

Supplementary Materials for

A reception-based panosensor platform to detect cancer biomarkers

Zvi Yaari, Yoona Yang, Elana Apfelbaum, Christian Cupo, Alex H. Settle, Quinlan Cullen, Winson Cai, Kara Long Roche, Douglas A. Levine, Martin Fleisher, Lakshmi Ramanathan, Ming Zheng, Anand Jagota, Daniel A. Heller*

*Corresponding author Email: hellerd@mskcc.org

Published 19 November 2021, *Sci. Adv.* 7, eabj0852 (2021)
DOI: [10.1126/sciadv.abj0852](https://doi.org/10.1126/sciadv.abj0852)

This PDF file includes:

Figs. S1 to S7
Tables S1 and S2

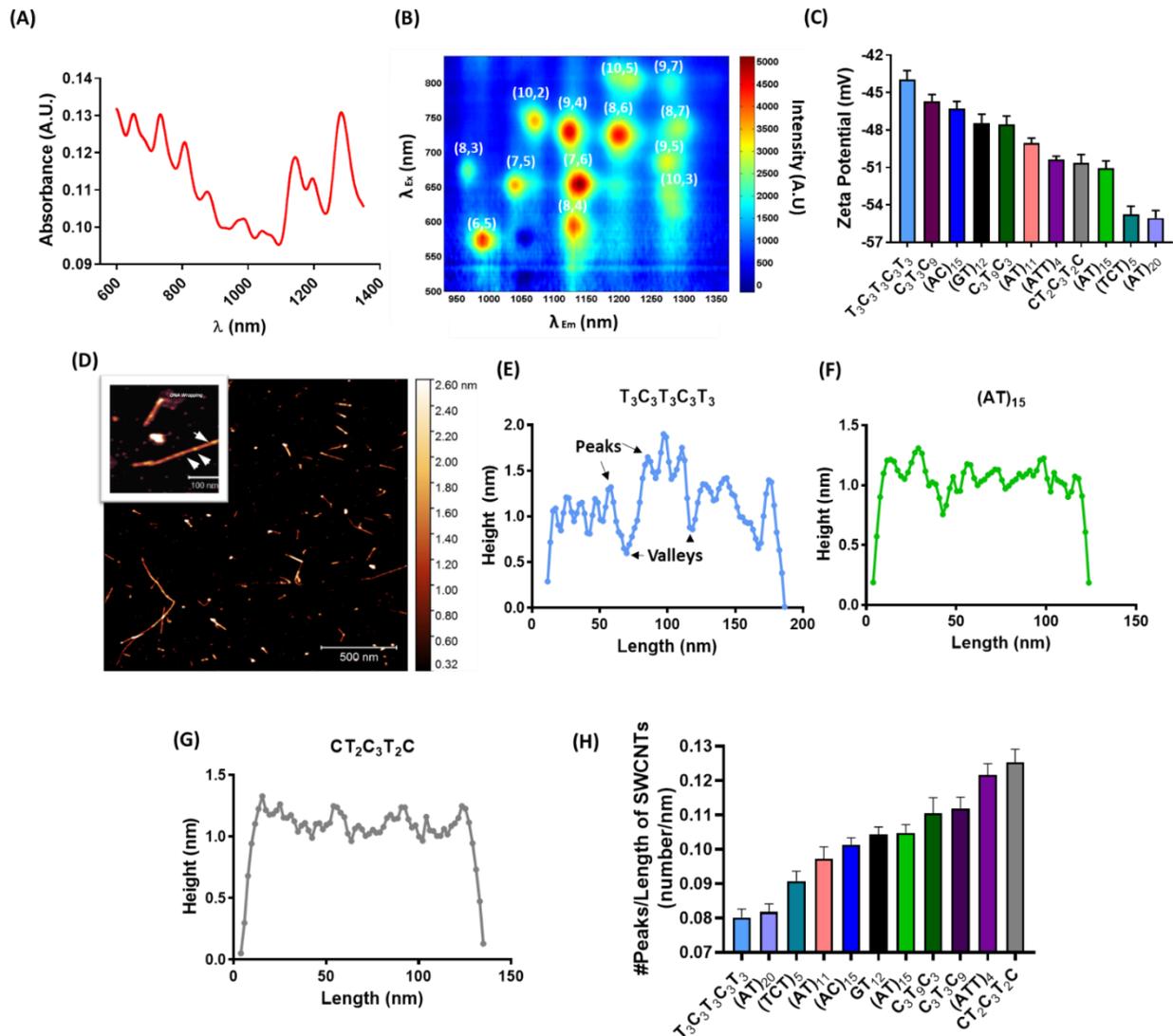


Fig. S1. Characterization of DNA-SWCNT complexes. (A) Representative absorbance spectrum of a DNA-SWCNT complex, (AC)₁₅-SWCNT. (B) 3D photoluminescence plot of the (AC)₁₅-SWCNT complex. (C) Zeta potentials of the DNA-SWCNT complexes; data were averaged over 7 acquisitions. (D) AFM image of (AC)₁₅-SWCNTs. White arrows denote height maxima ascribed to ssDNA. (E) Height profile along the length of an individual T₃C₃T₃C₃T₃-SWCNT complex. (F) Height profile along the length of an individual (AC)₁₅-SWCNT complex. (G) Height profile along the length of an individual CT₂C₃T₂C-SWCNT complex. (H) The average density of DNA peaks per unit length on the DNA-SWCNT complexes, quantified from AFM images; data were averaged over 20 acquisitions.

