Optimized detection of somatic allelic imbalances specific for homologous recombination deficiency improves the prediction of clinical outcomes in ovarian cancer

Fernando Perez-Villatoro[1,2], Jaana Oikkonen[1], Julia Casado[1,2], Anastasiya Chernenko[1,2], Doga C. Gulhan (Aff), Manuela Tumiati[1], Sakari Hietanen[3], Johanna Hynninen[3], Ulla-Maija Haltia[4], Hannele Laivuori (Aff), Panagiotis Konstantinopoulos[5], Sampsa Hautaniemi[1], Liisa Kauppi[1], Anniina Färkkilä[1,2,4,5]

[1] *Research Program for Systems Oncology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland*

[2] *iCAN digital precision cancer medicine flagship*

[3] *Department of Obstetrics and Gynecology, University of Turku and Turku University Hospital, Turku,Finland.*

[4] *Department of Obstetrics and Gynecology, Helsinki University Hospital, Helsinki, Finland.*

[5] *Dana-Farber Cancer Institute, Harvard Medical School, Boston, USA*

Correspondence:
Anniina Färkkilä, MD PhD, Docent in Translational Gynecologic Oncology
Principal Investigator, Academy of Finland Clinical Research Fellow
Specialist in Obstetrics and Gynecology

Research Program for Systems Oncology &
Department of Obstetrics and Gynecology
University of Helsinki and Helsinki University Hospital
00014 University of Helsinki
anniina.farkkila@helsinki.fi
+358505967786

ABSTRACT

Specific patterns of genomic allelic imbalances (AIs) have been associated with Homologous recombination DNA-repair deficiency (HRD). We performed a systematic pan-cancer characterization of AIs across tumor types, revealing unique patterns in ovarian cancer. Using machine learning on a multi-omics dataset, we generated an optimized algorithm to detect HRD in ovarian cancer (ovaHRDscar). ovaHRDscar improved the prediction of clinical outcomes in three independent validation cohorts (PCAWG, HERCULES, TERVA). Characterization of 98 spatiotemporally distinct tumor samples indicated ovary/adnex as the preferred site to assess HRD. In conclusion, ovaHRDscar improves the detection of HRD in ovarian cancer with the premise to improve patient selection for HR-targeted therapies.

59

60INTRODUCTION

61Ovarian cancer is one of the most lethal gynecological malignancies[1], and high-grade serous 62ovarian, peritoneal or fallopian tube cancer (HGSC) is the most common subtype causing 6370-80% of all ovarian cancer-related deaths[2]. The majority of the HGSCs are diagnosed at 64an advanced stage, and the cornerstone of the treatment is first line cytoreductive surgery 65combined with DNA damaging chemotherapy. However, nearly all patients will develop a 66progressive disease and the 5-year survival is only 40%.

67

68Around half of HGSC tumors are considered to harbor genomic alterations leading to defects 69in the homologous recombination (HR) DNA repair pathway[3]. HR-deficient (HRD) tumors are 70markedly sensitive to DNA damaging therapies such as platinum and Poly-ADP Ribose 71Polymerase (PARP) inhibitors[4]. Recently, large clinical trials have reported significant 72benefits from PARP inhibitors especially in patients with HRD tumors[5,6]. The HRD test used 73in the clinical trials is test based on allelic imbalances (AIs); 1) Large scale transitions 74(LSTs), which are two AIs of at least 10Mbs separated by no more of 3Mbs[7]; 2) Loss of 75heterozygosity (LOH) regions bigger than 15Mbs[8]; 3) Telomeric allelic imbalances (TAIs), 76which are AIs that extend to the telomeres[9]. However, the defining criteria for these AIs were 77originally designed using mostly breast cancer samples, therefore the details of the genomic 78instability occurring in HRD ovarian cancers have not been clearly defined. Herein, we show 79that ovarian cancers harbor distinct patterns of AIs as compared to other cancer types 80including triple negative breast cancer (TNBC). Using an iterative machine learning 81approach and The Cancer Genome Atlas (OVA-TCGA) multi-omics dataset, we present an 82optimized algorithm (ovaHRDscar) for HRD-specific AIs in ovarian cancer. We show that 83ovaHRDscar improves the detection of HRD and prediction of therapy responses and clinical 84outcomes in three independent clinical datasets as compared to previous algorithms. 85OvaHRDScar thus shows promise as a fully optimized algorithm that can be used to improve 86patient selection of for PARP inhibitor therapies in HGSC

87RESULTS

88

89**Systematic PanCancer characterization shows unique features of allelic imbalances in**

90**ovarian cancer**

91To characterize the features of AIs across different cancer types, we analyzed AIs from

92SNP-array data in eighteen most represented cancer types available in the TCGA (**Fig. 1a**).

93We calculated the number of allelic imbalances and their median size per each sample.

94Interestingly, ovarian cancers had the highest number of AIs per sample (**Fig. 1b**), but with

95the lowest median size (**Fig. 1c**). Concordantly, ovarian cancers showed the highest levels

96of LOH events per sample (**Supplementary Fig. 1a**) although with one of the lowest median

97sizes (**Supplementary Fig. 1b).** We next performed hierarchical clustering using the median

98length and number of AIs per sample, and the skewness of the size distribution of the AIs

99per cancer type. The cancer types were clustered in two main clusters, one of those

100conformed by six cancer types (BLCA, STAD, LUSC, BRCA, LUAD, OV) with higher

101amounts of AIs but with a lower median length (upper cluster: **Fig. 1d**) and the second

102cluster conformed by the remaining twelve types of cancers (lower cluster: **Fig. 1d**). The

103same main clusters were generated when using only LOH events (**Supplementary Fig. 1c**).

104

105Then, we compared specifically differences in AIs between ovarian cancer and TNBC, as

106both cancers were used to define the HRD test used in clinical assays[10] and are enriched in

107*BRCA* gene mutations[11]. We observed that ovarian cancer samples had a higher number of

108AIs as compared to TNBC (U test, p = 0.01, **Fig. 1e**). The difference in abundance of AIs

109between ovarian cancer and TNBC was evident specifically in *BRCA*wt tumors (U test, p =

1103.2e-04, **Fig. 1f**), but not in *BRCA*mut tumors (**Fig. 1g**). Interestingly, ovarian cancers had

111lower levels of LOH events than TNBCs (U test, p = 0.01, **Supplementary Fig. 1d**), also

112among the *BRCA*mut samples (U test, p = 0.02, **Supplementary Fig. 1e**) but not in the

113*BRCA*wt samples (**Supplementary Fig. 1f**). We then compared the numbers of AIs of

114different sizes between ovarian cancer and TNBC. When compared in all samples, ovarian
115cancers showed higher levels of AIs of different sizes while TNBC had higher level of LOH
116events (**Fig. 1h**). Interestingly, LOH events longer than 30Mb were in higher proportion in
117TNBC in *BRCA*wt samples and *BRCA*mut samples. These results thus highlight the different
118characteristics of AI events in ovarian cancer, especially in the *BRCA*wt tumors, as
119compared to other cancer types.
120
121**Machine-learning optimized ovaHRDscar improves the detection of HRD in ovarian**
122**cancer**
123The specific AIs that are quantified in the HRD test (HRD-AIs) were defined using mostly
124breast cancer samples or in a mixture with ovarian cancer samples[10]. To optimize the size
125selection criteria (cutoffs) of the HRD-AIs in ovarian cancer, we utilized ovarian cancer
126samples from the TCGA (OVA-TCGA) and its associated genomic and methylomics data.
127Importantly, in contrast to previous studies that focused on distinguishing specific AIs
128associated with *BRCA1/2* mutation status, we aimed to identify AIs overrepresented in
129samples carrying associated HRD alterations (mutations, gene deletions, promoter
130hypermethylation) compared to the HR proficient (HRP) samples (**Fig. 2a**). Via a systematic
131approach utilizing prior knowledge and deep multi-omics data, we identified 93 HRD
132samples harboring a somatic or germline mutation, gene deletion or promoter
133hypermethylation in the *BRCA1/2* or *RAD51* paralog genes (**Fig. 2a**). In contrast, we
134identified 34 HRP samples that do not harbor any of the alterations used to select HRD
135samples, nor deletion in any other HR gene. A detailed description of the genomic
136alterations in the samples is reported in **Supplementary Table 1**. Overall, the HRD samples
137had a higher amount of all AIs than the HRP samples quantified using SNP array data
138(**Supplementary Fig. 2a**). Importantly, the HRD samples had a notably higher proportion of
139AIs with specific sizes that span from 1Mbs to 30Mbs (**Supplementary Figs. 2b, c**), while
140the HRP samples contained a higher proportion of AIs and LOH events smaller than 1Mbs.
141

142We next applied a decision-tree[12] machine learning classifier in an iterative approach to 143identify the specific sizes of LOH, LST and TAI events overrepresented in the HRD samples 144(**Fig. 2b**). This approach quantified the AIs of a given size, then evaluates independently 1) 145the difference in abundance of the selected AIs between the HRD versus the HRP samples 146and 2) the classification accuracy of HRD from the HRP samples by machine learning 147decision trees that took abundance of AIs as the splitting criteria. Optimization of LOH, LST 148and TAI events identified distinct AIs in ovarian cancer as compared to the HRD score used 149previously in clinical trials (hereafter called Telli2016). Importantly, the iterative approach 150increased the accuracy of classification of the HRD/HRP samples from 84.9% with the 151previous cutoffs to 89.4% with the optimized cutoffs for LOH events (**Fig. 2c**). For LSTs, the 152accuracy increased from 87.0% to 90.6% when using the optimized cutoffs (**Supplementary** 153**Fig. 2d**). We also assessed the classification accuracy of LST conformed by three 154consecutive AIs, however this produced a lower accuracy (**Supplementary Fig. 2e**). 155Similarly, for TAI events, the accuracy increased from 72.2% to 80.6% (**Supplementary Fig.** 156**2f**). As the largest improvement in accuracy of TAI occurred after including TAIs larger than 1571MBs, and inclusion of bigger sizes did not increase considerably the accuracy, we decided 158to ignore TAIs smaller than 1Mb rather than selecting TAIs with a specific size.

