nature portfolio

Corresponding author(s): Kwang Pyo Kim

Last updated by author(s): Jul 18, 2024

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	To assess the tumor microenvironment in five NSCLC subtypes by comparing subtype-specific genes with a diverse set of cell type-specific genes, using the dataset 'local_extended.rds' obtained from https://luca.icbi.at.
Data analysis	No custom code was used or developed for the analyses presented in this study. Standard workflows and open-source R packages and software were used (see the Methods). The codes used for the analyses included in our manuscript were uploaded to GitHub repository with instructions for users: https://github.com/joonan-lab/PDIAMOND-NSCLC.

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

Genomic and transcriptomic raw data generated in this study are available in Korean Nucleotide Archive (KoNA, https://kobic.re.kr/kona) with the accession ID KAP210028. Mass spectrometry-based global, phosphoproteome, and acetylome data are available in the National Cancer Institute's Proteomic Data Commons with

the accession IDs PDC000500, PDC000501, and PDC000502, as well as in K-BDS (Korea BioData Station, https://kbds.re.kr) with the accession IDs KAP503860, KAP504787, and KAP504788.
Now, we are uploading data in EGA (for genomics and transcriptomics), JPOST and PRIDE (for proteomics).

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	This study does not involve human participants directly. Instead, we utilize samples from a biobank, which were obtained with appropriate IRB approval. Therefore, there are no issues related to "Human participants, their data, or biological material." Also our analysis was aimed at identifying and characterizing subtypes of NSCLC, we did not differentiate by sex in this study.
Reporting on race, ethnicity, or other socially relevant groupings	This study does not involve human participants directly. Instead, we utilize samples from a biobank, which were obtained with appropriate IRB approval. Therefore, there are no issues related to "Human participants, their data, or biological material." Also our analysis was aimed at identifying and characterizing subtypes of NSCLC, we did not differentiate by race, ethnicity, or other socially relevant groupings in this study. In our study, we collected samples from Korean patients with non-small cell lung cancer. The analyses were not performed on other race, ethnicity, or other socially relevant groupings.
Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

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	We finalized the number of available samples as described below. Tumors and normal tissues adjacent to the tumor (NAT) were harvested under the Clinical Proteomic Tumor Analysis Consortium (CPTAC) guidelines from Asan Medical Center, Korea, with the approval of the Institutional Review Board of Asan Medical Center (Approval number: 2019-1210). In brief, we retrieved the 408 NSCLC cases whose cold ischemic time of fresh frozen tumor and NAT tissue sample was less than 15 minutes from the NSCLC cases of the bioresource center of Asan Medical Center deposited from January 2010 to March 2019. To investigate the impact of proteogenomic findings on post-operative therapy and metastasis, we preferentially selected 137 NSCLC patients showing lymph node metastasis at the pathologic examination of resection specimens. We further included 113 cases with absent lymph node metastasis at the time of surgical resection. With the further exclusion of 21 cases showing inadequate nucleic acid quality metrics, 229 patients were finally enrolled in the study.
Data exclusions	See above "Sample size" section.
Replication	*Describe the measures taken to verify the reproducibility of the experimental findings. If all attempts at replication were successful, confirm this OR if there are any findings that were not replicated or cannot be reproduced, note this and describe why.*
Randomization	In our study, we grouped patients based on a clinical diagnosis and the multi omics data produced.
Blinding	In our study, blinding was not relevant as we were performing multi omics data-based patient classification.

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

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.