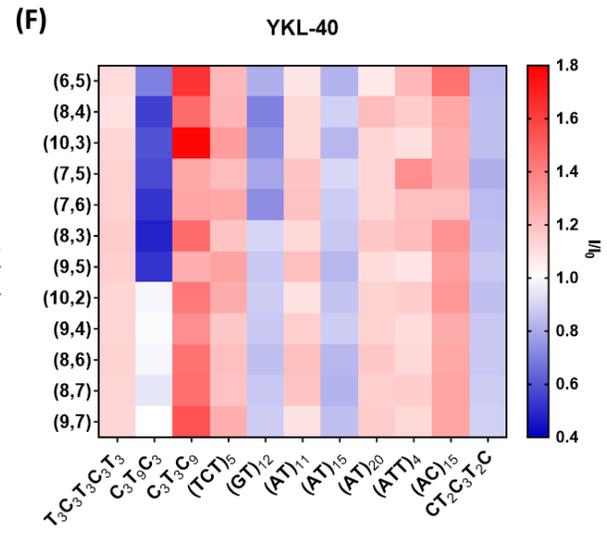
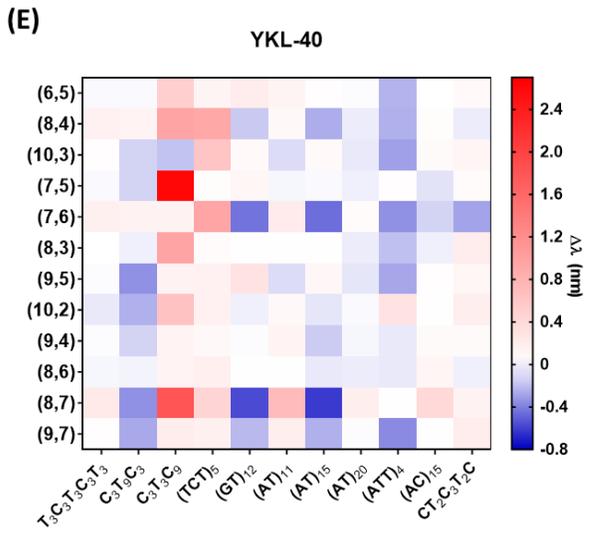
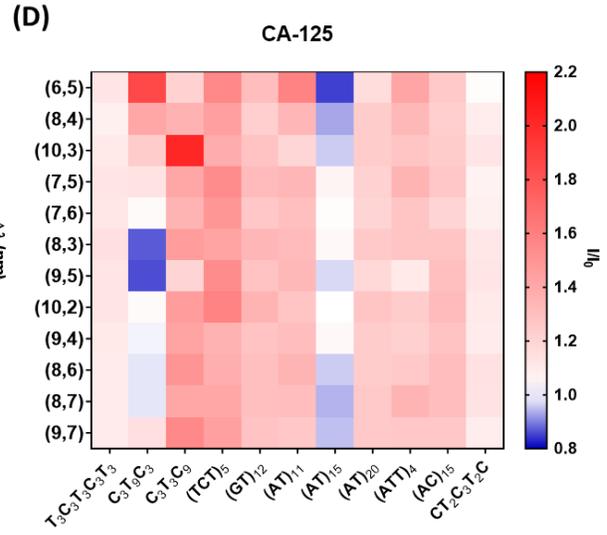
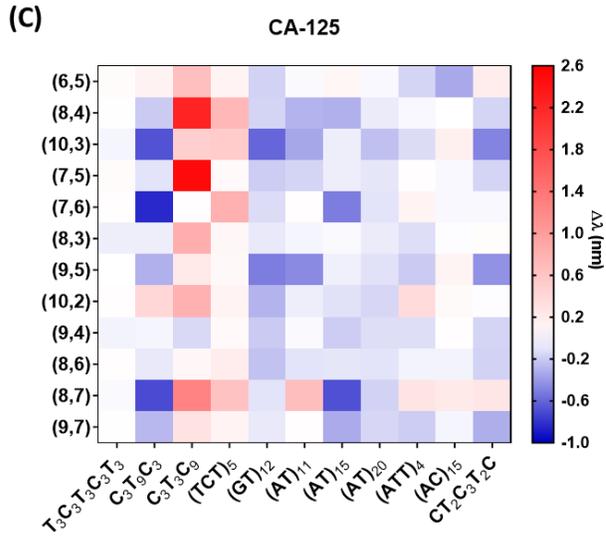
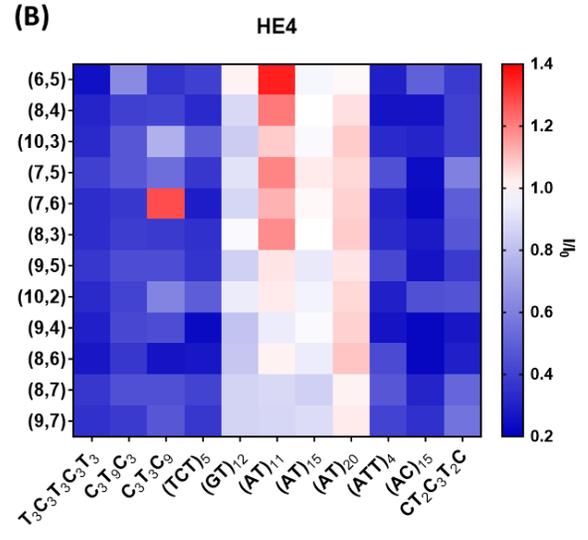
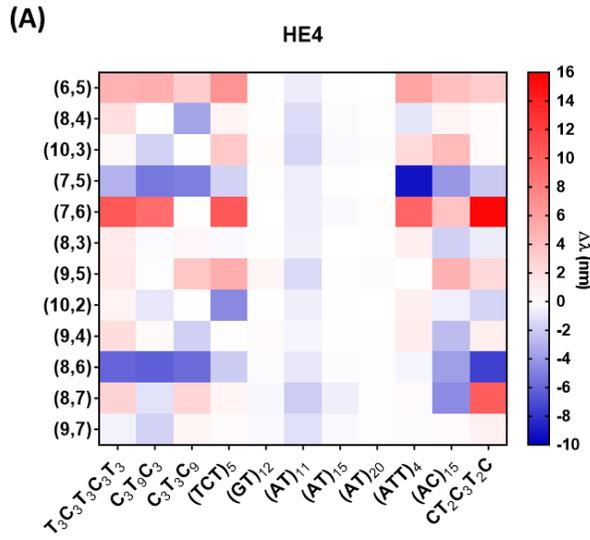


Fig. S2. Optical responses of DNA-SWCNT complexes to cancer biomarkers. (A) Heatmap of wavelength shifting responses of DNA-SWCNT complexes upon incubation with 100 nM HE4 in 10 % FBS in PBS. (B) Heatmap of intensity changes of DNA-SWCNT complexes upon incubation with 100 nM HE4 in 10% FBS in PBS. (C) Heatmap of wavelength shifts of DNA-SWCNTs complexes upon incubation with 100nM CA-125 in 10 % FBS in PBS. (D) Heatmap of intensity changes of DNA-SWCNT complexes upon incubation with 100 nM CA-125 in 10% FBS in PBS. (E) Heatmap of wavelength shifts of DNA-SWCNTs complexes upon incubation with 100 nM YKL-40 in 10% FBS in PBS. (F) Heatmap of intensity changes of DNA-SWCNTs complexes upon incubation with 100 nM YKL-40 in 10 % FBS in PBS.

