














# Integrated radiogenomics models predict response to neoadjuvant chemotherapy in high grade serous ovarian cancer

Received: 27 May 2022

Accepted: 20 September 2023

Published online: 24 October 2023

 Check for updates

Mireia Crispin-Ortuzar <sup>1,2,18</sup> ✉, Ramona Woitek <sup>2,3,4,18</sup>,  
Marika A. V. Reinius <sup>2,5,6</sup>, Elizabeth Moore<sup>5</sup>, Lucian Beer <sup>2,3,7</sup>, Vlad Bura<sup>2,3</sup>,  
Leonardo Rundo<sup>2,3,8</sup>, Cathal McCague <sup>2,3,6</sup>, Stephan Ursprung <sup>2,3</sup>,  
Lorena Escudero Sanchez <sup>2,3</sup>, Paula Martin-Gonzalez<sup>2,5</sup>, Florent Mouliere <sup>5,9</sup>,  
Dineika Chandrananda <sup>5</sup>, James Morris<sup>5</sup>, Teodora Goranova<sup>5</sup>, Anna M. Piskorz<sup>5</sup>,  
Naveena Singh<sup>10,20</sup>, Anju Sahdev <sup>11</sup>, Roxana Pintican<sup>12,13</sup>, Marta Zerunian<sup>14</sup>,  
Nitzan Rosenfeld <sup>2,5</sup>, Helen Addley<sup>2,3,6</sup>, Mercedes Jimenez-Linan<sup>2,6</sup>,  
Florian Markowitz <sup>2,5,19</sup>, Evis Sala<sup>2,3,6,15,16,17,19</sup> & James D. Brenton <sup>2,5,6,19</sup>

High grade serous ovarian carcinoma (HGSOC) is a highly heterogeneous disease that typically presents at an advanced, metastatic state. The multi-scale complexity of HGSOC is a major obstacle to predicting response to neoadjuvant chemotherapy (NACT) and understanding critical determinants of response. Here we present a framework to predict the response of HGSOC patients to NACT integrating baseline clinical, blood-based, and radiomic biomarkers extracted from all primary and metastatic lesions. We use an ensemble machine learning model trained to predict the change in total disease volume using data obtained at diagnosis ( $n = 72$ ). The model is validated in an internal hold-out cohort ( $n = 20$ ) and an independent external patient cohort ( $n = 42$ ). In the external cohort the integrated radiomics model reduces the prediction error by 8% with respect to the clinical model, achieving an AUC of 0.78 for RECIST 1.1 classification compared to 0.47 for the clinical model. Our results emphasize the value of including radiomics data in integrative models of treatment response and provide methods for developing new biomarker-based clinical trials of NACT in HGSOC.

High-grade serous ovarian carcinoma (HGSOC) is a major therapeutic challenge as it typically presents with advanced, multi-site metastatic disease. Neoadjuvant chemotherapy (NACT) followed by delayed primary surgery (DPS) is now the most frequent treatment strategy for advanced HGSOC<sup>1,2</sup>. However, response is variable, and 39% of patients do not obtain any objective benefit from neoadjuvant carboplatin and paclitaxel<sup>3–5</sup>. Variability in response is partly driven by the complexity of HGSOC, which spans a large range of scales—from macroscopic metastatic tumour volumes observed on radiological imaging to microscopic tumour-immune microenvironments and sub-

microscopic genomic diversity<sup>6–8</sup>. Patient care would be substantially improved if different sub-populations could be identified before treatment is started, for example by identifying likely non-responders who could receive immediate primary surgery.

So far, predictive studies have focused on individual data streams, such as clinical features<sup>9,10</sup>, CA-125<sup>11–13</sup>, computed tomography (CT) imaging<sup>14,15</sup>, and circulating tumour DNA (ctDNA)<sup>16</sup>. The superior predictive power of integrative models for complex endpoints is well demonstrated in several cancer types<sup>17–19</sup>, but training such models requires large, well-annotated, multi-omic datasets, which have not

been available for NACT treatment of HGSOc. In particular, radiological imaging is the only data source that non-invasively captures the spatial heterogeneity of metastatic disease, but has so far been underused. Existing radiomics studies focus only on correlations<sup>14,15</sup>, rather than combined predictive power, and do not consider NACT.

Here, we present *IRON*, an integrative radiogenomic framework to predict the volumetric response of heterogeneous, multi-site ovarian cancer to NACT. We used two independent, highly annotated data sets including clinical, chemotherapy treatment, CA-125, ctDNA, and radiomics features extracted from all primary and metastatic disease at diagnosis. The features were used as an input to an ensemble machine learning model trained to predict volume shrinkage during NACT. We validated the approach on an external training data set and demonstrated that radiomics features are essential to obtain significant predictive power. Our method is easily generalised and is applicable to other cancers with heterogeneous, multi-site disease.

## Results

This study is based on data from two prospective observational studies from Addenbrooke's Hospital ('NeOv',  $n = 92$ ) and the Barts Health NHS Trust ('Barts',  $n = 42$ ), respectively (Fig. 1a, Supplementary Fig. S1). Patients show variability in treatment regimes and response patterns (Fig. 1b). The NeOv dataset was randomly divided into a training set ( $n = 72$ ) and a hold-out internal validation set ( $n = 20$ , Supplementary Fig. S1a). The training data set was used for the exploratory analyses and to train the machine learning models. The hold-out set was only used for performance assessment after all predictive models had been fully trained. The Barts data were used as an independent, external validation data set.

### Response patterns to NACT are heterogeneous

All primary and metastatic lesions identified on pre- and post-NACT CT scans were segmented and labelled (Fig. 1c). As expected, tumour in the omentum and pelvic/ovarian locations accounted for the majority of the disease burden at baseline and were the most frequent tumour locations (Fig. 1d, e). There were marked differences in response at the same anatomic sites between patients (Supplementary Fig. S3). Omental disease showed significantly better response than pelvic disease (Fig. 1f, g, h). Despite anatomic differences in response, use of response evaluation criteria in solid tumours (RECIST 1.1) was strongly correlated with total volume change (Fig. 1f). The volume of omental disease at baseline was higher ( $p = 0.05$ ) in responders assessed by RECIST (complete or partial response; median = 85 cm<sup>3</sup>) compared to non-responders (stable disease or progression; median = 31 cm<sup>3</sup>).

While the number of disease locations at baseline was significantly correlated with response, there was no correlation with disease volume at baseline, either overall or in specific anatomic locations (Table 1). Taken together, these detailed volumetric data indicate that multivariable predictors are required to predict response to NACT rather than simple knowledge about disease burden and its anatomic distribution.

### ctDNA and CA-125 correlate with different types of disease burden

For all patients in the training and hold-out validation cohorts, ctDNA was assessed at baseline. We compared *TP53* mutant allele fraction (MAF), trimmed median absolute deviation from copy number neutrality based on shallow whole genome sequencing (t-MAD) and IchorCNA<sup>20–23</sup>. In addition, we explored the use of computed haploid genome equivalents per millilitre (hGE/ml)<sup>24</sup>. All measures were highly correlated ( $p < 0.0001$ , Supplementary Fig. S1c) and therefore, *TP53* MAF was included in univariable analyses.

Total disease burden at baseline (total volume, number of lesions, and summed RECIST 1.1 diameters) correlated significantly with both CA-125 and *TP53* MAF (Table 2). Neither baseline CA-125 nor *TP53* MAF

correlated significantly with response, but did show a significant positive correlation with the summed RECIST 1.1 diameters post-chemotherapy (Table 1).

Baseline CA-125 correlated with omental disease and pelvic/ovarian disease volume measured before chemotherapy. Similarly, baseline *TP53* MAF correlated significantly with pelvic/ovarian disease volume measured both before and after chemotherapy. However, ctDNA measurements did not correlate with omental disease at either time point (Table 2). This suggests that high *TP53* MAF at baseline could be a specific indicator for high disease burden in the ovaries or pelvis, which tends to show poorer response (Fig. 1f).

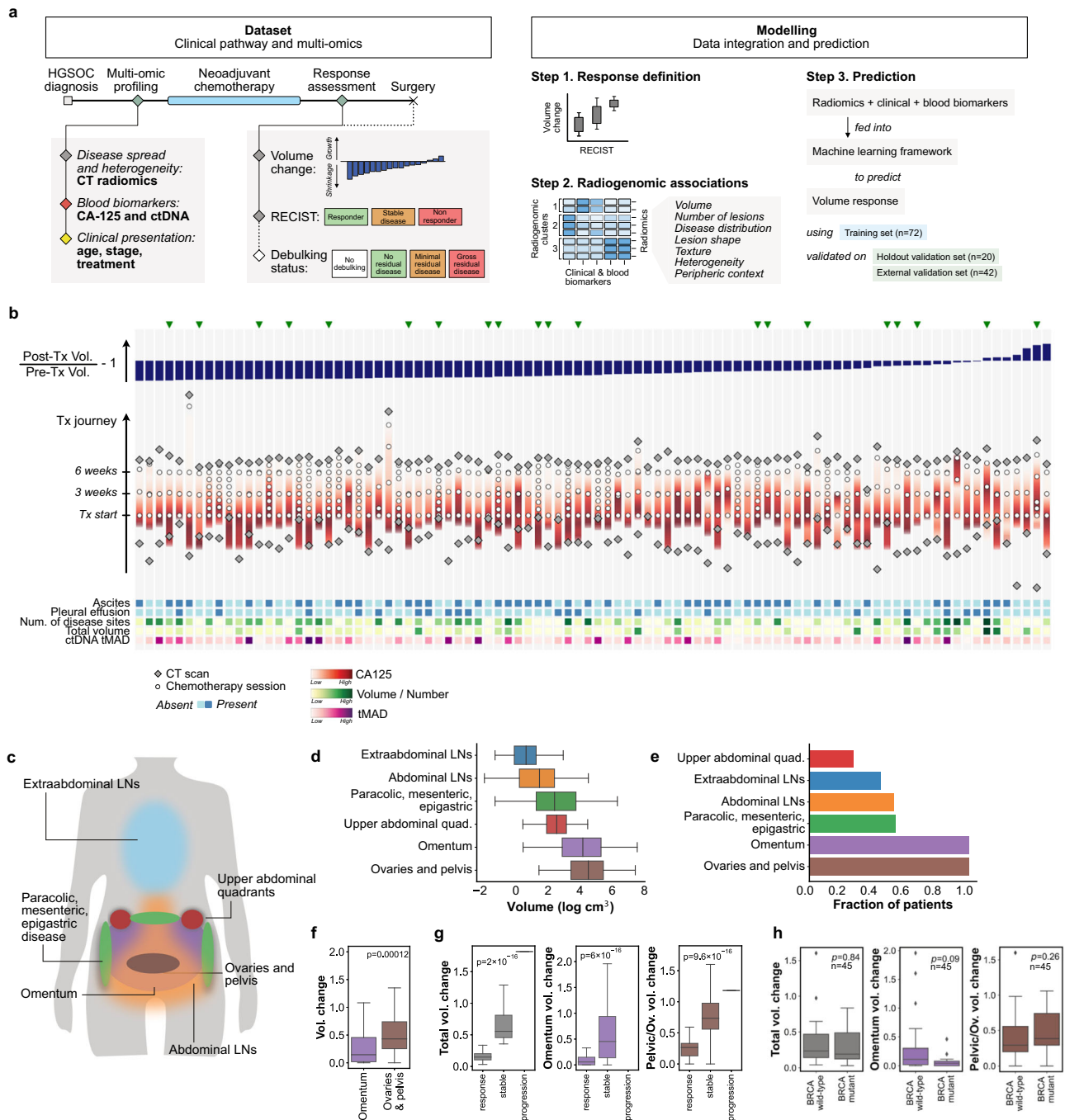
### Radiomics features correlate with clinical and biological characteristics

