nature research

Corresponding author(s):	Xuerui Yang
Last updated by author(s):	Jul 3, 2021

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>.

_					
V:	t၁	ŤΙ	ist	٦.	\sim

For a	I statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.
n/a	Confirmed
	\overline{X} The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
	🛛 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
	🔀 A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
	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. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted Give <i>P</i> 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
\boxtimes	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.
Sof	tware and code

Policy information about <u>availability of computer code</u>

Data collection No software was used for collection of the public data.

Data analysis

cscMap is a bioinformatics pipeline to search for the RNA chimeras resulted from fusions of the transcripts encoded by the two opposite DNA strands. The cscMap pipeline has been provided as a supplementary file and deposited in github (https://github.com/xryanglab/cscMap). Other softwares used in the present study include tophat 2.0, bowtie 2.0, samtools, bedtools and python 2.7.

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 <u>availability of data</u>

 $All \ manuscripts \ must \ include \ a \ \underline{data \ availability \ statement}. \ This \ statement \ should \ provide \ the \ following \ information, \ where \ applicable:$

- 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 authors declare that the main data supporting the findings of this study are available within the article and its Supplementary Information files. Most of the human rRNA-depleted total RNA-seq datasets were obtained from the ENCODE project. The datasets from other species were obtained from the GEO database. The details of all the datasets used in the present study are provided in Supplementary Data 1.

— ·	1 1			· C·			4.0	
FIP.	IM	-sr	ነውር	cific	re	nai	rtir	١ø
1 10	ı	- J		,,,,	1 C	$\rho \circ i$	CII	כאי

