

1 **Deep learning predicts HRD and platinum response from histology slides in**
2 **breast and ovarian cancer**

3 Erik N. Bergstrom¹⁻³, Ammal Abbasi¹⁻³, Marcos Díaz-Gay¹⁻³, Loïck Galland⁴⁻⁷, Scott M.
4 Lippman¹, Sylvain Ladoire⁴⁻⁷, and Ludmil B. Alexandrov^{1-3*}

5
6 **Affiliations:**

7 ¹Moore's Cancer Center, UC San Diego, La Jolla, CA, 92037, USA

8 ²Department of Cellular and Molecular Medicine, UC San Diego, La Jolla, CA, 92093, USA

9 ³Department of Bioengineering, UC San Diego, La Jolla, CA, 92093, USA

10 ⁴Department of Medical Oncology, Centre Georges-François Leclerc, Dijon, France

11 ⁵Platform of Transfer in Biological Oncology, Centre Georges-François Leclerc, Dijon, France

12 ⁶University of Burgundy-Franche Comté, France

13 ⁷Centre de Recherche INSERM LNC-UMR1231, Dijon, France

14

15 *Correspondence should be addressed to L2alexandrov@health.ucsd.edu.

16 **ABSTRACT**

17 Breast and ovarian cancers harboring homologous recombination deficiencies (HRD) can benefit
18 from platinum-based chemotherapies and PARP inhibitors. Standard diagnostic tests for
19 detecting HRD utilize molecular profiling, which is not universally available especially for
20 medically underserved populations. Here, we trained a deep learning approach for predicting
21 genomically derived HRD scores from routinely sampled hematoxylin and eosin (H&E)-stained
22 histopathological slides. For breast cancer, the approach was externally validated on three
23 independent cohorts and allowed predicting patients' response to platinum treatment. Using
24 transfer learning, we demonstrated the method's clinical applicability to H&E-images from high-
25 grade ovarian tumors. Importantly, our deep learning approach outperformed existing genomic
26 HRD biomarkers in predicting response to platinum-based therapies across multiple cohorts,
27 providing a complementary approach for detecting HRD in patients across diverse
28 socioeconomic groups.

29

30 **One-Sentence Summary:** A deep learning approach outperforms molecular tests in predicting
31 platinum response of HRD cancers from histological slides.

32 **MAIN TEXT**

33 Precision oncology aims to personalize cancer therapy by first identifying and, subsequently,
34 targeting molecular defects in tumors within each individual (1). Many cancers harbor failures of
35 specific DNA repair pathways and utilizing synthetic lethal relationships amongst peripheral
36 pathways has proven an effective treatment approach (2). Previous mechanistic studies and
37 clinical trials have shown that breast and ovarian cancers harboring homologous recombination
38 DNA repair deficiency (HRD) are highly sensitive to platinum salts and poly (ADP-ribose)
39 polymerase (PARP) inhibitors (3). Historically, HRD has been associated with germline
40 mutations in specific genes leading to an increased cancer risk with the most notable
41 susceptibility genes being *BRCA1* (4) and *BRCA2* (5). In addition to germline variants, somatic
42 mutations and epigenetic dysregulation can also lead to HRD (6). Importantly, HRD cancers
43 exhibit characteristic patterns of somatic mutations (6-10) and gene expression (11, 12), and
44 these patterns have been leveraged as predictive biomarkers for targeted response to platinum
45 therapy and PARP inhibitors. Notably, the pattern of single-base substitution signature 3 (SBS3),
46 part of the Catalogue of Somatic Mutations in Cancer (COSMIC) catalog of mutational
47 signatures (13), was previously utilized as a clinical biomarker for detecting HRD (7, 14).

48
49 In the United States, the FDA has approved two HRD companion diagnostic (CDx) tests for
50 patients with ovarian and metastatic breast cancer (15). Myriad myChoice® CDx and
51 FoundationOne® CDx determine HRD by quantifying overall genomic instability in
52 combination with *BRCA1/2* status (16, 17). Additionally, multiple research and CLIA-certified
53 HRD diagnostic tests have been developed (18) and utilized to characterize the prevalence of
54 HRD across different solid tumors (19-21). These studies have identified HRD as commonly

55 found in multiple refractory human cancers, including triple-negative breast cancer, high-grade
56 serous ovarian cancer, and pancreatic adenocarcinoma, as well as established the role of HRD-
57 targeted therapies with platinum salts and PARP inhibitors (16, 22-24). Currently, all existing
58 HRD diagnostic tests intrinsically rely on DNA and/or RNA profiling leading to clinical-
59 workflow bottlenecks largely attributed to the availability of sufficient tissue samples for
60 molecular assays as well as to time to decision making and overall cost (25-27). For example, the
61 cost of an FDA-approved or a CLIA-certified HRD test is several thousand dollars (27) and
62 results can take from 3 to 6 weeks (25). In turn, this has precluded the widespread utilization of
63 molecular diagnostics in standard therapy and clinical trials (1) with a disproportionately high
64 effect on patients from underserved populations (15).

65
66 While the utilization of sequencing-based diagnostics is limited, tumor biopsies are routinely
67 processed in clinical practice for the diagnosis of solid-tumors by light-microscopic
68 morphological review of tissue stained with hematoxylin and eosin (H&E) (26). Combined with
69 recent advances in computational pathology, deep learning artificial intelligence (AI)-based
70 models allow for both prognostic and diagnostic predictions using only digital H&E slides (28).
71 Here we introduce DeepHRD, a deep learning AI platform for detecting HRD from digitalized
72 H&E slides. We train and validate DeepHRD models on data from The Cancer Genome Atlas
73 (TCGA) project and demonstrate their ability to detect HRD using external datasets. Importantly,
74 using independent samples, we demonstrate that DeepHRD outperforms existing clinical
75 genomic biomarkers in predicting response to platinum-based therapies.