159

160Via systematic optimization, we observed the following AIs to be characteristic of HRD in 161ovarian cancer: 1) LOH > 15Mb and <50Mb, 2) for LSTs AI > 12Mb with a space between 162them <1Mb and 3) TAI >1Mb. The sum of these events is hereafter called the ovaHRDscar 163levels. Then, we used the distribution of ovaHRDscar levels to determine the optimal cutoff 164value to separate HRD and HRP samples. The value that best separated the two 165distributions was 54 (**Fig. 2d**), meaning that values higher than 54 correspond to HRD, with 166a higher accuracy for HR status classification (BA=0.89) as compared to the Telli2016 cutoff 167of 42 (BA=0.79).

168

169**ovaHRDscar algorithm shows concordance in WGS data and correlates with**
170**mutational signature 3**

171In order to illustrate the relationships of ovaHRDscar with other known signatures associated
172with HRD, we annotated OVA-TCGA samples according to mutations, gene deletions and
173promoter hypermethylation patterns previously reported associated with HRD[3] (**Fig. 2e**). On
174average, samples with somatic mutations in *BRCA1*, *BRCA2*, *PTEN*, or somatic mutations or
175gene deletions in any gene belonging to the Fanconi Anemia (FA), or the HR pathways
176showed high ovaHRDscar levels. Likewise, samples that contained hypermethylation in the
177promoter regions of the genes *BRCA1* or *RAD51C*, or germline mutations in *BRCA1* or
178*BRCA2* had on average high ovaHRDscar levels. As predicted, samples harboring an
179amplification in *CCNE1* (**Supplementary Fig. 2g**) had significantly lower levels of
180ovaHRDscar. Remarkably, samples with *EMSY* amplification and *CDK12* somatic mutation
181also presented with low levels. Thus, in addition to the *BRCA1/2* or *RAD51* paralogs, all the
182groups of samples with any disruption in genes related to HRD, except those positive for
183*EMSY* amplification and *CDK12* somatic mutation, contained on average significantly higher
184levels of ovaHRDscar as compared to samples with *CCNE1* amplification (**Supplementary**
185**Fig. 2g**).

186

187Next, we quantified the ovaHRDscar levels on whole genome sequencing (WGS) data of
188HGSC samples from the Pan-Cancer Analysis of Whole Genomes project (PCAWG)[13]. First,
189we evaluated the level of agreement of the ovaHRDscar levels between WGS and SNP-
190array data. For this we selected 41 OVA-TCGA samples that were also included in the
191PCAWG project. The ovaHRDscar levels were highly concordant between WGS and SNP-
192arrays (Lin's concordance correlation coefficient, ccc = 0.915; **Supplementary Fig. 2h**), this
193is consistent with a previous report in breast cancer samples that quantified AIs in the two
194platforms[14]. Next, we tested the correlation of ovaHRDscar with the single base substitution
195signature 3 (SBS3), which has been associated with HRD in human cancers samples[15]. We
196found that the ovaHRDscar levels detected in WGS positively correlated with the proportion

197of SBS3 in WGS (Pearson, r'=0.37, p= 8.4e-05; **Fig. 2f**). In addition, we observed a 198significant correlation of ovaHRDscar from SNP arrays with SBS3 inferred from whole 199exome sequencing (WES) data using SigMA, a likelihood-based approach to identify 200samples enriched in SBS3[16] in OVA-TCGA (Pearson r'=0.5, p<2.2e-16; **Supplementary Fig.** 201**2i**). The SBS3 proportions also correlated with the number of HRD-AIs using the Telli2016 202algorithm in PCAWG (**Supplementary Fig. 2j**) and in OVA-TCGA (**Supplementary Fig. 2k**). 203The ovaHRDscar algorithm detected a total 149 samples as HRD (55%) and the SigMA tool 204classified 151 samples (56%) as SBS3+; in comparison, the Telli2016 algorithm identified 205220 samples (81%) (**Fig. 2g**). Importantly, ovaHRDscar showed an improved agreement 206with SigMA (agreement 77% and cohen's kappa = 0.53) as compared to the Telli2016 207algorithm (agreement 69.3% and cohen's kappa = 0.34; **Fig. 2g**).

208

209**ovaHRDscar improves the prediction of PFS and OS compared to previous algorithms**

210Next, we measured the association of HR-status classification by ovaHRDscar with 211progression free survival (PFS, see methods) in HGSC patients treated with platinum-based 212chemotherapy from the TCGA and from an independent validation dataset (HERCULES). 213We compared the performance of ovaHRDscar versus *BRCA1/2* deficiency status, the 214Telli2016 algorithm, and to a HRD/HRP cutoff value of 63 HRD-AIs as proposed by Tacaya 215et al[17]. As *BRCA1/2* mutations can affect patient outcomes, we assessed the performances 216of ovaHRDscar in the TCGA also after excluding the samples used in the training set, even 217though clinical outcomes were not utilized for the optimization. *BRCA1/2* mutation or 218deletion status (*BRCA*mut/del) was not significantly associated with PFS (log-rank p=0.72; 219**Fig. 3a**). For OVA-TCGA (**Fig. 3a to 2c**), we found that ovaHRDscar positivity was 220associated with prolonged PFS (log-rank p=6.2e-05; **Fig. 3c**) with smaller 95% confidence 221intervals and a lower p-value as compared to the Telli2016 algorithm (log-rank p=0.0024; 222**Fig. 3b**); the median PFS was 19.3 months for the ovaHRDscar positive patients and 17.8 223months for the Telli2016 positive patients. Consistently, the ovaHRDscar positive patients 224had a longer PFS (log-rank p=0.0085; **Supplementary Fig. 3c**), while the Telli2016

225algorithm was not statistically significant in predicting PFS in the HERCULES cohort (log-
226rank p=0.051; **Supplementary Fig. 3b**).
227
228We next used Cox proportional hazard models to assess the performance of the HRD
229classification status algorithms for predicting PFS outcomes. We adjusted for the presence
230of residual tumor after the primary debulking surgery, as that has been shown to be a strong
231independent prognostic factor for PFS[18]. We found that ovaHRDcar positivity was associated
232most significantly with a prolonged PFS in OVA-TCGA when adjusting for residual tumor
233(Wald test p=9.2e−08, **Fig. 3d**) outperforming the Telli2016 algorithm (Wald test p=2.9e−07)
234and the Tacaya algorithm (Wald test p=5.2e−06), also after excluding the samples used in
235the optimization (middle panel, **Fig. 3d**) and when not adjusting for the residual tumor
236(**Supplementary Fig. 3d**). In contrast to the previous algorithms, ovaHRDscar positivity was
237associated with longer PFS also in the independent HERCULES cohort (HR 0.359, Wald
238test p=0.0061).The results were similar in the most challenging patient population after
239excluding the *BRCA*mut/del samples from the cohorts (Lower panel, **Fig. 3d**). We inspected
240the performance of the HRD-classification algorithms in a second independent prospective
241cohort (TERVA) with tumor-only SNP array profiling performed with a panel of normals (see
242methods). Importantly, ovaHRDscar positivity most significantly predicted longer PFS using
243log-rank test and cox proportional hazard model (**Supplementary Fig. 3e to 3g**). For most
244of the comparisons these results were consistent in the non-adjusted Cox regression models
245(**Supplementary Fig. 3e**).
246
247Subsequently, we next explored the association of ovaHRDscar with overall survival (OS) in
248HGSC patients in the OVA-TCGA cohort, and in an independent AU-OVA cohort in PCAWG.
249The clinical data in the prospective cohorts (HERCULES, TERVA) were not mature enough
250for OS evaluation. We found that both in the OVA-TCGA and PCAWG datasets,
251ovaHRDscar was superior to BRCA1/2 deficiency status, or to Tell2016 or Tacaya2020
252algorithms (**Fig. 3g, Supplementary Fig. 3j**). In Cox regression analysis for OS using the

253patient age at diagnosis as a covariable, ovaHRDscar had improved significance for
254predicting OS (**Fig. 3h**). Of note, the hazard-ratio values were lower in OVA-TCGA using the
255Telli2016 algorithm however the corresponding HR was not significant in the PCAWG
256cohort. Consistent with the PFS analysis, the ovaHRDscar hazard-ratios and significance
257were superior to the Tacaya2020 algorithm in all comparisons. These results were
258concordant also using a non-adjusted Cox regression analysis (**Supplementary Fig. 3k**).
259Additionally, we compared the performance of ovaHRDscar to the CHROD algorithm that
260uses structural variations and a random forest implementation to classify the HR status[19].
261We found that ovaHRDscar showed an improved prediction of OS as compared to the
262CHORD algorithm using log-rank test (**Supplementary Fig. 3l, m**) and using cox
263proportional hazard models (**Supplementary Fig. 3n**) in the PCAWG cohort.
264
265**Low intra-patient variation of ovaHRDscar in spatiotemporal tumor profiling**
266We next explored whether the anatomical site or timing of sample retrieval causes a
267potential bias in the HR status classification in ovarian cancer. For this, we explored the
268concordance of ovaHRDscar levels in the HERCULES prospective cohort, which includes 89
269tumor samples of different anatomical sites and different treatment phases (treatment naive,
270after neoadjuvant chemotherapy, or at relapse) from 33 HGSC patients (**Fig. 4a**). Consistent
271with the TCGA dataset, the ovaHRDscar levels corresponded with the known genomic
272predictors of HR status (**Fig. 4b**). Importantly, we found that the levels were similar in paired,
273anatomically matched samples obtained before and after neoadjuvant chemotherapy, and
274also in primary (treatment naive) as compared to relapsed tumors (**Fig. 4c**). Further, the
275intra-patient levels variability was not explained by the differences in tumor purity (minimum
27630%, see methods) (**Supplementary Fig. 4b** and **Supplementary Fig. 4c**). We observed
277some intra-patient variation between the anatomical sites (**Fig. 4a**), however it was lower
278than the inter-patient variation (**U test p=1.95e-38; Supplementary Fig. 4a**). To determine
279the optimal anatomical sampling site, we next assessed the HR-status per patient in
280treatment-naive primary samples and compared ovaHRDscar calculated from specific