Mycoplasma contamination

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. If you selection Deciding Decid	Tiera specific reporting						
Life sciences study design All studies must disclose on these points even when the disclosure is negative. Sample size Total IDM-cog attackers for due of 27 human camples and 20 mours earpies from the ENCODE project, which offers advantages of high sequences are all contents of the cognitive	Please select the or	ne below that	is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.				
Life sciences study design All studies must disclose on these points even when the disclosure is negative. Sample size Sample size Total RNA-reg distances include 271 human samples and 20 mouse earnples from the ENCODE project, which offers advantages of high securing dipolic large sample size, issues type variety, and strand-specific requestive with time paired and roads. The distances from the GEO distances, including people in (SEE 74815 and GES 748201, yeast (SEE 110133 and GES 97877). In (III (SEE 38878 and GES 748201, yeast (SEE 110133 and GES 97878), CES 12815 and GES 748201, yeast (SEE 110133 and GES 97878), and unit provides downloaded from the GEO distances, including people in (SEE 74815) and GES 748201, yeast (SEE 110133 and GES 148200), and unit provides distances in the CES 14816 people were also downloaded from the GEO distance, including people in (SEE 74815) and GES 748201, the sample were also distance includes bytem epithelial relifiers (MET 280, MET 280,	X Life sciences		Behavioural & social sciences				
All studies must disclose on these points even when the disclosure is negative. Sample size Total RRA-seq datasets include 271 horms samples and 20 moure samples from the ENCODE project, which offers advantages of high sequencing depths, layer samples were toly, and stand seconds sequencing with language of might sequencing depths, layer samples were sold and seconds sequencing with language of might sequencing depths, layer samples were sold seconds of the sequencing with language of might sequencing depths, layer sample search (SCR 11014). And SCR 1472 (SCR 1472 and SCR 1472 and	For a reference copy of t	the document wit	h all sections, see <u>nature.com/documents/nr-reporting-summary-flat.pdf</u>				
All studies must disclose on these points even when the disclosure is negative. Sample size Total RRA-seq datasets include 271 horms samples and 20 moure samples from the ENCODE project, which offers advantages of high sequencing depths, layer samples were toly, and stand seconds sequencing with language of might sequencing depths, layer samples were sold and seconds sequencing with language of might sequencing depths, layer samples were sold seconds of the sequencing with language of might sequencing depths, layer sample search (SCR 11014). And SCR 1472 (SCR 1472 and SCR 1472 and							
All studies must disclose on these points even when the disclosure is negative. Sample size Total RRA-seq datasets include 271 horms samples and 20 moure samples from the ENCODE project, which offers advantages of high sequencing depths, layer samples were toly, and stand seconds sequencing with language of might sequencing depths, layer samples were sold and seconds sequencing with language of might sequencing depths, layer samples were sold seconds of the sequencing with language of might sequencing depths, layer sample search (SCR 11014). And SCR 1472 (SCR 1472 and SCR 1472 and	lifo soion		udu daalaa				
Total RNA-seq datasets include 271 human samples and 20 mouse samples from the ENCODE project, which offers advantages of high sequencing depths, large sample and, tasset type variety, and strand-specific sequencing with long painted-end reads. The datasets from other sequencing depths, large sample as one times of the sequencing depths (according paths) in (2015), less (2015) and 05059170), C. ellegans of the sequencing with long painted-end reads. The datasets from other species were obtained to misse the control of the sequence of the control of the sequence of the control of the c	Life scier	ices st	uay design				
sequencing depths, large sample size, its sue type variety, and strand-specific sequencing with long parted-end reads. The datasets from the species were obtained from the GO database, including parted file (ISCF/3615 and GSF39273) and GSF39373). Intitly (SSE83877 and GSE93373) and GSF39373), furtily (SSE83877 and GSE933737), furtily (SSE83877 and GSE933737), furtily (SSE83877 and GSE93373), furtily (SSE83737 and GSE93373), furtily (SSE93737), furtily (SSE937	All studies must dis	sclose on thes	e points even when the disclosure is negative.				
