nature biomedical engineering

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41551-019-0351-1

In the format provided by the authors and unedited.

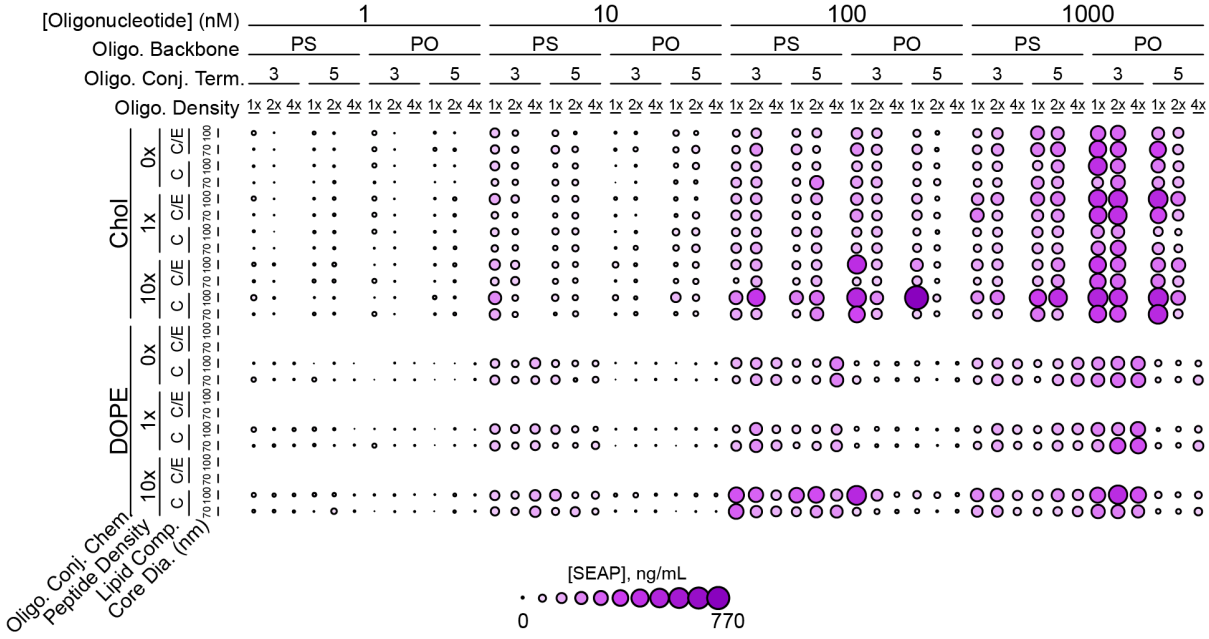
Exploration of the nanomedicine-design space with high-throughput screening and machine learning

Gokay Yamankurt[1,2,3,7], Eric J. Berns[4,7], Albert Xue[5], Andrew Lee[5*], Neda Bagheri[5*], Milan Mrksich[3,4,6*] and Chad A. Mirkin[2,3*]

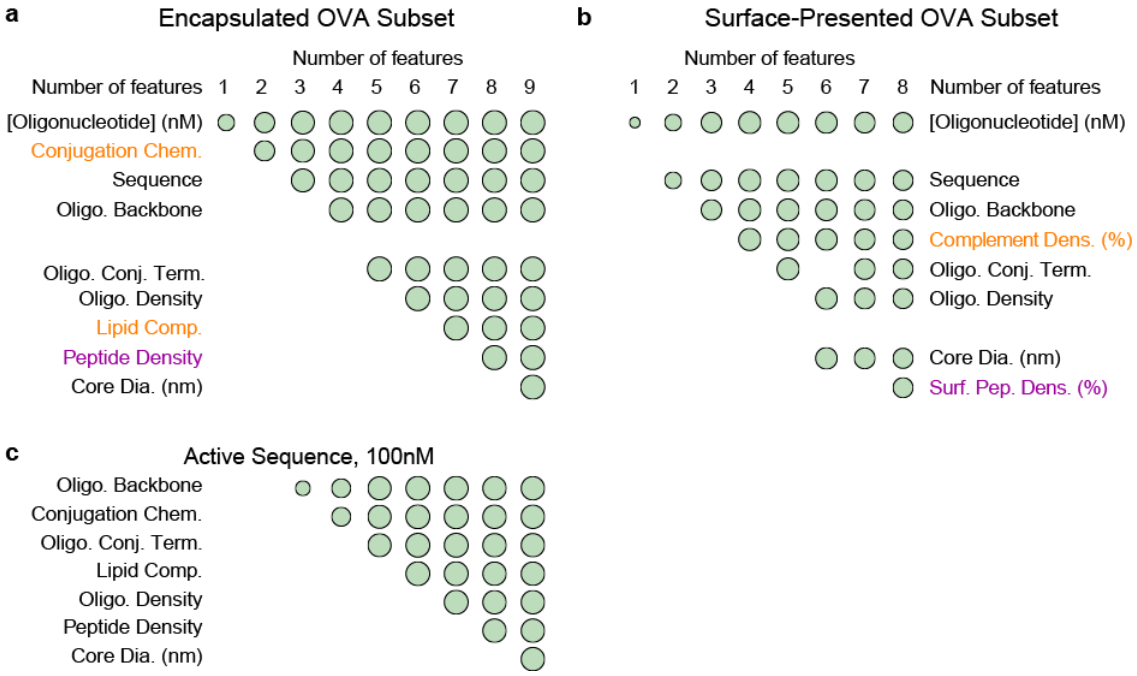
[1]Interdisciplinary Biological Sciences Graduate Program, Northwestern University, Evanston, IL, USA. [2]International Institute for Nanotechnology, Northwestern University, Evanston, IL, USA. [3]Department of Chemistry, Northwestern University, Evanston, IL, USA. [4]Department of Biomedical Engineering, Northwestern University, Evanston, IL, USA. [5]Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL, USA. [6]Department of Cell and Molecular Biology, Northwestern University, Chicago, IL, USA. [7]These authors contributed equally: Gokay Yamankurt, Eric J. Berns. *e-mail: andrew.lee3@northwestern.edu; n-bagheri@northwestern.edu; milan.mrksich@northwestern.edu; chadnano@northwestern.edu