281anatomical sites (or average if many samples per site) including 1) all samples (All) 2) 282ovarian (OVA) 3) ovarian or adnexal (OVA/ADN) 4) omentum, and OVA/ADN if omentum 283sample not available (OME-OVA/ADN) 5) OVA/ADN, and then omentum if OVA/ADN not 284available (OVA/ADN-OME). Overall, the level of agreement for the HR-status classification 285between the prioritization of anatomical sites ranged from 94% and 97% respectively 286(**Supplementary Fig. 4d**). Importantly, ovaHRDscar status calculated primarily from ovarian 287or adnexal tumor was the strongest predictor of PFS improvement (**Fig. 4d**, **Supplementary 288Fig. 4e**). Consistently, prioritizing ovarian tumors (OVA) classified all tumors harboring 289*CCNE1* amplification as HRP (**Supplementary Fig. 4d**).

290

291DISCUSSION

292HRD tumors exhibit a distinct clinical phenotype with superior responses to platinum-based 293chemotherapies and sensitivity to PARP inhibitors. Herein, we established ovaHRDscar, a 294systematically optimized biomarker for HRD to improve patient selection and clinical 295outcomes in ovarian cancer.

296

297Several genomic approaches have been utilized to detect HRD, including 1) identification of 298single genetic mutations leading to predicted HRD[21], 2) profiles of DNA repair deficiency 299gene expression[22,23], 3) specific mutational patterns accumulated due to HRD[8,9,24] or 4) 300structural genomic imbalances[7,25]. These genomic features have been implemented either 301alone, or in combinations in the search for optimal HRD detection which has profound 302therapeutic implications[26]. It is now becoming accepted that benefit from HR-directed 303therapeutic such as PARP inhibitors extend beyond the identification of HRD via individual 304genetic mutations[27]. This is due to the facts that and genes such as *BRCA1/2* and *RAD51*, 305can be altered beyond mutations e.g. via hypermethylation or gene deletions[3,28], and not all 306genomic events leading to HRD have yet been defined[29]. Allelic imbalances are indicative of 307the genetic consequences of HRD and, although not dynamically reflective of tumors

308functional HRD status, have shown promise as a biomarker predictive for the magnitude of
309benefit from PARP inhibitors, especially in the front-line setting[30,31]. The HRD test used in the
310ovarian cancer clinical trials was, however, optimized using breast cancer samples or a
311mixture of breast cancer and ovarian cancer samples using *BRCA1/2* mutation as the sole
312denominator of HRD, and *BRCA*wt status as HRP[8,9,10]. Importantly, also European Society of
313Medical Oncology indicated an urgent need to optimize the HRD test in ovarian cancer to
314improve the identification of especially the HRP tumors[23]. Via a pan-cancer characterization
315of AIs, we observed marked differences in the AI patterns of ovarian cancer as compared to
316other cancer types including TNBC, especially among the *BRCA*wt tumors. This prompted
317us to systematically identify the genomic footprints of HRD-AIs in ovarian cancer using
318carefully annotated multi-omics data from the TCGA and an iterative machine learning and
319statistical approach.
320
321ovaHRDscar levels were concordant with tumor genetic mutations and copy number events
322predictive of HRD status both in the TCGA dataset, and in an external validation cohort
323(HERCULES). We found significantly lower levels of ovaHRDscar in tumors with *CCNE1*
324amplification, proposed as mutually exclusive with HRD and associated with poor clinical
325outcomes[32]. In line with a previous report[17], tumors with *CDK12* mutation showed overall low
326levels of ovaHRDscar. Further, ovaHRDscar showed a higher concordance with mutational
327signature 3 as compared to the Telli2016 algorithm. Importantly, ovaHRDscar can be applied
328to detect HRD in HGSC samples using WGS or SNP-arrays, making it an attractive
329biomarker for the clinical setting.
330
331A dichotomous thresholding of a predictive HRD biomarker is needed for therapeutic
332decision making. In Telli2016, the cutoff for the total number of events was derived from a
333mixture of breast and ovarian cancer samples[10]. More recently, Tacaya et al set out to
334improve the HRD test by adjusting the cutoff value in ovarian cancer[17], however only
335*BRCA*mut status was used for separating the HRD from HRP samples and the same

336genomic features of HRD-AIs were used as in Telli et al. In ovaHRDscar, after optimization
337of both the genomic footprints of HRD-AIs, and the cutoff, we identified more samples as
338being HRP, and reliably separated HRD from HRP with improved accuracy as compared to
339the previous algorithms. In survival analyses, our fully optimized ovaHRDscar outperformed
340the two previous algorithms, indicating that both the identification of the genomic footprints of
341HRD-AIs, and the selection of the optimal cutoff are needed to improve HRD detection in
342ovarian cancer.
343
344HRD tumors are known to have superior responses to platinum-based chemotherapy and
345prolonged overall survival[33]. Consistently, ovaHRDscar improved the prediction of PFS and
346OS for platinum-based chemotherapy in the OVA-TCGA, also after excluding the patients
347used in the training set, even though clinical outcomes were not used in the optimization.
348ovaHRDscar significantly predicted PFS and OS also among only the *BRCA*wt tumors.
349Importantly, ovaHRDscar improved the prediction of clinical outcomes in two independent
350patient cohorts, and in multivariable models after adjusting for clinical covariables, indicating
351that ovaHRDscar reliably captures the phenotypic clinical behavior of HRD ovarian cancers.
352ovaHRDscar levels correlated with genomic features and responses to platinum-based
353chemotherapy, however none of the clinical cohorts included patients treated with PARP
354inhibitors, and thus the potential predictive biomarker role of ovaHRDscar would require
355prospective validation in larger patient series.
356
357Finally, as HGSC is characterized by marked intratumor heterogeneity, we to aimed to
358assess whether the anatomical site of tumor sampling or the exposure to chemotherapy
359affect HRD assessment. Our analysis of 98 samples collected from different anatomical sites
360and treatment phases indicated that ovaHRDscar levels remain similar in spatiotemporally
361distinct samples from each patient, including anatomically site-matched samples collected
362before and after neoadjuvant chemotherapy. ovaHRDscar can thus be reliably assessed
363during routine clinical practice also after neoadjuvant chemotherapy, given that the tumor

364purity remains higher than 30%. Interestingly, the ovaHRDscar levels were also similar
365between treatment naïve and relapsed tumors, reflecting the nature of HRD-AIs as historical
366consequences rather than dynamic readouts of functional HRD. Overall, the interpatient
367variation was higher than the intrapatient variation, however in four out of 21 (19%) of
368patients the HRD category was dependent on the sampling site/time. The survival and
369genetic analyses indicated that ovarian or adnexal sites followed by omentum could be the
370preferred sites for HRD testing, warranting future validation in larger cohorts.
371
372In conclusion, ovaHRDscar shows promise as a fully optimized, robust, clinically feasible
373assay for both outcome prediction and selection of patients to HR-directed therapies. With
374the fully documented, publicly available algorithms and optimization pipeline our method can
375be implemented clinically and applied to other tumor types for optimized patient selection
376and to improve outcomes in cancer.

377MATERIAL AND METHODS

378**Data set collection and classification**

379For PanCan samples, Allele-specific copy number segments were obtained from the
380Genomics Data Commons (GDC) portal (https://portal.gdc.cancer.gov/). TNBC were defined
381as those breast cancers marked as negative by *er_status_by_ihc*, *pr_status_by_ihc* and
382*her2_status_by_ihc* in the clinical file "*nationwidechildrens.org_clinical_patient_brca.txt*"
383available in the GDC. For TNBC, samples were considered with *BRCA*mut if reported in
384Knijnenburg et al 2018[34] with gene deletion, gene mutation or gene silencing in *BRCA*1 or
385*BRCA*2 genes; while *BRCA*wt those reported with information of no alterations.
386
387For OVA-TCGA analysis: Allele-specific copy number segments, DNA methylation, gene-
388level CNV profiles and clinical information data was obtained from the GDC data portal.
389Genes were considered with a "strong signal of deletion" if those were reported (labeled by -
3902) by Taylor et al 2018[35] and with signal of deletion those reported in the GDC portal. Gene

391promoter hypermethylation was considered when the probes upto 1500bp downstream the 392transcription start site had an average beta value ≥0.75. The catalog of mc3 somatic 393mutations was obtained from the PanCanAtlas-GDC data portal 394(https://gdc.cancer.gov/about-data/publications/pancanatlas). Somatic mutations were 395classified according to the recommendations of the American College of Medical Genetics 396and Genomics[36] using the web-tool VarSome[37] only pathogenic somatic mutations were 397considered in the analysis. For germline mutations we selected those labeled as pathogenic 398and prioritized by Huang et al 2018[38]. Genes were considered part of the HR pathway, or 399other associated pathways according to the Kyoto Encyclopedia of Genes and Genomes 400database[39]. Complementary clinical information was obtained from the PanCanAtlas-GDC 401data portal. For PCAWG: Allele-specific copy number segments, mutational drivers and 402clinical information was obtained from the International Cancer Genome Consortium data 403portal (https://dcc.icgc.org/pcawg).

