nature portfolio

Corresponding author(s): Tiannan Guo, Yi Zhu, Han Wen, Peijie Zhou

Last updated by author(s): Jun 17, 2026

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	NA
Data analysis	Raw DIA data files were analyzed with DIA-NN (version 1.8.1). Using the SMILES representations of the drug molecules, we employed the R-package ChemmineR (Cao et al., Bioinformatics, 2014). Data analysis and visualization were performed in R version 4.2.3 using the following packages: stats 4.2.3, ggplot2 3.4.2, ggpubr 0.6.0, ggnewscale 0.4.9, ggalluvial 0.12.5, ggrepel 0.9.3, ggsci 4.2.0, ChemmineR 3.50.0, ChemmineOB 1.36.0, rcdk 3.8.1, Rtsne 0.16, umap 0.2.10.0, pheatmap 1.0.12, dplyr 1.1.4, patchwork 1.3.2.9000, reshape2 1.4.4, DoseFinding 1.1-1, UpSetR 1.4.0 and openxlsx 4.2.5.2. The project data analysis codes are deposited in GitHub (https://github.com/guomics-lab/PTV-1)

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 are available upon request for non-commercial purposes. The proteomics data are deposited in ProteomeXchange Consortium (https://www.iprox.org/). Project ID: IPX0007409000.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	All TNBC patients enrolled in this study were female, with age information detailed in Table S13.
Reporting on race, ethnicity, or other socially relevant groupings	Data on race, ethnicity, or other socially relevant characteristics were not collected in this study
Population characteristics	All specimens were diagnosed as triple-negative breast cancer (TNBC) based on immunohistochemical assessment of ER, PR, and HER2 status, and graded according to WHO classification guidelines by two experienced pathologists. Patients were followed until August 2025.
Recruitment	All specimens were diagnosed as triple-negative breast cancer (TNBC) based on immunohistochemical assessment of ER, PR, and HER2 status, and graded according to WHO classification guidelines by two experienced pathologists. Patients were followed until August 2025. For each patient, one 5 µm tumor slice with >70% tumor area was sectioned from FFPE blocks for proteomic analysis.
Ethics oversight	Formalin-fixed paraffin-embedded (FFPE) breast tumor specimens were collected from 501 patients who underwent surgery between January 2012 and December 2019 (TNBC-HMU).This study was approved by the Ethics Committee of Harbin Medical University Cancer Hospital (KY2025-33).

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 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	No statistical methods were used to predetermine sample sizes. The scale of each dataset was determined by experimental feasibility, clinical specimen availability, and the biological objectives of each component. Multi-time-point ProteinTalks Dataset (mtPTDS). To capture the temporal dynamics of proteomic response, protein expression was profiled at 11 time points (0, 2, 4, 6, 8, 10, 12, 24, 36, 48, and 60 h) across three TNBC cell lines (HCC1806, HCC1143, and MDA-MB-453-1) treated with 55 single agents and 57 drug combinations. In total, mtPTDS comprised 2528 proteomic samples, including 41 pooled samples, 84 technical replicates, and 49 biological replicates used for quality control and reproducibility assessment. After accounting for these QC and replicate samples, 2354 independent proteomic samples remained. PTDS. Total 18 breast cancer cell lines were treated with 63 FDA-approved drugs and 59 combinations at 6, 24, and 48 hours post-treatment in biological triplicate, generating 16,311 perturbation proteomic DA-MS data files and 23,482 cytotoxicity measurements. The number of cell lines and drugs was determined by the throughput capacity of the DIA-MS platform and the goal of representing diverse cancer contexts. PTNC. we performed cell viability assays to evaluate the drug sensitivity of four TNBC cell lines (BT549, HCC1806, HS578T, and MDA-MB-468) to 81 anticancer compounds. For each drug-cell line pair, cells were treated at four concentration points, and three independent biological replicate experiments were performed. Drug sensitivity labels were derived from the resulting dose-response measurements. To ensure label reliability, we excluded drug–cell line pairs with inconsistent results across replicates or ambiguous sensitivity classification. After filtering, PTNC contained 2076 drug-dose response measurements, corresponding to 173 high-confidence drug-cell line combinations, which were used as an independent perturbation proteomics dataset for downstream analyses. PTPC was generated as an additional independent perturbation proteomics dataset across five cancer cell lines (A375, HCT116, NCI-H1975, NCI-H2170, and Panc 10.05). Cells were treated with 708 repurposed compounds, and proteomic profiles were collected at 6 and 24 hrs post-treatment. The dataset comprised 2307 proteomic samples in total, including 90 pooled samples, 86 technical replicates, and 110 biological

replicates, which were used for quality control and assessment of technical and biological reproducibility. After accounting for these QC and replicate samples, PTPC contained 2021 independent proteomic samples for model evaluation.
TNBC-HUM. To enable clinical translation and validation model predictions in human tumors, 501 treatment-naïve FFPE breast tumor specimens from TNBC patients who subsequently received chemotherapy and subjected to proteomic profiling. The cohort size was determined by clinical specimen availability within the predefined project timeframe and is consistent with the scale of previously published proteomic studies in cancer cohorts.

Data exclusions

Overall, in PTDS, 689 proteomics raw files with identification counts below 1000 were discarded. In TNBC-HMU, after excluding 244 patients with missing prognostic records, 501 patients with complete prognostic data and high-quality proteomic profiles were retained for subsequent analysis.

Replication

We confirm that all attempts at replication for the experiments reported in this study were successful. For the large-scale perturbation proteomics and drug-response datasets, reproducibility was assessed using biological replicates, technical replicates, pooled quality-control samples and predefined quality-control criteria before downstream analysis. Samples or drug-cell line pairs that did not meet quality-control or reproducibility criteria were excluded from subsequent analyses, as described below.