Table of Contents

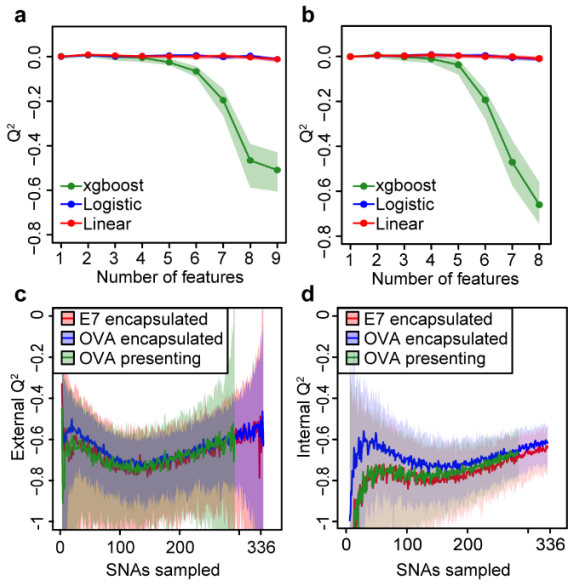
Supplementary Figures



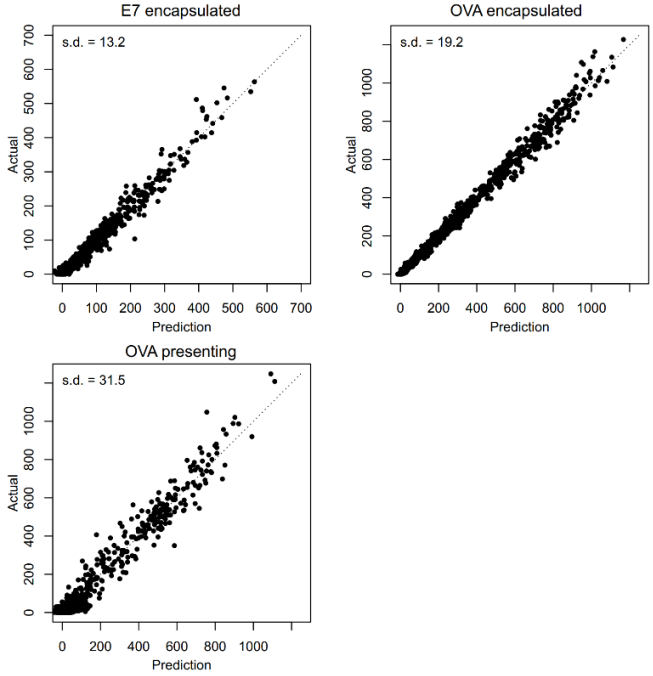
Supplementary Figure 1. Data for E7 subset. A dimension-stacking plot of the active-sequence SNAs in the encapsulated E7 subset, showing the SEAP concentration for each combination of design properties. Larger and darker circles indicate greater SEAP concentration.



