nature portfolio

Corresponding author(s):	Eric VIVIER
Last updated by author(s):	Nov 2, 2022

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our <u>Editorial Policies</u> and the <u>Editorial Policy Checklist</u>.

~				
5	tat	ŀις	ŤΙ.	\sim

For	all statistical analyses, confirm that the following items are present in the 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
	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.
X	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 <i>Give P values as exact values whenever suitable.</i>
\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
\times	Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), 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

Meso Quickplex SQ 120 system and MSD Workbench 3.0.18 software: MesoScale Discovery (used to measure cytokines concentrations) Cellometer Auto T4 plus cell counter: Nexcelom Bioscience (used to count cells)

TopCount NXT™ (Microplate Scintillation and Luminescence Counter): Perkin Elmer (used to detect radioactivity)

FACS CANTO II n°2: BD Bioscience(used to read fluorescence)

BD LSR Fortessa™ X-20 n°2: BD Bioscience (used to detect fluorescence)

Biacore T200 apparatus :Biacore GE Healthcare & Cytiva, Uppsala (Catalog No. 28975001)(used to acquire affinity data)

BD FACSDiva v8.0 software (for flow cytometry data acquisition)

MACSQuant® Analyzer from Miltenyi Biotec (used to read fluorescence)

Data analysis

VenturiOne® v6.1 software (Applied Cytometry Inc.) (for flow cytometry analysis)

FlowJo v10.5.2 software (for flow cytometry data analysis)

Kaluza analysis v1.3 (for flow cytometry data analysis)

GraphPad Prism v8.0.2 and v8.3.0 (for graphics and statistical analysis)

 $\label{thm:continuous} \textit{GraphPad Prism v8.3.0 (for graphics and statistical analysis of in vivo data)}$

Biacore T200 Evaluation software v3.0 and v3.1 (to analyze affinity data)

Phoenix v1.4 (including WinNonLin v 6.4) Pharsight (Certara Inc.) (for PK and TK analysis)

Excel 2019 (for MSD analysis)

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 Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All data supporting the results are available in the main text or the supplementary materials. The detailed molecular organization and the sequences of the NKCE used in the present study can be found in Supplementary Information Figure 1 and Figure 2, and in patent WO2016207273 and WO2022144836A1. Recombinant proteins were built from sequences found at https://www.ncbi.nlm.nih.gov.

Human research participants

Policy information about	studies involving numan	i research participants a	nd Sex and Gender in Research.
,		'	

Reporting on sex and gender	N/A
Population characteristics	N/A
Recruitment	N/A

Ethics oversight N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one be	low that is the best fit for your research. I	you are not sure, read the appropriate sections before making your selection.
X Life sciences	Behavioural & social sciences	Ecological, evolutionary & environmental sciences

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

Life sciences study design

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

Sample size	For in vivo studies (mouse model), the sample size was set to a minimum of 10 mice per groups to ensure a proper statistical analysis of data
	Sample size determination was not applicable to any other studies than in vivo studies.

Data exclusions Data were excluded when technical issues or aberrant values were identified. Excluded values are specified in the manuscript tables

Replication Experiments were performed on a minimum of 3 (and up to 10) individual healthy donor or patient samples to validate reproducibility of findings. For in vivo experiments, the dose-efficacy studies were repeated two to four times (pooled data are represented). All replications were successful with good consistency between healthy donors, patient samples and attempts.

Randomization N/A. None of the experimental methods used in this study necessitated any randomization of sample groups.

Blinding N/A. None of the experimental methods used in this study necessitated any blinding.

Behavioural & social sciences study design

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

Study description

Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).

Research sample

State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.

Sampling strategy	Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.
Data collection	Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.
Timing	Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.
Data exclusions	If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.
Non-participation	State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.

If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.

Ecological, evolutionary & environmental sciences study design

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

Randomization

Research sample

Data exclusions

Reproducibility

Randomization

Blinding

Location

Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, Study description hierarchical), nature and number of experimental units and replicates.

> Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.

Sampling strategy Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.

Data collection Describe the data collection procedure, including who recorded the data and how.

Timing and spatial scale Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken

> If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.

> Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.

> Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.

Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.

Did the study involve field work?

Field work, collection and transport

Field conditions Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).

State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).

Access & import/export

Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).

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

Antibodies

Antibodies used

Human NK cell activation panel CD107a/b-APC cocktail (Anti-CD56-PE-Vio770/Anti-CD107a-APC/Anti-CD107b-APC/Anti-CD3-VioBlue)/Miltenyi/130-095-212 (dilution: 1/10) Anti-CD56-PEVio770/Miltenyi/130-100-676 (dilution: 1/50)

Anti-NKp46-PE/Beckman Coulter /IM3711 (dilution: 1/50)
Anti-CD16-PE/BD Biosciences/556619 (dilution: 1/20)
Anti-CD16-APC-Cy7/BD Biosciences/557758 (dilution: 1/20)
Anti-CD32-PE/Beckman Coulter/IM1935 (dilution: 1/10)
Anti-CD64-PE/Beckman Coulter/IM3601U (dilution: 1/10)

Anti-CD64-PerCp-Cy5.5/BD Biosciences/561194 (dilution: 1/20)

Anti-CD123-PE/Biolegend/306006 (dilution: 1/20) Anti-CD123-APC/Biolegend/396706 (dilution: 1/20) Anti-CD69-FITC/Miltenyi/130-113-523 (dilution: 1/200) Anti-CD45-APC/Miltenyi/130-110-771 (dilution: 1/50) Anti-CD33-BB515/BD Biosciences/564588 (dilution: 1/40)

Anti-CD33-BB515/BD Biosciences/564588 (dilution: 1/40)
Anti-CD3-BV510/BD Biosciences/740187 (dilution: 1/40)
Anti-CD56-PECy7/BD Biosciences/557747 (dilution: 1/40)

Anti-CD107a-APC/Miltenyi/130-095-510 (dilution: 1/10) Anti-CD107b-APC/Miltenyi/130-103-960 (dilution: 1/10) Anti-CD107a-PE/Miltenyi/130-111-621 (dilution: 1/50) Anti-CD107b-PE/Miltenyi/130-118-818 (dilution: 1/50)

Anti-TNF-BUV395/BD Biosciences/563996 (dilution: 1/40) Anti-IFNγ-BV605/Biolegend/502536 (dilution: 1/40)

Anti-human MIP1β-PE/BD Biosciences/550078 (dilution: 1/40) Anti-CD3-Pacific Blue/BD Biosciences/624033 (dilution: 1/50) IgE-Vioblue/Miltenyi Biotec/130-117-931 (dilution: 1/50)

Anti-CD45-BV510 (cynomolgus)/BD Biosciences/563530 (dilution: 1/20) Anti-CD14-FITC (cynomolgus)/Miltenyi Biotec/130-110-518 (dilution: 1/50) Anti-CD123-PE (cynomolgus)/BD Biosciences/554529 (dilution: 1/20)

Anti-CD33-PE Vio770 (cynomolgus)/Miltenyi Biotec/130-113-350 (dilution : 1/50)

Anti-CD193-APC (cynomolgus)/Biolegend/310708 (dilution: 1/20) Anti-CD203c-APC (cynomolgus)/Invitrogen/17-2039-42 (dilution: 1/20) Anti-CD14-VioBlu/Miltenyi Biotech/130-110-524 (dilution: 1/35) IgE-APC/Miltenyi Biotech/130-117-930 (dilution: 1/35)

Anti-🛮 TCR-APC-Vio770/Miltenyi Biotech/130-113-536 (dilution: 1/20) REA-control-VioBlu/Miltenyi Biotech/130-104-609 (dilution: 1/12)

