nature portfolio

https://doi.org/10.1038/s41551-022-00860-y

Supplementary information

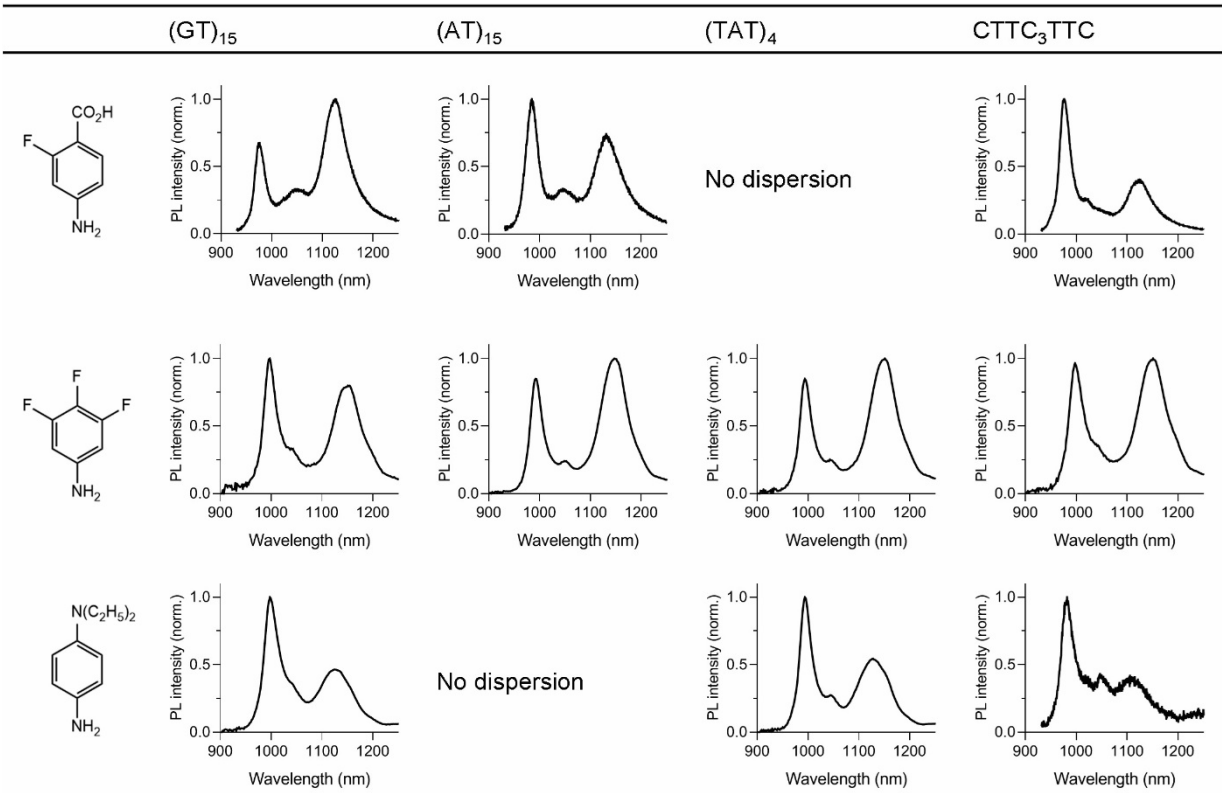
Detection of ovarian cancer via the spectral fingerprinting of quantum-defect-modified carbon nanotubes in serum by machine learning