Replication Randomization For all the experimental, the samples were randomly allocated into groups. Blinding Blinding was not relevant as no direct comparison between experimental groups was done. The study is mainly analysis of the public data. Reporting for specific materials, systems and methods We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each materials system or method listed is relevant to your study. If you are not sure if a list lem applies to your research, read the appropriate section before selecting a response. Materials & experimental systems Mathodies Mathodies Antibodies Antibodies Antibodies Sheep anti-DiG antibody (Roche, 11333089001), Donkey Anti-Sheep IgG H&L fluorophore-conjugated secondary antibody (Abcam, ab150180) Validation Sheep anti-DiG antibody was validated by the manufacturer and used in at least 25 peer-reviewed research articles. Eukaryotic cell lines Cell lines Cell line source(s) All the cells were authenticated with the methods provided by ATCC, including Karyotyping, DNA barcoding, and PCR assays	Sample size	sequencing de species were (GSE79375 an downloaded f (GSE60052 an epithelial cell	Jencing depths, large sample size, tissue type variety, and strand-specific sequencing with long paired-end reads. The datasets from other cles were obtained from the GEO database, including zebrafish (GSE73615 and GSE74929), yeast (GSE110413 and GSE99170), C. elegans E79375 and GSE79375), fruitfly (GSE83877 and GSE101603), and E. coli (GSE41190). Datasets from some human samples were also impleaded from the GEO database, including MCF7 cells (GSE94372 and GSE89888), PC3 cells (GSE65112 and GSE48230), and lung tissues E60052 and GSE52248). The whole genome DNA-seq data was obtained from the GEO database (GSE48216). The dataset includes 9 breast helial cell lines (MCF10A, MCF10F, MCF12A, HCC1143, HCC1143BL, HCC1187, T4, T47D, T47D_KBluc). The analyses were done on a per-				
Randomization For all the experiments, the samples were randomly allocated into groups. Blinding Blinding was not relevant as no direct comparison between experimental groups was done. The study is mainly analysis of the public data. Reporting for specific materials, systems and 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 Methods // a Involved in the study Antibodies ChIP-seq Eukaryotic cell lines Plow cytometry Palaecrotology and archaeology MRI-based neuroimaging Antibodies Plow cytometry Chinical data Dual use research participants Chinical data Dual use research of concern Antibodies Antibodies Antibodies Sheep anti-DIG antibody (Roche, 11333089001), Donkey Anti-Sheep IgG H&L fluorophore-conjugated secondary antibody (Abcam, ab150180) Validation Sheep anti-DIG antibody was validated by the manufacturer and used in at least 25 peer-reviewed research articles. Eukaryotic cell lines Cell line source(s) All the cell lines (KS62, MCF7, NCI-H460, AS49, HUH7, PC3) were from ATCC. Authentication All the cells were authenticated with the methods provided by ATCC, including Karyotyping, DNA barcoding, and PCR assays	Data exclusions	No data was e	excluded from the analyses.				
Reporting for specific materials, systems and methods We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each materia 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 Methods n/a Involved in the study Antibodies Eukaryotic cell lines Sheep anti-DIG antibody (Roche, 11333089001), Donkey Anti-Sheep IgG H&L fluorophore-conjugated secondary antibody (Abcam, ab150180) We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each materia systems and methods used in many studies. Here, indicate whether each materia systems and methods used in many studies. Here, indicate whether each materia systems and methods used in the study. Materials & experimental systems Methods n/a Involved in the study n/a Involved in the study MRI-based neuroimaging MRI-based neuroimaging MRI-based neuroimaging Antibodies Antibodies Antibodies Sheep anti-DIG antibody (Roche, 11333089001), Donkey Anti-Sheep IgG H&L fluorophore-conjugated secondary antibody (Abcam, ab150180) Validation Sheep anti-DIG antibody was validated by the manufacturer and used in at least 25 peer-reviewed research articles. Eukaryotic cell lines Cell line source(s) All the cell lines (K562, MCF7, NCI-H460, A549, HUH7, PC3) were from ATCC. Authentication All the cells were authenticated with the methods provided by ATCC, including Karyotyping, DNA barcoding, and PCR assays	Replication	All the experi	mental data were obtained with at least 3 biological replicates. All attempts at replication were successful.				
