nature portfolio

Corresponding author(s): Ke-Da Yu

Last updated by author(s): Jan 26, 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 Editorial Policies and the Editorial Policy Checklist.

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	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.*
☒	☐	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. *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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	no specific commercial, open or custom code for data collection
Data analysis	ASCAT (v2.4.3), GISTIC2.0 (v2.0.22), R (v4.0.3), SPSS Statistics 20.0, GraphPad 8.0.2, CaseViewer (v2.4),FlowJo(v10), ImageJ (v1.8.0) RNA-seq:the libraries were sequenced with the Illumina sequencing platform

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

Data availability. RNA-seq data generated in this study are deposited in the Sequence Read Archive database under the accession number PRJNA713612 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA713612/). FUSCC TNBC sequence data were available in the NCBI Gene Expression Omnibus (OncoScan array; GSE118527: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE118527) and Sequence Read Archive (whole exome sequencing and RNA-seq; SRP157974: https://trace.ncbi.nlm.nih.gov/Traces/sra/?study=SRP157974). The expression data, CNA data, and clinical data of the TCGA cohort were downloaded from the cBioPortal website (https://www.cbioportal.org/). The expression data were then transformed according to the log2(RSEM+1) method. The METABRIC expression data were downloaded from the cBioPortal website (https://www.cbioportal.org/). The expression data of the SMC cohort were available in the GEO database (GSE113184).

Kaplan-Meier survival plots were generated online with the Kaplan-Meier plotter database (https://kmplot.com/analysis/), and hazard ratios with 95% confidence intervals and log-rank P values were calculated. The transcription factor binding site prediction were performed online with the JASPAR database (https://jaspar.genereg.net/). A public STAT3 ChIP-seq data were available in the GEO database (GSE152203: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE152203). Source data are provided with this paper.

Field-specific reporting

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.

☒ 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	1. The FUSCC TNBC used for integrated analysis included 302 women with TNBC. This cohort was from Fudan University Shanghai Cancer Center (FUSCC) TNBC cohort (Sequence Read Archive (SRA) dataset: SRP157974; Gene Expression Omnibus (GEO) dataset: GSE118527) .Among these 465 patients, 302 patients who had both RNA-seq data and copy number data were included in our study for screening candidate genes. 2. Other sample sizes determined for experiments are included in the relevant figure legends.
Data exclusions	no data were excluded
Replication	The reproducibility for each analysis was confirmed by at least three independent experiments.
Randomization	For animal experiments, Female 6-week-old NOD.CB17-Prkdc scid/JSlac mice underwent randomization before cell injection. For the treatment groups, when the tumor volumes reached 50–100 mm3, mice underwent the second randomization before beding assigned to vehicle or Stattic (10 mg/kg) treatment.
Blinding	Data acquisition in the studies, was conducted in a blinded manner. For animal experiments, tumor volumes were measured without knowing the cell type injected (by assigning numerical IDs). For the other experiments, samples were also assigned numerical IDs to process in a blind fashion.

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.*
Randomization	*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.

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*
Research sample	*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*
Data exclusions	*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.*
Reproducibility	*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.*
Randomization	*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.*
Blinding	*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? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*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).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

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

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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

Antibodies

Antibodies used

ENSA (Abcam, Cat# ab180513) used at 1:10000 for immunoblot. ENSA (Proteintech, Cat# 14518-1-AP) used at 1:100 for IHC. MVK (Affinity, Cat # DF3238) used at 1:1000 for immunoblot. FDPS (Proteintech, Cat# 16129-1-AP) used at 1:3000 for immunoblot, at 1:200 for IHC. ACAT2 (Abcam, Cat# ab131215) used at 1:3000 for immunoblot. SREBP2 (Abcam, Cat# ab30682) used at 1:1000 for immunoblot, at 1:100 for IHC. SREBP2 (Proteintech, Cat# 28212-1-AP) used at 1:200 for IHC. LSS (Abcam, Cat# ab140124) used at 1:3000 for immunoblot. FDFT1 (Abcam, Cat# ab195046) used at 1:3000 for immunoblot. HMGCR (ab174830, Cat# ab174830) used at 1:3000 for immunoblot. HMGCS1 (Proteintech, Cat# 17643-1-AP) used at 1:1000 for immunoblot. FLAG (Sigma-Aldrich, Cat# F1804) used at 1:1000 for immunoblot. p-STAT3 (Tyr705) (Cell Signaling, Cat# 9145) used at 1:2000 for immunoblot. p-STAT3 (Ser 727) (Cell Signaling, Cat# 9134) used at 1:1000 for immunoblot.STAT3 (Cell Signaling, Cat# 9139) used at 1:1000 for immunoblot, at 4μg for ChIP. PP2AC (Cell Signaling, Cat# 2038) used at 1:2000 for immunoblot. GAPDH (Proteintech, Cat# 60004-1-Ig) used at 1:10000 for immunoblot. alpha tubulin (Proteintech, Cat# 11224-1-AP) used at 1:3000 for immunoblot. mouse IgG2a isotype (Cell Signaling, Cat# 61656) used at 4μg for ChIP. HRP-linked anti-mouse antibodies (Proteintech, Cat# SA00001-1) used at 1:5000 for Immunoblot. HRP-linked anti-rabbit antibodies (Proteintech, Cat# SA00001-2) used at 1:5000 for Immunoblot.