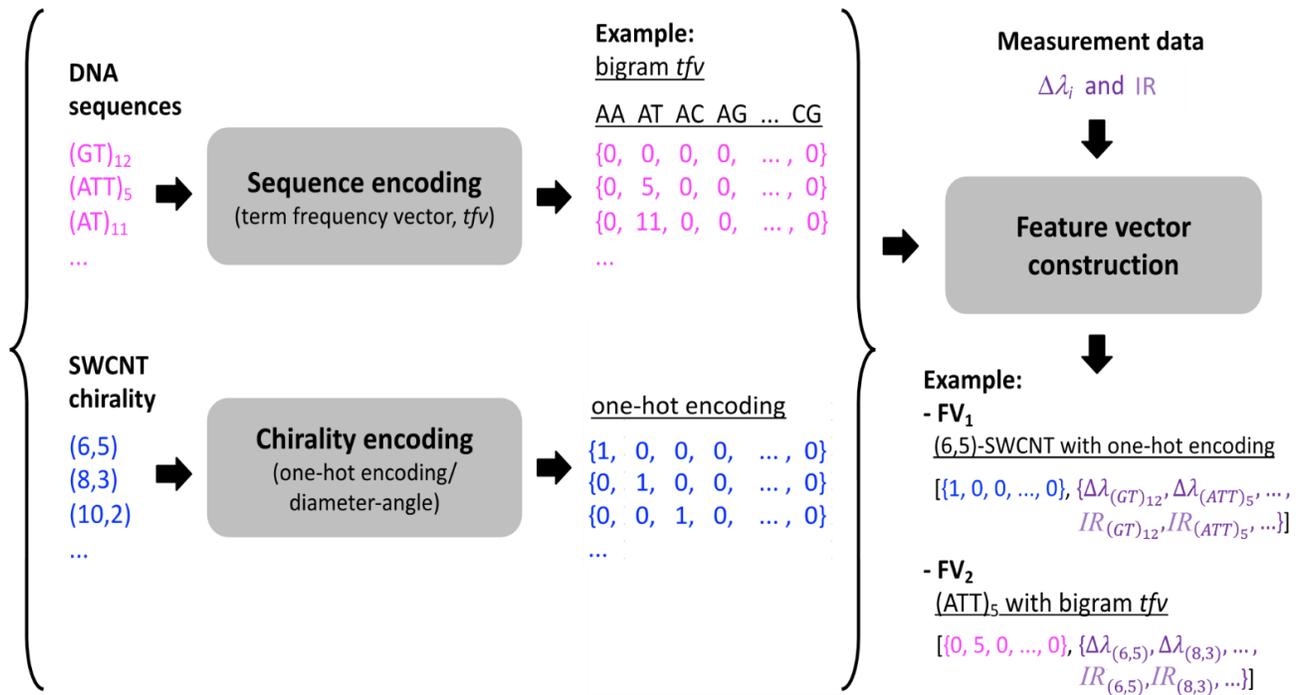


Fig. S3. Scheme for input feature construction. The DNA sequence was encoded to a numeric vector using the term-frequency method,⁴⁷ and SWCNT chirality was transformed into numeric vectors using a one-hot encoding technique. The numeric vectors could be then combined with a set of measurement data ($\Delta\lambda_i$, IR) using the FV₁ and FV₂ construction methods. This figure presents an example of a feature vector for a single example obtained by the FV₁ and FV₂ methods.