In the format provided by the authors and unedited

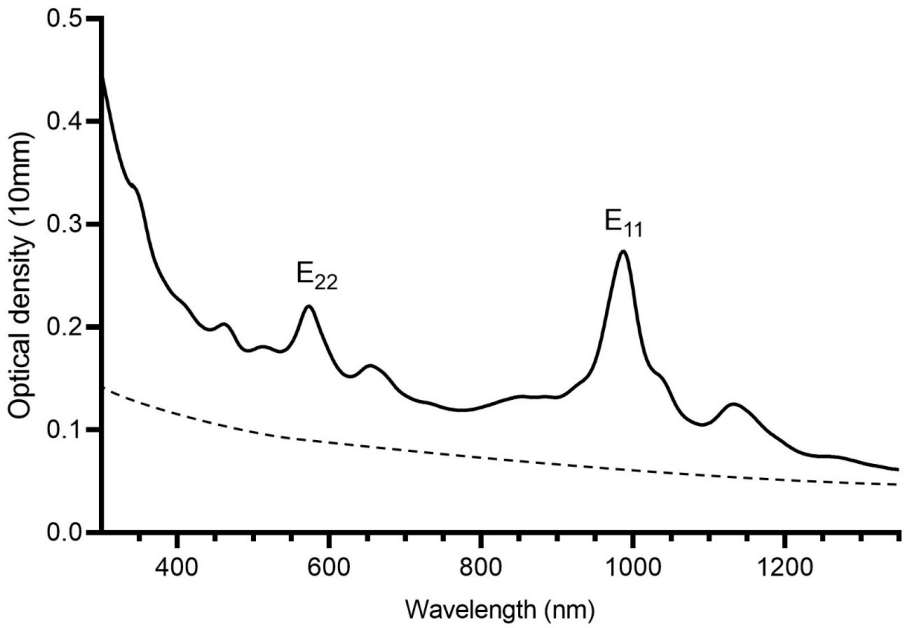
Contents

Supplementary Figures

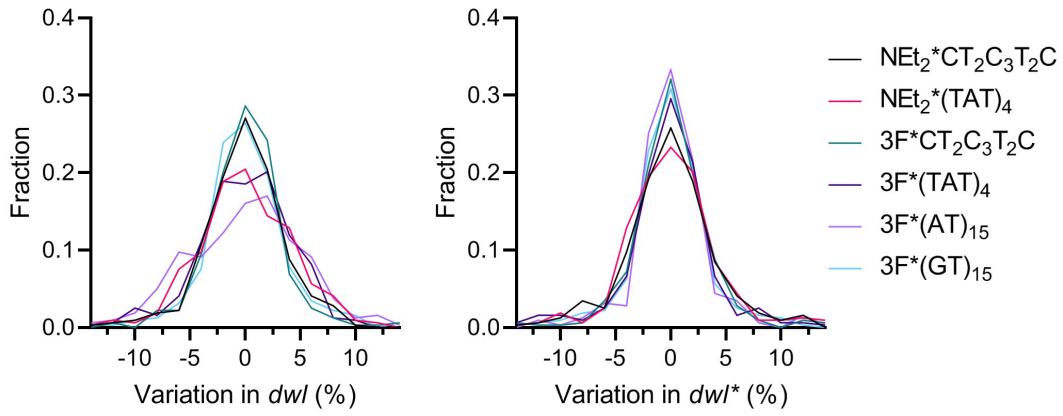
Supplementary Tables



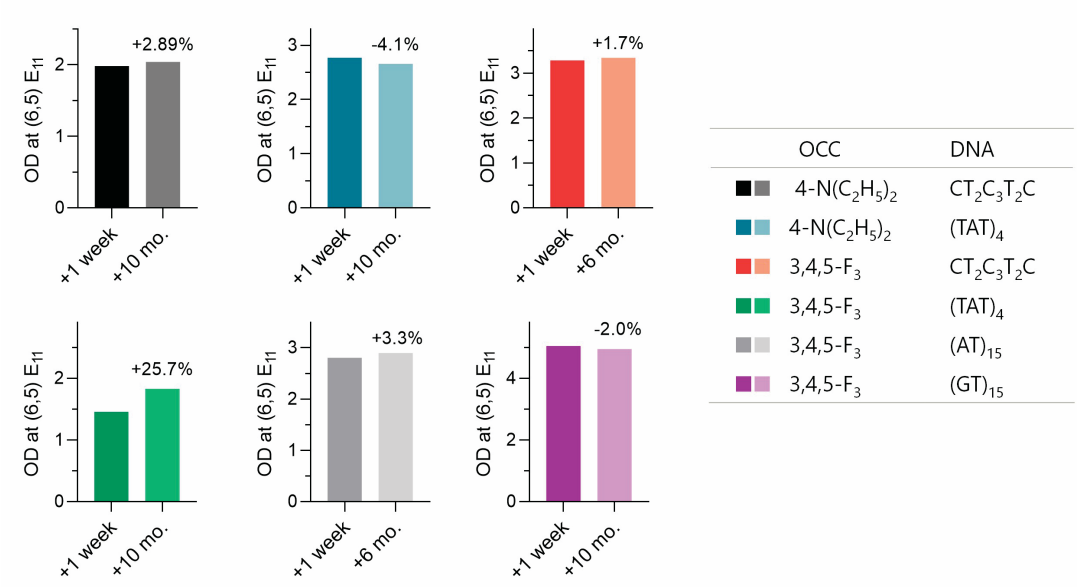
Supplementary Fig. 1 l Fluorescence spectra of OCC-DNA complexes in PBS. The excitation wavelength was 575 nm



