nature research

Corresponding author(s): John-William Sidhom

Last updated by author(s): Dec 1, 2020

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 our <u>Editorial Policies</u> and the <u>Editorial Policy Checklist</u>.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.							
n/a	Со	nfirmed					
	\boxtimes	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement					
	\boxtimes	A statement on 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.					
	\boxtimes	A description of all covariates tested					
	\boxtimes	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons					
	\boxtimes	A full description of the statistical parameters 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.					
\boxtimes		For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings					
\boxtimes		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					
		Our web collection on <u>statistics for biologists</u> contains articles on many of the points above.					

Software and code

Policy information about availability of computer code									
Data collection	No code was utilized to collect data. Data was collected from previously published datasets.								
Data analysis	All analysis was completed with the use of DeepTCR and code is open source and available on github.								

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 and reviewers. 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 <u>data availability statement</u>. 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

All data analyzed in this manuscript is available in previously published work as well as available on the DeepTCR github repository.

Field-specific reporting

Life sciences

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

Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see <u>nature.com/documents/nr-reporting-summary-flat.pdf</u>

Life sciences study design

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

Sample size	Sample size was for all data was determined by previously published work. No sample size was determined by the authors of this work.
Data exclusions	(N/A
Replication	For all analyses, Monte-Carlo simulations were utilized with pseudo-random seeding.
Randomization	(N/A
Blinding	(N/A

Reporting for specific materials, systems and methods

Methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	n/a	Involved in the study
\boxtimes	Antibodies	\boxtimes	ChIP-seq
\boxtimes	Eukaryotic cell lines	\boxtimes	Flow cytometry
\boxtimes	Palaeontology and archaeology	\boxtimes	MRI-based neuroimaging
\boxtimes	Animals and other organisms		
\boxtimes	Human research participants		
\boxtimes	Clinical data		
\boxtimes	Dual use research of concern		