nature portfolio

Corresponding author(s): Laurence Josset

Last updated by author(s): Apr 21, 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	No software was used for data collection
Data analysis	The code is available on https://github.com/genepii/Multi_omics_analyses_Verdy. A versioned release of the code has also been archived on Zenodo55. The codes used for 16S rRNA-seq analysis and human transcriptomics are available on https://github.com/benjjneb/dada2 and on https://github.com/zhxiaokang/RASflow respectively. Statistical analyses were conducted using R software, version 4.0.5 (R Foundation for Statistical Computing)

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 dehosted fastq have been deposited on the SRA database under accession number PRJNA1373503. The raw and processed data, are available on https://

github.com/genepii/Multi_omics_analyses_Verdy and in the Supplementary Data. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

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	The sex of preterm neonates at birth was determined and used as a variable for analyses.
Reporting on race, ethnicity, or other socially relevant groupings	No socially relevant categorization variables were applied.
Population characteristics	The clinical samples selected for this study were tracheal aspirates from hospitalized preterm neonates (<30 weeks gestational age), collected as part of the standard clinical routine.
Recruitment	Very premature infants were enrolled over a period of one year (September 2019-September 2020) at the Hospices Civils de Lyon, admitted to the neonatal intensive care unit of Centre Hospitalier Nord. Given the established impact of gestational age on the progression of BPD (bronchopulmonary dysplasia), only infants born before 30 weeks of gestation were included in the study. A total of 52 infants were followed until hospital discharge and classified as having BPD according to the criteria proposed by Jobe and Bancalari. BPD was considered positive when oxygen was required at an equivalent of 36 weeks while severity was graded according to the level of oxygen requirement. Premature neonates were excluded from the study if they died or were transferred to another hospital before reaching 36 weeks of gestational age.
Ethics oversight	The samples used in this study were collected as part of a non-interventional research study which was approved by the local ethics committee of the Lyon University Hospital on 12/09/2019 (AGORA N°19-122). Consent was obtained from the parents and a letter of information was sent to each family.

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	The sample size was not determined methodologically. Given the pilot nature of the study and the limited availability of samples, we enrolled all premature neonates born at <30 weeks gestational age over one year in a neonatal intensive care unit.
Data exclusions	Premature neonates were excluded from the study if they died or were transferred to another hospital before reaching 36 weeks of gestational age.
Replication	To validate the generalizability of the model with four endotypes, we performed an internal validation using caret R package (v7.0) with cross-validation 10-fold and random forest classification.
Randomization	Randomization was not relevant for this study. Inclusions were performed over one year, and all neonates who met the inclusion criteria were retained.
Blinding	Blinding was not relevant for this study

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

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	The samples used in this study were collected as part of a non-interventional research study was approved by the local ethics committee ("Comité Ethique") of the Lyon University Hospital on 12/09/2019. (AGORA N° 19-122) Consent was obtained from the parents and a letter of information was sent to each family. All ethical regulations relevant to human research participants were followed.
Study protocol	The full protocol can be accesssed upon request
Data collection	Data were collected over a period of one year (September 2019-September 2020) at the Hospices Civils de Lyon, admitted to the neonatal intensive care unit of Centre Hospitalier Nord
Outcomes	Bronchopulmonary dysplasia was considered positive when oxygen was required at an equivalent of 36 weeks while severity was graded according to the level of oxygen requirement.

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.*