To capture the radiological complexity of the disease we defined several collections of radiomics features (Supplementary Data 1). Volumes and number of lesions were calculated for each of the relevant anatomical sites, in order to model the effect of tumour location. Shape features, first-order histogram statistics and texture features ('intensity radiomics') were calculated for each lesion and averaged over the whole disease. Intra-lesion heterogeneity was assessed by contracting the lesion contours and calculating the ratio of radiomics features before and after the contraction ('rim radiomics'). Similarly, the external context of the lesions was assessed by calculating the ratio of radiomics features before and after dilating the contours ('peripheral radiomics'). Finally, we also defined a series of binary variables to describe additional radiological findings ('semantic features'): ascites and pleural effusion were assessed manually, and we used a previously developed automated tissue-specific sub-segmentation tool to identify hypodense (cystic/necrotic spaces) and hyperdense (calcifications) lesion parts within the manual segmentations<sup>25</sup>.

We found that imaging features grouped into six distinct clusters (Fig. 2a). Cluster 1 was associated with baseline CA-125 levels ( $r_{\text{median}} = 0.34$ ) and contained mostly lesion volume metrics. This is consistent with previous work suggesting that CA-125 correlates with lesion volume<sup>26</sup>. Clusters associated with ctDNA features were generally dominated by features quantifying lesion heterogeneity and context. Cluster 5 was primarily associated with ctDNA *TP53* status ( $r_{\text{median}} = 0.25$ ), and contained predominantly peripheral radiomics features, which quantify lesion context. Cluster 6 was associated with ctDNA *TP53* MAF and tMAD ( $r_{\text{median}} = 0.20$  and  $0.19$ , respectively), and contained predominantly rim ratio radiomics features, which provide information on intra-lesion heterogeneity.

Cluster 4 was highly correlated to stage ( $r_{\text{median}} = 0.36$ ), and was composed of a mixture of features related to shape, volume, and number of lesions, which quantify the disease burden. This is consistent with the definition of FIGO stage, which relies on the assessment of the extent and spread of the disease<sup>27</sup>. Cluster 2 was associated mostly with age ( $r_{\text{median}} = 0.23$ ) and contained almost exclusively rim ratio radiomics features, which provide information about intra-tumour heterogeneity. The remaining group (cluster 3) was formed by a heterogeneous mixture of features, and did not associate with any biological or clinical feature.

These results indicate that some of the information from global biomarkers such as stage, CA-125 or ctDNA can also be captured in multi-lesion radiomic features that quantify the extent, spread, heterogeneity and context of the disease. In addition, as shown in Fig. 2b, clusters 1, 2, 3 and 5 contain imaging features that are significantly correlated with volumetric response to treatment after multiple comparison correction. Cluster 3 has negligible associations with either biological or clinical features, demonstrating that radiomics features can also contribute unique information to integrated radiogenomic predictive models.



**Fig. 1 | Structure of the study and main characteristics of the training cohort.** **a** Key time points and variables in the dataset (left) and steps of the modelling strategy (right). See also Supplementary Fig. S1 for additional information. **b** Treatment courses of all 92 patients in the NeOV cohort, ordered by decreasing volumetric tumour response following NACT. Patients analysed in the hold-out validation set were randomly selected and are indicated with a green triangle. Treatment journeys progress vertically (bottom to top) and are aligned at the time of the first chemotherapy cycle. Additional biomarkers obtained at baseline are depicted in the bottom heatmap. **c** Sites of primary and metastatic disease in HGSOc. **d** Distribution of tumour volumes by site for patients in the training cohort. **e** Distribution of tumour sites by patient. **f** Volume changes of the omental

and pelvic/ovarian disease for all patients in the training cohort. *p* value obtained from the two-sided Mann-Whitney *U* test. **g** Total and site-specific volume change stratified by RECIST 1.1 response status for the training cohort. *p* value obtained from the point biserial correlation coefficient, two-tailed. **h** Total and site-specific volume change stratified by *BRCA* mutation status. These figures are restricted to the *n* = 45 patients in the training cohort for whom the *BRCA* mutation status was known. *p* value obtained from the two-sided Mann-Whitney *U* test. Boxes indicate the upper and lower quartiles, with a line at the median. Outliers are shown as circles and identified via the interquartile range rule. Source data are provided as a Source Data file.

### Integrative ensemble model predicts volume response to neoadjuvant chemotherapy

We built an integrative radiogenomic machine learning framework to predict response to neoadjuvant chemotherapy, called *IRON* (Integrated Radiogenomics for Ovarian Neoadjuvant therapy). The

framework is a robust ensemble of three machine learning pipelines, each of which includes a classifier (elastic net, support vector regression, or random forest) preceded by collinearity reduction and feature selection steps (see Methods). To further prevent overfitting, models were trained using a 5-fold cross-validation setup that was repeated

**Table 1 | Spearman correlation coefficients ( $r_s$ ) and corresponding two-sided  $p$  values ( $p$ ) between baseline measurements of tumour burden and different assessments of treatment response**

		Pre-chemotherapy									
		TP53 MAF	CA-125	Oment. vol.	Pelvis/ovaries vol.	LN vol.	Total vol.	Number lesions	Summed diameter	Ascites	Pleural effusion
Volume change	$r_s$	-0.05	-0.30	-0.07	0.12	-0.13	0.01	<b>-0.38</b>	-0.08	-0.15	-0.03
	$p$	0.93	0.07	0.90	0.62	0.60	0.96	<b>0.02</b>	0.90	0.55	0.93
Summed diam. change	$r_s$	-0.06	-0.28	0.04	0.24	-0.16	0.15	<b>-0.34</b>	0.02	-0.03	-0.01
	$p$	0.93	0.08	0.93	0.16	0.55	0.55	<b>0.04</b>	0.93	0.93	0.96

$p$  values are adjusted for multiple correction.

Significant correlations ( $p < 0.05$ ) are highlighted in bold.

**Table 2 | Spearman correlation coefficients ( $r_s$ ) and corresponding two-sided  $p$  values ( $p$ ) for pre- and post-NACT measurements of tumour burden versus blood biomarkers measured at baseline**

		Pre-chemotherapy							
		Oment. vol.	Pelvis/ovaries vol.	LN vol.	Total vol.	Number lesions	Summed diameter	Ascites	Pleural effusion
Baseline TP53 MAF	$r_s$	0.04	<b>0.37</b>	0.26	<b>0.37</b>	<b>0.32</b>	<b>0.48</b>	0.05	0.13
	$p$	0.77	<b>0.005</b>	0.06	<b>0.005</b>	<b>0.02</b>	<b>0.0003</b>	0.77	0.36
Baseline CA-125	$r_s$	<b>0.41</b>	<b>0.38</b>	0.21	<b>0.48</b>	<b>0.28</b>	<b>0.45</b>	<b>0.28</b>	0.23
	$p$	<b>0.002</b>	<b>0.005</b>	0.13	<b>0.0003</b>	<b>0.04</b>	<b>0.0009</b>	<b>0.04</b>	0.10
		Post-chemotherapy							
		Oment. vol.	Pelvis/ovaries vol.	LN vol.	Total vol.	Number lesions	Summed diameter	Ascites	Pleural effusion
Baseline TP53 MAF	$r_s$	-0.04	<b>0.30</b>	0.13	<b>0.29</b>	0.19	<b>0.41</b>	0.03	0.13
	$p$	0.79	<b>0.03</b>	0.36	<b>0.03</b>	0.17	<b>0.003</b>	0.79	0.36
Baseline CA-125	$r_s$	0.10	0.23	0.10	0.22	0.20	<b>0.32</b>	-0.07	-0.16
	$p$	0.50	0.10	0.50	0.11	0.15	<b>0.02</b>	0.63	0.24

$p$  values are adjusted for multiple correction.

Significant correlations ( $p < 0.05$ ) are highlighted in bold.

five times with different seeds. To capture the behaviour of the whole metastatic disease, we defined the response metric as the relative change in total disease volume, which was also correlated with surgical debulking status (Supplementary Fig. S5).