Validation

All antibodies are validated by the manufacturer and information can be found at the following websites: https://www.abcam.cn/ensa-antibody-epr80082-ab180513.html?#description_protocols, https://www.ptgcn.com/products/ENSA-Antibody-14518-1-AP.htm, http://www.affbiotech.cn/goods-2387-DF3238-MVK_Antibody.html, https://www.ptgcn.com/products/FDPS-Antibody-16129-1-AP.htm, https://www.abcam.com/acat2acetyl-coa-acetyltransferase-antibody-epr8417-ab131215.html, https://www.abcam.com/srebp2-antibody-ab30682.html, https://www.ptgcn.com/products/SREBF2-Antibody-28212-1-AP.htm, https://www.abcam.com/lss-antibody-epr6703-ab140124.html, https://www.abcam.com/fdft1-antibody-epr16481-ab195046.html, https://www.abcam.com/hmgcr-antibody-epr1685n-ab174830.html, https://www.ptgcn.com/products/HMGCS1-Antibody-17643-1-AP.htm, https://www.sigmaaldrich.com/DE/de/product/sigma/f1804, https://www.cellsignal.com/products/primary-antibodies/phospho-stat3-tyr705-d3a7-xp-rabbit-mab/9145, https://www.cellsignal.com/products/primary-antibodies/phospho-stat3-ser727-antibody/9134, https://www.cellsignal.com/products/primary-antibodies/stat3-124h6-mouse-mab/9139, https://www.cellsignal.com/products/primary-antibodies/pp2a-c-subunit-antibody/2038, https://www.ptgcn.com/products/TUBA1B-Antibody-11224-1-AP.htm, https://www.cellsignal.com/products/primary-antibodies/mouse-e5y6q-mab-igg2a-isotype-control/61656, https://www.ptgcn.com/products/HRP-conjugated-Affinipure-Goat-Anti-Mouse-IgG-H-L-secondary-antibody.htm, https://www.ptgcn.com/products/HRP-conjugated-Affinipure-Goat-Anti-Rabbit-IgG-H-L-secondary-antibody.htm.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

The TNBC cell lines BT549, MDA-MB-231, MCF7, T47D, SKBR3 and BT474 and embryonic kidney cells (HEK293T) were obtained from the American Type Culture Collection (ATCC) .

Authentication

The cell were authenticated by ATCC service-STR profiling.

Mycoplasma contamination

The cell lines were regularly confirmed to be negative for mycoplasma contamination with a Mycoplasma Detecting Kit (Vazyme).

Commonly misidentified lines
(See ICLAC register)

no

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 organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals

Female 6-week-old NOD.CB17-Prkdc scid/JSlac mice were used for the in vivo mouse xenograft models. Female 5-week-old nu/nu mice were used for the establishment of mini-PDX models. Mice were exposed to 12h light, 12h darkness cycle at temperature of 21 ±3 ℃ and an average of 55% humidity.

Wild animals	the study did not involve any wild animals
Field-collected samples	the study did not involve any samples collected from the field
Ethics oversight	The animal protocols were approved by the Animal Welfare Committee of Shanghai Medical College at Fudan University.The mini-PDX study protocol was approved by the Institutional Ethics Committee of Shanghai LIDE Biotech.

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

Human research participants

Policy information about studies involving human research participants

Population characteristics	For IHC, 8 pairs of primary TNBC tissues and adjacent normal tissues and primary TNBC specimens from 138 female patients (average age: 54 years), who underwent surgery at FUSCC from 2010-2013 were obtained from the Department of Pathology, Fudan University Shanghai Cancer Center. Patient-derived organoids used in the current study were derived from post-surgery specimens of three female patients who underwent surgery at the Department of breast, Fudan University Shanghai Cancer Center. Mini-Patient-derived xenograft models used in the current study were derived from fresh surgical tumor specimens acquired from seven female breast cancer patients (average age: 55 years) at FUSCC.
Recruitment	We used available samples for this study. there is no self-selection bias or other biases are present.
Ethics oversight	Acquisition of all clinical samples was approved by the Ethics Committee of FUSCC (Protocol number: 050432-4-1911D)

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 ICMJEguidelines for publication of clinical research and a completedCONSORT 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.*

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
☐	☐	Public health
☐	☐	National security
☐	☐	Crops and/or livestock
☐	☐	Ecosystems
☐	☐	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☐	☐	Demonstrate how to render a vaccine ineffective
☐	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☐	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☐	☐	Increase transmissibility of a pathogen
☐	☐	Alter the host range of a pathogen
☐	☐	Enable evasion of diagnostic/detection modalities
☐	☐	Enable the weaponization of a biological agent or toxin
☐	☐	Any other potentially harmful combination of experiments and agents

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ 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 publication.*	*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 submission	*Provide a list of all files available in the database submission.*
Genome browser session (e.g. UCSC)	*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 used.*
Data quality	*Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.*
Software	*Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.*

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	For cell cycle analysis, a total of 1×106 cells were fixed with precooled 70% ethanol overnight and then processed using the Cell Cycle and Apoptosis Analysis Kit (Yeasen, #40301ES50) according to the manufacturer's instructions. For the cell apoptosis assay, 5×105 cells were collected and incubated with annexin V-fluorescin isothiocyanate (FITC) and propidium iodide (PI) staining solution from the Annexin V-FITC/PI Apoptosis Detection Kit (Yeasen, #40302ES50).

Instrument	Beckman Cytomics FC 500 BD FACSCanto II
Software	FlowJo v10 software.
Cell population abundance	For Flow Cytometry experiments, more than 80% cells were reserved after FFS/SSC gating.
Gating strategy	For apoptosis assay, cell debris are ruled out by gating with the FFS/SSC to exclude debris and cell aggregates; For cell cycle analysis, FSC-A/FSC-H followed by FL6-A PE-A/FL6-H PE-H were used to gate cells of interest.

☒ 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 angle.*
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 & inference

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: ☐ Whole 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 connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity

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 predictive analysis

Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.