Supplementary Fig. 2 l Absorption spectrum of the ss(GT)$_{15}$ wrapped 3,4,5-trifluoroaryl OCC functionalized OCC-DNA complex. The quantification of SWCNT concentration was performed after subtracting the background signal (dashed line). The optical density of the (6,5) SWCNT E_{11} peak was approximated as 5 μg/mL based on a previous report by Zheng et al.[1]



Supplementary Fig. 3 l Frequency distribution of variation in E_{11} and E_{11}- wavelength shifts. The variation was calculated as the difference between the centre wavelength of each measurement from the average of the triplicate measurements.



Supplementary Fig. 4 l Long term stability of OCC-DNAs. Optical density of (6,5) E_{11} bands of the OCC-DNAs, measured 1 week and 6-10 months after synthesis. The OCC-DNAs in 1XPBS were stored at 4 °C until measured.

Supplementary Table 1 | Patient demographics and pathology in the model study populations. The median of age and serum biomarker levels are shown. The number of serum samples for each biomarker are specified. Mean ± standard deviation is shown in parentheses.

Training and validation set, N = 210

	Number	Age	Hysterectomy	Serum biomarker level				
				CA125 (U/mL)	HE4 (pM)	YKL40 (pM)	Bilirubin (mg/dL)	Creatinine (mg/dL)
HGSOC	49 in total Stage 1: 1 Stage 3: 13 Stage 4: 35	64 (63.84 ± 12.48)	39	749 N=49 (1621 ± 3234)	232.6 N=47 (664.9 ± 1148)	55879 N=47 (141829 ±187234)	0.14 N=49 (0.3644 ±0.1464)	0.8 N=49 (0.9022 ±0.3244)
Remission in HGSOC	7	61 (61.71 ± 6.13)	4	6.7 N=7 (9.371 ± 5.989)	47.55 N=6 (46.80 ± 9.901)	30474 N=7 (49391 ± 43293)	0.4 N=7 (0.4857 ± 0.1464)	0.8 N=7 (0.7714 ± 0.0951)
Other gynecologic cancer	51 in total Uterine: 9 Endometerial: 32 Mullerian: 2 Other ovarian: 8	68 (64.53 ± 10.96)	33	18 N=51 (793.8 ±1720)	90.2 N=49 (349.3 ± 896.3)	64041 N=49 (133288 ± 208597)	0.4 N=46 (0.4826 ± 0.48)	0.8 N=49 (0.8816 ± 0.2991)
Non-gynecologic cancer	29 in total Breast: 23 Peritoneal: 2 Others: 2	57 (57.03 ±10.73)	7	12.9 N=29 (778.2 ± 3242)	45.90 N=22 (87.64 ± 125.3)	43252 N=23 (61719 ± 53850)	0.4 N=21 (0.4143 ± 0.2308)	0.8 N=21 (0.7667 ± 0.091)
Remission in other cancers	18	62.5 (62.94 ± 12.86)	13	9.5 N=18 (11.21 ± 7.845)	60 N=17 (56.99 ± 18.65)	35739 N=17 (55864 ± 51451)	0.4 N=13 (0.4385 ± 0.1758)	0.8 N=13 (0.8077 ± 0.1188)
Healthy	56	NA	NA	12 N=56 (13.50 ± 6.739)	36.45 N=48 (39.20 ± 9.897)	24982 N=59 (17167 ± 2235)	0.3 N=54 (0.4056 ± 0.2750)	0.8 N=54 (0.8389 ± 0.1123)