We used three distinct datasets: a training set (NeOV  $n=72$ ), a hold-out validation set (NeOV  $n=20$ ) and an independent validation set (Barts  $n=42$ ). We trained four models by successively adding clinical and molecular features: (i) age, FIGO stage, and treatment; (ii) CA-125; (iii) radiomics features; and (iv) ctDNA, (Fig. 3a and Supplementary Data 1). All features used in the model were extracted from data obtained at diagnosis. Once trained, the four models were frozen and tested on the hold-out and external validation sets.

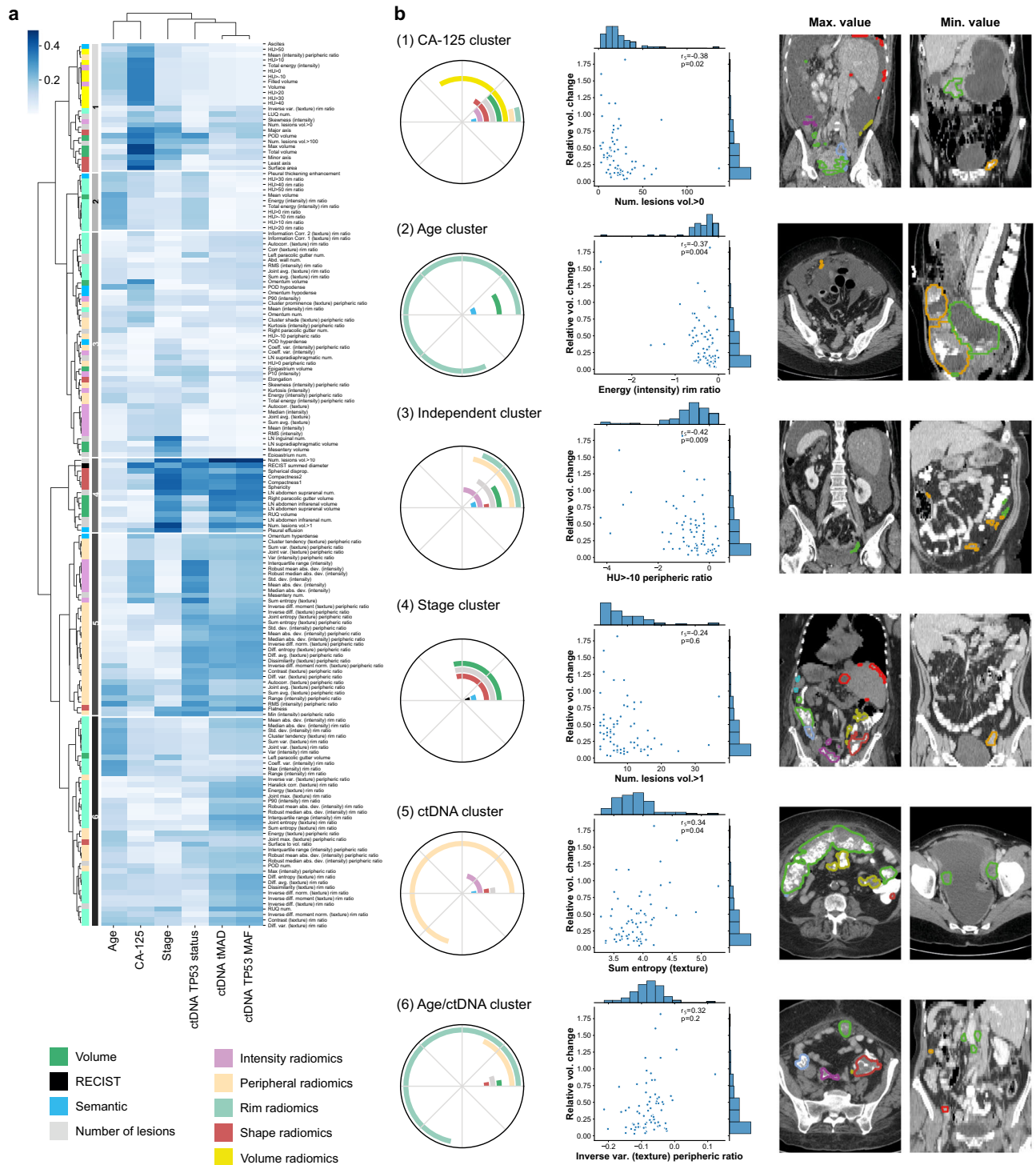
On the hold-out validation set, adding successive features resulted in a gradual MSE reduction of 15% for the model without ctDNA, and 14% after integrating ctDNA (Fig. 3b and Supplementary Table S8). Only the two models that included radiomic features were able to produce response scores that were significantly correlated with the observed volume response (Spearman  $r=0.5$ ,  $p=0.02$ , Fig. 3b and Supplementary Table S8). Although the models were not trained to predict RECIST 1.1, we observed that the scores predicted by the integrated radiomics model were able to correctly rank the three RECIST 1.1 response groups ( $p=0.02$ , Fig. 4). Similarly, we found that the model was able to accurately predict which patients were most likely to undergo delayed primary surgery, but it was not able to predict surgical debulking status (Supplementary Fig. S5).

We tested the generalisability of the models in an independent external cohort (Barts cohort,  $n=42$ ). As this cohort did not have ctDNA testing performed, we imputed ctDNA measurements to the training set averages. The clinical and CA-125 models did not achieve significant performances in terms of Spearman correlation with

response. The radiomics and ctDNA models both achieved significant performance at a similar level ( $r=0.32$ ,  $p=0.04$  and  $r=0.32$ ,  $p=0.02$  respectively, Fig. 3b and Table S8). For volume shrinkage thresholds corresponding to clinically relevant criteria (RECIST, WHO, spherical-volume and ellipsoidal-volume response), the full model achieved an average area under the receiver-operating characteristic curve (AUC) of 0.7, compared to an average of 0.5 for the clinical and CA-125 models (Supplementary Fig. S7). When assessed against the RECIST classification labels, the full model achieved an AUC of 0.8, similar to the radiomics model (0.78), compared to 0.47 and 0.50, respectively for the clinical and CA-125 models.

We studied the relative contribution of the features used by the final models in two different ways. Firstly, we evaluated the selection frequency for each feature (Fig. 5a). We found that the treatment regimen and the number of sessions of chemotherapy before the second scan were consistently selected across models. We also found that CA-125 was used in models that did not include radiomics, but was dropped from radiomics integration models. Semantic features, in particular pleural thickening and the presence of a hyperdense region in an omental lesion, were selected across all the relevant models. The only volumetric features selected consistently were mean volume and the volume of the infrarenal lymph nodes. A small and consistent number of radiomics features was also selected, most of them belonging to the category of features quantifying lesion context. We found that volume-related radiomic features play an important but not essential role, as the model is still predictive if they are removed during inference. Secondly, we quantified relative feature importances within the models (Fig. 5). We found that most of the models integrating a large number of features tend to be more dense, with features sharing similar, lower importance levels. Features that tended to have larger importance were





**Fig. 2 | Analysis of the correlations between radiomics and non-imaging features.** **a** Spearman correlation coefficients between imaging (rows) and clinical and biological features (columns), both clustered using a hierarchical approach, using the training cohort. **b** Composition and characteristics of the six identified imaging feature clusters. Polar plots indicate the relative contribution of the different classes of imaging features. Scatter plots show the feature of each cluster with the

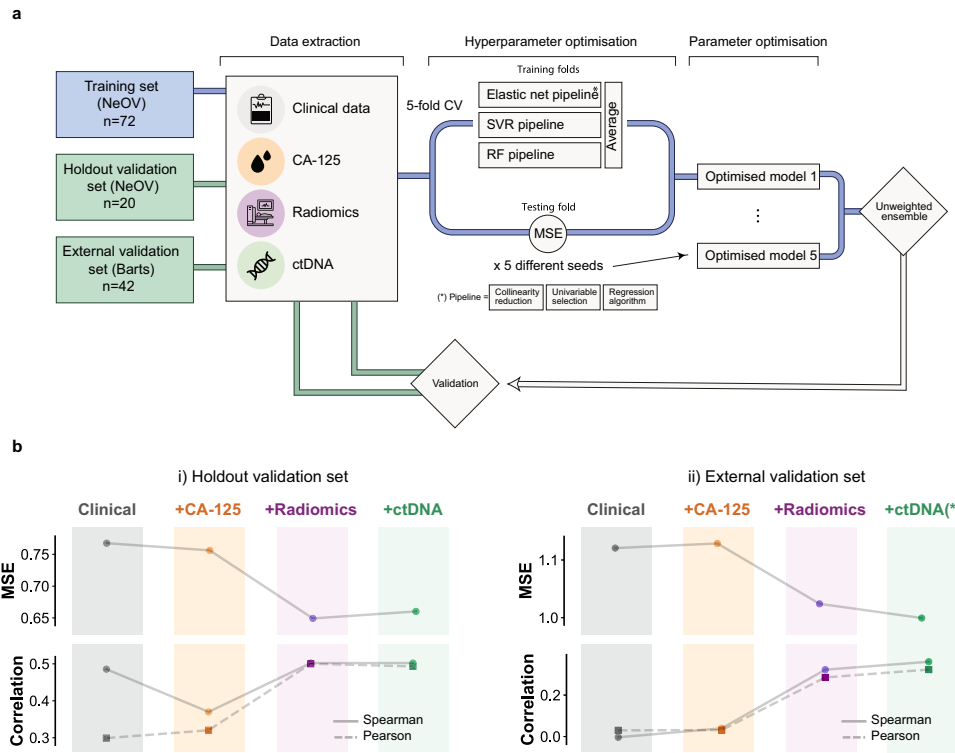
highest Spearman correlation with volumetric treatment response. Each feature is illustrated by displaying one slice from the patient with the maximum value (left), and one from the patient with the minimum value (right). Source data are provided as a Source Data file. *p* values are two-sided and corrected for multiple comparisons.

generally consistent with those that had the highest selection frequency, as can be observed by comparing the two panels in Fig. 5.

