nature portfolio

Corresponding author(s): Souad Rahmouni and Michel georges

Last updated by author(s): 14 Nov 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	All third party programs used are referred to in Materials and Methods, including version. All custom-made script are made freely available via https://github.com/helene-peree/CEDAR
Data analysis	All third party programs used are referred to in Materials and Methods, including version. All custom-made script are made freely available via https://github.com/helene-peree/CEDAR

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 data are available within the constraints imposed by GDPR (General Data Protection Regulation). Processed data (nominal, permutational, and conditional cis-eQTL summary statistics, as well as gene expression data) have been deposited in open access on Zenodo under accession numbers: https://zenodo.org/records/21457760, https://zenodo.org/records/21495473, https://zenodo.org/records/21535308, https://zenodo.org/records/21538172, https://zenodo.org/

records/21538768, https://zenodo.org/records/21538968, https://zenodo.org/records/21539705, https://zenodo.org/records/21539811, https://zenodo.org/records/21539922, https://zenodo.org/records/21540023, https://zenodo.org/records/21540192, https://zenodo.org/records/21540504, https://zenodo.org/records/21540671, https://zenodo.org/records/21540784. Seurat objects used for analysing 10x Genomics 3' GEM and FLEX single-cell RNA-seq data were deposited in open access on Zenodo under accession number: https://zenodo.org/records/21373443. Source data for figures are provided with this paper.

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	Information of participants, including biological sex are included. The CEDAR cohort is nearly perfectly balanced for gender (55% women, and 45% of men).
Reporting on race, ethnicity, or other socially relevant groupings	The CEDAR cohort is composed of white European individuals only. Indeed, the available IBD GWAS data, that are used is this work,, are primarily based on the analysis of individuals of European descent, and the co-localisation methods (f.i. theta-based method developed at the UAG) require the disease and eQTL information to be collected in populations from the same geographic region, otherwise they do not work.
Population characteristics	Data collected from the electronic medical records (EMR) included age at sampling, ancestry, sex, weight, BMI, and alcohol history, declared ethnicity, family history of disease, surgical and medication history, blood type when available and known allergies (STable 1).
Recruitment	Peripheral blood was collected from 251 healthy European subjects at the academic hospital of the University of Liège (CHU) between October 2018 and November 2022. Gut biopsies were obtained from 60 healthy adults between June 2019 and December 2021. Donors were recruited through the gastroenterology department of the university hospital of Liège as part of a screening campaign for colon cancer.
Ethics oversight	Written informed consent was obtained from all participants prior to donation in agreement with the recommendations of the declaration of Helsinki for experiments involving human subjects. The experimental protocol was approved by the Ethics committee of CHU Liège (reference number: 2017/214).

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	Blood: sample size of 200 is relatively standard (even if at the low side) for cis-eQTL analysis given that (i) cis-eQTL explaining > 10% of the variance are common, and (ii) that the proportion of alternative hypotheses is high (hence FDR-based approaches effective). Gut: sample size of 60 is low. Yet, previous studies have proven effective with even lower sample size (f.i. Nat Genet 50, 493-497, 2018), and we compensate by having biopsies at three locations per individual hence increasing sample size to 180. The sample size for eQTL analyses was largely dictated by the limited resources available to conduct RNA-Seq. Thus, for blood we analysed 200 individuals while for gut we analysed 60 individuals. eQTL vary in the magnitude of their effects on gene expression. These numbers are sufficient to detect the largest eQTL effects (with adequate power), but not to detect the eQTL with smaller effects. We have to expect that as the result of the "winner's curse", the reported effects of some of the reported eQTL will be overestimated. This should not affect the conclusions of this work, including the colocalization results.
Data exclusions	We only excluded samples presenting evidence for technical issues including genotype cal rate, and (unresolved) mismatches between RNA and DNA genotypes. The applied procedures/criteria are fully described in M&M.
Replication	Replication, sensu stricto, was not applied to this study. We evaluate the statistical significance of our findings using p-values, corrected for multiple testing, and false discovery rate.
Randomization	The statistical models used included correction for defined covariates (such as sequencing batch, sex, age, BMI, etc…) as well as correction ofor hidden confounders by adding "expression principal components" in the analyses. As linkage disequilibrium patterns differ depending on the population (and would therefore impact colocalization analyses), we performed "genotype principal component analyses" and excluded individuals that would not fall within the group of indivdiuals with European ancestry.
Blinding	The only analysis conducted where "blinding" was applied is the evaluation (from the literature) whether genetic perturbation of a gene in a rodent model affected the susceptibility to colitis. This was done for a collection of candidate genes (from eQTL colocalization) and for an equally sized collection of control genes. The people who performed these searches received the full set of genes without knowing (i.e. blind to) whether the genes was candidate or control.

