nature portfolio

Corresponding author(s): Antonio Salas

Last updated by author(s): Dec 10, 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	The sequencing was performed using a 150 paired-end configuration in a Novaseq6000 platform or downloaded from the NCBI Gene Expression Omnibus (GEO) database under accession numbers GSE261482 and GSE103119 and from the ArrayExpress repository under accession numbers E-MTAB-12793, E-MTAB-14564 and E-MTAB-14588. Original code was uploaded to GItHub (https://github.com/Sandraviz/Mycoplasma-pneumoniae_-Transcriptomic-Signatures) and is linked to the DOI: 10.5281/zenodo.14258748
Data analysis	Quality control of raw data was carried out using FastQC (https://www. bioinformatics. babraham. ac. uk/projects/fastqc), alignment and read counting were performed using STAR (https://github.com/alexdobin/STAR), alignment filtering was done with SAMtools (https://www.htslib.org/) and read counting was carried out using FeatureCounts (https://subread.sourceforge.net/featureCounts.html). All data analysis from the processed counts were carried out in R version 4.3.2 using open-source packages as detailed in the Methods section.

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

The data of the cohort used in the discovery phase is available in the NCBI Gene Expression Omnibus (GEO) under accession number GSE103119. The EUCLIDS cohort and the new generated data from the cohorts of EUCLIDS,PERFORM and DIAMONDS used in the validation phase are avaliable in GEO or ArrayExpress under accession numbers GSE261482, E-MTAB-12793, E-MTAB-14564 and E-MTAB-14588.

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	Sex (biological attribute) was reported by the clinicians during the recruitment phase. This variable was considered in the statistical analysis when necessary, as detailed in the Methods section.
Reporting on race, ethnicity, or other socially relevant groupings	n/a
Population characteristics	n/a
Recruitment	Harmonised procedures for patient recruitment, classification, clinical data and sample collection, processing and storage were followed across the participating centers. For every patient, a prospective and standardized case report form (CRF) was used to record patient demographic characteristics, presence of comorbidities, and clinical signs and symptoms at the time of the initial presentation at the healthcare center, and at each timepoint of research sampling (presentation, 48-72 hours, and convalescence).
Ethics oversight	Informed Consent Forms (ICF): We have obtained proper informed consent from all participants, or their guardians, prior to sample collection. Ethical Approvals: We have secured all required ethical approvals. Each participating country's Ethics Committee (EC) approval is available upon request.

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	Samples were collected as part of three European projects (EUCLIDS, PERFORM and DIAMONDS). All samples meeting the study criteria were selected; therefore, no sample size calculation was performed.
Data exclusions	From the EUCLIDS dataset with accession number GSE261482, we exclude one sample with Mycoplasma pneumoniae pneumonia.
Replication	We validated the gene expression signatures obtained in the discovery phase using an independent dataset of newly generated data (validation cohort), as detailed in the Methods section.
Randomization	Randomization was performed for the spliting of the discovery dataset into training and test. We used the caret (version 6.0.94) package and mantain the proportion of conditions between subsets (training/test).
Blinding	Blinding was not performed in this study since all samples sequenced are of knonw origin.

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	n/a
Study protocol	Study protocol DIAMONDS: registro ISRCTN12394803. https://doi.org/10.1186/ISRCTN12394803 Study protocol EUCLIDS: see DOI: 10.1016/S2352-4642(18)30113-5 Study protocol PERFORM: see DOI: 10.1093/cid/ciad615
Data collection	Data collection DIAMONDS: The DIAMONDS consortium includes 28 partners, with clinical recruitment planned in 11 European countries and 3 non-European countries (African and two Asian populations to provide a global perspective). Data collection involves obtaining written informed consent and gathering information from clinical notes on presenting symptoms, investigation results, final diagnoses, and hospital resource usage. Each patient will receive a unique DIAMONDS study number for identification, and all samples will be labeled with this number. A patient identification log, containing hospital number and patient name, will be maintained and only accessible to the local clinical research team. Data will be collected using paper or electronic clinical record forms and uploaded in anonymized form to the study database at the following timepoints: - At recruitment, early in the illness - During all research sampling - At convalescence, after the final diagnosis A universal Case Record Form (CRF) will be completed at recruitment and first sample collection, regardless of patient admission status. Additional data will be collected on specific CRFs based on the patient's clinical background. Data collection PERFORM: see DOI: 10.1093/cid/ciad615 Data collection EUCLIDS: see DOI: 10.1016/S2352-4642(18)30113-5
Outcomes	Outcomes DIAMONDS. To ensure clear and consistent assignment of clinical syndromes and phenotypes for biomarker research, comprehensive clinical data was collected for each patient. Each case underwent review by a local clinical research panel, including a pediatric infectious diseases consultant. Standardized training was provided to clinical staff across all centers to ensure uniform data collection and patient categorization. The research team also organized mandatory virtual phenotyping sessions and conducted regular quality control reviews. Patients were classified based on a predefined list of common pediatric conditions, potentially receiving multiple diagnoses. Using a specific algorithm, patients were categorized into groups such as bacterial infections, viral infections, inflammatory conditions, and other types of infections. Diagnosis certainty determined further subcategories. For instance, 'definite bacterial' required identifying a bacterial pathogen in a sterile site, while 'definite viral' needed virus identification, matching syndrome, absence of bacteria, and specific inflammatory markers. Other patients were classified into five 'uncertain' categories based on clinical and microbiological findings. Additional categories included non-bacterial non-viral infections, inflammatory conditions, trivial illnesses, and uncertain infections or inflammation. Outcomes PERFORM: see DOI: 10.1093/cid/ciad615 Outomes EUCLIDS: see DOI: 10.1016/S2352-4642(18)30113-5

Plants

Seed stocks	n/a
Novel plant genotypes	n/a
Authentication	n/a