**Discussion**

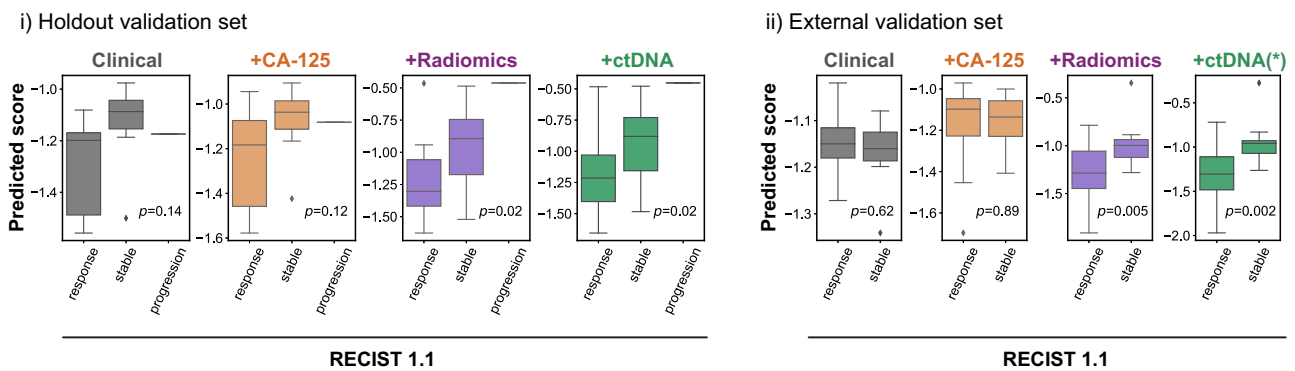
The clinical presentation of HGSOc is with complex, highly heterogeneous disease that is invariably metastatic throughout the abdomen.

The accompanying genomic and cellular heterogeneity has impeded therapeutic progress and strongly suggests that understanding response to treatment must involve the integration of data from different sources and scales. We have shown that an integrated radiogenomic machine learning model based on baseline multi-scale data predicts volumetric response to NACT (*p* = 0.04, external validation



**Fig. 3 | Training scheme and validation results for the *IRON* machine learning framework.** **a** Schematic of the machine learning framework for model training and validation. **b** Validation of the discriminative power of predictive models in the hold-out (left) and external validation cohorts (right), for models containing, from left to right, clinical, clinical+CA-125, clinical+CA-125+radiomics, and clinical+CA-125+radiomics+ctDNA features, respectively. The metrics are mean square error

(MSE, top) and Spearman (continuous) or Pearson (dashed) correlation (bottom). The magnitude of the mean squared error (MSE) is comparable to the standard deviation of the volumetric response (Supplementary Table S9). ctDNA values in the external validation cohort were imputed using training set averages. Source data are provided as a Source Data file.



**Fig. 4 | Using the *IRON* framework to predict responding, stable and progressive disease according to RECIST 1.1.** Validation of the ability of predictive models to describe RECIST 1.1 response in the hold-out (left,  $n = 20$ ) and external (right,  $n = 42$ ) validation cohorts, for models containing, respectively, clinical, clinical+CA-125, clinical+CA-125+radiomics, and clinical+CA-125+radiomics+ctDNA

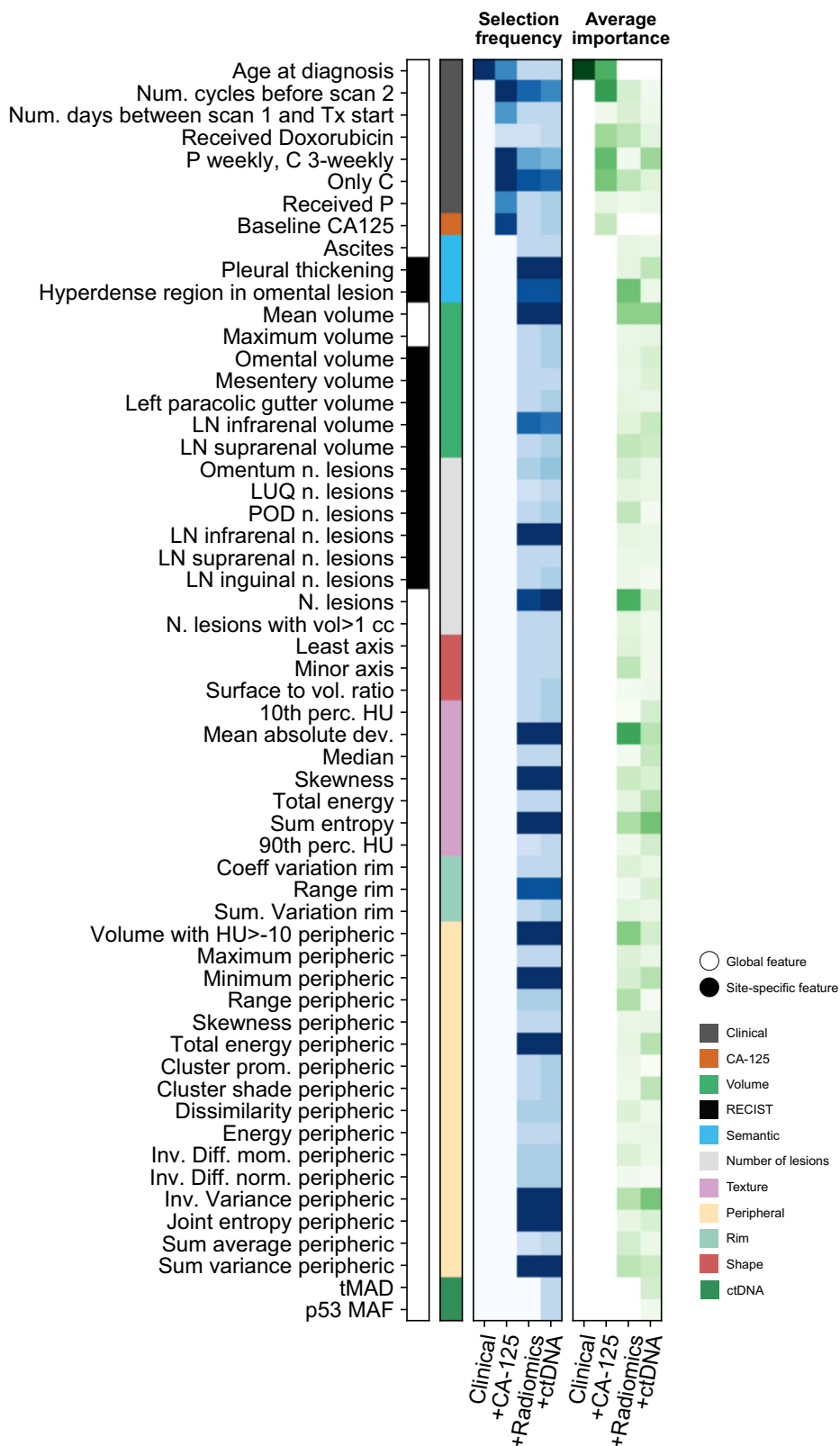
features. Boxes indicate the upper and lower quartiles, with a line at the median. Outliers are shown as circles and identified via the interquartile range rule.  $p$  values are obtained using the point biserial coefficient and are two-tailed. ctDNA values in the external validation cohort were imputed using training set averages. Source data are provided as a Source Data file.

cohort). Our results show that radiomics features are critical for the prediction, and that models that do not include radiomics fail to predict response. We also revealed site-specific patterns of differential response and correlation of radiomics features with ctDNA detection.

The field of radiomics has grown exponentially in recent years, showing great promise across tumour sites and endpoints<sup>28</sup>. At the same time, radiomics has been criticised for lack of robustness and reproducibility, as well as lack of biological interpretability<sup>29,30</sup>. Our

study shows that both problems can be overcome by the right design choices.

We made robustness a design priority for our predictive framework, *IRON*, which included strong feature selection, model ensembles on multiple levels, and repeated re-shufflings of the data to avoid biases. In addition, we trained the model on a dataset with heterogeneous imaging parameters<sup>31</sup>; and we included new families of imaging features based on ratios between different volumes of interest,



**Fig. 5 | Feature importances in the fully integrated IRON radiogenomic model.** Importances of the features used by the predictive models. The first (blue) heatmap illustrates the selection frequency. The heatmap shows the number of times that a given feature was selected in a model. The different columns correspond to different models with increasing, cumulative numbers of input features. As the optimisation is repeated five times, the range of the selection frequency is 0–15

(three algorithms in the ensemble times five repetitions). The first (green) heatmap illustrates the averaged, normalised feature importances for the elastic net and random forest components of the models. Importances are defined from the feature coefficients for the elastic net regression, and from impurity-based Gini importances for random forest. Source data are provided as a Source Data file.

which were designed to partially cancel out such biases. We also curated an independent cohort from a different institution for validation, which we were able to perform successfully.

Our feature definitions were also designed to improve interpretability: the ratio features are not only robust to imaging parameters, but they also helped us to explore internal heterogeneity (rim features) and the external context of the lesions (peripheral features). Interestingly, we found that ratio features were the most frequently selected in the final models. This result is in line with previous work that found that qualitative radiological features describing the edges of peritoneal disease (nodular, diffuse, or mixed) were significantly associated with CLOVAR subtype<sup>32</sup> and BRCA mutation status<sup>33</sup>. The concept of peritumoural radiomics has been explored before in breast, lung, and liver cancer<sup>34–37</sup>. The periphery of tumours and the adjacent peritumoural stroma are characterised by molecular and cellular changes directly related to tumour biology. Neoangiogenesis and fibroblast invasion have been demonstrated in the tumour periphery of ovarian cancer<sup>38</sup>, whereas characteristic immune cell populations have been identified in the peritumoural stroma<sup>39,40</sup>. Our findings motivate the development of more detailed studies focusing on the boundary regions of ovarian cancer lesions, demonstrating that data-driven radiomics analyses can support biological hypothesis generation.

