nature portfolio

Corresponding author(s): David Enard,Holger Heyn

Last updated by author(s): Oct 2, 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

scRNAseq and snRNAseq annotated data was downloaded in the AnnData format from https://developmental.cellatlas.io/fetal-immune for the Developmental Atlas, and from https://www.tissueimmunecellatlas.org/ for the Adult atlas and from https://5locationslung.cellgeni.sanger.ac.uk/ for the Lung atlas. Macrophage activation differentially expressed genes were obtained from the study by Panousis et al. 2023, available as Supplementary material.

To interrogate the spatial distribution of immune cells we downloaded the Lung cell atlas Cell2location output from https://5locationslung.cellgeni.sanger.ac.uk/. We used the non-negative matrix factorisation (NMF) factor loadings were used As in Madissoon et al., and we normalised the mean cell-type loadings per cell type across all the ST samples that passed QC (N = 11) to get a weighted cell-type loading representing the relative abundance of a given cell type in the different tissues (eg. parenchyma, epithelia) across the ST slides. Then, we produced a heatmap using the ComplexHeatmap R package (version 2.10.0) clustering the cell-type weighted values in order to group cell types with a similar tissue distribution.

To run ABC-MK we obtained human divergence and polymorphic data using human hg38 assembly and Ensembl release 109 annotations. Divergence counts were obtained by estimating fixed substitution in the human branch based on human, chimpanzee, gorilla, and orangutan alignments. Transcript sequences were obtained by aligning human CDS transcripts from Ensembl v109 on panTro6, gorGor6, and ponAbe3 assemblies from UCSC using pblat. Best reciprocal orthologs were realigned using MACSE v2. The number of non-synonymous and synonymous fixations on the human branch was estimated using Hyphy v2.5.

Polymorphic data were obtained from 1000GP phase 3 high-coverage phased data across seven African ancestry populations. We annotated non-synonymous and synonymous polymorphism using VEP and estimated ancestral and derived allele frequencies using Ensembl v109 human ancestral allele information from EPO multi-alignments. In total, 17,538 orthologues were included in the analysis, counting sites in

multiple isoforms only once. To test for long-term adaptation while avoiding confounding signals due to gene age, we filter differentially expressed genes with the ortholog gene list described in Di et al. This list contains Ensembl v99 human genes with orthologs in a minimum of 251 out of 261 mammalian genome assemblies resulting in a total of resulting 13,495 orthologs genes.

We obtained the Phylostrata values for human genes from the study of Litman and Stein 2019, where the authors determine consensus ages for human protein-coding and noncoding genes, using publicly available databases. The PS data was downloaded from the supplementary material provided with their original paper.

Data analysis

Differantially expressed genes (DEG) were obtained for each dataset using Scanpy package for the Developmental, Adult and Lung atlases and DESeq2 for the activated Macrophages. We considered genes upregulated in one cell type versus the rest of the cells in the same compartment using the wilcoxon test and selected the genes with a Benjamini- Hochberg FDR corrected p-value < 1x10−4 and a minimum log2fc value of 1. We obtanied top 500 DEG of each cell type (or all genes in case there were fewer than 500) to run ABC-MK and the resampling software.

We measured the long-term rate of protein adaptation using ABC-MK, an Approximate Bayesian Computation extension of the MK19 test. Such a rate is usually estimated as the proportion of adaptive non-synonymous substitution and can be derived from classical MK-test by the quantity α. ABC-MK use a generic Approximate Bayesian Computation (ABC) to infer the rate of adaptation: first, samples the parameter values from prior distributions; second, simulates a model using these random parameter values; and third, calculates informative summary statistics and compares them to observed empirical data.

For each analysis we subset 1e5 values from prior distributions to simulate and estimate α(x) summary statistics, which allow flexibility for the underlying DFE and BGS strength. Posterior distributions were estimated using ABCreg[31] comparing ABC-MK summary sta24 tistics and the empirical datasets. We set tolerance in ABCreg to record 2,500 acceptances from the rejection method and post-adjusted parameters using local linear regression. We excluded the datasets accounting for less than 500 polymorphic sites following recommendation from Murga-Moreno et al. 2024, since ABC-MK is not intended to run MK-test analyses at the gene level or small pools of genes. In addition to remove HLA and HIST genes because extreme nucleotide polymorphism level can lead to artefact biassed MK-results as well as to avoid spurious signals due balancing selection.