Test set, N = 54

	Number	Age	Hysterectomy	Serum biomarker level				
				CA125 (U/mL)	HE4 (pM)	YKL40 (pM)	Bilirubin (mg/dL)	Creatinine (mg/dL)
HGSOC	7 in total Stage 2: 1 Stage 3: 2 Stage 4: 4	67 (66.86 ± 4.53)	3	294 N=7 (1487 ± 3340)	175.7 N=7 (197.6 ± 200.8)	96224 N=7 (119873 ± 99934)	0.3 N=7 (0.3 ± 0.1)	0.8 N=7 (0.7571 ± 0.1397)
Other gynecologic diseases	5 in total Endometrial adenocarcinoma: 1 Mature cystic ovarian teratoma: 1 Uterine fibroids: 2 Adenomyosis: 1	48 (44.40 ± 9.00)	0	NA	NA	NA	NA	NA
Non-gynecologic cancer	32 in total Breast: 30 Gastrointestinal and mediatized to gynecologic organs: 2	64 (62.81 ± 12.60)	0	43.5 N=32 (161 ± 231.7)	90.25 N=32 (364.9 ± 1388)	82543 N=32 (115627 ± 85356)	0.5 N=32 (1.013 ± 1.588)	0.8 N=32 (0.8875 ± 0.3160)
Healthy	10	NA	NA	15 N=10 (16.17 ± 7.07)	31.40 N=10 (30.67 ± 4.95)	22924 N=9 (35729 ± 22752)	0.4 N=10 (0.57 ±0.501)	0.8 N=10 (0.77 ± 0.675)

Supplementary Table 2 | Loadings for principal component analysis plot in Fig 2d of the main text.

The table is available as a supplementary spreadsheet file.

Supplementary Table 3 | Model performance of all machine learning algorithms and OCC-DNA sensor array combinations tested for binary HGSOC classification using the training and validation set of samples (N_{sa}=215). OCC-DNA sensor responses of 2-hour incubation were used if not specified. The F_1-score, positive predictive value (PPV), sensitivity, and sensitivity at 98% specificity were the averaged values in 10-fold cross-validation. The spectral variables used to train each algorithm are *Δwl* and *Δint* for N_f=2, *dwl, dwl**, *dint*, and *dint** for N_f=4, and *dwl, dwl**, *dint*, *dint**, *Δwl,* and *Δint* for N_f=6. Abbreviations: Logi = Logistic regression, DT = Decision tree, ANN = Artificial neural network, RF = Random forest, SVM = Support vector machine.

The table is available as a supplementary spreadsheet file.

Supplementary Table 4 | Holm-Šídák's multiple comparison t-tests of Fig 3d. Effect of weighted harmonic means of PPV and sensitivity in the model optimization process to improve the sensitivity at 98% specificity. The weighing factor, b denotes the importance of sensitivity with respect to PPV in F-scoring. The analysis compares the statistical significance of the improvement in sensitivity of the 10-fold cross-validation of nanosensor arrays, comprised of up to 6 OCC-DNAs, at varying b.

The table is available as a supplementary spreadsheet file.

Supplementary Table 5 | Multivariate regression of sensor responses to medications. Overall, the coefficients of a few medications were statistically significant. The low value of adjusted R^2 indicates that 1) multivariate models based on medication failed to explain the variation of spectroscopic variables, and 2) the medications have low explanatory power to account for the variation of sensor response. Healthy group was excluded because the medication information is not available. t-statistics in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

The table is available as a supplementary spreadsheet file.