Our study confirms previous observations that the presence of ctDNA is correlated with volume of disease at the start of treatment in HGSOC patients<sup>20</sup>. The integration of clinical data into our models also yielded important insights, as features describing the type of NACT and its timing were consistently selected by our models. The ICON8 study (to which part of the NeOv cohort was recruited) showed that PFS in patients with ovarian cancer undergoing NACT was unaffected by the administration regimen of paclitaxel (3-weekly as is standard or in a dose-dense weekly regimen)<sup>3</sup>. However, the feature describing the mode of administration (weekly versus 3-weekly) was consistently selected by our prediction models. There are several possible explanations for this, including the limited size of our dataset, or a possible non-linear interplay with some of the other variables in the model. Alternatively, the administration mode could potentially be a factor that affects response to NACT but not survival. This is still clinically important, since disease extent after NACT affects resectability of the disease at DPS and, therefore, perioperative morbidity and operating time<sup>41</sup>.

Previous studies have explored the value of radiogenomic data integration for HGSOC<sup>14,42</sup>. Their analyses focus on immediate primary surgery and localised biomarkers (pathology and genomics from a single biopsy; radiomics from single lesions), whereas we focus on neoadjuvant therapy and analyse global biomarkers capturing metastatic disease (circulating tumour DNA and multi-site radiomics). We enriched our data set by segmenting the full disease burden and including radiomics of all abdominopelvic tumour locations, reaching a median of 18 volumes of interest per scan in the training set. Our analysis showed that response to NACT in HGSOC varies spatially: lesions in the ovaries and/or pelvis had poorer response than those in the omentum. Our analysis also integrates imaging features for all of the abdominopelvic disease with blood-based biomarkers, including ctDNA and CA-125. We showed that global tumour burden and volume of pelvic and ovarian tumours at baseline are significantly correlated with both CA-125 and ctDNA *TP53* MAF. However, our analysis of volumetric data also showed that ctDNA signal was most strongly correlated with the volume of disease in the ovaries and/or pelvis, suggesting that the simultaneous reading of ctDNA and CA-125 at baseline could play a role in helping determine disease spread and response in the diagnostic setting. This variation could be due to potential differences in ctDNA release kinetics or to genomic heterogeneity between disease sites.

Our study has several limitations. Firstly, the response metric (relative total volume change) treats the disease as a single entity,

ignoring the subtle interplay between the pelvis/ovaries and the omentum. Cell-of-origin analysis via mutation, methylation or fragmentomics to infer contribution from specific cell types to the circulating free DNA pool could help to predict specific volume reductions. Secondly, the importance of some data streams is driven by the specific features that were included in the model. Indeed, given the correlation between ctDNA and total disease volume, and the higher dimensionality of the collections of clinical and radiomics features, it is not surprising that we did not find any additional value in adding ctDNA based biomarkers to our integrated models. Our study was limited to ctDNA *t-MAD*, *TP53* MAF and *TP53* mutation status. A more sophisticated analysis of ctDNA, applied on larger multi-institutional datasets, and measured longitudinally across different time points, may find that there are other complementarities between ctDNA and imaging data<sup>43–46</sup>. Another important limitation of our results is the size of the datasets used. This may have impacted the added value of ctDNA features and reduced the strength of the radiomic correlations, which are often borderline significant. Our conclusions need to be validated in larger cohorts as the next step in their development.

Our dataset also lacked detailed quantification of known biomarkers such as *BRCA1/2* or *CCNE1* mutations, homologous recombination deficiency, or molecular subtypes as evaluated from gene expression or copy number data, which are likely to be important factors and should be incorporated into future models<sup>47–55</sup>. For a small subset of samples in the hold-out validation cohort that did have *BRCA1/2* status information, we did not find that it had an effect on either the response or the errors of the prediction (Supplementary Fig. S6). Detailed investigations regarding histopathological features such as the presence of necrosis, extent of tumoural fibrosis, or quantification of reactive stroma, were also beyond the scope of this study. Adding extensive molecular profiling and computational pathology in the future will not only potentially improve the performance of the model, but also enable detailed biological interpretation of the main drivers behind response.

One potential barrier for clinical implementation of this work is that manual segmentation for volumetry is time consuming. However, accurate and highly adaptable deep learning models for automatic segmentation are being rapidly developed across cancer types, facilitating the integration of volumetry into clinical workflows as well as large-scale radiomics computation<sup>56,57</sup>.

In conclusion, our study is a proof-of-principle for the integration of radiogenomic data to describe and predict the response of HGSOC patients to NACT. We demonstrate that the systematic multi-scale integration of standard-of-care biomarkers provides critical predictive power and important insights into the complex spatial configuration of the disease. Further clinical development of *IRON* could have significant impacts as a stratification tool in clinical and experimental settings—for example avoiding delays in surgery for patients who are unlikely to respond to chemotherapy—and could bring forward a new generation of clinical trials for HGSOC, with rapid, effective endpoints that improve and expedite the discovery of new therapies.

## Methods

### Patient cohorts

This study was performed in accordance with the principles of the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. Written informed consent was obtained from all patients prior to any study related procedures.

Two patient cohorts were used in the study. The main cohort (the 'NeOV' cohort) was randomly split into a training set and a hold-out set. The training set was used to train the machine learning models, and as a discovery dataset for univariable analyses. The hold-out set was set aside and used only to validate the model predictions. A second dataset (the 'Barts' cohort) was used for external validation.



For both data sets, patients had a confirmed histopathological diagnosis of HGSOc and were treated with neoadjuvant chemotherapy before delayed primary surgery. All patients within the main data set were treated at Cambridge University Hospitals NHS Foundation Trust between 2009 and 2020 and were recruited into a prospective clinical study approved by the local research ethics committee (REC reference numbers: 08/H0306/61). All patients within the Barts data set were treated at Barts Health NHS Trust between 2009 and 2018 and recruited into prospective clinical study approved by the local research ethics committee (IRAS reference numbers: 243824).

Patients were identified and included in the analysis based on the availability of at least two CT scans at baseline and after NACT. Additionally, for the NeOV cohort at least one baseline plasma sample for ctDNA assessments was required.

### Clinical data

Data regarding patient demographics, treatment, and disease were collected from the patient electronic medical records including notes from multidisciplinary team discussions (MDTs). PFS was defined as the time between histopathological diagnosis and first radiological evidence of progression or recurrence. Where progression date was unclear from radiology reports alone, (e.g. successive imaging studies with subtle/mixed changes), clinical interpretation of progression was incorporated in PFS date calling (e.g. documentation of breaking bad news to patients, treatment decisions for subsequent line therapy). OS was defined as the time from diagnosis to death. Stage was determined using the International Federation of Gynaecology and Obstetrics (FIGO) criteria for ovarian cancer<sup>27</sup>.

**Management.** The management of all patients in the study including indications for surgery were discussed and decided upon within MDTs as per the UK National Health Service (NHS) guidelines. Surgeries were performed through a midline laparotomy by a team specialised in surgical gynaecology aiming to achieve total macroscopic tumour clearance. Overall,  $n=107$  patients were treated with platinum-based chemotherapy in combination with paclitaxel, while 23 received carboplatin as a monotherapy. Patients were considered to have received weekly chemotherapy if the average time interval between doses ranged between 6 and 10 days, 3-weekly if the average time interval was between 18 and 24 days, and irregular otherwise. Of those receiving combination therapy, 81 patients received carboplatin and paclitaxel 3-weekly, 18 patients received carboplatin three-weekly and paclitaxel weekly, 2 patients received carboplatin and paclitaxel weekly, and the rest received the treatment at irregular intervals. In addition, 4 patients were treated with Doxorubicin. Supplementary Table S3 shows the breakdown for the three datasets.

**BRCA status.** Germline *BRCA1* and *BRCA2* mutational status was determined for 45 patients in the training cohort and 15 patients in the internal hold-out cohort. The remaining cases reflect historical practice of not testing patients above the age of 70.

### ctDNA

Blood samples were collected before initiation of treatment with chemotherapeutic agents. DNA was extracted from plasma (1.2–4 ml) using QIAvac 24 Plus vacuum manifold and the QIAamp Circulating Nucleic Acid kit (Qiagen), or with QIAasymphony (Qiagen) as per manufacturer instructions. DNA quantification was performed using Qubit dsDNA broad-range or high-sensitivity assay kits and the Qubit Fluorometer (Thermo Fisher Scientific). Tagged-amplicon deep sequencing DNA libraries were prepared as described by ref. 58. Following purification with AMPure XP magnetic beads (Beckman Coulter Life Sciences), 10nM libraries were quantified using Agilent

Bioanalyzer and Agilent DNA 1000 kit or Agilent TapeStation and ScreenTape D1000 (Agilent Technologies) according to manufacturer instructions, and pooled for sequencing on MiSeq, HiSeq 2500 or HiSeq 4000 (Illumina).

Shallow whole genome DNA libraries (10 million reads per sample) were prepared using the ThruPLEX DNA-Seq kit (Takara) and purified with AMPure XP magnetic beads (Beckman Coulter Life Sciences). 10nM libraries were quantified using Agilent D5000 ScreenTape System or Roche KAPA library quantification kits and pooled for sequencing on HiSeq 4000 (Illumina) in paired-end 150-base pair mode. On removing adaptor sequences, shallow whole genome sequence reads were aligned to the 1000 Genomes Project version of the unmasked human reference genome GRCh37 using the BWA-MEM alignment software<sup>59</sup>. Somatic copy number analysis was performed using CNAclinic<sup>60</sup> to generate trimmed Median Absolute Deviation from copy-number neutrality (t-MAD) scores as previously described<sup>21</sup>.

Demultiplexed TAm-Seq reads were aligned to the GRCh37 reference genome by amplicon, and mutations called where non-reference alleles met probability criteria in both replicates, as previously described<sup>58</sup>. Samples lacking mutation calls were manually curated using the Integrative Genomics Viewer.