REA-control-violaly/Miltenyi Biotech/130-104-609 (dilution: 1/12)

REA-control-APC Vio770 / Miltenyi Biotech/130-104-618 (dilution: 1/12)

Polyclonal Anti-asialo-GM1/Biolegend/Poly21460

Validation

For flow cytometry studies, primary antibodies targeting NHP antigens were titrated using cynomolgus PBMCs under the same conditions as those used in the study. Primary antibodies targeting human antigens were titrated using human PBMCs, or human purified NK cells or human AML cell lines according to their use in the corresponding studies and under the same conditions as those used in the studies.

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s) KG-1a, Kasumi-6, GDM-1, MOLM-13 and THP-1 AML cell lines were purchased at ATCC. M-07e, EOL-1, Kasumi-1, F36-P, NB-4, OCI-AML2, MV4-11, OCI-AML3, and SKM-1 AML cell lines were purchased at DSMZ.

Authentication None of the cell lines used were authenticated

Commonly misidentified lines (See ICLAC register)

N/A. No commonly misidentified lines were used in the study.

Palaeontology and Archaeology

Specimen provenance

Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information). Permits should encompass collection and, where applicable, export.

Specimen deposition

Indicate where the specimens have been deposited to permit free access by other researchers.

Dating methods

If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.

Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight

Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other research organisms

Policy information about <u>studies involving animals</u>; <u>ARRIVE guidelines</u> recommended for reporting animal research, and <u>Sex and Gender in</u> Research

Laboratory animals

Mus musculus /NOD.Cg-Prkdcscid/J; commonly referred to as SCID

For in vivo efficacy study, mice were 13 to 15 weeks-old and weighed from 17.8 to 24.9 grams at the start of the study. For comparative study in the absence of presence of NK cell, mice were 10 to 12 weeks-old and weighed from 16.5 to 24.4 grams at the start of the study. They were housed on a 12 hours light/dark cycle. Environmental conditions including animal maintenance, room temperature ($22^{\circ}C \pm 2^{\circ}C$), relative humidity ($55\% \pm 15\%$) and lighting times were recorded by the supervisor of laboratory animal sciences and welfare and the records were archived.

Macaca fascicularis (Mauritius) / Cynomolgus non-human primates / males and females were used for the study.

Wild animals

N/A. No wild animals were used in this study

Reporting on sex

For repeated dosing in non human primate, sex were considered in the study design and an equal number of females (4 animals) and males (4 animals) were included in the study. Data are reported desaggregated for each animal. All the findings presented in the study apply for both sex.

Field-collected samples

N/A. No samples were collected in fields

Ethics oversight

All animal procedures were approved by the Sanofi Animal Care and Use Committee,

followed the French and European regulations on care and protection of the Laboratory Animals, and in accordance with the standards of the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration

Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.

Study protocol

Note where the full trial protocol can be accessed OR if not available, explain why.

Data collection	Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.
Outcomes	Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.
Outcomes	Describe flow you pre-defined printary and secondary outcome measures and flow you assessed these measures.

1 Circy	IIIIOIIIIGCIOII	about	uuu i	GOC I	Cocar cri	01	-
,							

used.

Data quality

10	4 4	·u	1.	u.	J

olicy illioi mation about <u>ut</u>	ar use research or concern
Hazards	
Could the accidental, deli in the manuscript, pose a	perate or reckless misuse of agents or technologies generated in the work, or the application of information presented threat to:
No Yes Public health National security Crops and/or livest Ecosystems Any other significa Experiments of concer Does the work involve an No Yes Demonstrate how Confer resistance t Enhance the virule Increase transmiss Alter the host rang Enable evasion of of Enable the weapor	nt area n y of these experiments of concern: to render a vaccine ineffective to therapeutically useful antibiotics or antiviral agents ince of a pathogen or render a nonpathogen virulent bility of a pathogen
ChIP-seq	
Data deposition	
Confirm that both raw	and final processed data have been deposited in a public database such as <u>GEO</u> .
Confirm that you have	deposited or provided access to graph files (e.g. BED files) for the called peaks.
Data access links May remain private before public	For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.
Files in database submiss	on Provide a list of all files available in the database submission.
Genome browser session (e.g. <u>UCSC</u>)	Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.
Methodology	
Replicates	Describe the experimental replicates, specifying number, type and replicate agreement.
Sequencing depth	Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.
Antibodies	Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.