Supplementary Table 6 | SVM model performance of OCC-DNA sensor array combinations tested for HGSOC biomarker prediction using the training and validation set of samples. OCC-DNA sensor responses at 2-hour incubation with serum samples were used. Four spectra variables, *dwl* + *dwl** + *dint* + *dint** of each OCC-DNA nanosensor were used (N_f = 4). The R^2, F-score, accuracy, and sensitivity were the averaged values in 10-fold cross-validation. For the binary classification models, we divided the samples into two groups (abnormal vs. normal levels) based on the clinical references (CA125: 50 U/mL, HE4: 150 pM, YKL40: 1650 pM).

The table is available as a supplementary spreadsheet file.

Supplementary Table 7 | Hyperparameter space of each machine learning algorithm. The range was iteratively tightened by Bayesian hyperparameter optimization.

Machine learning algorithm	Parameter space
Logistic regression	'C': loguniform(log(10 ** (-2)), log(1e7)) 'solver': ['newton-cg', 'lbfgs', 'liblinear', 'sag', 'saga']
Decision tree	'max_depth': range(1, 6) 'max_features': range(1, Num) 'criterion': ['gini', 'entropy'] 'min_samples_split': loguniform(log(1e-5), log(0.5)) 'min_samples_leaf': loguniform(log(1e-5), log(0.5))
Artificial neural network	'hidden_layer_sizes': range(2, (Num + 3)) 'activation': ['relu', 'logistic', 'tanh'] 'solver': ['sgd', 'adam'] 'alpha': loguniform(log(1e-4), log(0.1)) 'learning_rate_init': loguniform(log(1e-4), log(0.1))
Random forest	'max_depth': range(1, 20) 'max_features': range(1, Num) 'n_estimators': range(100, 400) 'criterion': ['gini', 'entropy'] 'min_samples_split': loguniform(log(1e-5), log(0.5)) 'min_samples_leaf': loguniform(log(1e-5), log(0.5))
Support vector machine & Support vector regression	'C': loguniform(log(10 ** (-2)), log(1e7)) 'kernel': 'rbf' 'gamma': loguniform(log(10 ** (-3)), log(1e3))

Num = the_number_of_feature_vectors_in_the_array = (the_number_of_spectroscopic_variables)(the_number_of_OCC – DNA_nanosensor_in_the_array)

Appendix 1

PROBAST (**P**rediction model study **R**isk **O**f **B**ias **AS**sessment **T**ool)
Version of 15/05/2019 – Page 7
For more information, please see www.probast.org

Step 1: Specify your systematic review question

Criteria	**Specify your systematic review question**
Intended use of model:	Diagnosis of high-grade serous ovarian cancer using a machine perception nanosensor array
***Participants** including selection criteria and setting:*	Female adults patients diagnosed with ovarian cancer or other diseases, and healthy female blood donors
***Predictors**:*	Near-infrared fluorescence response of carbon nanotube based sensors
Outcome to be predicted:	High-grade serous ovarian cancer

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	**PROBAST boxes to complete**	**Tick as appropriate**	**Definition for type of prediction model study**
Development only	Development	X	Prediction model development without external validation. These studies may include internal validation methods, such as bootstrapping and cross-validation techniques.
Development and validation	Development and validation	✓	Prediction model development combined with external validation in other participants in the same article.
Validation only	Validation	X	External validation of existing (previously developed) model in other participants.

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants			
A. Risk of Bias			
Describe the sources of data and criteria for participant selection: Fluorescence responses of nanosensors in patient serum were collected using a custom-built near-infrared plate reader. Patient demographics (age and sex), medication, hysterectomy, diagnosis, and biomarker levels were acquired by chart review.			
		Dev	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		Y	Y
1.2 Were all inclusions and exclusions of participants appropriate?		PN	PN
Risk of bias introduced by selection of participants	**RISK:** *(low/ high/ unclear)*	**high**	**high**
Rationale of bias rating: Because the medical record of healthy donors was not available, it is possible that selection bias was introduced to the healthy control. A large fraction of breast cancers in the non-HGSOC group of the test set may introduce systematic bias. Specimens obtained from symptomatic individuals at diagnosis were used for the development and assessment of the technology.			
B. Applicability			
Describe included participants, setting and dates: Serum samples were collected from 269 patients between August 2018 and March 2021 in the Memorial Sloan Kettering Cancer Center, USA.			
Concern that the included participants and setting do not match the review question	**CONCERN:** *(low/ high/ unclear)*	**low**	**low**
Rationale of applicability rating: Included patients appear representative of the population specified in the review question			