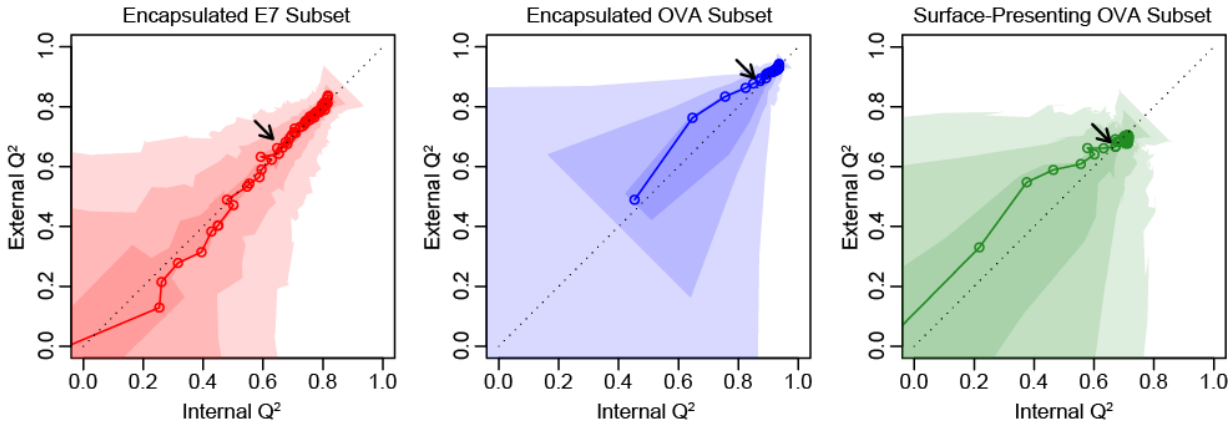
Supplementary Figure 2. **Comparison of parameter importance for different subsets.** Highest Q^2 scoring property combinations are shown across different number of properties for the **a,** encapsulated OVA subset, **b** surface-presenting OVA subset, and **c** encapsulated OVA subset with active sequence and 100 nM. Bubble areas correspond to Q^2 values from Fig. 6. Orange and purple properties denote exclusive and shared properties between the two subsets, respectively.



Supplementary Figure 3. **Models with randomized data.** Models built with randomized data show Q^2 values of zero or below, indicating that the models are specific. **a,** The Q^2 of the highest performing SNA property combinations across different numbers of properties for encapsulated OVA and **b**, surface-presented OVA subsets. **c,** Xgboost Q^2 performance when selecting and training on a random SNA subsample and testing predictions on the unselected SNAs or **d,** cross-validating within the selected subsample. All plots have 90% confidence intervals.



Supplementary Figure 4. **Error in machine learning models.** Standard deviation of the error between the predicted and the actual values of SEAP concentrations for all SNAs from the xgboost model. The standard deviation was 13.2, 19.2 and 31.5 ng/mL for the E7 encapsulated, OVA encapsulated and the surface-presented OVA subset. The error, which are very small compared to the activities of >1000 ng/mL indicate a very good fit.



Supplementary Figure 5. **Relationship between internal and external Q^2.** The non-observable external Q^2 (predicting immune activity of non-synthesized SNAs from a synthesized subsample) is plotted against the observable internal Q^2 (cross-validating within the synthesized subsample) for all three subsets. The median line and 90%, 50% and 20% confidence intervals are shown.

Supplementary Tables

Supplementary Table 1. Multi-factor ANOVAs of 3 SNA subsets.

	Encapsulated OVA Subset			Encapsulated E7 Subset			Surface-Presented OVA Subset		
Factor	d.f.	F	P	d.f.	F	P	d.f.	F	P
Concentration	3	1240	<1E-220	3	412	2.5E-220	3	183	4.7E-106
Sequence	1	381	2.0E-79	1	261	3.6E-56	1	246	1.2E-52
Conj. Chem.	1	338	4.3E-71	1	103	6.0E-24	N/A	N/A	N/A
Backbone	1	22.6	2.1E-06	1	3.64	0.056	1	241	8.5E-52
Conj. Term.	1	32.6	1.3E-08	1	3.34	0.068	1	2.73	0.099
Oligo. Dens.	2	5.59	0.0038	2	11.5	1.0E-05	2	2.23	0.11
Antigen Dens.	2	0.945	0.39	2	33.2	5.6E-15	2	0.673	0.51
Lipid Comp.	1	2.17	0.14	1	0.0839	0.77	N/A	N/A	N/A
Core Diameter	1	0.0248	0.87	1	20.4	6.6E-06	1	0.0218	0.88
Comp. Dens.	N/A	N/A	N/A	N/A	N/A	N/A	2	1.34	0.26

Supplementary Table 2. Oligonucleotide sequences used in this study. Sp18 refers to the spacer 18 modifier (Glen Research, Sterling, VA) and X is either cholesteryl-TEG or thiol modifier. Thiol modified is converted to DOPE as described in methods. SH refers to thiol modifier.

Name	**Sequence (5' -> 3')**	**Backbone**
ODN1826-3' Mod	TCC ATG ACG TTC CTG ACG TT-Sp18-Sp18-**X**	PO and PS
ODN1826-5' Mod	**X**-Sp18-Sp18-TCC ATG ACG TTC CTG ACG TT	PO and PS
GpC-ODN1826-3' Mod	TCC ATG AGC TTC CTG AGC TT-Sp18-Sp18-**X**	PO and PS
GpC-ODN1826-5' Mod	**X**-Sp18-Sp18-TCC ATG AGC TTC CTG AGC TT	PO and PS
Compliment ODN1826-3' Mod	AAC GTC AGG AAC GTC ATG GA-Sp18-SH	PO
Compliment ODN1826-5' Mod	SH-Sp18-AAC GTC AGG AAC GTC ATG GA	PO