76

77 RESULTS

78 We implemented DeepHRD, a weakly supervised convolutional neural network architecture that
79 uses multiple instance learning (MIL; **Fig. 1**) for predicting HRD status from digital H&E slides
80 (29-31). Specifically, for training DeepHRD, a soft label is assigned to each digital whole-slide
81 H&E image (WSI) based on an HRD score derived using sequencing or genotyping data from
82 the same cancer sample (**Supplementary Materials**). Further, all partitioned regions of that
83 WSI, termed, tiles, are assigned a weak label based upon the sample's classification. It is
84 assumed that all tiles within a negatively labeled sample are homologous recombination
85 proficient (HRP), whereas at least one tile must exhibit an HRD phenotype within a positively
86 labeled sample. These assumptions allow the model to be trained using only a single
87 classification label for an entire WSI without the need for detailed manual annotations from a
88 pathologist, which currently do not exist for characterizing HRD.

89
90 DeepHRD is based on a multi-resolution decision designed to mimic the standard diagnostic
91 protocol used by pathologists, which performs an initial prediction on a low magnification (*i.e.*,
92 5x magnification) and then automatically selects regions of interest (ROI) to perform a
93 secondary prediction on an enhanced magnification within ROIs (*i.e.*, 20x magnification; **Fig.**
94 **1a**) (30). Once fully trained, the model generates HRD predictions directly from digital tissues
95 slides without the need for genomic profiling (**Fig. 1b**). Further, DeepHRD maps individual tile
96 predictions back to the original WSI, which allows visualizing the relative importance of tissue
97 regions for the obtained predictions (**Fig. 1b**). The final model encompasses an ensemble of five
98 identical architectures, with each producing multi-resolution prediction scores. The average of
99 these scores is used to make a final prediction for each tissue slide. Importantly, DeepHRD

100 estimates epistemic uncertainty using Bayesian dropout during inference of a tissue slide to
101 calculate confidence intervals for the final model prediction (**Supplementary Materials**). The
102 confidence intervals are subsequently used to provide a computational diagnostic
103 recommendation (**Fig. 1b**).

104

105 DeepHRD models were trained and internally validated using data from 1,008 TCGA breast
106 cancers (32) with flash frozen (FF) slides and 1,055 TCGA breast cancers with formalin-fixed
107 paraffin-embedded (FFPE) slides (**fig. S1**). All samples had whole-exome sequencing and
108 microarray genotyping data for calculating a genomic HRD score (**fig. S1**). We trained
109 DeepHRD breast cancer models by separating the samples with: (i) 70% used for training; (ii)
110 15% for adjusting training parameters; and (iii) 15% held-out for testing the final model (**Fig. 1a**;
111 **fig. S1**). Two independent models were trained, one for FF and one for FFPE tissue slides
112 (**Supplementary Materials**). Prior to training, the number of HRD and HRP samples per breast
113 cancer subtype were balanced to prevent learning subtype specific histological features (**fig. S1**).

114 Each trained DeepHRD breast cancer model allows making a patient-level prediction using only
115 a single FF or FFPE digital slide. Specifically, DeepHRD predicts whether a breast cancer is
116 HRD or HRP, and it overlays an HRD probability mask to the digital slide, thus, allowing
117 subsequent pathological investigations (**Fig. 1b**).

118

119 The DeepHRD breast cancer FF model exhibited an overall performance with an AUC of 0.81
120 ([0.77-0.85] 95% Confidence Interval (CI); **Fig. 2a**) on the held-out TCGA samples. The
121 generalizability of the FF model was externally validated by applying it to 116 primary breast
122 cancer slides from the Clinical Proteomic Tumor Analysis Consortium (37) and 419 primary

123 breast tumors from the Molecular Taxonomy of Breast Cancer International Consortium (38)
124 (**fig. S2a**) resulting in an AUC of 0.76 ([0.71-0.82] 95% CI; **Fig. 2a**). Notably, while HRD is
125 enriched in luminal B, basal-like, and Her2 enriched breast cancers (**fig. S1a**), DeepHRD was
126 able to distinguish HR deficiency and proficiency across all subtypes (**Fig. 2b**). The DeepHRD
127 breast cancer FFPE model exhibited an AUC of 0.81 ([0.77-0.86] 95% CI; **Fig. 2c**) on the held-
128 out TCGA samples, which was identical to the flash frozen model. These results indicate that the
129 fixation procedure and differences in staining coloration have minimal effects on the
130 performance of predicting HRD status directly from breast cancer tissue slides.
131
132 Importantly, the FFPE model was capable of distinguishing metastatic breast cancers (MBCs),
133 part of an independent clinical cohort, that had a complete response to platinum chemotherapy
134 ($n=9$) from MBCs having only a partial or no response to treatment ($n=68$) with an AUC of 0.76
135 ([0.54-0.93] 95% CI; **Fig. 2c**; **fig. S2b**). Additionally, clinical response to platinum-based therapy
136 and progression-free survival were assessed using the Response Evaluation Criteria in Solid
137 Tumors, version 1.1 (RECIST 1.1; **fig. S2b**) (36). Separating the MBCs treated with platinum
138 based upon DeepHRD's prediction revealed a median progression-free survival of 14.4 months
139 for HRD patients and 3.9 months for HRP patients (p -value=0.0019, log-rank test). The model's
140 predictive value was consistent after correcting for breast cancer subtype, age of diagnosis, and
141 the genomic HRD score with a hazard ratio of 0.47 ([0.27-0.83] 95% CI; q -value=0.0087; **Fig.**
142 **2d**). Further, DeepHRD captured 7 of the 9 complete responders to platinum treatment. In
143 comparison, neither the separation based upon *BRCAl/2* mutations nor detecting the HRD-
144 associated signature SBS3 resulted in a significant difference in progression-free survival (q -
145 value=0.13 and $q=0.34$, respectively; **Fig. 2d**). While the small sample size of *BRCAl/2* mutated

146 tumors (~8% of MBCs) influenced the significance levels compared to wild-type tumors, the
147 predictions from DeepHRD captured 4-fold more platinum sensitive samples. Lastly, the tissue
148 slides from the MBC were digitalized using a Hamamatsu Photonics Nanozoomer system, while
149 all other cohorts were digitalized using an Aperios ScanScope system, further demonstrating the
150 generalizability of DeepHRD.