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
[ ]	[x] Antibodies
[x]	[ ] Eukaryotic cell lines
[x]	[ ] Palaeontology and archaeology
[x]	[ ] Animals and other organisms
[x]	[ ] Clinical data
[x]	[ ] Dual use research of concern
[x]	[ ] Plants

Methods

n/a	Involved in the study
[x]	[ ] ChIP-seq
[ ]	[x] Flow cytometry
[x]	[ ] MRI-based neuroimaging

Antibodies

Antibodies used

TCRγδ-PE/Dazzle 594, clone B1, BioLegend 331226
CD45RA-BV650, clone HI100, BD 3963
CD3-APC/Fire 750, clone SK7, BioLegend 344840
CD4-BV711, clone SK3, BD 563028
CD8-BV786, clone RPA-T8, BD 563823
CD127-BV421, clone HIL-7R-M21, BD 562436
CD183(CXCR3)-PE, clone G025H7, BioLegend 353706
CD196(CCR6)-APC, clone G034E3, BioLegend 353416
TCRVα7.2-BV605, clone 3C1, BioLegend 351720
CD161-FITC, clone HP-3G10, BioLegend 339906
CD194(CCR4)-PE-Cy7, clone L291H4, BioLegend 359410
CD25-BB700, clone M-A251, BD 566447
CD14-AF700, clone M5E2, BD 557923
CD16-AF700, clone 3G8, BioLegend 302026
CD19-AF700, clone SJ25C1, BioLegend 363034
CD3-APC/Fire 750, clone SK7, BioLegend 344840
CD19-BB700, clone SJ25C1, BD 566396
CD27-BB515, clone M-T271, BD 564642
IgD-APC, clone IA6-2, BioLegend 348222
CD20-BV711, clone 2H7, BD 563126
CD38-PE/Dazzle 594, clone HIT2, BioLegend 303538
CD14-BV421, clone MϕP9, BD 563743
CD15-AF700, clone HI98, BioLegend 301920
CD16-BV786, clone 3G8, BD 563690
CD56-BV650, clone NCAM16.2, BD 564057
CD11c-PE, clone 3,9, BioLegend 301606
HLA-DR-PE-Cy7, clone L243, BioLegend 307616
CD123-BV605, clone 7G3, BD 564197
TotalSeq-B anti-human Hashtag antibodies,

Validation

For cell sorting, the antibodies selected were based on the recommendation of the inertnational immunology consortium (Maecker HT, McCoy JP, Nussenblatt R. Standardizing immunophenotyping for the Human Immunology Project. Nat Rev Immunol 12:191-200 (2012). doi: 10.1038/nri3158). The protocol used for cell hashing was based on the paper from Stoeckius M et al. Genome Biol 19: 224 (2018). https://doi.org/10.1186/s13059-018-1603-1 . The protocole was validated on PBMCs in our laboratory before beeingapplied to gut biopsies.

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.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Peripheral blood mononuclear cells (PBMC) were isolated from fresh blood using SepMate™-50 (IVD) collection tubes (StemCell Technologies #85460) and Lymphoprep™ density gradient medium (StemCell Technologies #07861). After two washing steps, PBMC were stained with two panels of antibodies for 30 minutes at 4°C (STable 28). Cells suspended in PBS were then filtered using a 100 µm CellTrics filter (Sysmex #04004-2328). Cell sorting was performed on a FACS Aria III instrument (BD Biosciences) calibrated using CS&T beads (BD Biosciences). Fluorescence compensations were performed using CompBeads (BD Biosciences #552843).

Instrument

FACS Aria III

Software

BD FACSDiva™ Software

Cell population abundance

Purity of the sorted cell populations was evaluated using flow cytometry by assessing the percentage of the target cell population before and after sorting and ranged from 89% (for NKT cells) to 99,9%(for Neutrophils). Suplementary Figure 1 shows the % of each sorted cell population before and after cell sorting.

