confidence intervals are provided for all HRs  
 98 within the Kaplan-Meier plots.