151
152 Ovarian cancer patients have traditionally received first-line platinum chemotherapies making
153 them ideal to evaluate whether HRD predictions from tissue slides may have a direct clinical
154 benefit. To test whether DeepHRD can be used for other cancer types, we trained an independent
155 FF ovarian cancer model by performing transfer learning on the TCGA ovarian cancer cohort
156 ($n=589$) using the pretrained weights and biases generated from the FF breast cancer model with
157 the convolutional weights and biases frozen during training (**Fig. 3a; fig. S1**). A similar training
158 approach employed for the breast cancer models was utilized for the ovarian cancer model
159 (**Supplementary Materials**). To assess the ability of the DeepHRD ovarian model to separate
160 individuals benefiting from treatment with platinum chemotherapy, the model was applied to a
161 held-out set of 66 high-grade serous ovarian cancers that received treatment with first-line
162 platinum chemotherapy. Patients predicted to be HRD had a median survival of 4.6 years, while
163 those predicted to be HRP had a median survival of 3.2 years with a hazard ratio of 0.45 ([0.22-
164 0.90] 95% CI; q -value=0.024) after correcting for the stage of the cancer, age, and the genomic
165 HRD score (**Fig. 3b**). In comparison, we observed a worse separation when using a base model
166 without transfer learning (HR=0.53 [0.26-1.07] 95% CI; q -value=0.076; **Fig. 3c**), suggesting that
167 the transfer learning provides a benefit when attempting to train AI-based approaches on smaller
168 datasets. Consistent with the breast cancer cohort, neither separation based on mutations in

169 *BRCAl/2* nor detecting the HRD-associated signature SBS3 resulted in a significant difference in
170 survival (q-values>0.10; **Fig. 3c**).

171 **DISCUSSION**

172 The development of DeepHRD prediction models for breast and ovarian cancers demonstrates
173 the practicality of deploying AI-based guidance into clinical diagnostics and precision medicine
174 workflows. Results across multiple external cohorts indicate that the platform is applicable to
175 routinely sampled tissue blocks and generalizable across different cancers, digital scanning
176 systems, and tissue fixation procedures. DeepHRD's performance was consistent across primary
177 and metastatic breast cancers and, by incorporating transfer learning, the model was also
178 applicable to serous ovarian cancer. Since HRD is a complementary biomarker guiding the use
179 of platinum therapies and an FDA-approved companion diagnostic for the use of PARP
180 inhibitors (*15-17*), the performance of our DeepHRD platform has direct implications for the
181 treatment of other cancer types with known HR-deficiencies (*19*), notably within pancreatic
182 adenocarcinomas. Despite clear benefit from first-line platinum therapy in HRD-positive patients
183 with this refractory disease, there is a 3 to 6 week turn-around for genomic testing which is not
184 appropriate for an advanced pancreatic cancer with a median progression-free survival of 6
185 months (*25*). Further, access to HRD genomic testing is even more limited in developing
186 countries (*15*). With increasing clinical evidence for the treatment of HRD tumors with platinum
187 salts and PARP inhibitors and the limitations of existing genomic tests, there is a need to develop
188 novel frameworks to guide the current standard-of-care for HRD tumors (*23*).

189
190 Recently, deep learning AI approaches have demonstrated the ability to detect genomic
191 alterations directly from H&E images, including biomarkers related to patient outcome, which
192 could be leveraged for pre-screening tests. While demonstrating a high-concordance in
193 predicting genomic HRD, DeepHRD was also capable of directly predicting individual patient

194 outcome to HRD-targeted therapy using response and progression-free survival based on
195 RECIST criteria. Furthermore, DeepHRD provided better prediction of clinical response and
196 progression-free survival to platinum therapies than existing genomic biomarkers. Notably, our
197 approach captured patients with *BRCA1/2* wild-type tumors who responded to platinum therapy,
198 thus, identifying 4-fold more responders than *BRCA1/2* mutation-testing alone (**Fig. 2d**). These
199 results demonstrate that molecular assays, traditionally used for assessing HRD in a clinical
200 setting, can be substituted and/or complemented with AI-based deep learning models for
201 predicting clinical response from conventional diagnostic histopathological slides.

202
203 While there has been a recent explosion of deep learning methods applied to digital pathology
204 (28), the immediate translation into clinical practice has been limited by the high costs associated
205 with acquiring infrastructure for routinely capturing digital H&E slides (27, 39). With the
206 development of tertiary scanning services seeking to alleviate these overhead costs, a recent
207 study has shown a potential alternative approach for utilizing analogous deep learning AI-models
208 to make predictions from photographs taken directly by hand-held devices (40). In coordination
209 with the optimization of lightweight deep learning architectures, using images from a
210 smartphone attached to the ocular lens of a conventional light microscope promises inexpensive,
211 efficient, and accurate deep-learning read-outs within seconds of preparing an H&E slide. By
212 relying on smartphone microscopy images, this transition would provide AI-based diagnostic
213 solutions for equitable and efficient clinical management for cancer patients across globally
214 diverse socioeconomic groups.