Gating strategy

After exclusion of debris (SSC-A and FSC-A) and doublets using (FSC-A and FSC-H), we targeted monocytes (classical, non-classical and intermediate), T lymphocytes (including mucosal-associated invariant T cells (MAIT) cells, naive and memory regulatory T cells, naive and memory CD8+ T cells, naive and memory CD4+ T cells, Th1, Th2, Th17 and Th1/17 helper T cells), B lymphocytes (naive, memory and plasmocytes), dendritic cells (plasmacytoid and myeloid), natural killer cells (NK and NKT) and innate lymphoid cells (ILC). Definition of sorted cells was as follow:

Name	Abbreviation	Definition
Naïve CD4 T cells	nT4	CD3+/CD4+/CD8-/CD45RA+
memory CD4 T cells	mT4	CD3+/CD4+/CD8-/CD45RA-
Naïve CD8 T cells	nT8	CD3+/CD4-/CD8+/CD45RA+
Memory CD8 T cells	mT8	CD3+/CD4-/CD8+/CD45RA-
Naïve regulatory T cells	nTreg	CD3+/CD4+/CD8-/CD25+/CD127-/CD45RA+
Memory regulatory T cells	mTreg	CD3+/CD4+/CD8-/CD25+/CD127-/CD45RA-
Th1	Th1	CD3+/CD4+/CD8-/CCR6-/CXCR3+
Th2	Th2	CD3+/CD4+/CD8-/CCR6-/CXCR3-/CCR4+
Th17	Th17	CD3+/CD4+/CD8-/CCR6+/CXCR3-/CCR4+
Th1-17	Th1-17	CD3+/CD4+/CD8-/CCR6+/CXCR3+
γδ T cells	gd T	CD3+/TCR γδ+
MAIT cells	MAIT	CD3+/TCR Vα7.2+
NKT cells	NKT	CD3+/CD56+
NK cells	NK	CD3-/CD56+
Naïve B cells	nB	CD3-/CD19+/CD27-/IgD+
Memory B cells	mB	CD3-/CD19+/CD27+/IgD-/CD20+
Plasmocytes	pB	CD3-/CD19+/CD27+/IgD-/CD20-/CD38+
Classical monocytes	cMO	CD3-/CD19-/CD14+/CD16-

Non-classical monocytes	ncMO	CD3-/CD19-/CD14-/CD16+
Intermediate monocytes	iMO	CD3-/CD19-/CD14+/CD16+
Myeloid dendritic cells	mDC	CD3-/CD19-/CD14-/CD16-/HLADR+/CD123-/CD11c+
Plasmacytoid dendritic cells	pDC	CD3-/CD19-/CD14-/CD16-/HLADR+/CD123+/CD11c-
Innate Lymphoid cells	ILC	CD3-/CD19-/CD14-/CD16-/CD127+/CD161+
Granulocytes	GRAN	CD15+
Neutrophils	NEUT	CD15+/Siglec8-
Eosinophils	EOS	CD15+/Siglec8+

Antibodies used for scRNA-seq

TotalSeq™-A0251 anti-human Hashtag 1 Antibody, BioLegend 394601
TotalSeq™-A0252 anti-human Hashtag 2 Antibody, BioLegend 394603
TotalSeq™-A0253 anti-human Hashtag 3 Antibody, BioLegend 394605
TotalSeq™-A0254 anti-human Hashtag 4 Antibody, BioLegend 394607
TotalSeq™-A0255 anti-human Hashtag 5 Antibody, BioLegend 394609
TotalSeq™-A0256 anti-human Hashtag 6 Antibody, BioLegend 394611
TotalSeq™-A0257 anti-human Hashtag 7 Antibody, BioLegend 394613
TotalSeq™-A0258 anti-human Hashtag 8 Antibody, BioLegend 394615
TotalSeq™-A0259 anti-human Hashtag 9 Antibody, BioLegend 394617
TotalSeq™-A0260 anti-human Hashtag 10 Antibody, BioLegend 394619
TotalSeq™-A0262 anti-human Hashtag 12 Antibody, BioLegend 394623
TotalSeq™-A0263 anti-human Hashtag 13 Antibody, BioLegend 394625
TotalSeq™-B0251 anti-human Hashtag 1 Antibody, BioLegend 394631
TotalSeq™-B0252 anti-human Hashtag 2 Antibody, BioLegend 394633
TotalSeq™-B0253 anti-human Hashtag 3 Antibody, BioLegend 394635
TotalSeq™-B0254 anti-human Hashtag 4 Antibody, BioLegend 394637
TotalSeq™-B0255 anti-human Hashtag 5 Antibody, BioLegend 394639
TotalSeq™-B0256 anti-human Hashtag 6 Antibody, BioLegend 394641
TotalSeq™-B0257 anti-human Hashtag 7 Antibody, BioLegend 394643
TotalSeq™-B0258 anti-human Hashtag 8 Antibody, BioLegend 394645
TotalSeq™-B0259 anti-human Hashtag 9 Antibody, BioLegend 394647
TotalSeq™-B0260 anti-human Hashtag 10 Antibody, BioLegend 394649
Human TruStain FcX™ (Fc Receptor Blocking Solution), BioLegend 422301

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.