Reporting for specific materials, systems and methods We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each materia 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 Methods Mathodies Mathodies Mal involved in the study Antibodies Dalaeontology and archaeology MRI-based neuroimaging MRI-based neuroimaging MRI-based neuroimaging Antibodies Antibodies Antibodies Antibodies Antibodies Sheep anti-DIG antibody (Roche, 11333089001), Donkey Anti-Sheep IgG H&L fluorophore-conjugated secondary antibody (Abcam, ab150180) Validation Sheep anti-DIG antibody was validated by the manufacturer and used in at least 25 peer-reviewed research articles. Eukaryotic cell lines Cell line source(s) All the cells were authenticated with the methods provided by ATCC, including Karyotyping, DNA barcoding, and PCR assays	Randomization	For all the exp	periments, the samples were randomly allocated into groups.				
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 National Involved in the study Natibodies Natibodies National Experimental Systems Nethods National Involved in the study National Experimental Systems Nethods National Experimental Exp	Blinding	Blinding was r	not relevant as no direct comparison between experimental groups was done. The study is mainly analysis of the public data.				
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 National Involved in the study Natibodies Natibodies National Experimental Systems Nethods National Involved in the study National Experimental Systems Nethods National Experimental Exp	-						
waterials & experimental systems Materials & experimental systems Methods	Reportin	g for s	pecific materials, systems and methods				
n/a Involved in the study							
n/a Involved in the study	Materials & exi	nerimental	systems Methods				
Antibodies Antibodies used Sheep anti-DIG antibody (Roche, 11333089001), Donkey Anti-Sheep IgG H&L fluorophore-conjugated secondary antibody (Abcam, ab150180) Validation Sheep anti-DIG antibody was validated by the manufacturer and used in at least 25 peer-reviewed research articles. Eukaryotic cell lines Cell line source(s) All the cell lines (K562, MCF7, NCI-H460, A549, HUH7, PC3) were from ATCC. Authentication All the cells were authenticated with the methods provided by ATCC, including Karyotyping, DNA barcoding, and PCR assays							
Eukaryotic cell lines Eukaryotic cell lines Antibodies Antibodies used Sheep anti-DIG antibody (Roche, 11333089001), Donkey Anti-Sheep IgG H&L fluorophore-conjugated secondary antibody (Abcam, ab150180) Validation Sheep anti-DIG antibody was validated by the manufacturer and used in at least 25 peer-reviewed research articles. Eukaryotic cell lines Policy information about cell lines Cell line source(s) All the cells were authenticated with the methods provided by ATCC, including Karyotyping, DNA barcoding, and PCR assays		,					
Antibodies Antibodies Antibodies Antibodies Antibodies used Sheep anti-DIG antibody (Roche, 11333089001), Donkey Anti-Sheep IgG H&L fluorophore-conjugated secondary antibody (Abcam, ab150180) Validation Sheep anti-DIG antibody was validated by the manufacturer and used in at least 25 peer-reviewed research articles. Eukaryotic cell lines Policy information about cell lines Cell line source(s) All the cell lines (K562, MCF7, NCI-H460, A549, HUH7, PC3) were from ATCC. Authentication All the cells were authenticated with the methods provided by ATCC, including Karyotyping, DNA barcoding, and PCR assays	_ _						
Antibodies Antibodies Antibodies Antibodies Antibodies used Sheep anti-DIG antibody (Roche, 11333089001), Donkey Anti-Sheep IgG H&L fluorophore-conjugated secondary antibody (Abcam, ab150180) Validation Sheep anti-DIG antibody was validated by the manufacturer and used in at least 25 peer-reviewed research articles. Eukaryotic cell lines Policy information about cell lines Cell line source(s) All the cell lines (K562, MCF7, NCI-H460, A549, HUH7, PC3) were from ATCC. Authentication All the cells were authenticated with the methods provided by ATCC, including Karyotyping, DNA barcoding, and PCR assays			ology MRI-based neuroimaging				
Clinical data Dual use research of concern Antibodies Antibodies Antibodies used Sheep anti-DIG antibody (Roche, 11333089001), Donkey Anti-Sheep IgG H&L fluorophore-conjugated secondary antibody (Abcam, ab150180) Validation Sheep anti-DIG antibody was validated by the manufacturer and used in at least 25 peer-reviewed research articles. Eukaryotic cell lines Cell line source(s) All the cell lines (K562, MCF7, NCI-H460, A549, HUH7, PC3) were from ATCC. Authentication All the cells were authenticated with the methods provided by ATCC, including Karyotyping, DNA barcoding, and PCR assays		nd other organis					