### Imaging protocol

Although adnexal masses are most frequently detected on ultrasound (US), CT is a universal tool for the detailed staging of patients with adnexal masses and is crucial for treatment planning as it maps the disease for surgical debulking and helps stratify patients for primary debulking surgery or neoadjuvant chemotherapy. US suffers from high observer-dependency and does not provide images covering the entire pelvis and abdomen as CT does. Therefore, CT was chosen as the imaging modality for this study. Clinically requested contrast-enhanced venous phase CT scans covering the abdomen and pelvis (with or without the chest depending on the clinical request and imaging findings) were either acquired at Cambridge University Hospitals NHS Foundation Trust (CUHNHSFT) or in other institutions across the UK and then imported into the picture archiving and communication system (PACS) at CUHNHSFT. Therefore, different manufacturers and scanning protocols were used. Baseline scans were acquired between 0 and 14 weeks before initiation of neoadjuvant chemotherapy and post-treatment scans were acquired for response assessment after 1.6–5.8 months of treatment. All scans were initially identified on the local PACS and then fully anonymised for further study-related processing.

### Data processing

The code and data used in this study are available at <https://github.com/micrisor/OvarianIntegration><sup>61</sup>.

**Image segmentation and labelling.** On axial images reconstructed with a slice thickness of typically 5 mm (Supplementary Table S4), and pixel spacings ranging between 0.053 and 0.095, using abdominal soft tissue window settings, all cancer lesions were segmented semi-automatically by a board-certified radiologist with ten years of experience in clinical imaging, using Microsoft Radiomics (project InnerEye; Microsoft, Redmond, WA, USA). The volumes of interest (VOIs) were annotated for their anatomic location: omentum, right upper quadrant, left upper quadrant, epigastrium, mesentery, right paracolic gutter, left paracolic gutter, ovaries & pelvis, infrarenal abdominal lymph nodes, suprarenal abdominal lymph nodes, inguinal lymph nodes, supradiaphragmatic lymph nodes, other chest lymph nodes, parenchymal liver metastases, and lung metastases. Cystic and solid tumour parts were included in these segmentations. Automated sub-segmentation of hyperdense/calcified, hypodense/cystic or fatty and intermediately dense/solid tissue was performed for omental lesions and lesions of the ovaries and pelvis using a previously

described and validated technique<sup>25</sup>. Baseline and follow-up CT scans were evaluated according to RECIST 1.1 for response assessment<sup>62</sup>. Pleural effusions and ascites were assessed semiquantitatively (0 = none, 1 = trace, 2 = less than 5 cm when measured perpendicularly to chest/abdominal wall, 3 = 5 cm or more when measured perpendicularly to chest/abdominal wall).

**Radiomics features.** VOIs drawn manually were split into connected components using MATLAB's *bwlabeln* function with a three-dimensional connectivity of 26, which assumes that voxels are connected if their faces, edges, or corners touch. Voxels with intensities below -100 HU were removed from the radiomics calculations. Radiomics features were extracted using the IBSI-calibrated CERR Radiomics toolbox<sup>63</sup> (December 2018 version, GitHub hash: 5974376be7103d5c3831690c62aa721fc784d949), including shape, intensity-volume histogram, first-order, and Haralick texture features (see Supplementary Data 1 for the full list). Intensity-volume histogram features, inspired by the Vx features commonly extracted in radiotherapy dose-volume histograms, corresponded to the volumes spanned by voxels above a certain intensity value (denoted 'HU > x' in Supplementary Data 1). To calculate Haralick texture features for each lesion, co-occurrence matrices for 100 grey levels (up to a maximum of 1000 HU) were computed independently for each direction along 2D slices and averaged. To calculate the rim and peripheral radiomics features, for each VOI two copies were created by eroding and dilating the contours by 0.4 cm along the 2D slices. The value of the margin was chosen in order to capture slices of at least 1 cm diameter. Erosion and dilation were achieved by convolving the contour with a circular mask of the desired margin. Ratio features were computed by dividing the results obtained from the eroded and standard volumes (rim features), or the dilated and standard volumes (peripheral features). Shape features were not included in the ratios. Once standard and ratio features were calculated for each lesion, a single value was extracted for the whole patient by taking the unweighted mean of all lesions.

**Other imaging features.** In contrast to texture features, which we averaged across lesions, we did use the volume and the number of lesions in each of the anatomic locations as individual features. We also computed the mean, maximum, and total volume, as well as the number of lesions with volume bigger than 0, 1, 10 and 100 cm<sup>3</sup>. In addition, we defined four binary features that indicated whether or not there were hypodense or hyperdense regions in either omental or pelvic/ovarian lesions. Ascites and pleural effusion were used as defined by the radiologist, as explained above (section 14, Radiological image analysis).

**Clinical features.** Chemotherapy regimens were extracted from the clinical records. We recorded whether the patient had received any Carboplatin, Paclitaxel or Doxorubicin in three binary variables. Mean periods were calculated by averaging the time intervals between sessions. We defined weekly regimen as having a mean period of 6–10 days, both included; and 3-weekly as having a mean frequency of 18–24 days, both included. Typical combinations included weekly Paclitaxel and 3-weekly Carboplatin; both weekly; and both three-weekly (Supplementary Fig. S1). These combinations were, therefore, encoded in binary variables. We also recorded whether patients had received Carboplatin only in a binary variable. FIGO stage was encoded by assigning ordinal numbers to in order of aggressiveness, from 1 for stage 1A to 10 for stage 4B. The exact mapping is listed in Supplementary Table S5. CA-125 values were also extracted from the clinical records, with the measurement closest to the beginning of treatment being used for analysis. Performance status was only available for a

subset of the patients in the training set, and it did not correlate with volumetric response (Supplementary Fig. S2). It was, therefore, not included in the predictive models.

### Statistical analysis

**Tumour burden correlations.** Correlations were calculated using Spearman correlation coefficient and *p* values were corrected using the Benjamini–Hochberg procedure as implemented in the scikit-learn Python package<sup>64</sup>. The total tumour burden affecting lymph nodes was calculating by combining the lesions found in infrarenal, suprarenal, inguinal, supradiaphragmatic, and chest lymph nodes. The difference between the relative volume change in the omentum and in the ovaries/pelvis was quantified using the Mann-Whitney *U* test (two-sided). The difference in relative volume change for patients with different BRCA status was also quantified using the Mann-Whitney *U* test (two-sided).

**Imaging clusters.** To identify clinically or biologically meaningful clusters of radiomics features, we clustered their Spearman correlation coefficients used a hierarchical clustering approach. The optimal number of imaging clusters was obtained by maximising their correlation with any of the clinical and biological features. To do this, we calculated the maximum Spearman correlation coefficient between each cluster and any of the biological/clinical features, and averaged the result across all clusters (Supplementary Fig. S4a). The metric reached a plateau at six clusters, which was therefore chosen as the optimal number (Supplementary Fig. S4b). Clusters were cross-checked using non-negative matrix factorisation. We matched the number of clusters to 6, used a coordinate descent solver and an initialisation based on non-negative random matrices, following the scikit-learn implementation. The most predictive feature in each cluster (Fig. 2b) was chosen in terms of its Spearman correlation with volumetric treatment response. *p* values were corrected using the Benjamini-Hochberg procedure.

### Machine learning models

**Endpoint.** Models were trained to predict the relative volumetric response,

$$\text{Response} = \log \left( \frac{\sum_i^N v_i^1}{\sum_i^N v_i^0} \right), \quad (1)$$

where  $v_k^0$  is the volume of the *k*th lesion found at the pre-treatment time point,  $v_k^1$  is the volume of the *k*th lesion found at the post-treatment time point, *N* is the number of lesions at the pre-treatment time point, and *N'* is the number of lesions at the post-treatment time point.

**Training.** We created a machine learning framework to predict response to chemotherapy, evaluated on the basis of relative total volume change. All features used in the model were extracted from data obtained at diagnosis. We used the NeOv training set to train and optimise the models. Once trained and frozen, we evaluated the models in the internal hold-out set and in the external validation set. We used an increasing number of features, in order of general availability. We started with clinical features (age, stage, and 9 treatment features); then added baseline CA-125; then imaging features (164 features); and finally ctDNA (3 features). For each combination we retrained the framework and derived a new model. The full list of features can be found in Supplementary Data 1.

The predictions were based on an unweighted ensemble regressor<sup>65</sup>. The ensemble included three different machine learning algorithms: an elastic net, a support vector regressor with a radial basis function (RBF) kernel, and a random forest, all of them coded in Python using the scikit-learn package<sup>64</sup>. Each algorithm was embedded

in a scikit-learn pipeline with three pre-processing steps, namely collinearity reduction, z-score standardisation, and univariable feature selection. Collinearity reduction removed all features with a mutual Pearson correlation above 0.95, retaining only the one with the highest correlation with the response variable. The feature selection step removed all features that were not ranked within the top  $k$  according to their  $F$  value. The scores produced by the three pipelines were averaged to form the prediction.

We used a five-fold cross validation setup to optimise model hyperparameters in the training set, covering the hyperparameter ranges shown in Supplementary Table S6. The optimisation was based on a randomised search in the hyperparameter space to optimise mean square error (MSE). Once the optimal hyperparameters were found (Supplementary Table S7), we determined model parameters by re-fitting the model to the entire training set. To increase model robustness, we repeated this process five times with five different cross-validation seeds. The five resulting optimal models were combined to form the final ensemble, in which the prediction is simply the average of the five predicted scores. In this regard, our ensemble setup has two different tiers: the randomisation tier (five seeds), and the algorithmic tier (three regressors acting in parallel for each seed). The modelling framework is completely agnostic to cancer type and is generalisable to other applications.

**Validation.** We validated the models on the hold-out internal validation set ( $n = 20$ ) and the external validation set ( $n = 42$ ). The external validation set did not have ctDNA data available, so the corresponding features were set to a constant value corresponding to the training set averages as a trivial imputation mechanism. To quantify the calibration of the models, we computed the MSE. To quantify the discriminative power of the models, we computed the Spearman correlation coefficient and  $p$  value between the predicted and observed response scores. To estimate a potential classification power on the external validation cohort, we evaluated the volumetric thresholds corresponding to RECIST 1.1 criteria (30% reduction), WHO criteria (50% reduction), spherical-volume (65% reduction), and ellipsoidal volume (30% reduction) and computed the corresponding area under the receiver-operating characteristic curve (AUC). Finally, we evaluated whether the predicted scores were also able to rank patients into the different RECIST 1.1 categories using the  $p$  value associated with the point biserial correlation coefficient and the AUC for responder vs. stable disease.