215

216 **FIGURE LEGENDS**

217 **Fig. 1. Multi-resolution convolutional neural network architecture to detect homologous**

218 **recombination deficiency from histopathological tissue slides. a)** Training a DeepHRD model

219 for detecting homologous recombination deficiency (HRD) from whole-slide images (WSIs). For

220 each WSI, a single prediction score is estimated based on the detection of HRD. Specifically,

221 each WSI undergoes preprocessing and quality control **(1)**. This module consists of tissue

222 segmentation, filtering for non-focused tissue, and final tiling of regions that contain tissue at 5x

223 magnification. All tiles for a single image are processed through the first multiple instance

224 learning (MIL) ResNet18 convolutional neural network **(2)**. This architecture uses the average of

225 the top 25 predicted tile scores as the WSI predicted score. Dropout is incorporated into the fully

226 connected layers in the feature extraction module to reduce overfitting during training. The same

227 dropout technique is also incorporated during inference to simulate Monte Carlo dropout used to

228 calculate confidence intervals in the final WSI prediction. The tile feature vectors from the

229 penultimate layer of the feature extraction are used to automatically select regions of interest

230 (ROI) from the original WSI for additional assessment **(3)**. The feature vectors are reduced in

231 dimensions using principal component analysis and a custom k-means clustering module is used

232 to determine the optimal number of clusters per sample. The selected tiles are then resampled at

233 a 20x magnification **(4)**. These sets of tiles are used to train a second MIL-ResNet18 model **(5)**

234 using an identical architecture to the one previously used in **(2)**. The average predictions across

235 both models are aggregated for a single WSI **(6)**. The resulting distribution of scores are used to

236 calculate confidence intervals and establish a threshold of confidence for a final prediction. **b)**

237 Using a trained DeepHRD model for HRD prediction from a single whole-slide image.

238 DeepHRD produces a final prediction score for individual patient biopsies, with a
239 computational-based diagnosis for subsequent clinical action.

240

241 **Fig. 2. DeepHRD for detecting homologous recombination deficiency and predicting**
242 **response to treatment in primary and metastatic breast cancer. a)** The receiver operating
243 characteristic curves (ROCs) for classifying homologous recombination deficiency (HRD) in the
244 TCGA held-out set and the independent set of primary breast cancers, encompassing the
245 independent CPTAC and METABRIC primary breast cancer cohorts. **b)** Representative TCGA
246 tissue slides are shown for both HRD and homologous recombination proficient (HRP) samples
247 across multiple breast cancer subtypes along with the resulting predictions for each segmented
248 tile at 5x and 20x resolutions. **c)** ROCs for formalin-fixed paraffin-embedded (FFPE) diagnostic
249 model in the TCGA held-out set and for classifying metastatic breast cancer (MBC) patients who
250 are complete responders to platinum therapy. **d)** Kaplan-Meier survival curves for MBC patients
251 treated with platinum chemotherapy separated by DeepHRD model predictions (*left*), *BRCA1/2*
252 mutation status (*middle*), and SBS3 activity as detected by SigMA (*right*). Q-values are corrected
253 after considering breast cancer subtype, age at diagnosis, and the standard-of-care binary HRD
254 classification score ≥ 42 (*i.e.*, HRD score). Cox regression showing the \log_{10} -transformed hazard
255 ratios are shown with their 95% confidence intervals (*bottom*). Q-values less than or equal to
256 0.05 are annotated with * while q-values above 0.05 are annotated with n.s. (*i.e.*, non-
257 significant).

258

259 **Fig. 3. DeepHRD transfer learning in ovarian cancer for predicting response to platinum**
260 **treatment. a)** Schematic demonstrating the transfer learning method to train an ovarian

261 homologous recombination deficiency (HRD) model from whole-slide H&E image (WSI) using
262 a pretrained breast DeepHRD model. The pretrained flash-frozen breast model is used to initiate
263 the weights and biases of all parameters in the ovarian model. HRD-scores are calculated from
264 SNP6 genotyping microarray by deriving loss of heterozygosity (LOH), large-scale transitions
265 (LST), and telomeric allelic imbalance (TAI). **b)** Kaplan-Meier survival curves comparing the
266 outcomes of patients treated with platinum chemotherapy split by the prediction of the
267 DeepHRD transfer learning model. **c)** Kaplan-Meier survival curves comparing the outcomes of
268 platinum-treated patients split by the base model predictions with no transfer learning applied
269 (*left*), *BRCA1/2* mutation status (*middle*), and SBS3 activity as detected by SigMA (*right*). Q-
270 values are corrected after considering ovarian cancer stage, age at diagnosis, and the standard-of-
271 care binary HRD classification score ≥ 63 (*i.e.*, HRD score). Cox regression showing the log₁₀-
272 transformed hazard ratios are shown with their 95% confidence intervals (*bottom*). Q-values less
273 than or equal 0.05 are annotated with * while q-values above 0.05 are annotated with n.s. (*i.e.*,
274 non-significant).
275

276 **Acknowledgements**

277 The results shown are in part based upon data generated by the TCGA Research Network:
278 <http://cancergenome.nih.gov/>. Additional data used in this publication were generated by the
279 Clinical Proteomic Tumor Analysis Consortium (CPTAC) and the Molecular Taxonomy of
280 Breast Cancer International Consortium (METABRIC). The computational analyses reported in
281 this manuscript have utilized the Triton Shared Computing Cluster at the San Diego
282 Supercomputer Center of UC San Diego.

283

284 **Funding**

285 This work was funded by a US National Institutes of Health grant R01ES032547 and UC San
286 Diego start-up funding to LBA.

287

288 **Author contributions**

289 ENB and LBA designed the overall study. ENB performed all analyses with help from AA,
290 MDG, LG, SML, SL, and LBA. Specifically, AA assisted in the calculation and interpretation of
291 the genomically-derived HRD scores. MDG assisted in the processing of the digital images. LG
292 and SL assisted in the analysis and interpretation of the metastatic breast cancer cohort. SML
293 assisted in the interpretation and analysis of the survival and clinical associations. ENB and LBA
294 wrote the manuscript with help and input from all other authors. All authors read and approved
295 the final manuscript.