404

405**HRD-AI optimization**

406The identification of HRD-AIs[10] was performed using an R package based on a script 407previously reported[40]. Samples were classified as HRD and HRP according to Fig. 1A. LOH 408events of different minimum sizes i and maximum sizes L, Large scale transitions (LSTS: 409two allelic imbalances of a size S separated by a genomic distance no longer than M) of 410different S and M values and telomeric allelic imbalances (TAI) of minimum size t were 411selected by an R script. The mean difference in abundance of the selected AIs between 412HRD and HRP samples, was calculated using Mann Whitney U test. Decision trees were 413used to classify samples as HRD and HRP with the R package 'rpart' according to the 414abundance of the selected AIs (Figure 1B). Accuracy of the classifier was assessed using 415balanced accuracy. The values i, L, S, M, t, with the lower U test p-value and best accuracy 416for classification were selected for ovaHRDscar.

417

418**Survival analysis**

419Survival plots, log-rank test and cox regression models were performed with an R script
420using the package 'survminer' and 'survival'. Data from patients treated by cisplatin or
421carboplatin was selected. For PCAWG, data from all patients was used (no treatment
422information available). Only data from primary samples (treatment naive) were used.
423*BRCA*mut/del status included pathogenic mutations, germline mutations and gene deletions
424with a "strong signal of deletion" as described above. Residual tumor after surgery was
425categorized as present or absent. For indicated cox regressions, the residual tumor status or
426patient age at diagnosis was used as a covariable. Progression free survival (PFS) and
427overall survival (OS) was defined as in Liu et al 2018[41]. The CHROD signature HR status
428classification for PCAWG samples was adopted from Nguyen et al 2020[19].
429
430**Prospective HERCULES and TERVA data analysis**
431The tumor samples were prospectively collected in the HERCULES (http://www.project-
432hercules.eu) and TERVA projects (https://www.healthcampusturku.fi/innovation-new/terva-
433project/). The Ethics Committee of the Hospital District of Southwest Finland, approved both
434studies ETMK Dnro: 145 /1801/2015. All patients gave their informed consent. For
435HERCULES, paired fresh tumor samples and blood normal samples were sequenced using
436Illumina-Hiseq WGS. Alignment to the human genome GRCh38 was done using the burrows
437wheeler aligner BWA-MEM[42]. Mutations were detected using GATK4-Mutect2 approach[43].
438Mutect2 was used for detection of allele specific copy numbers. Calculation of tumor purity
439was estimated in two ways: 1) Based in somatic copy-number profiles using the software
440Ascat[44] 2) Based in *TP53* variant allele frequency (*TP53*-VAF), purity was estimated using
441the formula: *2 / ((CN / TP53-VAF) - (CN - 2)),* where CN corresponds to the copy-number
442value in the corresponding somatic mutations in *TP53*; subsequently the higher purity value
443was selected. For the TERVA samples tumor-only profiling, the tumor samples were
444genotyped using the Infinium™ Global Screening Array-24 v2.0. B allele frequency and
445LogR ratios per sample probe was calculated using Illumina-GenomeStudio. Ascat software
446was used for the detection of allele specific copy numbers, *ascat.predictGermlineGenotypes*

module was performed adjusting parameters according to a panel of 200 normal germline blood samples.

Statistics

The statistics analysis was performed in R. Difference in mean abundances was calculated using Mann Whitney U test. The distribution Skewness of AI sizes in different types of cancers was performed using the package DescTools. Agreement was calculated using the Cohen kappa test. Concordance was measured using Lin's concordance correlation coefficient. Level of correlations was assessed using Pearson correlations. P value less than 0.05 was considered statistically significant.

Code Availability

The code used to perform the AI optimization is available on Github: (https://github.com/farkkilab/findHRDsignatures), ovaHRDscar implementation is available as and R package on Github (https://github.com/farkkilab/ovaHRDscar).

Data availability

Data available... ?

ACKNOWLEDGMENTS

Sigrid Jusélius Foundation (AF), Finnish cancer society (AF), Academy of Finland (AF), Paolo Foundation (AF), The Finnish Medical Foundation (AF), Finnish Cultural Foundation (AF), Instrumentarium Foundation (AF), University of Helsinki (AF), the European Union's Horizon 2020 research and innovation program under grant agreement No 667403 for HERCULES.

473AUTHOR CONTRIBUTION

474

475COMPETING INTEREST

476The authors declare no competing interests

477

478Supplementary Information is available for this paper

479

480REFERENCES

4811 Reid, B. M., Permuth, J. B. & Sellers, T. A. Epidemiology of ovarian cancer: a review.
482 *Cancer Biol Med* **14**, 9-32, doi:10.20892/j.issn.2095-3941.2016.0084 (2017).

4832 Torre, L. A. *et al.* Ovarian cancer statistics, 2018. *CA Cancer J Clin* **68**, 284-296,
484 doi:10.3322/caac.21456 (2018).

4853 Konstantinopoulos, P. A., Ceccaldi, R., Shapiro, G. I. & D'Andrea, A. D. Homologous
486 Recombination Deficiency: Exploiting the Fundamental Vulnerability of Ovarian
487 Cancer. *Cancer Discov* **5**, 1137-1154, doi:10.1158/2159-8290.CD-15-0714 (2015).

4884 Lord, C. J. & Ashworth, A. PARP inhibitors: Synthetic lethality in the clinic. *Science*
489 **355**, 1152-1158, doi:10.1126/science.aam7344 (2017).

4905 Ray-Coquard, I. *et al.* Olaparib plus Bevacizumab as First-Line Maintenance in
491 Ovarian Cancer. *N Engl J Med* **381**, 2416-2428, doi:10.1056/NEJMoa1911361
492 (2019).

4936 Gonzalez-Martin, A. *et al.* Niraparib in Patients with Newly Diagnosed Advanced
494 Ovarian Cancer. *N Engl J Med* **381**, 2391-2402, doi:10.1056/NEJMoa1910962
495 (2019).

4967 Popova, T. *et al.* Ploidy and large-scale genomic instability consistently identify
497 basal-like breast carcinomas with BRCA1/2 inactivation. *Cancer Res* **72**, 5454-5462,
498 doi:10.1158/0008-5472.CAN-12-1470 (2012).

4998 Abkevich, V. *et al.* Patterns of genomic loss of heterozygosity predict homologous
500 recombination repair defects in epithelial ovarian cancer. *Br J Cancer* **107**, 1776-
501 1782, doi:10.1038/bjc.2012.451 (2012).
5029 Birkbak, N. J. *et al.* Telomeric allelic imbalance indicates defective DNA repair and
503 sensitivity to DNA-damaging agents. *Cancer Discov* **2**, 366-375, doi:10.1158/2159-
504 8290.CD-11-0206 (2012).
50510 Telli, M. L. *et al.* Homologous Recombination Deficiency (HRD) Score Predicts
506 Response to Platinum-Containing Neoadjuvant Chemotherapy in Patients with Triple-
507 Negative Breast Cancer. *Clin Cancer Res* **22**, 3764-3773, doi:10.1158/1078-
508 0432.CCR-15-2477 (2016).
50911 Stefansson, O. A. *et al.* Genomic profiling of breast tumours in relation to BRCA
510 abnormalities and phenotypes. *Breast Cancer Res* **11**, R47, doi:10.1186/bcr2334
511 (2009).
51212 Breiman, L. *Classification and regression trees*. (Wadsworth International Group,
513 1984).
51413 Consortium, I. T. P.-C. A. o. W. G. Pan-cancer analysis of whole genomes. *Nature*
515 **578**, 82-93, doi:10.1038/s41586-020-1969-6 (2020).
51614 de Luca, X. M. *et al.* Using whole-genome sequencing data to derive the homologous
517 recombination deficiency scores. *NPJ Breast Cancer* **6**, 33, doi:10.1038/s41523-020-
518 0172-0 (2020).
51915 Alexandrov, L. B. *et al.* The repertoire of mutational signatures in human cancer.
520 *Nature* **578**, 94-101, doi:10.1038/s41586-020-1943-3 (2020).
52116 Gulhan, D. C., Lee, J. J., Melloni, G. E. M., Cortes-Ciriano, I. & Park, P. J. Detecting
522 the mutational signature of homologous recombination deficiency in clinical samples.
523 *Nat Genet* **51**, 912-919, doi:10.1038/s41588-019-0390-2 (2019).
52417 Takaya, H., Nakai, H., Takamatsu, S., Mandai, M. & Matsumura, N. Homologous
525 recombination deficiency status-based classification of high-grade serous ovarian
526 carcinoma. *Sci Rep* **10**, 2757, doi:10.1038/s41598-020-59671-3 (2020).

52718 Polterauer, S. *et al.* Prognostic value of residual tumor size in patients with epithelial
528 ovarian cancer FIGO stages IIA-IV: analysis of the OVCAD data. *Int J Gynecol*
529 *Cancer* **22**, 380-385, doi:10.1097/IGC.0b013e31823de6ae (2012).
53019 Nguyen, L., J, W. M. M., Van Hoeck, A. & Cuppen, E. Pan-cancer landscape of
531 homologous recombination deficiency. *Nat Commun* **11**, 5584, doi:10.1038/s41467-
532 020-19406-4 (2020).
53320 Barretina-Ginesta, M. P. DNA damaging agents in ovarian cancer. *EJC Suppl* **15**, 67-
534 72, doi:10.1016/j.ejcsup.2020.06.001 (2020).
53521 Wagle, N. *et al.* High-throughput detection of actionable genomic alterations in
536 clinical tumor samples by targeted, massively parallel sequencing. *Cancer Discov* **2**,
537 82-93, doi:10.1158/2159-8290.CD-11-0184 (2012).
53822 Kang, J., D'Andrea, A. D. & Kozono, D. A DNA repair pathway-focused score for
539 prediction of outcomes in ovarian cancer treated with platinum-based chemotherapy.
540 *J Natl Cancer Inst* **104**, 670-681, doi:10.1093/jnci/djs177 (2012).
54123 Konstantinopoulos, P. A. *et al.* Gene expression profile of BRCAness that correlates
542 with responsiveness to chemotherapy and with outcome in patients with epithelial
543 ovarian cancer. *J Clin Oncol* **28**, 3555-3561, doi:10.1200/JCO.2009.27.5719 (2010).
54424 Wang, Y. K. *et al.* Genomic consequences of aberrant DNA repair mechanisms
545 stratify ovarian cancer histotypes. *Nat Genet* **49**, 856-865, doi:10.1038/ng.3849
546 (2017).
54725 Polak, P. *et al.* A mutational signature reveals alterations underlying deficient
548 homologous recombination repair in breast cancer. *Nat Genet* **49**, 1476-1486,
549 doi:10.1038/ng.3934 (2017).
55026 Swisher, E. M. *et al.* Rucaparib in relapsed, platinum-sensitive high-grade ovarian
551 carcinoma (ARIEL2 Part 1): an international, multicentre, open-label, phase 2 trial.
552 *Lancet Oncol* **18**, 75-87, doi:10.1016/S1470-2045(16)30559-9 (2017).