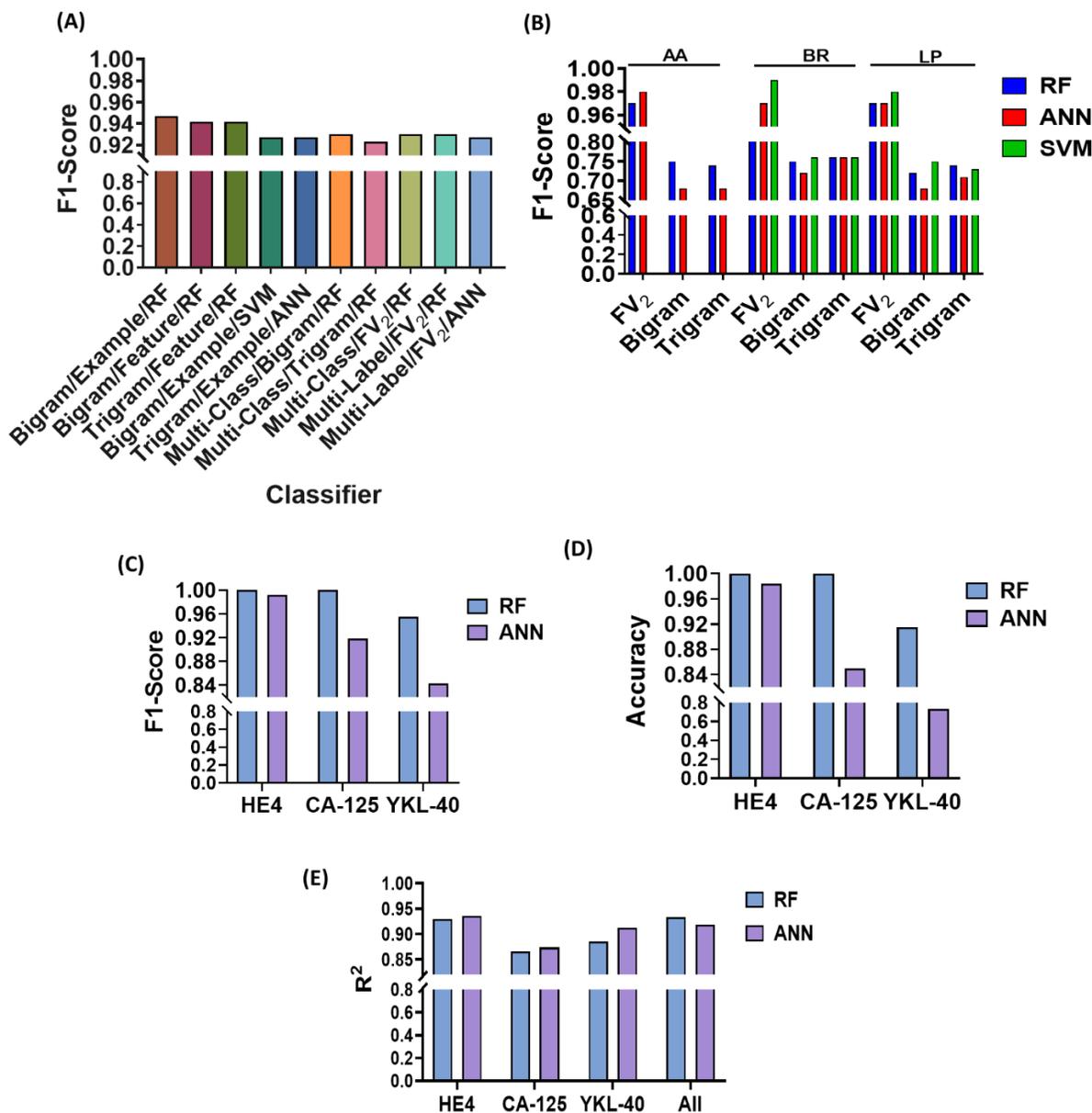


Fig. S4. Comparison of machine learning models. (A) F1-scores of different ML algorithms and FV sets. ‘Example’ denotes complete data of examples after removing incomplete sets of specific examples. ‘Feature’ denotes the complete data set of features after removing outliers results of specific features. ‘Multi-Label’ denotes the classification of single biomarkers in a mixture. ‘Multi-Class’ denotes the classification of single or mixed biomarkers with other analytes. (B) F1-scores of different ML algorithms and FV sets using different multi-label classifiers such as adaptive algorithm (AA), binary relevance (BR), and label powerset (LP). (C) F1-scores of each biomarker classification using random forest (RF) and artificial neural network (ANN) models. (D) Accuracy of each biomarker classification using RF and ANN models. (E) R² values from regression analysis denoting the accuracy of quantitative prediction of the concentration of each biomarker, as well as the combined accuracy.