296

297

298

299 **Competing interests**

300 LBA is a compensated consultant and has equity interest in io9, LLC and Genome Insight. His
301 spouse is an employee of Biotheranostics, Inc. SML is a co-founder and has equity interest in io9,
302 LLC. AA and LBA declare U.S. provisional patent application with serial numbers 63/366,392 for
303 detecting homologous recombination deficiency from genomics data. ENB, SML, and LBA
304 declare U.S. provisional patent application with serial numbers 63/269,033 for artificial
305 intelligence architecture for predicting cancer biomarkers. All other authors declare they have no
306 known competing financial interests or personal relationships that could have appeared to
307 influence the work reported in this paper.

308

309 **Data and materials availability**

310 All datasets utilized in this study were previously generated and publicly available. These data can
311 be access through the accession codes listed in **Supplementary Materials**. The sources code for
312 DeepHRD will be made publicly available upon the publication of this study.

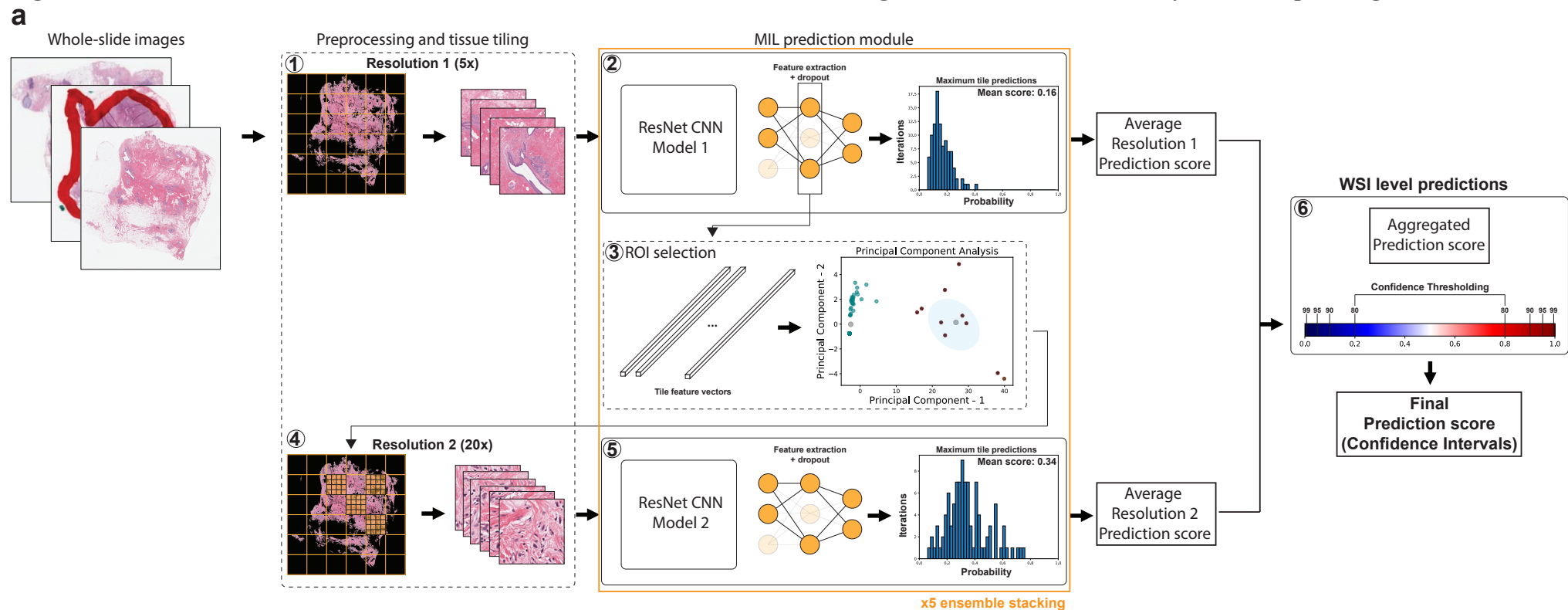
313 REFERENCES

- 314 1. R. L. Schilsky, D. L. Longo, Closing the Gap in Cancer Genomic Testing. *N Engl J Med*
315 **387**, 2107-2110 (2022).
- 316 2. J. Setton *et al.*, Synthetic Lethality in Cancer Therapeutics: The Next Generation. *Cancer*
317 *Discov* **11**, 1626-1635 (2021).
- 318 3. R. Prakash, Y. Zhang, W. Feng, M. Jasin, Homologous recombination and human health:
319 the roles of BRCA1, BRCA2, and associated proteins. *Cold Spring Harb Perspect Biol* **7**,
320 a016600 (2015).
- 321 4. E. M. John *et al.*, Prevalence of pathogenic BRCA1 mutation carriers in 5 US
322 racial/ethnic groups. *JAMA* **298**, 2869-2876 (2007).
- 323 5. K. E. Malone *et al.*, Prevalence and predictors of BRCA1 and BRCA2 mutations in a
324 population-based study of breast cancer in white and black American women ages 35 to
325 64 years. *Cancer Res* **66**, 8297-8308 (2006).
- 326 6. P. Polak *et al.*, A mutational signature reveals alterations underlying deficient
327 homologous recombination repair in breast cancer. *Nat Genet* **49**, 1476-1486 (2017).
- 328 7. H. Davies *et al.*, HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on
329 mutational signatures. *Nat Med* **23**, 517-525 (2017).
- 330 8. S. Nik-Zainal *et al.*, Landscape of somatic mutations in 560 breast cancer whole-genome
331 sequences. *Nature* **534**, 47-54 (2016).
- 332 9. L. B. Alexandrov *et al.*, Signatures of mutational processes in human cancer. *Nature* **500**,
333 415-421 (2013).
- 334 10. L. B. Alexandrov *et al.*, The repertoire of mutational signatures in human cancer. *Nature*
335 **578**, 94-101 (2020).
- 336 11. G. Peng *et al.*, Genome-wide transcriptome profiling of homologous recombination DNA
337 repair. *Nat Commun* **5**, 3361 (2014).
- 338 12. M. J. Larsen *et al.*, Classifications within molecular subtypes enables identification of
339 BRCA1/BRCA2 mutation carriers by RNA tumor profiling. *PLoS One* **8**, e64268 (2013).
- 340 13. J. G. Tate *et al.*, COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids*
341 *Res* **47**, D941-D947 (2019).
- 342 14. D. C. Gulhan, J. J. Lee, G. E. M. Melloni, I. Cortes-Ciriano, P. J. Park, Detecting the
343 mutational signature of homologous recombination deficiency in clinical samples. *Nat*
344 *Genet* **51**, 912-919 (2019).
- 345 15. N. Y. L. Ngoi, D. S. P. Tan, The role of homologous recombination deficiency testing in
346 ovarian cancer and its clinical implications: do we need it? *ESMO Open* **6**, 100144
347 (2021).
- 348 16. M. L. Telli *et al.*, Homologous Recombination Deficiency (HRD) Score Predicts
349 Response to Platinum-Containing Neoadjuvant Chemotherapy in Patients with Triple-
350 Negative Breast Cancer. *Clin Cancer Res* **22**, 3764-3773 (2016).
- 351 17. G. M. Frampton *et al.*, Development and validation of a clinical cancer genomic profiling