For PTDS, 18 breast cancer cell lines were treated with 63 FDA-approved drugs and 59 drug combinations at 6, 24 and 48 hours post-treatment in biological triplicate, generating 16,311 perturbation proteomic DA-MS data files and 23,482 cytotoxicity measurements. PTDS includes 618 pool samples, 1238 technical replicates. Quality-control analyses confirmed that the data quality of the QC samples was suitable for downstream analysis, as shown in Extended Data Fig. 1.

For mtPTDS, the dataset comprised 2528 proteomic samples, including 41 pooled samples, 84 technical replicates and 49 biological replicates, which were used for quality control and reproducibility assessment. After excluding QC and replicate samples, 2354 independent proteomic samples remained for downstream analysis. Quality-control analyses confirmed that the QC samples met the criteria required for downstream analysis, as shown in Extended Data Fig. 2.

For PTNC, each drug-cell line pair was tested at four concentration points, with three independent biological replicate experiments. Drug sensitivity labels were derived from the resulting dose-response measurements. To ensure label reliability and reproducibility, drug-cell line pairs showing inconsistent replicate results or ambiguous sensitivity classification were excluded. After filtering, PTNC contained 2076 drug-dose response measurements corresponding to 173 high-confidence drug-cell line combinations, which were used as an independent perturbation proteomics dataset for downstream analyses. The quality-control results are provided in Table S7.

For PTPC, the dataset comprised 2307 proteomic samples in total, including 90 pooled samples, 86 technical replicates and 110 biological replicates, which were used to assess technical and biological reproducibility. After excluding QC and replicate samples, PTPC contained 2021 independent proteomic samples for model evaluation. Quality-control analyses confirmed that the QC samples were suitable for downstream analysis, as shown in Extended Data Fig. 5.

For TNBC-HUM, to enable clinical translation and validate model predictions in human tumors, 501 treatment-naïve FFPE breast tumor specimens from patients with TNBC who subsequently received chemotherapy were subjected to proteomic profiling. The dataset also included 41 technical replicates, 46 biological replicates, 35 pooled samples and 43 mouse liver samples for quality control and reproducibility assessment. Quality-control analyses confirmed that the QC samples met the criteria required for downstream analysis, as shown in Extended Data Fig. 9. The cohort size was determined by clinical specimen availability within the predefined project timeframe and is consistent with the scale of previously published proteomic studies in cancer cohorts.

For the remaining biological functional validation experiments, at least three biological replicates or independent experiments were performed unless otherwise stated. The corresponding statistical parameters, including the number of replicates and statistical tests used, are reported in the relevant figure legends. No findings reported in the manuscript failed replication or were found to be irreproducible.

Randomization

The dataset, including various cell line and drug perturbations, is randomly divided into training, validation, and test sets with a split ratio of 0.7:0.2:0.1, resulting in 11426, 3264, and 1632 samples repectively.

Blinding

In the test set, samples with actual labels are input into the model for prediction, and the model's accuracy is evaluated by comparing the predicted values with the actual labels.

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
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

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

Antibodies

Antibodies used

Antibodies used: anti-ZFYVE19 (supplier name: Proteintech, catalog number: 23163-1-AP, lot number: 00188157. This antibody is goat anti-rabbit IgG polyclonal, therefore no clone number is applicable.), anti-PALLD (supplier name: Proteintech, catalog number: 10853-1-AP, lot number: 00048282. This antibody is goat anti-rabbit IgG polyclonal, therefore no clone number is applicable.), and anti-RCOR3 (supplier name: Abcepta, catalog number: AP55371, lot number: ASBF03049875. This antibody is goat anti-rabbit IgG polyclonal, therefore no clone number is applicable.).

Validation

The ZFYVE19 polyclonal antibody has been tested and validated by IHC using paraffin-embedded human liver cancer tissue at a dilution of 1:200 (40× magnification) (https://www.ptgcn.com/products/pictures/pdf/23163-1-AP.pdf).
The PALLD polyclonal antibody has been tested and validated by IHC using paraffin-embedded human colon cancer tissue at a dilution of 1:200 (40× magnification) (under 40× lens) (https://www.ptgcn.com/products/pictures/pdf/10853-1-AP.pdf). Supporting IHC staining data are provided in previous publications (Gilam et al. Nat Commun, 2016; Collazo et al., Cancer Res, 2014).
The RCOR3 polyclonal antibody has been tested and validated by western blot using lysates from 3T3 cells (https://www.abcepta.com/products/pdf_download/AP93461-KBTB4-Rabbit-Polyclonal-Antibody).

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)

Nineteen cell lines (BT20, BT549, DU4475, HCC1143, HCC1395, HCC1806, HCC1937, HCC38, HCC70, Hs578T, MDA-MB-436, MDA-MB-453-1, MDA-MB-468, HCC1187, A375, HCT116, NCI-H1975, NCI-H2170, and Panc 10.05) were purchased from ATCC. MDA-MB-453-2, MDA-MB-231, T47D and MCF7 were purchased from Meisen. Detailed information is listed in Table S1A. Two MDA-MB-453 cell lines from different sources are distinguished as MDA-MB-453-1 (ATCC) and MDA-MB-453-2 (Meisen).

Authentication

All cell lines were authenticated by short tandem repeat (STR) profiling, most recently performed on June 12, 2026.

Mycoplasma contamination

Mycoplasma contamination testing was negative, assessed by PCR-based detection using Genetic Testing Biotechnology (GTB)'s Mycoplasma Detection Kit, which detects over 500 mycoplasma species with a detection limit of 0.1 ng template DNA. The method was performed as previously described

Commonly misidentified lines
(See ICLAC register)

Our lines are not included in commonly misidentified lines

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.