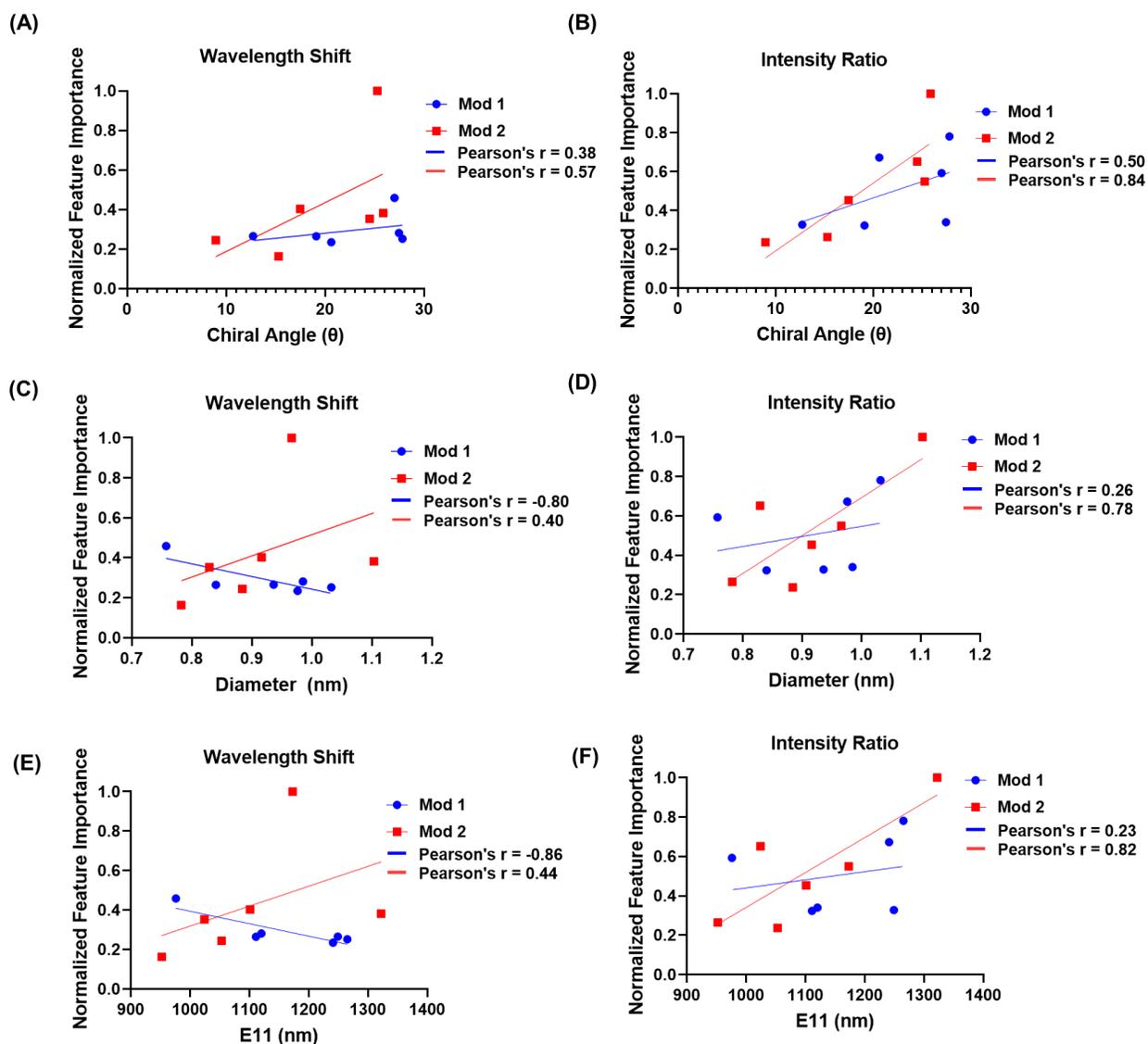


Fig. S5. Feature Importance Analysis. (A) Normalized feature importance values of wavelength shift and (B) intensity change of 12 SWCNT species/chiralities, plotted by nanotube chiral angle. (C) Normalized feature importance values of wavelength shift and (D) intensity change of 12 SWCNT species/chiralities, plotted by nanotube diameter. (E) Normalized feature importance values of wavelength shift and (F) intensity change of 12 SWCNT species/chiralities, plotted by nanotube emission wavelength.