352 test based on massively parallel DNA sequencing. *Nat Biotechnol* **31**, 1023-1031 (2013).
- 353 18. E. H. Stover, K. Fuh, P. A. Konstantinopoulos, U. A. Matulonis, J. F. Liu, Clinical assays
354 for assessment of homologous recombination DNA repair deficiency. *Gynecol Oncol*
355 **159**, 887-898 (2020).
- 356 19. A. L. Heeke *et al.*, Prevalence of Homologous Recombination-Related Gene Mutations
357 Across Multiple Cancer Types. *JCO Precis Oncol* **2018**, (2018).

- 358 20. H. Wen *et al.*, Homologous recombination deficiency in diverse cancer types and its
359 correlation with platinum chemotherapy efficiency in ovarian cancer. *BMC Cancer* **22**,
360 550 (2022).
- 361 21. L. Nguyen, W. M. M. J. A. Van Hoeck, E. Cuppen, Pan-cancer landscape of homologous
362 recombination deficiency. *Nat Commun* **11**, 5584 (2020).
- 363 22. B. M. Norquist *et al.*, Mutations in Homologous Recombination Genes and Outcomes in
364 Ovarian Carcinoma Patients in GOG 218: An NRG Oncology/Gynecologic Oncology
365 Group Study. *Clin Cancer Res* **24**, 777-783 (2018).
- 366 23. F. Perez-Villatoro *et al.*, Optimized detection of homologous recombination deficiency
367 improves the prediction of clinical outcomes in cancer. *NPJ Precis Oncol* **6**, 96 (2022).
- 368 24. E. Rodler *et al.*, Cisplatin with veliparib or placebo in metastatic triple-negative breast
369 cancer and BRCA mutation-associated breast cancer (S1416): a randomised, double-
370 blind, placebo-controlled, phase 2 trial. *Lancet Oncol* **24**, 162-174 (2023).
- 371 25. W. Park *et al.*, Genomic Methods Identify Homologous Recombination Deficiency in
372 Pancreas Adenocarcinoma and Optimize Treatment Selection. *Clin Cancer Res* **26**, 3239-
373 3247 (2020).
- 374 26. P. A. Ascierto, C. Bifulco, G. Palmieri, S. Peters, N. Sidiropoulos, Preanalytic Variables
375 and Tissue Stewardship for Reliable Next-Generation Sequencing (NGS) Clinical
376 Analysis. *J Mol Diagn* **21**, 756-767 (2019).
- 377 27. J. N. Kather *et al.*, Deep learning can predict microsatellite instability directly from
378 histology in gastrointestinal cancer. *Nat Med* **25**, 1054-1056 (2019).
- 379 28. A. Echle *et al.*, Deep learning in cancer pathology: a new generation of clinical
380 biomarkers. *Br J Cancer* **124**, 686-696 (2021).
- 381 29. G. Campanella *et al.*, Clinical-grade computational pathology using weakly supervised
382 deep learning on whole slide images. *Nat Med* **25**, 1301-1309 (2019).
- 383 30. J. Li *et al.*, A multi-resolution model for histopathology image classification and
384 localization with multiple instance learning. *Comput Biol Med* **131**, 104253 (2021).
- 385 31. T. Lazard *et al.*, Deep learning identifies morphological patterns of homologous
386 recombination deficiency in luminal breast cancers from whole slide images. *Cell Rep*
387 *Med* **3**, 100872 (2022).
- 388 32. N. Cancer Genome Atlas, Comprehensive molecular portraits of human breast tumours.
389 *Nature* **490**, 61-70 (2012).
- 390 33. Z. Sztupinszki *et al.*, Migrating the SNP array-based homologous recombination
391 deficiency measures to next generation sequencing data of breast cancer. *NPJ Breast*
392 *Cancer* **4**, 16 (2018).
- 393 34. H. Takaya, H. Nakai, S. Takamatsu, M. Mandai, N. Matsumura, Homologous
394 recombination deficiency status-based classification of high-grade serous ovarian
395 carcinoma. *Sci Rep* **10**, 2757 (2020).
- 396 35. Z. Sztupinszki *et al.*, Comparative Assessment of Diagnostic Homologous
397 Recombination Deficiency-Associated Mutational Signatures in Ovarian Cancer. *Clin*
398 *Cancer Res* **27**, 5681-5687 (2021).
- 399 36. L. Galland *et al.*, Efficacy of platinum-based chemotherapy in metastatic breast cancer
400 and HRD biomarkers: utility of exome sequencing. *NPJ Breast Cancer* **8**, 28 (2022).
- 401 37. N. C. I. C. P. T. A. C. (CPTAC), The Clinical Proteomic Tumor Analysis Consortium
402 Breast Invasive Carcinoma Collection (CPTAC-BRCA) (Version 1) [Data set]. *The*
403 *Cancer Imaging Archive*, (2020).