55327 Miller, R. E. *et al.* ESMO recommendations on predictive biomarker testing for
554 homologous recombination deficiency and PARP inhibitor benefit in ovarian cancer.
555 *Ann Oncol* **31**, 1606-1622, doi:10.1016/j.annonc.2020.08.2102 (2020).
55628 Golmard, L. *et al.* Contribution of germline deleterious variants in the RAD51
557 paralogs to breast and ovarian cancers. *Eur J Hum Genet* **25**, 1345-1353,
558 doi:10.1038/s41431-017-0021-2 (2017).
55929 Radhakrishnan, S. K., Jette, N. & Lees-Miller, S. P. Non-homologous end joining:
560 emerging themes and unanswered questions. *DNA Repair (Amst)* **17**, 2-8,
561 doi:10.1016/j.dnarep.2014.01.009 (2014).
56230 Ray-Coquard, I. *et al.* Olaparib plus Bevacizumab as First-Line Maintenance in
563 Ovarian Cancer. *N Engl J Med* **381**, 2416-2428, doi:10.1056/NEJMoa1911361
564 (2019).
56531 Gonzalez-Martin, A. *et al.* Niraparib in Patients with Newly Diagnosed Advanced
566 Ovarian Cancer. *N Engl J Med* **381**, 2391-2402, doi:10.1056/NEJMoa1910962
567 (2019).
56832 Etemadmoghadam, D. *et al.* Synthetic lethality between CCNE1 amplification and
569 loss of BRCA1. *Proc Natl Acad Sci U S A* **110**, 19489-19494,
570 doi:10.1073/pnas.1314302110 (2013).
57133 Norquist, B. M. *et al.* Mutations in Homologous Recombination Genes and Outcomes
572 in Ovarian Carcinoma Patients in GOG 218: An NRG Oncology/Gynecologic
573 Oncology Group Study. *Clin Cancer Res* **24**, 777-783, doi:10.1158/1078-0432.CCR-
574 17-1327 (2018).
57534 Knijnenburg, T. A. *et al.* Genomic and Molecular Landscape of DNA Damage Repair
576 Deficiency across The Cancer Genome Atlas. *Cell Rep* **23**, 239-254 e236,
577 doi:10.1016/j.celrep.2018.03.076 (2018).
57835 Taylor, A. M. *et al.* Genomic and Functional Approaches to Understanding Cancer
579 Aneuploidy. *Cancer Cell* **33**, 676-689 e673, doi:10.1016/j.ccell.2018.03.007 (2018).

58036 Richards, S. *et al.* Standards and guidelines for the interpretation of sequence
581 variants: a joint consensus recommendation of the American College of Medical
582 Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**,
583 405-424, doi:10.1038/gim.2015.30 (2015).
58437 Kopanos, C. *et al.* VarSome: the human genomic variant search engine.
585 *Bioinformatics* **35**, 1978-1980, doi:10.1093/bioinformatics/bty897 (2019).
58638 Huang, K. L. *et al.* Pathogenic Germline Variants in 10,389 Adult Cancers. *Cell* **173**,
587 355-370 e314, doi:10.1016/j.cell.2018.03.039 (2018).
58839 Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic*
589 *Acids Res* **28**, 27-30, doi:10.1093/nar/28.1.27 (2000).
59040 Marquard, A. M. *et al.* Pan-cancer analysis of genomic scar signatures associated
591 with homologous recombination deficiency suggests novel indications for existing
592 cancer drugs. *Biomark Res* **3**, 9, doi:10.1186/s40364-015-0033-4 (2015).
59341 Liu, J. *et al.* An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-
594 Quality Survival Outcome Analytics. *Cell* **173**, 400-416 e411,
595 doi:10.1016/j.cell.2018.02.052 (2018).
59642 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler
597 transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
59843 Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and
599 heterogeneous cancer samples. *Nat Biotechnol* **31**, 213-219, doi:10.1038/nbt.2514
600 (2013).
60144 Van Loo, P. *et al.* Allele-specific copy number analysis of tumors. *Proc Natl Acad Sci*
602 *U S A* **107**, 16910-16915, doi:10.1073/pnas.1009843107 (2010).
603

604FIGURE LEGENDS

605**Figure 1. Pan-cancer characterization of AIs reveal unique patterns in ovarian cancer.**
606**a** Types of cancer with more than 200 samples selected from TCGA, their corresponding
607number of samples are shown in the bars. **b** Number of AIs bigger than 3Mb and smaller

608than 50Mb per sample, ovarian cancer shows the highest abundance of AIs. **c** Median 609length of AIs (bigger than 3Mb and smaller than 50Mb) per sample, ovarian cancer shows 610the lowest median size per sample. **d** Hierarchical clustering for the types of cancer using 611as variables the median length, median number of AIs per sample and the skewness of the 612distribution of AIs sizes. **e** Ovarian cancer samples show a higher amount of AIs than TNBC 613(U test, p=0.01). **f** Comparison of *BRCA*mut samples (somatic deletion, promoter 614hypermethylation and gene deletions of *BRCA*1/2 genes), shows not significantly higher 615amounts (U test) of AIs in ovarian cancer than in TNBC. **g** Comparison of *BRCA*wt samples, 616shows higher amounts of AIs in ovarian cancer than in TNBC (U test, p=1.1e-04). **h** 617Difference in abundance for AIs of specific sizes between ovarian cancer and TNBC, dot 618sizes correspond to U test p values and dot colors to the fold change (ratio of ovarian-TNBC 619median abundance of AIs minus one), only dots for corresponding significantly differences 620are shown (U test, $p < 0.05$).

621

622

623**Figure 2. Optimization of size selection for AIs associated with HRD in ovarian cancer 624and relationship with HRD biomarkers. a** Selection criteria for HRD, HRP and Undefined 625samples in OVA-TCGA. **b** A schematic of the approach used to optimize the size selection 626criteria (cutoff) of HRD-AIs in ovarian cancer samples. Iteratively, AIs of different sizes were 627taken, in each iteration it was evaluated: 1) the difference of abundance (U test) of the 628selected AIs between HRD versus HRP samples, 2) the accuracy of a decision tree classifier 629for HR status taking as splitting criteria the abundance of the selected AIs; all results were 630stored and the best cutoffs were selected for the ovaHRDscar. **c** For LOH events, the size of 631dots represent the decision tree balanced accuracy (BA) when using the corresponding AIs 632sizes, colors correspond to the statistical (U test, p value) difference in abundance of AIs 633between HRD versus HRP samples; black box correspond to the selection criteria proposed 634by Telli2016, blue box correspond to the best BA value. **d** Density distribution of HRD-AIs 635according to Telli2016 and ovaHRDscar algorithms. The red dashed line represents the

636cutoff established to define HRD or HRP samples for Telli2016 (>= 42) and for ovaHRDscar
637(>= 54), the blue dashed line in the left panel represents the cutoff proposed in Tacaya et al.
6382020 (>= 63). The balanced accuracy (BA) for classifying the HR status is shown for
639Telli2016 and ovaHRDscar. **e** Levels of ovaHRDscar in OVA-TCGA samples harboring
640different genetic or epigenetic alterations associated with HRD in ovarian cancer[3]. In the
641outer ring of the pie chart every line represents a sample; in the center of the pie chart the
642colors correspond to the average number of HRD-AIs per genetic or epigenetic alteration.
643Somatic mutations included gene deletions. **f** The proportion of single base substitution
644signature 3 (SBS3) is correlated (Pearson r'=0.37) with the number of ovaHRDscar in
645PCAWG samples. **g** The SBS3 status inferred by SigMA has a higher agreement with
646ovaHRDscar (agrement=77%, cohen's kappa = 0.53) than with the Telli2016 algorithm
647(agrement=69.3%, cohen's kappa = 0.34), in the pie charts and table + and - correspond
648respectively to the number of HRD positive and HRD negative samples identified under each
649criteria. In the button is shown the number of samples selected under the corresponding
650criteria and the level of agreement.
651
652**Figure 3. HRD detected by ovaHRDscar is associated with higher PFS and OS in**
653**HGSC patients treated with platinum. a to c** Kaplan-Meier plots of PFS in OVA-TCGA
654patients stratified with different criteria, in **a** patients were stratified according to
655*BRCA*mut/del status with no significant difference in their PFS probability over time (log-
656rank, p=0.72); in **b** patients were stratified according to Telli2016 (log-rank, p=0.0024); in **c**
657patients were stratified using the ovaHRDscar algorithm, HRD patients had a improvement
658in PFS time over HRP (log-rank, p=6.2e-05). **d** Cox regression results for the association of
659PFS with several genomic HR status classification criteria (*BRCA*mut/del, Telli2016,
660Tacaya2020, ovaHRDscar) in two independent cohorts, for every regression it was used as
661covariable the residual tumor status (yes or no) after surgery. Three panels are shown: at left
662the OVA-TCGA cohort, at the middle OVA-TCGA cohort excluding the HRD and HRP
663samples used for the AIs optimization, at the right the Hercules WGS prospective cohort.