To measure a significant increase in adaptation, we build a null distribution accounting for 1,000 control datasets matching the size of the analysed immune genes subset for each cell population, following the resampling approach described in Enard et al. 2020 and Di et al. 2021. We evaluated each cell population estimation in accordance with its corresponding null empirical distribution. Thereby, we ensured statistically significant outcomes surpassing the empirical distribution to arise primarily from positive selection related to the cell populations' immune function, rather than other confounding factors

We followed Di et al. 2022 to repeatedly perform ABC-MK analyses in size-matched DEG sets, but creating 1000 datasets of random genes, representing a mix of both the DEG and control datasets from the resampling. Such strategy serves as an unbiased measure of Type I error, which allowed to estimate the number of cell types below an error threshold by chance. We also provide False Discovery Rate (FDR) values from multiple testing corrections using the Benjamini-Hochberg method.

We performed GO term analyses using the Biological Process (BP) ontology for each set of DE genes used for the ABC-MK test, using the R packages cluster- Profiler (version 4.2.2), enrichplot (version 1.14.2), GO.db and org.Hs.eg.db (BioConductor version 3.14).

To test for PS enrichment, we used the hypergeometric test considering all the genes with PS values as the population (N = 27, 974) and the DE genes of each cell type (obtained as described above) as the sample (n ≤ 500). For each PS (1-19) this test uses the hypergeometric distribution to determine the statistical significance of having drawn a sample of k genes of a given PS (out of n genes of the sample), from a gene population of size N containing K number of PS genes. We considered significant p-values ≤ 1x10−3. The test was performed using the phyper() function of the R stats package (R version 4.1.0).

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

We provide the scripts necessary to replicate our analysis in the following Github repository: https://github.com/irepansalvador/Immune Adaptation Atlas 2023.git.

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	*Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design; whether sex and/or gender was determined based on self-reporting or assigned and methods used.* *Provide in the source data disaggregated sex and gender data, where this information has been collected, and if consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected.* *Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.*
Reporting on race, ethnicity, or other socially relevant groupings	*Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status).* *Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.)* *Please provide details about how you controlled for confounding variables in your analyses.*
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

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Quantification of the rate of adaptation at the cell type level since humans split from chimpanzees.
Research sample	Transcriptome reference maps of healthy human organs and tissues at different developmental and 661 african ancestry individuals.
Sampling strategy	The top 500 DEG were obtained to normalise differences in the number of differentially expressed genes between cell populations, while maximising the number of genes required to perform accurate MK-test estimations in human lineage. African ancestry individuals were chosen to minimize spourius signals on MK-test estimation due to demographic changes.
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	HLA and HIST genes from the 500 DEG since extreme nucleotide polymorphism level can lead to artefact biassed MK-results as well as to avoid spurious signals due balancing selection
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	To highlight evolutionary patterns associated with an increased adaptation rate on DEG on immune human cell lines, we compare each dataset with 1000 control datasets while matching the total number of analysed genes for each line. The procedure involves adding control genes by pairs to the control set as it grows, ensuring that it has the same range of confounding factor values as the gene set being tested. Control datasets represent null distribution, accounting for the same average values of multiple confounding factors that can also determine the rate of adaptive evolution compared to the rest of the genome. This approach avoids matching set genes individually with control genes, which would limit the potential gene pool. Instead, we progressively match the gene set

with pair controls and set a 5% matching limit over the gene set average confounding values. Additionally, we restrict each control gene from appearing more than three times in each control set as we increase the control pool.

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

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

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)

State the source of each cell line used and the sex of all primary cell lines and cells derived from human participants or vertebrate models.

Authentication

Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.

Mycoplasma contamination

Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.

Commonly misidentified lines
(See ICLAC register)

Name any commonly misidentified cell lines used in the study and provide a rationale for their use.

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.