nature research

Corresponding author(s): Chad A. Mirkin, Milan Mrksich, Neda Bagheri, Andrew Lee

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	AB Sciex ToF/Tof Series Explorer
Data analysis	MATLAB was used for ANOVA; other statistical tests were all implemented in the open source R environment. All machine learning was implemented in the R environment. The dimensional stacking code is publicly available at http:github.com/xuebert/dimensional_stacking, and is also implemented in the R environment.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study are available within the paper and its Supplementary Information. All data generated in this study are available from the corresponding authors upon reasonable request.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Each SNA condition was tested in two wells (biological replicates), and each of these samples was tested on two SAMDI spots (technical replicates). This method was used to ensure a very high likelihood of obtaining a measurement for each condition and to reduce variability from noise.
Data exclusions	One subset of data was discarded and the experiment repeated for that subset because the level of SEAP activity generally for all conditions was well above typical observations. This effect was traced to the contamination of lipid stock used for liposome synthesis with endotoxins.
Replication	While most SNAs were tested once (with two biological replicates), some of the SNAs in each subset, along with a number of control wells, were present in multiple subsets. Hence, they were synthesized and tested independently in separate experiments on different days, thus enabling a verification of trends across experiments.
Randomization	A homogenous suspension of cells was distributed to all of the wells tested in each subset.
Blinding	The investigators were not blinded; however, the data was obtained by mass spectrometry and analysed with algorithms by defined numerical rules. Therefore, blinding was not necessary.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	RAW Blue cells were purchased from Invivogen (San Diego, CA).
Authentication	RAW Blue cells were used without further authentication testing after purchasing.
Mycoplasma contamination	Cell lines tested negative for Mycoplasma with the MycoAlert Plus Kit (Lonza, Basel, Switzerland).

Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.