DOMAIN 2: Predictors

A. Risk of Bias

List and describe predictors included in the final model, e.g. definition and timing of assessment:

Predictors were defined as a difference in sensor response acquired from patient serum and phosphate buffered silane (PBS). Specifically, the E_{11} peak position feature, *dwl*, was defined as the wavelength difference between the E_{11} peak in the patient sample, wl, and PBS, wl_0, *dwl* = wl – wl_0. The E_{11} peak intensity feature, *dint*, was normalized as *dint* = int/int_0, where int and int_0 are the E_{11} peak intensity in serum and PBS, respectively. Similarly, we defined E_{11}^- peak related features, *dwl** and *dint**, indicating the relative E_{11}^- peak position and intensity.

The nanosensors were incubated in 20% serum at room temperature for 2 hours and in a cold room (4 ^oC) after the spectral acquisition at 2-hour time point. Data were taken at three time points during incubation: 2 hours, 24 hours, and 72 hours.

		Dev	Val
2.1 Were predictors defined and assessed in a similar way for all participants?		Y	Y
2.2 Were predictor assessments made without knowledge of outcome data?		Y	Y
2.3 Are all predictors available at the time the model is intended to be used?		Y	Y
Risk of bias introduced by predictors or their assessment	**RISK:** *(low/ high/ unclear)*	**low**	**low**

Rationale of bias rating:

They appear to match the review question.

B. Applicability

Concern that the definition, assessment or timing of predictors in the model do not match the review question	**CONCERN:** *(low/ high/ unclear)*	**low**	**low**

Rationale of applicability rating:

They appear to match the review question.

DOMAIN 3: Outcome

A. Risk of Bias

Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination:

Diagnoses were identified from a chart review of each patient; all diagnoses included histology and were confirmed by a gynecologic oncology attending physician.

There is no time interval between predictor assessment (spectral acquisition at 2, 24, and 72 hours of incubation) and outcome determination.

		Dev	Val
3.1 Was the outcome determined appropriately?		Y	Y
3.2 Was a pre-specified or standard outcome definition used?		Y	Y
3.3 Were predictors excluded from the outcome definition?		Y	Y
3.4 Was the outcome defined and determined in a similar way for all participants?		Y	Y
3.5 Was the outcome determined without knowledge of predictor information?		Y	Y
3.6 Was the time interval between predictor assessment and outcome determination appropriate?		Y	Y
Risk of bias introduced by the outcome or its determination	**RISK:** *(low/ high/ unclear)*	**low**	**low**

Rationale of bias rating:

They appear to match the review question.

B. Applicability

At what time point was the outcome determined:

Outcome was determined at the same moment as the measurement of the predictors which is in line with the diagnostic nature of the study

If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:

N/A

Concern that the outcome, its definition, timing or determination do not match the review question	**CONCERN:** *(low/ high/ unclear)*	**low**	**low**

Rationale of applicability rating:

The outcome is the same as that specified in the review question.

DOMAIN 4: Analysis

Risk of Bias

Describe numbers of participants, number of candidate predictors, outcome events and events per candidate predictor:

Participants: 269 waste samples were collected from female healthy blood donors and patients diagnosed with ovarian and other cancers under a Memorial Sloan Kettering Cancer Center Institutional Review Board approved protocol. From this sample set, 56 specimens were collected from patients diagnosed with high grade serous ovarian cancer, 71 specimens from healthy donors, 56 with other gynecologic diseases, 61 with non-gynecologic diseases, and 25 in remission.