- 404 38. C. Curtis *et al.*, The genomic and transcriptomic architecture of 2,000 breast tumours
405 reveals novel subgroups. *Nature* **486**, 346-352 (2012).
- 406 39. V. Baxi, R. Edwards, M. Montalto, S. Saha, Digital pathology and artificial intelligence
407 in translational medicine and clinical practice. *Mod Pathol* **35**, 23-32 (2022).
- 408 40. M. Y. Lu *et al.*, Data-efficient and weakly supervised computational pathology on whole-
409 slide images. *Nat Biomed Eng* **5**, 555-570 (2021).
- 410

Figure 1. Multi-resolution convolutional neural network architecture to detect homologous recombination deficiency from histopathological tissue slides



b

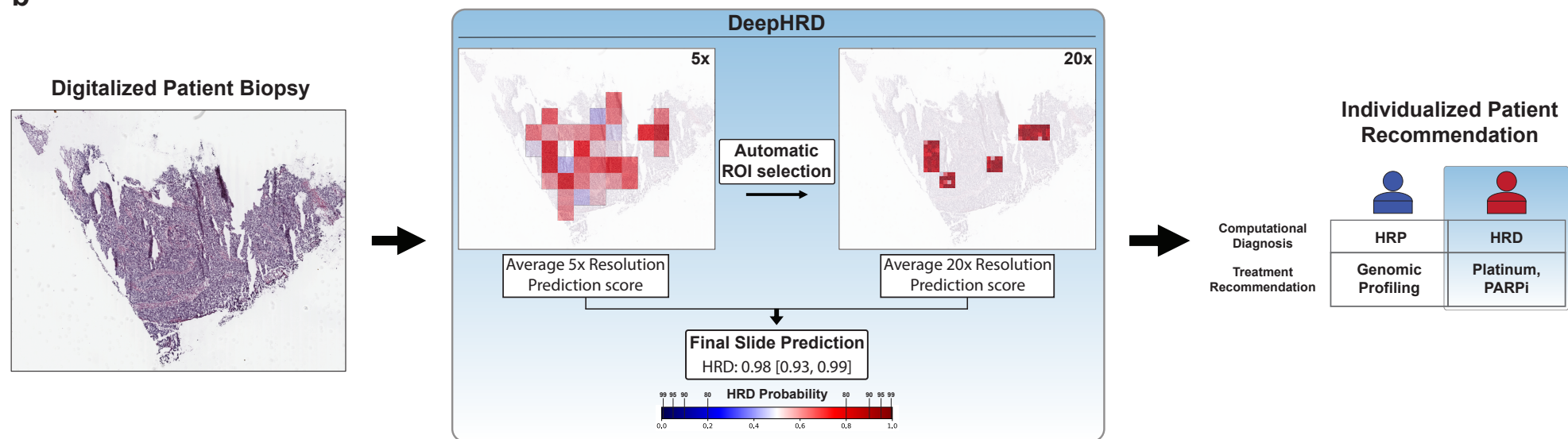


Figure 2. DeepHRD for detecting homologous recombination deficiency and predicting response to treatment in primary and metastatic breast cancer