664Inside each panel is shown for classification criteria: the number of patients (N) selected as
665positive or HRD and their corresponding proportion (Prop), the hazard-ratio (HR) for the cox
666regression and the p-value (Pval) of the regression. ovaHRDscars show higher significance
667(lower p values) in every cohort even after excluding patients with tumors harboring
668*BRCA*mut/del (bottom section). **e to g,** Kaplan-Meier plots of OS for OVA-TCGA patients
669selected using different criteria, ovaHRDscar show the higher significance (lower long-rank
670test p-values). **h** Cox regression results for the association of OS with several genomic HR
671status classification criteria (*BRCA*mut/del, Telli2016, Tacaya2020, ovaHRDscar) in two
672independent cohorts. Panel at the right corresponds to the AU-PCAWG samples, the left
673and center panels and their corresponding columns correspond to those described in Fig.
6743d. ovaHRDscars show higher significance (lower p values) in every cohort even after
675excluding patients with tumors harboring *BRCA*mut/del (bottom section).
676
677**Figure 4. Low intra-patient anatomical variation of ovaHRDscar. a** An overview of the
678samples and their corresponding ovaHRDscar levels per patient inside a prospective cohort
679analyzed by WGS (HERCULES). Per patient tumor samples were taken in three different
680treatment times and from different anatomical sites as described; in parentheses are
681displayed the corresponding number of samples. **b** Levels of ovaHRDscar in samples
682harburing different genetic or epigenetic alterations associated with HRD, in the outer ring of
683the pie chart every bar represents a sample carrying the corresponding alteration; in the
684center of the pie chart the colors correspond to the average levels of ovaHRDscar per
685alteration. **c** ovaHRDscar values between paired samples for each patient (connected dots)
686does not vary significantly (paired U test) over treatment phase. **d** Cox regression results for
687the association of PFS with HR status classification by Telli2016 and ovaHRDscar, the HR
688status per patient was assessed using three anatomical sample prioritization approaches: 1)
689average HRD-AIs per all samples 2) omentum, and OVA/ADN if omentum sample not
690available (OME-OVA/ADN) 3) OVA/ADN, and then omentum if OVA/ADN not available
691(OVA/ADN-OME). Last prioritization approach has a higher association with PFS.

692

693SUPPLEMENTARY FIGURE LEGENDS

694

695**Supplementary Figure 1. Pan-cancer characterization of LOH and unique patterns in**
696**ovarian cancer. a.** Boxplot for the number of LOH bigger than 3Mb and smaller than 50Mb
697per sample, ovarian cancer shows the higher abundance of LOH. **b** Median length of LOH
698(bigger than 3Mb and smaller than 50Mb) per sample. **c** Hierarchical clustering of the types
699of cancer using as variables the median length, median number of LOH per sample and the
700skewness of the distribution of LOH sizes. **d** Ovarian cancer samples show a lower amount
701of LOH than TNBC (U test, p=0.005). **e** Comparison of *BRCA*mut samples (somatic deletion,
702promoter hypermethylation and gene deletions of *BRCA*1/2 genes), ovarian cancer shows
703higher amounts of LOH than TNBC (U test, p=0.044). **f** Comparison of *BRCA*wt samples, not
704significant abundance difference of LOH (U test, NS) between ovarian cancer and TNBC.

705

706**Supplementary Figure 2. Descriptive statistics of HRD and HRP ovarian samples,**
707**optimization of HRD selection. a** Number of allelic imbalances in HRD and HRP samples
708in OVA-TCGA, in average HRD samples contain a higher number of allelic imbalances (U
709test, p value=1.99e-05). **b** The average proportion of segments (AI or normal) equal or
710greater than the given size in HRD (red) and HRP (blue) samples, the blue and red lines
711correspond to smoothing using cubic splines. **c** Density distribution of LOH events in HRD
712(red) and HRP (blue) samples. **d** LSTs specific AIs size cutoffs (blue boxes) uperform the
713reported by Telli et al 2016 (black box); the size of dots represent the decision tree balanced
714accuracy (BA) when using the corresponding cutoff, colors correspond to the statistical (U
715test p value) difference in abundance of AIs between HRD versus HRP samples. **e** Statistical
716significance for identifying HRD samples using three tandem allelic imbalances LSTs of a
717given minimum size (x axis) and distance between them smaller than 1 to 4 Mb (y axis). Dot
718sizes and colors resemble those in Supplementary Figure 1e. **f** The selection of TAIs >1Mb

719differentiate HRD from HRP samples better than using all size TAIs, the upper panel shows
720statistical (U test, p values) difference in abundance of AIs between HRD versus HRP
721samples when selecting the corresponding TAIs, dot sizes and colors in the lower panel
722correspond to the description in panel d and e. **g** ovaHRDscar levels per genomic alteration
723in OVA-TCGA, U test p values are shown for the comparison of levels between the
724corresponding alterations versus CCNE1 amplification. **h** Bland-Altman plot that shows the
725concordance (Concordance correlation coefficient, *CCC* = 0.915) between the number of
726ovaHRDscars detected by SNP-arrays and WGS in the intersecting samples from OVA-
727TCGA and PCAWG. **i** Correlation (Pearson, r=0.5) between the SBS3 exposure in WES
728data from OVA-TCGA inferred using SigMA versus the number of ovaHRDscars in SNP
729arrays. **j** Correlation (Pearson, r=0.38) between the SBS3 proportion in WGS data from
730PCAWG versus the number of scars using the Telli2016 approach. **k** Correlation (Pearson,
731r=0.51) between the SBS3 exposure in WES data from OVA-TCGA inferred using SigMA
732versus the number of scars using the Telli2016 approach in SNP arrays.
733
734**Supplementary Figure 3. HRD detected by ovaHRDscar is associated with better PFS**
735**and OS in ovarian cancer patients after normalizing with relevant covariables. a to c**
736Kaplan-Meier plots of PFS in Hercules cohort, the patients were stratified using different
737criteria; in the left panel (**a**) patients were stratified according to *BRCA*mut/del status with no
738significant difference in their PFS probability (p=0.14); in the middle panel (**b**) patients were
739stratified according to Telli et al, 2016 algorithm; in the right panel (**c**) patients were stratified
740using the ovaHRDscar algorithm and there is a significant difference (improvement) in PFS
741time (p = 0.024). **d** Hazard ratios (HR) for improvement of PFS in ovarian cancer patients
742using different criteria in two independent cohorts, the colors, rows and columns descriptions
743are the same as in Figure 2d. **e to f** Kaplan-Meier plots of PFS in the TERVA cohort,
744patients were stratified using different criteria; in the left panel (**e**) patients were stratified
745according to Telli et al, 2016 algorithm; in the right panel (**f**) samples were stratified using the
746ovaHRDscar. **g** Hazard ratios (HR) for improvement of PFS in the TERVA cohort, patients

747were stratified using different criteria the colors, rows and columns descriptions are the same
748as in Figure 2b. **h to j** Kaplan-Meier plots of OS in Hercules patients and stratified using
749different criteria; in the left panel (**h**) patients were stratified according to *BRCA*mut/del
750status with no significant difference (p=0.045), in the middle panel (**i**) patients were stratified
751according to Telli et al, 2016 algorithm with no significant difference (p=0.022), in the right
752panel (**j**) samples were stratified using the ovaHRDscar algorithm and there is a significant
753difference (improvement) in PFS time (p = 0.006). **k** Hazard ratios (HR) for improvement of
754OS in ovarian cancer, patients were stratified using different criteria inside two independent
755cohorts, for every cox regression it was used as covariable the patient age at diagnosis, the
756rows, colors and columns descriptions are the same as in Figure 2h. **l to m** Kaplan-Meier
757plots of OS in the PCAWG cohort, at the left panel (**l**) the patients were stratified according
758to CHORD signature reported by developers, at the right panel (**m**) the patients were
759stratified according to ovaHRDscar approach. **n** Hazard ratios (HR) for improvement of OS
760in the PCAWG cohort, patients were stratified using different CHROD and ovaHRDscar
761criteria; the dot colors, rows and columns descriptions are the same as in Figure 2d.
762
763**Supplementary Figure 4. Differences in ovaHRDcar intra patient and relationship with**
764**tumor purity.** **a** Difference in the ovaHRDscar values intra-patients and inter-patients in the
765Hercules prospective cohort (U test, p=1.95e-38)**.** **b** Schemed of the paired samples over
766treatment time, their purity and ovaHRDcar levels. **c** For each patient it was calculated the
767difference in ovaHRDscar and purity between every pair of their samples, no association
768between purity and ovaHRDscar difference (Pearson, p=0.4). **d** HR status classification
769using five different approaches to assess the ovaHRDscar values per patient (see results
770sections), per patient genomic alterations related with HR are shown. **e** Hazard ratios (HR)
771for improvement of PFS in the hercules cohort using different criteria to determine HR and
772under the five different approaches to assess the scar values (see results sections), for
773every cox regression it was used as covariable the residual tumor status (yes or no) after
774surgery, the colors, rows and columns descriptions are the same as in Figure 2b.