Candidate predictors: 24 = 6 nanosensor types * 4 different spectral parameters

Outcome events: 2, high-grade serous ovarian cancer vs. others

Number of events per candidate predictor: unclear

Describe how the model was developed (for example in regards to modelling technique (e.g. survival or logistic modelling), predictor selection, and risk group definition):

We created models based on Decision Tree, Logistic Regression, Artificial Neural Networks, Random Forest, and Support Vector Machine for binary classification. Hyperparameters for each model were optimized using Bayesian Optimization, implemented in the HyperOpt library. The loss function to minimize in the hyperparameter optimization was set to (1 – F-score). F-score is a measure of accuracy in binary classification and calculated from the harmonic mean of PPV and sensitivity.

To identify the optimal predictor combinations, we optimized the machine learning algorithms at varying sensor combinations, durations of incubation, and spectral variables within the sensor array.

Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants):

Model performance was evaluated using ten-fold cross-validation. In the cross-validation process, stratified shuffle split validation was used to randomly partition the data set into ten subsamples. Each partition, nine of the ten subsamples were used to train the model while a single subsample was used to test the trained model. The average F-score of the ten-fold cross-validation was used to assess model performance. The trained models were then tested with an independent set of patient sera, sampled from different patients and from a later time point (test set) as an external validation.

Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit, and whether they were adjusted for optimism:

The averaged F-score and clinical sensitivity (sensitivity at 98% specificity) were used to determine the model performance.

Describe any participants who were excluded from the analysis:

N/A

Describe missing data on predictors and outcomes as well as methods used for missing data:

N/A

		Dev	Val
4.1 Were there a reasonable number of participants with the outcome?		PY	PY
4.2 Were continuous and categorical predictors handled appropriately?		NI	NI
4.3 Were all enrolled participants included in the analysis?		Y	Y
4.4 Were participants with missing data handled appropriately?		NI	NI
4.5 Was selection of predictors based on univariable analysis avoided?		Y	
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		PY	Y
4.7 Were relevant model performance measures evaluated appropriately?		Y	Y
4.8 Were model overfitting and optimism in model performance accounted for?		Y	
4.9 Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?		NI	
Risk of bias introduced by the analysis	**RISK:** *(low/ high/ unclear)*	**unclear**	**low**

Rationale of bias rating:

Background chronic diseases could affect the blood constituents, introducing interferences in sensor response. Chemical and biological interferents could interact with the nanosensors, contributing to the fluorescence response. From a patient chart review, we identified chronic diseases and most-common medications administered to the patients. We found that 75% of HGSOC and 68% of other disease patients suffered from at least one chronic condition and the relative abundance between these disease groups were similar. Regarding the medications, we statistically assessed the contribution of each interferent to the sensor results using a multivariate regression model. The regression model determined a linear correlation between the sensor response and medication, using estimated parameters and errors. The adjusted R2 of the regression model ranged from -0.045 to 0.233, indicating weak linear correlations of each sensor response to medications. We confirmed that the sensor array platform accurately classified the disease status regardless of medication and chronic conditions, evidenced by high F-scores of the HGSOC prediction models. The analysis suggested no indications that such interferences reduced specificity in HGSOC detection by the sensor platform.

Step 4: Overall assessment

Overall judgement about risk of bias and applicability of the prediction model evaluation		
Overall judgement of risk of bias	**RISK:** *(low/ high/ unclear)*	**high**
Summary of sources of potential bias: In terms of participants, the medical record of healthy donors was not available, potentially introducing selection bias to the non-ovarian cancer group. A large fraction of breast cancers in the non-HGSOC group of the test set may introduce systematic bias in participant selection and the development of machine learning models. Specimens obtained from symptomatic individuals at diagnosis were used for the development and assessment of the technology. Comorbidities such as background chronic diseases could affect the blood constituents. Chemical and biological interferents could interact with the nanosensors, introducing interferences in sensor response.		
Overall judgement of applicability	**CONCERN:** *(low/ high/ unclear)*	**low**
Summary of applicability concerns: No concerns		

1. Zheng, M. & Diner, B.A. Solution redox chemistry of carbon nanotubes. *J. Am. Chem. Soc.* **126**, 15490-15494 (2004).