**Feature importance.** We evaluated feature importance in two different steps. First, we computed the frequency with which features were selected after the collinearity reduction and univariable selection steps. We repeated the process for each of the three algorithms and each of the five cross-validation seeds, which means that features could be selected between 0 and a maximum of 15 times, as seen in Fig. 5. The table in Fig. 5 displays only features that were chosen at least three out of five times in each cross-validation loops, for robustness. Second, we computed the importance of each individual feature within the regression algorithm. This was only possible for the elastic net, where we used the feature's coefficients, and the random forests, where we used impurity-based feature importances. The results were averaged across the five seeds and the two algorithms, with the table in Fig. 5 displaying only features that were chosen at least three out of five times in the cross-validation loops, as before. Finally, we assessed the effects of volume-related features, including the two most frequently selected volume features (mean volume and volume of infrarenal lymph nodes), and the two clusters with important contributions from volume features (cluster 1 and cluster 4). Features were removed at inference stage, and replaced with the corresponding average calculated from the training set. Removing features that drive the prediction would result in a significant change in performance, which can be

either positive or negative depending on whether the corresponding features in the training and testing sets are mutually calibrated.

## Data availability

The clinical, genomic, and radiomics data generated in this study and used to train predictive models have been deposited in a Github repository at <https://github.com/micrisor/OvarianIntegration><sup>61</sup> with <https://doi.org/10.5281/zenodo.8152137>. The remaining data are available within the Supplementary Information and Source Data files. Source data are provided with this paper.

## Code availability

The code and data used in this study are available at <https://github.com/micrisor/OvarianIntegration><sup>61</sup> with <https://doi.org/10.5281/zenodo.8152137>.

## References

- Vergote, I. et al. Neoadjuvant chemotherapy or primary surgery in stage IIIC or IV ovarian cancer. *N. Engl. J. Med.* **363**, 943–953 (2010).
- Kehoe, S. et al. Primary chemotherapy versus primary surgery for newly diagnosed advanced ovarian cancer (CHORUS): An open-label, randomised, controlled, non-inferiority trial. *Lancet* **386**, 249–257 (2015).
- Clamp, A. R. et al. Weekly dose-dense chemotherapy in first-line epithelial ovarian, fallopian tube, or primary peritoneal carcinoma treatment (ICON8): primary progression free survival analysis results from a GCIG phase 3 randomised controlled trial. *Lancet* **394**, 2084–2095 (2019).
- Van Meurs, H. S. et al. Which patients benefit most from primary surgery or neoadjuvant chemotherapy in stage IIIC or IV ovarian cancer? An exploratory analysis of the European Organisation for Research and Treatment of Cancer 55971 randomised trial. *Eur. J. Cancer* **49**, 3191–3201 (2013).
- Meyer, L. A. et al. Use and effectiveness of neoadjuvant chemotherapy for treatment of ovarian cancer. *J. Clin. Oncol.* **34**, 3854–3863 (2016).
- Schwarz, R. F. et al. Spatial and temporal heterogeneity in high-grade serous ovarian cancer: a phylogenetic analysis. *PLoS Med.* **12**, e1001789 (2015).
- Bashashati, A. et al. Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. *J. Pathol.* **231**, 21–34 (2013).
- Zhang, A. W. et al. Interfaces of malignant and immunologic clonal dynamics in ovarian cancer. *Cell* **173**, 1755–1769.e22 (2018).
- Cioffi, R. et al. Neoadjuvant chemotherapy in high-risk ovarian cancer patients: role of age. *Tumori* **105**, 168–173 (2019).
- Tajik, P. et al. The FIGO stage IVA versus IVB of ovarian cancer: prognostic value and predictive value for neoadjuvant chemotherapy. *Int. J. Gynecol. Cancer* **28**, 453–458 (2018).
- Batra, S., Arora, R. & Dave, K. Predictive value of changes in the serum CA-125 levels in patients undergoing interval debulking surgery after neoadjuvant chemotherapy in advanced epithelial ovarian carcinoma. *Int. J. Reproduct. Contracept. Obstetr. Gynecol.* **8**, 483 (2019).
- Zhang, D. et al. Serum CA125 levels predict outcome of interval debulking surgery after neoadjuvant chemotherapy in patients with advanced ovarian cancer. *Clin. Chim. Acta* **484**, 32–35 (2018).
- Liang, W.-F. et al. The added value of CA125 normalization before interval debulking surgery to the chemotherapy response score for the prognostication of ovarian cancer patients receiving neoadjuvant chemotherapy for advanced disease. *J. Cancer* **12**, 946–953 (2021).
- Lu, H. et al. A mathematical-descriptor of tumor-mesoscopic-structure from computed-tomography images annotates



- prognostic- and molecular-phenotypes of epithelial ovarian cancer. *Nat. Commun.* **10**, 764 (2019).
15. Vargas, H. A. et al. Radiogenomics of high-grade serous ovarian cancer: multireader multi-institutional study from the Cancer Genome Atlas Ovarian Cancer Imaging Research Group. *Radiology* **285**, 482–492 (2017).
  16. Sharbatoghli, M. et al. Prediction of the treatment response in ovarian cancer: a ctDNA approach <https://doi.org/10.1186/s13048-020-00729-1> (2020).
  17. Nicolajilwan, M. et al. Addition of MR imaging features and genetic biomarkers strengthens glioblastoma survival prediction in TCGA patients. *J. Neuroradiol.* **42**, 212–221 (2015).
  18. Grossmann, P. et al. Defining the biological basis of radiomic phenotypes in lung cancer. *eLife* **6**, e23421 (2017).
  19. Chaddad, A., Daniel, P., Sabri, S., Desrosiers, C. & Abdulkarim, B. Integration of radiomic and multi-omic analyses predicts survival of newly diagnosed IDH1 wild-type glioblastoma. *Cancers* **11**, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6721570/> (2019).
  20. Parkinson, C. A. et al. Exploratory analysis of TP53 mutations in circulating tumour DNA as biomarkers of treatment response for patients with relapsed high-grade serous ovarian carcinoma: a retrospective study. *PLoS Med.* **13**, e1002198 (2016).
  21. Moulriere, F. et al. Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci. Transl. Med.* **10**, eaat4921 (2018).
  22. Kim, Y. M. et al. Prospective study of the efficacy and utility of TP53 mutations in circulating tumor DNA as a non-invasive biomarker of treatment response monitoring in patients with high-grade serous ovarian carcinoma. *J. Gynecol. Oncol.* **30**, 32 (2019).
  23. Adalsteinsson, V. A. et al. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat. Commun.* **8**, 1–13 (2017).
  24. Scherer, F. et al. Distinct biological subtypes and patterns of genome evolution in lymphoma revealed by circulating tumor DNA. *Sci. Transl. Med.* **8**, 364ra155 (2016).
  25. Rundo, L. et al. Tissue-specific and interpretable sub-segmentation of whole tumour burden on CT images by unsupervised fuzzy clustering. *Comput. Biol. Med.* **120**, 103751 (2020).
  26. Cruickshank, D., Fullerton, W. T. & Kloppner, A. The clinical significance of pre-operative serum CA 125 in ovarian cancer. *BJOG Int. J. Obstetr. Gynaecol.* **94**, 692–695 (1987).
  27. Prat, J. & on Gynecologic Oncology, F. C. Staging classification for cancer of the ovary, fallopian tube, and peritoneum. *Int. J. Gynecol. Obstetr.* **124**, 1–5 (2014).
  28. Keek, S. A., Leijenaar, R. T., Jochems, A. & Woodruff, H. C. A review on radiomics and the future of theranostics for patient selection in precision medicine. *Br. J. Radiol.* **91**, 20170926 (2018).
  29. Traverso, A., Wee, L., Dekker, A. & Gillies, R. Repeatability and reproducibility of radiomic features: a systematic review. *Int. J. Radiat. Oncol. Biol. Phys.* **102**, 1143–1158 (2018).
  30. Sanduleanu, S. et al. Tracking tumor biology with radiomics: a systematic review utilizing a radiomics quality score. *Radiother. Oncol.* **127**, 349–360 (2018).
  31. Willeminck, M. J. et al. Preparing medical imaging data for machine learning. *Radiology* **295**, 4–15 (2020).
  32. Vargas, H. A. et al. Association between morphologic CT imaging traits and prognostically relevant gene signatures in women with high-grade serous ovarian cancer: a hypothesis-generating study. *Radiology* **274**, 742–751 (2015).
  33. Nougaret, S. et al. High-grade serous ovarian cancer: associations between BRCA mutation status, CT imaging phenotypes, and clinical outcomes. *Radiology* **285**, 472–481 (2017).
  34. Shan, Q. Y. et al. CT-based peritumoral radiomics signatures to predict early recurrence in hepatocellular carcinoma after curative tumor resection or ablation. *Cancer Imaging* **19**, 11 (2019).
  35. Braman, N. M. et al. Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. *Breast Cancer Res.* **19**, 57 (2017).
  36. Braman, N. et al. Association of peritumoral radiomics with tumor biology and pathological response to preoperative targeted therapy for HER2 (ERBB2)-positive breast cancer. *JAMA Netw. Open* **2**, e192561 (2019).
  37. Dou, T. H., Coroller, T. P., van Griethuysen, J. J., Mak, R. H. & Aerts, H. J. Peritumoral radiomics features predict distant metastasis in locally advanced NSCLC. *PLoS One* **13**, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6214508/> (2018).
  38. Granot, D. et al. In vivo imaging of the systemic recruitment of fibroblasts to the angiogenic rim of ovarian carcinoma tumors. *Cancer Res.* **67**, 9180–9189 (2007).
  39. de Lima, C. A., Jammal, M. P., Etchebehere, R. M., Murta, E. F. C. & Nomelini, R. S. Lymphocytes in peritumoral stroma: evaluation in epithelial ovarian neoplasms. *Immunol. Investig.* **49**, 397–405 (2020).
  40. Tuhkanen, H. et al. Genetic alterations in the peritumoral stromal cells of malignant and borderline epithelial ovarian tumors as indicated by allelic imbalance on chromosome 3p. *Int. J. Cancer* **109**, 247–252 (2004).
  41. Günakan, E. et al. Factors associated with surgical morbidity of primary debulking in epithelial ovarian cancer. *Obstet. Gynecol. Sci.* **63**, 64–71 (2020).
  42. Boehm, K. M. et al. Multimodal data integration using machine learning improves risk stratification of high-grade serous ovarian cancer. *Nat. Cancer* **3**, 723–733 (2022).
  43. Ulz, P. et al. Inferring expressed genes by whole-genome sequencing of plasma DNA. *Nat. Genet.* **48**, 1273–1278 (2016).
  44. Ulz, P. et al. Inference of transcription factor binding from cell-free DNA enables tumor subtype prediction and early detection. *Nat. Commun.* **10**, 1–11 (2019).
  45. Esfahani, M. S. et al. Inferring gene expression from cell-free DNA fragmentation profiles. *Nat. Biotechnol.* **40**, 585–597 (2022).
  46. Jiang, P. et al. Plasma DNA end-motif profiling as a fragmentomic marker in cancer, pregnancy, and transplantation. *Cancer Discov.* **10**, 664–673 (2020).
  47. Frey, M. K. & Pothuri, B. Homologous recombination deficiency (HRD) testing in ovarian cancer clinical practice: a review of the literature. *Gynecol. Oncol. Res. Pract.* **4**, 1–11 (2017).
  48. Sato, E. et al. Intraepithelial CD8+ tumor-infiltrating lymphocytes and a high CD8+/regulatory T cell ratio are associated with favorable prognosis in ovarian cancer. *Proc. Natl Acad. Sci. USA* **102**, 18538–18543 (2005).
  49. Goode, E. L. et al. Dose-response association of CD8+ tumor-infiltrating lymphocytes and survival time in high-grade serous ovarian cancer. *JAMA Oncol.* **3**, e173290–e173290 (2017).
  50. Patch, A. M. et al. Whole-genome characterization of chemoresistant ovarian cancer. *Nature* **521**, 489–494 (2015).
  51. Rosen, D. G. et al. Cyclin E expression is correlated with tumor progression and predicts a poor prognosis in patients with ovarian carcinoma. *Cancer* **106**, 1925–1932 (2006).
  52. Karst, A. M. et al. Cyclin E1 deregulation occurs early in secretory cell transformation to promote formation of fallopian tube-derived high-grade serous ovarian cancers. *Cancer Res.* **74**, 1141–1152 (2014).
  53. Bolton, K. L. et al. Association between BRCA1 and BRCA2 mutations and survival in women with invasive epithelial ovarian cancer. *JAMA* **307**, 382–390 (2012).
  54. Tan, D. S. et al. "BRCAness" syndrome in ovarian cancer: a case-control study describing the clinical features and outcome of patients with epithelial ovarian cancer associated with BRCA1 and BRCA2 mutations. *J. Clin. Oncol.* **26**, 5530–5536 (2008).