Peak calling parameters | Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files

Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.

Flow Cytometry

Plots

Confirm that:

The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

All plots are contour plots with outliers or pseudocolor plots.

A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Healthy human buffy coats were provided by the Etablissement Français du Sang (EFS, the French blood service, Marseille; (AC-2019-3428)). Peripheral mononuclear cells (PBMC) were isolated from buffy coats by Ficoll density gradient centrifugation. Human NK cell were purified from PBMC with a bead-based negative selection kit from STEMCELL Technologies or Miltenyi Biotec.

For non-human primate studies, whole blood (100µl) and bone marrow (50µl) were lyzed before to be stained with the dedicated antibody cocktails. After 10mn at ambient temperature, the samples were washed, centrifuged, and fixed before acquisition.

Instrument

FACS CANTO II n°2: BD Bioscience(used to read fluorescence)
BD LSR Fortessa™ X-20 n°2: BD Bioscience (used to detect fluorescence)
MACSQuant® Analyzer from Miltenyi Biotec (used to read fluorescence)

Software

VenturiOne® v6.1 software (Applied Cytometry Inc.) (for flow cytometry analysis)

FlowJo v10.5.2 software (for flow cytometry data analysis) BD FACSDiva v8.0 software (for flow cytometry data acquisition) Kaluza analysis v1.3 (for flow cytometry data analysis)

Cell population abundance

For in vitro experiments, human NK cells were purified from PBMCs by negative selection with kits from Miltenyi Biotec, with a mean of purity of about 90%. Only samples with NK cell purity superior to 80% were kept for the study.

Gating strategy

NK cells: Time Gate / Single cells (SSC-A SSC-W) / Living cells (livedead negative) / Leucocytes (CD45+)/ CD3- / CD56+ Human Basophils: Time Gate / Single cells (SSC-A FSC-A) / Living cells (livedead negative) / CD14- / IgE+ (i.e. Fcepsilon@RI+) / TCRalphabeta —

Non human primate Basophils: Single cells (SSC-A FSC-A) / viable cells (Zombie NIR negative) / CD45+ / CD193+/CD203c+ / IgE+ (i.e. Fcepsilon®RI+)

AML blasts: Time Gate / Single cells (SSC-A SSC-W) / Living cells (livedead neg) / Leucocytes (CD45+)/ CD33+

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type

Indicate task or resting state; event-related or block design.

Design specifications

Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.

Behavioral performance measures

State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).

Acquisition

Imaging type(s)

Specify: functional, structural, diffusion, perfusion.

Field strength

Specify in Tesla

Sequence & imaging parameters

Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness. orientation and TE/TR/flip anale.

Area of acquisition

State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.

Diffusion MRI Used	Not used
Preprocessing	
Preprocessing software	Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).
Normalization	If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.
Normalization template	Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.
Noise and artifact removal	Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).
Volume censoring	Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.
Statistical modeling & infere	nce
Model type and settings	Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).
Effect(s) tested	Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.
Specify type of analysis: WI	nole brain ROI-based Both
Statistic type for inference (See Eklund et al. 2016)	Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.
Correction	Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).
Models & analysis	
n/a Involved in the study Functional and/or effective Graph analysis Multivariate modeling or p	
Functional and/or effective conn	Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).
Graph analysis	Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).
Multivariate modeling and predic	Specify independent variables, features extraction and dimension reduction, model, training and evaluation