

Corresponding author(s):	Antonina Mitrofanova
Last updated by author(s):	Jul 26, 2019

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.

Statistics					
For all statistical analys	ses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.				
n/a Confirmed					
☐ ☐ The exact sar	\square The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement				
A statement	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly				
The statistica Only common	l test(s) used AND whether they are one- or two-sided 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 descrip	tion of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) in (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)				
For null hypo Give P values a	thesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted s 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				
	Our web collection on <u>statistics for biologists</u> contains articles on many of the points above.				
Software and	code				
Policy information abo	ut <u>availability of computer code</u>				
Data collection	Data were obtained from public repositories, including dbGap and GEO databases				
Data analysis	Code and executables are available as a part of SWEAVE document and are deposited at http://license.rutgers.edu/technologies/2019-121_pathchemo				
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. 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 abo All manuscripts must - Accession codes, ur - A list of figures that	ut <u>availability of data</u> include a <u>data availability statement</u> . This statement should provide the following information, where applicable: nique identifiers, or web links for publicly available datasets have associated raw data v restrictions on data availability				
Data were obtained from	n public repositories, including dbGap and GEO databases				
Field-spec	ific reporting				
Please select the one l	pelow that is 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				

Life sciences study design

Antibodies

Eukaryotic cell lines

Animals and other organisms Human research participants

Palaeontology

Clinical data

All studies must di	sclose on these points even w	hen the disclosure is negative.
Sample size	,	was defined using two-sample two-tailed t-test. Such reconstruction requires 3 samples for each phenotype. ach phenotype. Validation on test sets achieved significant power, thus the sample sizes were sufficient.
Data exclusions	No data were excluded	
Replication	Not applicable	
Randomization	Not applicable	
Blinding	Not applicable	
Reportin	g for specific	materials, systems and methods
		es of materials, experimental systems and methods used in many studies. Here, indicate whether each materia u are not sure if a list item applies to your research, read the appropriate section before selecting a response.
Materials & ex	perimental systems	Methods
n/a Involved in t	he study	n/a Involved in the study

ChIP-seq

Flow cytometry

MRI-based neuroimaging