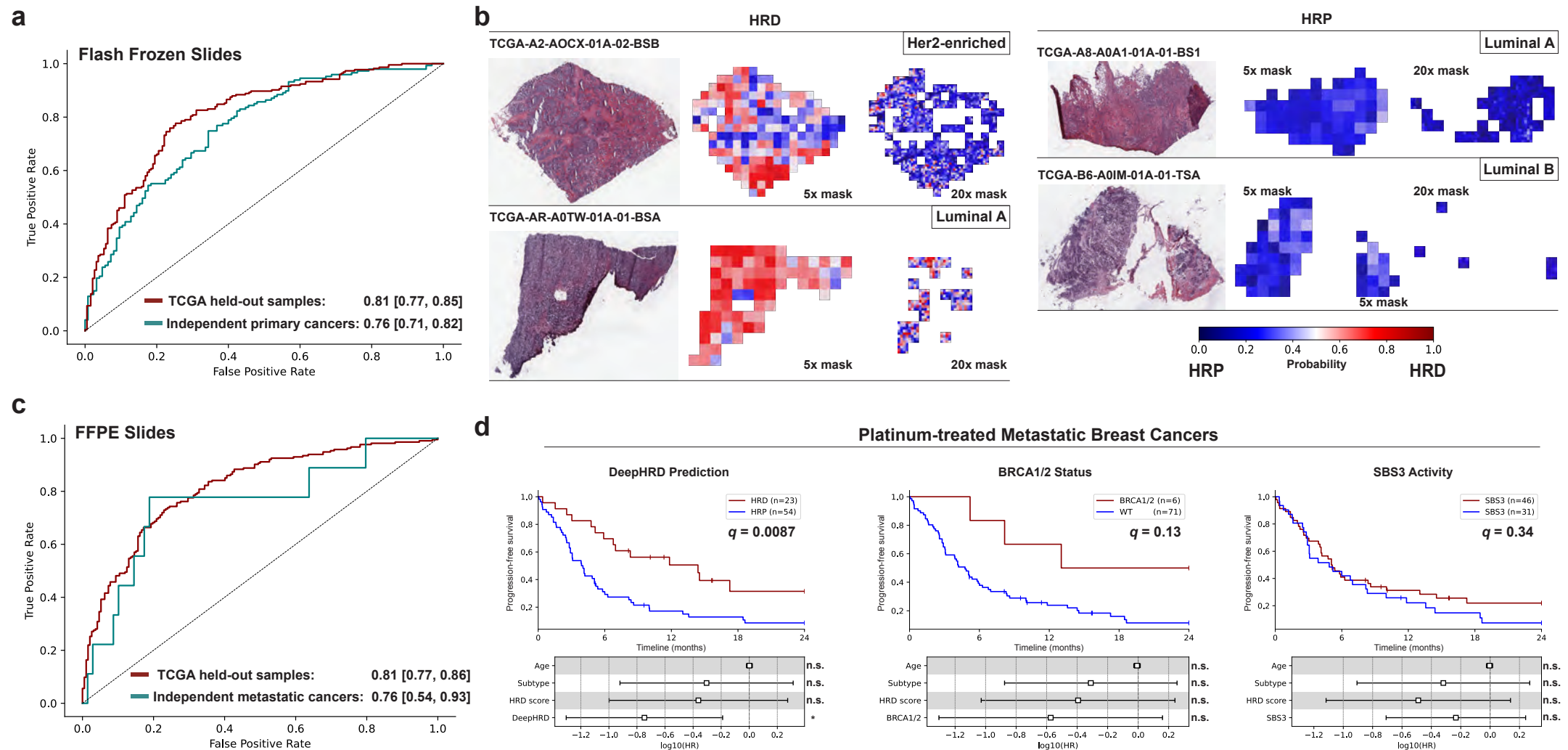
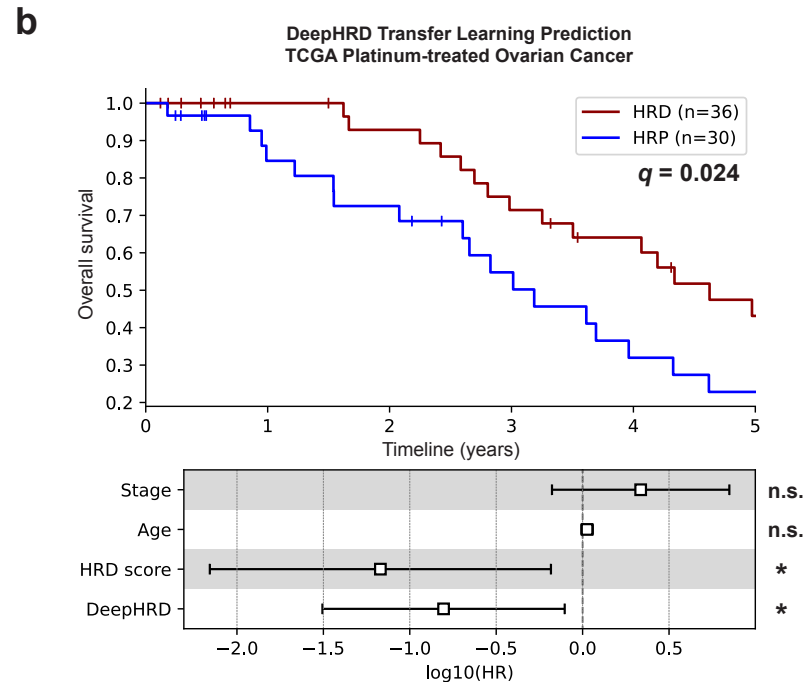
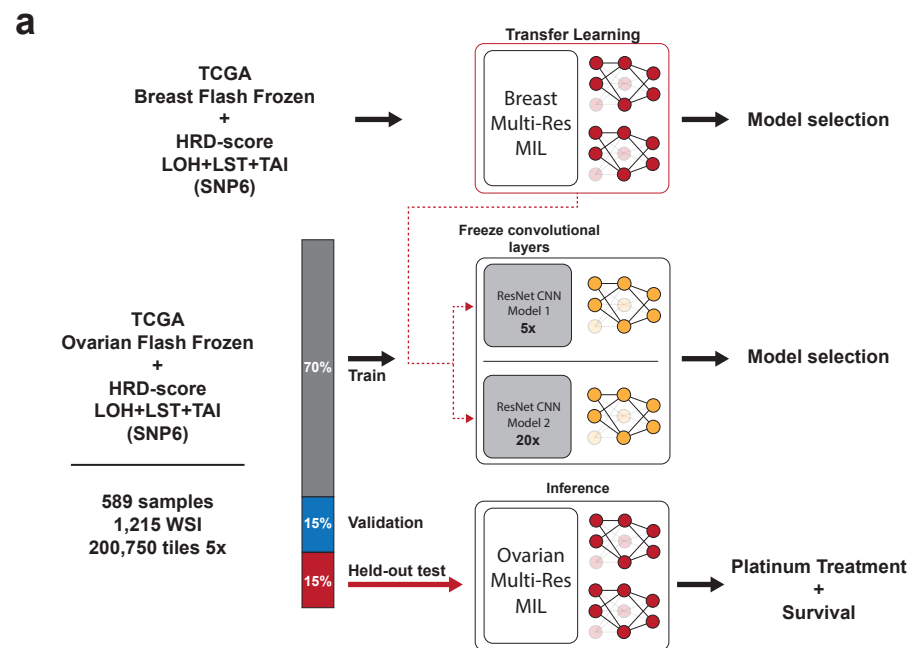


Figure 3. DeepHRD transfer learning in ovarian cancer for predicting response to platinum treatment



c

TCGA Platinum-treated Ovarian Cancer