Figure 1

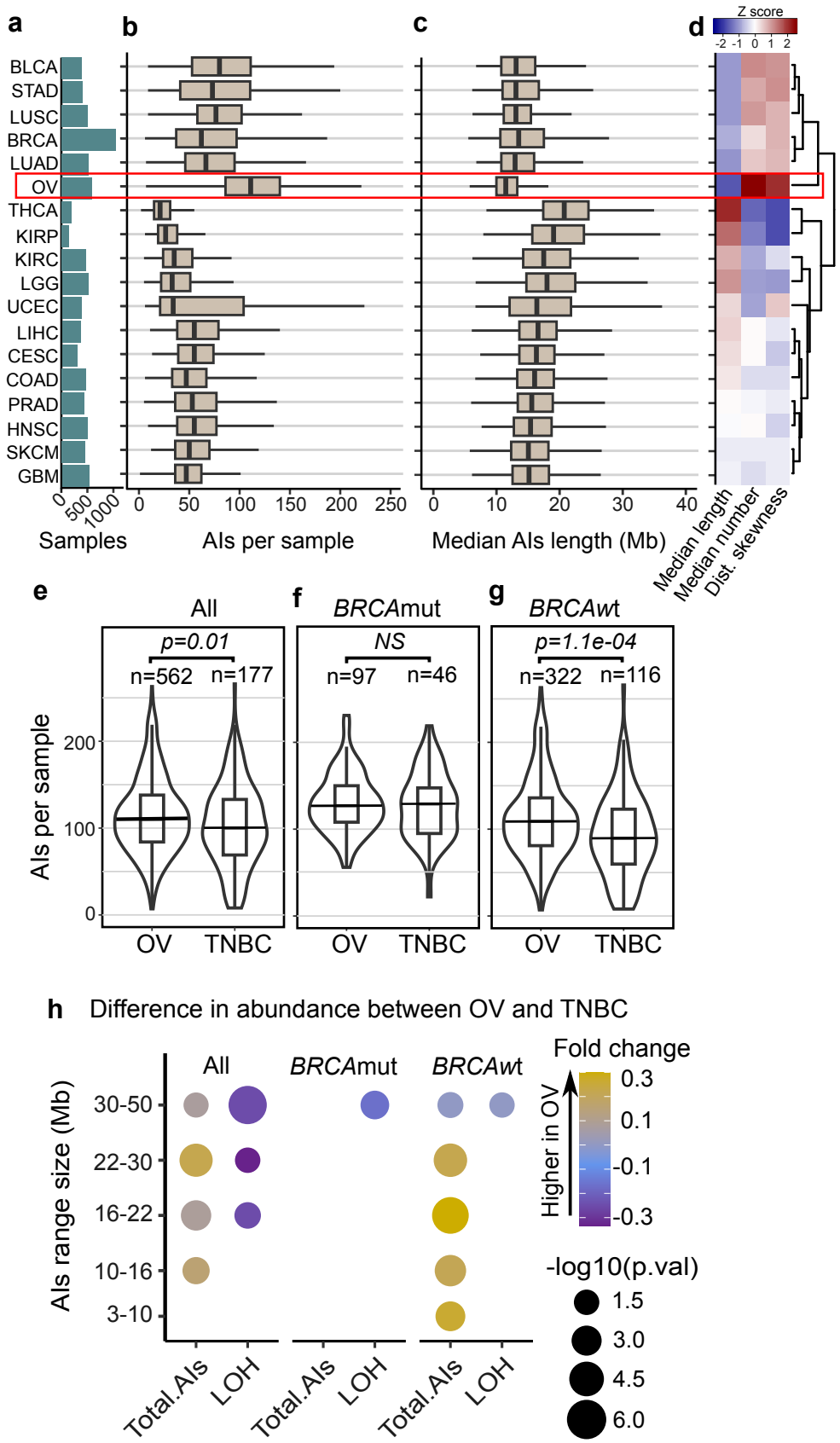
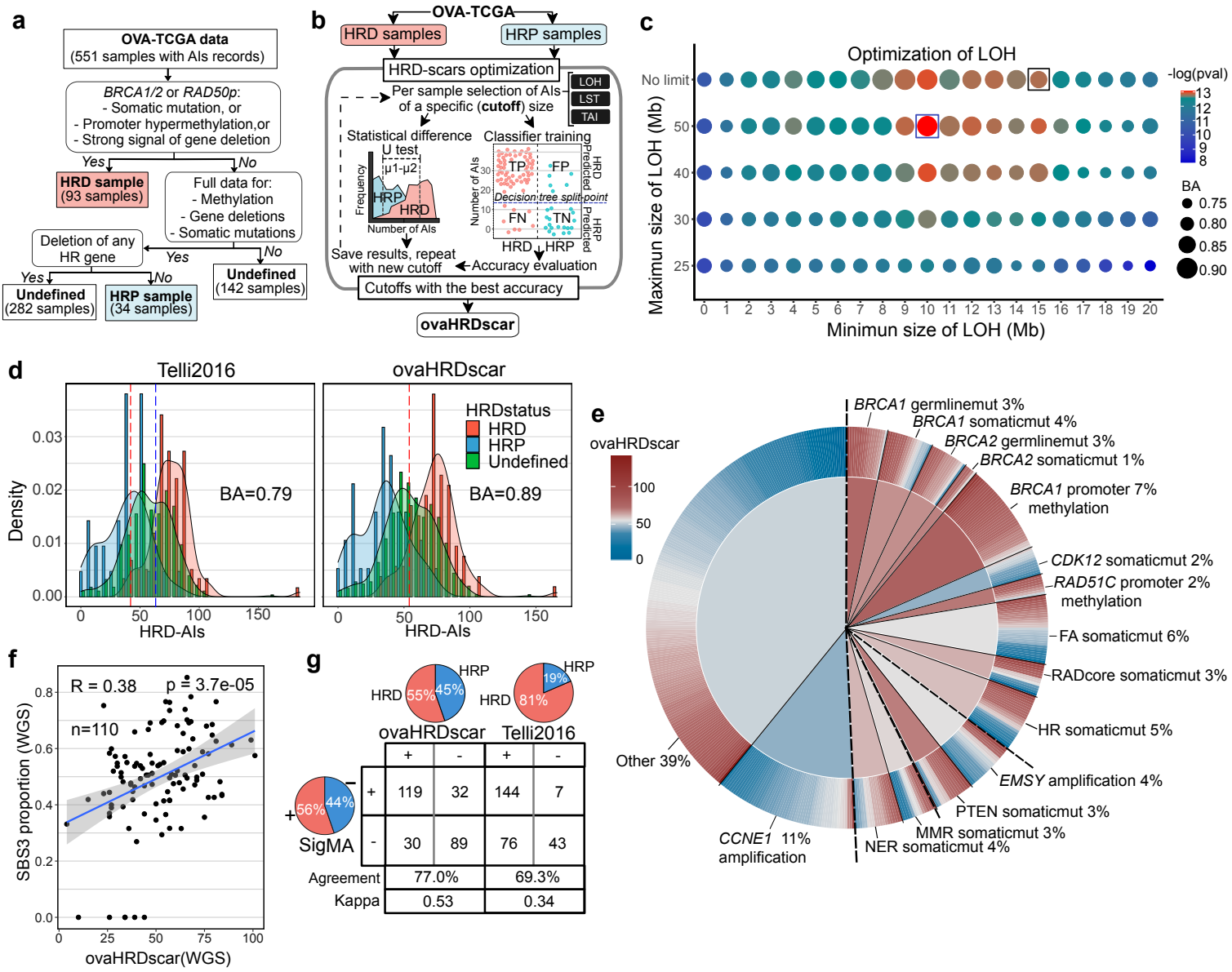
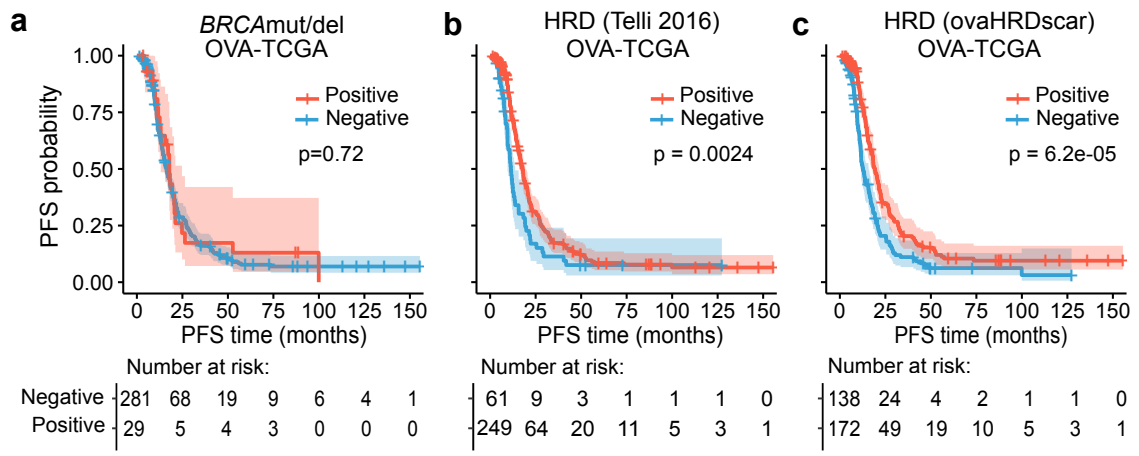

a
b
c
d
Z score
-2 -1 0 1 2
BLCA
STAD
LUSC
BRCA
LUAD
OV
THCA
KIRP
KIRC
LGG
UCEC
LIHC
CESC
COAD
PRAD
HNSC
SKCM
GBM
0 500 1000
Samples
0 50 100 150 200 250
AIs per sample
0 10 20 30 40
Median AIs length (Mb)
Median length
Median number
Dist. skewness
e
All
p=0.01
n=562 n=177
f
BRCAmut
NS
n=97 n=46
g
BRCAwt
p=1.1e-04
n=322 n=116
AIs per sample
0
100
200
OV TNBC
OV TNBC
OV TNBC
h Difference in abundance between OV and TNBC
All
BRCAmut
BRCAwt
AIs range size (Mb)
30-50
22-30
16-22
10-16
3-10
Total.AIs
LOH
Total.AIs
LOH
Total.AIs
LOH
Fold change
Higher in OV
0.3
0.1
-0.1
-0.3
-log10(p.val)
1.5
3.0
4.5
6.0