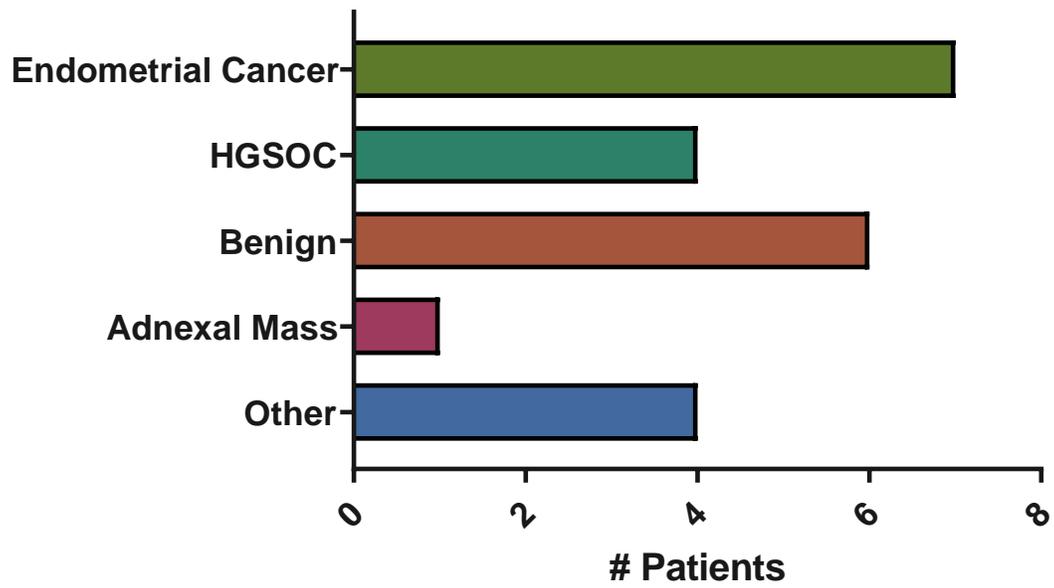


Fig. S6. Uterine lavage sample patient diagnoses. Diagnoses of patients providing uterine lavage samples in this study; HGSOc = high grade serous ovarian carcinoma.

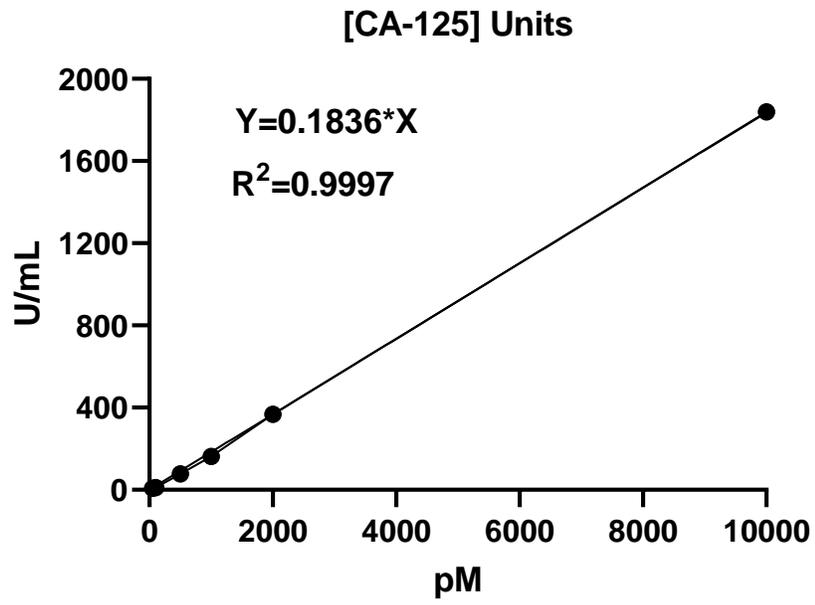


Fig. S7. Concentration unit conversion of CA-125. The linear curve of Units/mL vs. concentration in pM.

Table S1. Biomarkers - Physical Properties

Protein	HE4	CA-125	YKL-40
#Amino Acids	124	14,507	383
MW (Da)	12,993	1,519,175	42,625
pI	4.69	5.13	8.69
Grand Average Hydropathicity (GRAVY)	0.06	-0.321	-0.218
Aliphatic Index	70.81	66.01	78.49
# (+) Residues	8	818	40
# (-) Residues	11	1,263	35
Post Translation Modifications	N-linked glycosylation	Heavily O- linked glycosylation and some N-linked as well. Carbohydrate content estimated at ~24-28% with the total MW estimated ~3.5 million Da	N-linked glycosylation

Table S2. Data used to construct the plot in Figure 4B.

	Concentration (nM)		
	HE4	CA-125	YKL-40
A	100	0	0
B	0	100	0
C	0	0	100
D	30	70	30
E	70	70	30
F	30	70	70
G	70	70	70
H	30	30	30
I	70	30	30
J	30	30	70
K	70	30	70
L	30	0	0
M	70	0	0
N	0	30	0
O	0	70	0
P	0	0	30
Q	0	0	70