55. Cass, I. et al. Improved survival in women with BRCA-associated ovarian carcinoma. *Cancer* **97**, 2187–2195 (2003).
56. Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J. & Maier-Hein, K. H. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **18**, 203–211 (2021).
57. Buddenkotte, T. et al. Calibrating ensembles for scalable uncertainty quantification in deep learning-based medical image segmentation. *Comput. Biol. Med.* **163**, 107096 (2023).
58. Forshew, T. et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci Transl. Med.* **4**, 136ra68 (2012).
59. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, <http://arxiv.org/abs/1303.3997> (2013).
60. Chandrananda, D. CNA clinic, <https://github.com/sdchandra/CNAclinic> (2017).
61. IRON: Integrated Radiogenomics for Ovarian Neoadjuvant therapy. <https://doi.org/10.5281/zenodo.8152137>, <https://github.com/micrisor/OvarianIntegration> (2023).
62. Eisenhauer, E. A. et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur. J. Cancer* **45**, 228–247 (2009).
63. Apte, A. P. et al. Technical note: extension of CERR for computational radiomics: a comprehensive MATLAB platform for reproducible radiomics research. *Med. Phys.* **45**, 3713–3720 (2018).
64. Pedregosa, F. et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
65. Ju, C., Bibaut, A. & van der Laan, M. The relative performance of ensemble methods with deep convolutional neural networks for image classification. *J. Appl. Stat.* **45**, 2800–2818 (2018).

## Acknowledgements

We dedicate this work to the memory of Prof. Naveena Singh, whose invaluable contributions and enduring passion for the field have left an indelible mark, and she will be greatly missed. We acknowledge funding and support from Cancer Research UK and the Cancer Research UK Cambridge Centre [A22905, C9545/A29580, A25117, A17197, A19274, A20240], the Wellcome Trust Innovator Award [RG98755], The Mark Foundation for Cancer Research and the CRUK National Cancer Imaging Translational Accelerator (NCITA) [A27066]. Additional support was also provided by the National Institute of Health Research (NIHR) Cambridge Biomedical Research Centre [RC-1215-20014] and EPSRC Tier-2 capital grant [EP/P020259/1]. This research was also supported by Target Ovarian Cancer and the Medical Research Council through their Joint Clinical Research Training Fellowship for Dr Moore. Work in the Cancer Molecular Diagnostics Laboratory/Blood Processing Laboratory was supported by the NIHR Cambridge Biomedical Research Centre [BRC-1215-20014], Cancer Research UK Cambridge Centre and the Mark Foundation Institute for Integrated Cancer Medicine. Microsoft Radiomics was provided to Addenbrooke's Hospital (Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK) by the Microsoft InnerEye project. We thank all patients who participated in and donated samples and data to this study. We thank Karen Hosking and the OV04 study team for their help with clinical samples and data. We thank Oscar Rueda for useful discussions and support with data analysis. We thank staff from the Cancer Molecular Diagnostics Laboratory for performing blood collections. We thank the Cancer Research UK Cambridge Institute Compliance & Biobanking, Research Instrumentation and Cell Services, Genomics and Bioinformatics core facilities for their support with various aspects of this study. The views expressed are those of the

authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Author contributions

M.C.O., R.W., F.Ma, E.S. and J.D.B. conceived and designed the analysis. R.W., M.R., E.M., L.B., V.B., C.M., S.U., N.S., A.S., R.P., M.Z., H.A. and M.J.L. collected, processed, and annotated the data. M.C.O., R.W. and M.R. performed the main analysis. L.R., L.E.S., P.M.G., F.Mo., D.C., J.M., N.R., T.G. and AMP contributed to the analysis. M.C.O., R.W. and M.R. wrote the paper. All authors read and edited the paper.

## Competing interests

F.Mo., D.C., A.P., J.D.B. and N.R. are inventors of the patent “Enhanced detection of target DNA by fragment size analysis” (WO/2020/094775). N.R. and J.D.B. are inventors of the patent “TAm-Seq v2 method for ctDNA estimation” (WO 2016/009224A1). T.G. is an employee of and owns shares from AstraZeneca. F.Ma. and J.D.B. are founders and directors of Tailor Bio Ltd and own shares in Tailor Bio Ltd. J.D.B. has received honoraria and personal payments from AstraZeneca, GSK and Clovis Oncology. M.C.O. has received honoraria from GSK. E.S. is co-founder and shareholder of Lucida Medical Ltd. L.R. has received consulting fees from Lucida Medical Ltd. The remaining authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-41820-7>.

**Correspondence** and requests for materials should be addressed to Mireia Crispin-Ortuzar.

**Peer review information** *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

<sup>1</sup>Department of Oncology, University of Cambridge, Cambridge, UK. <sup>2</sup>Cancer Research UK Cambridge Centre, University of Cambridge, Cambridge, UK. <sup>3</sup>Department of Radiology, University of Cambridge, Cambridge, UK. <sup>4</sup>Centre for Medical Image Analysis and AI (MIAAI), Danube Private University, Krems, Austria. <sup>5</sup>Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK. <sup>6</sup>Cambridge University Hospitals NHS Foundation Trust,

Cambridge, UK. <sup>7</sup>Department of Biomedical Imaging and Image-guided Therapy, Medical University of Vienna, Vienna, Austria. <sup>8</sup>Department of Information and Electrical Engineering and Applied Mathematics, University of Salerno, Fisciano, SA, Italy. <sup>9</sup>Department of Pathology, Amsterdam UMC location Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. <sup>10</sup>Department of Cellular Pathology, Barts Health NHS Trust, London, UK. <sup>11</sup>Department of Radiology, Barts Health NHS Trust, London, UK. <sup>12</sup>“Iuliu Hatieganu” University of Medicine and Pharmacy, Cluj-Napoca, Romania. <sup>13</sup>Department of Radiology, County Clinical Emergency Hospital, Cluj-Napoca, Romania. <sup>14</sup>Department of Surgical and Medical Sciences and Translational Medicine, Sapienza University of Rome-Sant’Andrea University Hospital, Rome, Italy. <sup>15</sup>Dipartimento di Scienze Radiologiche ed Ematologiche, Università Cattolica del Sacro Cuore, Rome, Italy. <sup>16</sup>Dipartimento Diagnostica per Immagini, Radioterapia Oncologica ed Ematologia, Policlinico Universitario A. Gemelli IRCCS, Rome, Italy. <sup>17</sup>Western Balkans University, Tirana, Albania. <sup>18</sup>These authors contributed equally: Mireia Crispin-Ortuzar, Ramona Woitek. <sup>19</sup>These authors jointly supervised this work: Florian Markowitz, Evis Sala, James D. Brenton. <sup>20</sup>Deceased: Naveena Singh. ✉e-mail: [mc973@cam.ac.uk](mailto:mc973@cam.ac.uk)