Antibodies Antibodies Antibodies used Sheep anti-DIG antibody (Roche, 11333089001), Donkey Anti-Sheep IgG H&L fluorophore-conjugated secondary antibody (Abcam, ab150180) Validation Sheep anti-DIG antibody was validated by the manufacturer and used in at least 25 peer-reviewed research articles. Eukaryotic cell lines Policy information about cell lines Cell line source(s) All the cell lines (K562, MCF7, NCI-H460, A549, HUH7, PC3) were from ATCC. Authentication All the cells were authenticated with the methods provided by ATCC, including Karyotyping, DNA barcoding, and PCR assays	Human res	search participa	nts				
Antibodies Antibodies used Sheep anti-DIG antibody (Roche, 11333089001), Donkey Anti-Sheep IgG H&L fluorophore-conjugated secondary antibody (Abcam, ab150180) Validation Sheep anti-DIG antibody was validated by the manufacturer and used in at least 25 peer-reviewed research articles. Eukaryotic cell lines Policy information about cell lines Cell line source(s) All the cell lines (K562, MCF7, NCI-H460, A549, HUH7, PC3) were from ATCC. Authentication All the cells were authenticated with the methods provided by ATCC, including Karyotyping, DNA barcoding, and PCR assays	✓ Clinical data						
Antibodies used Sheep anti-DIG antibody (Roche, 11333089001), Donkey Anti-Sheep IgG H&L fluorophore-conjugated secondary antibody (Abcam, ab150180) Validation Sheep anti-DIG antibody was validated by the manufacturer and used in at least 25 peer-reviewed research articles. Eukaryotic cell lines Policy information about cell lines Cell line source(s) All the cell lines (K562, MCF7, NCI-H460, A549, HUH7, PC3) were from ATCC. Authentication All the cells were authenticated with the methods provided by ATCC, including Karyotyping, DNA barcoding, and PCR assays	Dual use research of concern						
Antibodies used Sheep anti-DIG antibody (Roche, 11333089001), Donkey Anti-Sheep IgG H&L fluorophore-conjugated secondary antibody (Abcam, ab150180) Validation Sheep anti-DIG antibody was validated by the manufacturer and used in at least 25 peer-reviewed research articles. Eukaryotic cell lines Policy information about cell lines Cell line source(s) All the cell lines (K562, MCF7, NCI-H460, A549, HUH7, PC3) were from ATCC. Authentication All the cells were authenticated with the methods provided by ATCC, including Karyotyping, DNA barcoding, and PCR assays							
Validation Sheep anti-DIG antibody was validated by the manufacturer and used in at least 25 peer-reviewed research articles. Eukaryotic cell lines Policy information about cell lines Cell line source(s) All the cell lines (K562, MCF7, NCI-H460, A549, HUH7, PC3) were from ATCC. Authentication All the cells were authenticated with the methods provided by ATCC, including Karyotyping, DNA barcoding, and PCR assays	Antibodies						
Eukaryotic cell lines Policy information about cell lines Cell line source(s) All the cell lines (K562, MCF7, NCI-H460, A549, HUH7, PC3) were from ATCC. Authentication All the cells were authenticated with the methods provided by ATCC, including Karyotyping, DNA barcoding, and PCR assays	Antibodies used						
Policy information about cell lines Cell line source(s) All the cell lines (K562, MCF7, NCI-H460, A549, HUH7, PC3) were from ATCC. Authentication All the cells were authenticated with the methods provided by ATCC, including Karyotyping, DNA barcoding, and PCR assays	Validation	Sheep anti-DIG antibody was validated by the manufacturer and used in at least 25 peer-reviewed research articles.					
Policy information about cell lines Cell line source(s) All the cell lines (K562, MCF7, NCI-H460, A549, HUH7, PC3) were from ATCC. Authentication All the cells were authenticated with the methods provided by ATCC, including Karyotyping, DNA barcoding, and PCR assays							
Cell line source(s) All the cell lines (K562, MCF7, NCI-H460, A549, HUH7, PC3) were from ATCC. Authentication All the cells were authenticated with the methods provided by ATCC, including Karyotyping, DNA barcoding, and PCR assays	Eukaryotic c	ell lines					
Authentication All the cells were authenticated with the methods provided by ATCC, including Karyotyping, DNA barcoding, and PCR assays	Policy information about <u>cell lines</u>						
, , , , , , , , , , , , , , , , , , , ,	Cell line source(s) All the cell lines (K562, MCF7, NCI-H460, A549, HUH7, PC3) were from ATCC.		All the cell lines (K562, MCF7, NCI-H460, A549, HUH7, PC3) were from ATCC.				
	Authentication	, , , , , , , , , , , , , , , , , , , ,					

All the cells were tested for mycoplasma contamination every 3 months, and the results have been negative.

Commonly misidentified lines (See ICLAC register)

No commonly misidentified cell lines were used in the study.

in | reporting summary