Figure 2



SigMA	ovaHRDscar +	ovaHRDscar -	Telli2016 +	Telli2016 -
+	119	32	144	7
-	30	89	76	43
Agreement	77.0%		69.3%	
Kappa	0.53		0.34	

Figure 3



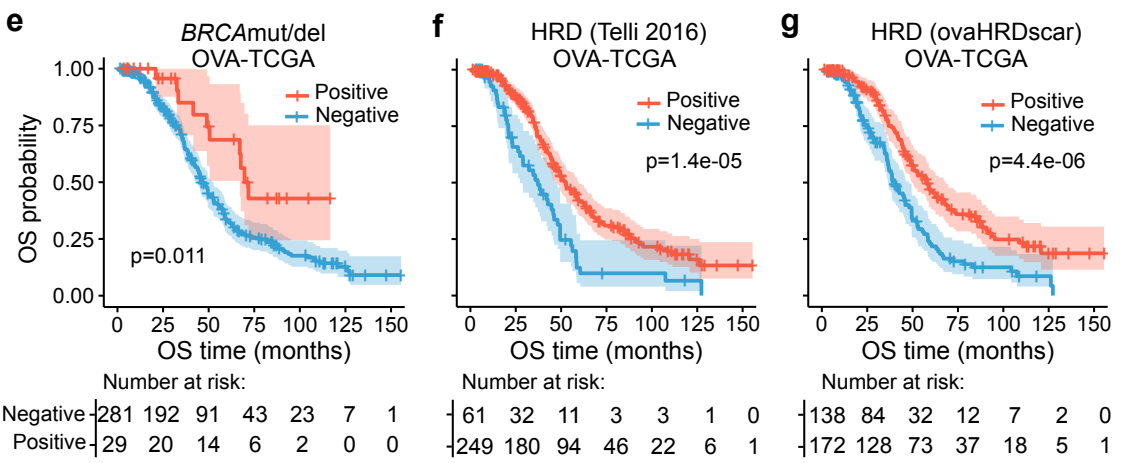
d *Cox model of PFS adjusted for residual tumor*

	OVA-TCGA n=310				OVA-TCGA* n=246				Hercules n=33			
All samples:	N	Prop	HR	Pval	N	Prop	HR	Pval	N	Prop	HR	Pval
*BRCA*mut/del	29	9%	0.855	0.0038	20	8%	0.671	0.015	5	15%	0.516	0.042
Telli2016	249	80%	0.467	1.6e−07	196	80%	0.479	1e−05	22	67%	0.682	0.037
Tacaya2020	146	47%	0.545	2.9e−07	108	44%	0.542	1.8e−05	14	42%	0.467	0.014
ovaHRDscar	172	55%	0.529	9.2e−08	131	53%	0.526	5.2e−06	16	48%	0.359	0.0061
*BRCA*mut/del excluded:	n=281				n=226				n=28			
Telli2016	220	78%	0.471	5.2e−07	176	78%	0.491	3e−05	17	61%	0.714	0.17
Tacaya2020	123	44%	0.537	1.1e−06	91	40%	0.533	3.9e−05	9	32%	0.509	0.093
ovaHRDscar	147	52%	0.523	3e−07	113	50%	0.521	1.1e−05	11	39%	0.379	0.046

Hazard-ratio (HR): 0.8 0.7 0.6 0.5 0.4

-log10(p.val): 3 4 5 6 7 | 2 3 4 5 | 0.9 1.2 1.5 1.8 2.1

*Excluding patients used for ovaHRDscar optimization



h *Cox model of OS adjusted for age at diagnosis*

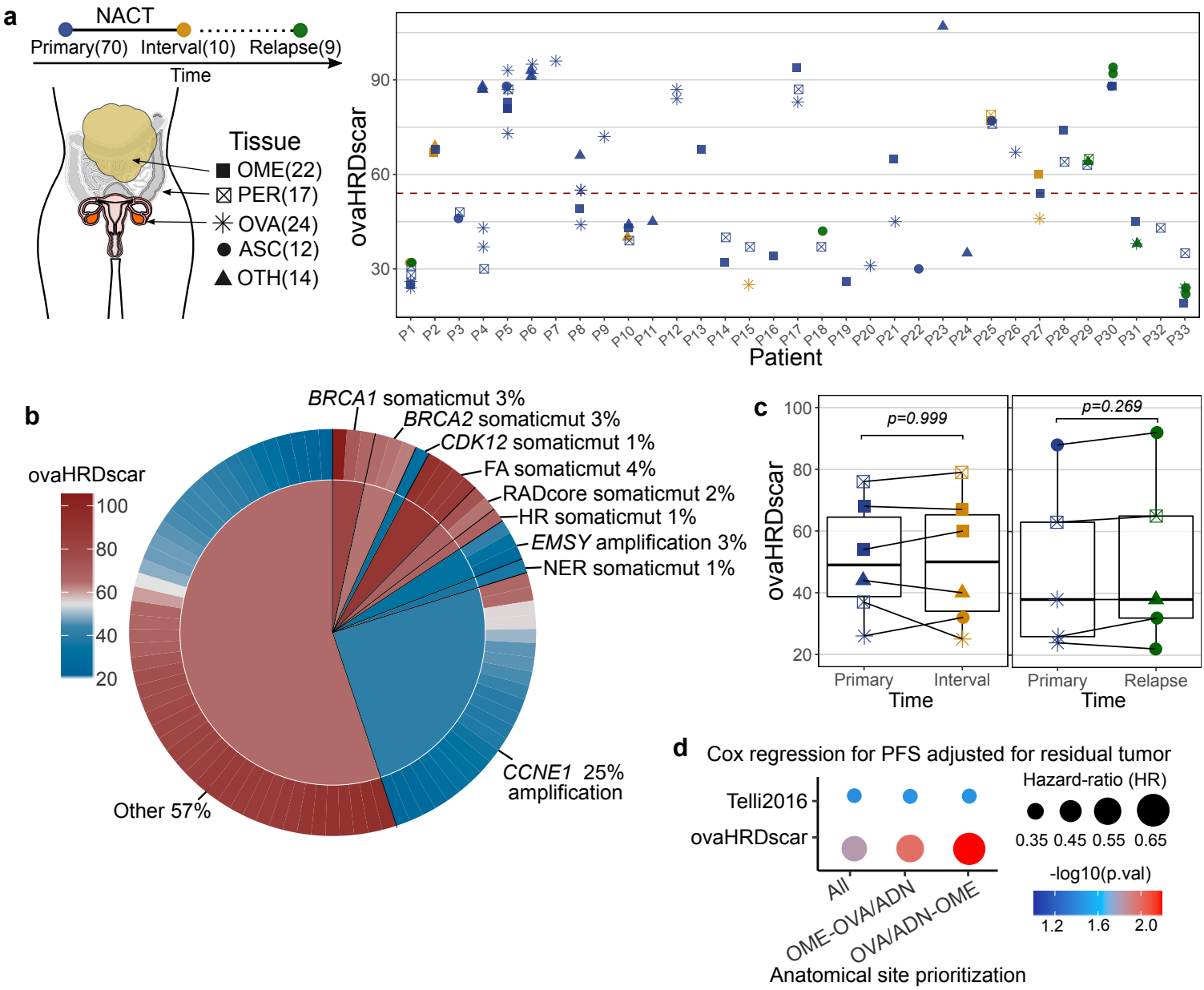
	OVA-TCGA n=310				OVA-TCGA* n=246				PCAWG n=60			
All samples:	N	Prop	HR	Pval	N	Prop	HR	Pval	N	Prop	HR	Pval
*BRCA*mut/del	29	9%	0.485	0.0014	20	8%	0.453	0.016	13	22%	0.502	0.097
Telli2016	249	80%	0.481	1.9e−06	196	80%	0.44	6.2e−06	43	72%	0.506	0.067
Tacaya2020	146	47%	0.546	4.3e−06	108	44%	0.499	2.5e−05	20	33%	0.602	0.16
ovaHRDscar	172	55%	0.525	1e−06	131	53%	0.483	6.2e−06	28	47%	0.474	0.026
*BRCA*mut/del excluded:	n=281				n=261				n=47			
Telli2016	220	78%	0.51	1.3e−05	176	78%	0.459	2e−05	30	64%	0.593	0.22
Tacaya2020	123	44%	0.563	2.4e−05	91	40%	0.504	5.6e−05	11	23%	1.07	0.71
ovaHRDscar	147	52%	0.528	3.2e−06	113	50%	0.481	9.2e−06	17	36%	0.599	0.25

Hazard-ratio (HR): 1.0 0.6 0.5 0.4

-log10(p.val): 3 4 5 6 | 2 3 4 5 | 0.4 0.8 1.2

*Excluding patients used for ovaHRDscar optimization

Figure 4


a
NACT
Primary(70)
Interval(10)
Relapse(9)
Time
Tissue
OME(22)
PER(17)
OVA(24)
ASC(12)
OTH(14)
ovaHRDscar
Patient
b
ovaHRDscar
BRCA1 somaticmut 3%
BRCA2 somaticmut 3%
CDK12 somaticmut 1%
FA somaticmut 4%
RADcore somaticmut 2%
HR somaticmut 1%
EMSY amplification 3%
NER somaticmut 1%
CCNE1 25% amplification
Other 57%
c
p=0.999
p=0.269
ovaHRDscar
Primary
Interval
Relapse
Time
d
Cox regression for PFS adjusted for residual tumor
Telli2016
ovaHRDscar
All
OME-OVA/ADN
OVA/ADN-OME
Anatomical site prioritization
Hazard-ratio (HR)
0.35 0.45 0.55 0.65
-log10(p.val)
1.2 1.6 2.0