nature portfolio

Corresponding author(s): Francois Pattou, Stefano Romeo

Last updated by author(s): 20/08/2024

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	Data collection into Clinical Research Form (CRF) proprietary of University Hospital of Lille, France (ABOS cohort), Campus Bio-Medico University Hospital, Rome, Italy (MAFALDA cohort), Helsinki University Hospital, Helsinki, Finland (Helsinki cohort), Obesity Clinic at Antwerp University Hospital, Edegem, Belgium (UZA cohort) and National Health Service records UK (UK Biobank); data export as xls or cvs; subsequent data processing in R
Data analysis	The cluster analysis was performed using the partitioning around medoids (pam) method in R (package 'cluster', version 2.1.4). Data from the UZA, MAFALDA, Helsinki and UK biobank cohorts were normalized using ABOS values for centering and scaling. Then, participants were allocated in the cluster they were most similar to after the exclusion of participants with BMI≤30 kg/m2; excessive alcohol consumption or having absolute standardized values of 5 or higher in at least one of the clustering traits, calculated as their Euclidean distance from the nearest cluster medoid derived from ABOS coordinates. Differential analysis of liver transcriptomic across the clusters was performed using moderated t-tests from the R Bioconductor package Limma version 3.52. Data were reported as median (interquartile range) for continuous variables and frequencies (percentages) for categorical variables. Clusters were compared using the Kruskal-Wallis test, Chi-squared test, or Fisher's exact test, as appropriate. Pathway enrichment on the transcriptome was performed with the R package ClusterProfiler (4.2.1), based on Gene Ontology Biological Process pathways. In the UK Biobank, clusters were compared using ANOVA, Kruskal–Wallis test, or Chi-square test as appropriate, adjusted for multiple testing separately for clinical data and genetic data, using Bonferroni method. Similarly, post-hoc comparisons were carried out with Bonferroni correction. The incidence of chronic liver disease, cardiovascular disease and type 2 diabetes were defined as the composite occurrence of the clinical event or event-related death during follow-up. Then, the cumulative incidence of the clinical outcomes was computed according to the Aalen-Johansen method for chronic liver disease, cardiovascular disease and type 2 diabetes, taking into account the competing occurrence of other-cause death, and of selected liver disease (only in the case of chronic liver disease; see above for ICD-10 codes). Cause-specific hazard ratios were calculated through Cox regressions, adjusted for age, sex, and alcohol intake. The proportional hazard assumption was verified through the inspection of the Schoenfeld residuals. Sensitivity analyses were

performed a) including only individuals with BMI≥27 kg/m2 and b) excluding those with harmful alcohol consumption (>50/60 g/day for women/men).Statistical analyses and graphical representations were performed using R statistical software version 4.2.2 (R Foundation for Statistical Computing, Vienna, Austria).

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 individual data analyzed in the current study are not publicly available due to national data protection laws and restrictions imposed by the ethics committee to ensure participant privacy. However, researchers can apply for access through an individual project agreement with the principal investigator at the University Hospital of Lille, France. The study protocol and meth-ods (NCT01129297) have been published and are available without restriction. Data access is conditional upon signing a data use agreement, which ensures data usage for the intended re-search purposes only. Researchers must submit a detailed request outlining their research objec-tives and methodology directed to the principal investigator of the ABOS study cohort (fran-cois.pattou@univ-lille.fr). Data will be available only to researchers affiliated with recognized institutions and for research that aligns with the original scope of the ABOS cohort study. Ac-cess will be granted approximately one month after the inter-institutional agreement for the in-dividual project is finalized and the study is registered on the Lille University Hospital site, in compliance with RGPD regulations. Data from UZA, MAFALDA and Helsinki cohorts are not publicly available due to governance limitations but are available for research by approval from principal investigators. All other data supporting the findings of this study are available within the article. UK Biobank data are publicly available to researchers through an open appli-cation via https://www.ukbiobank.ac.uk/register-apply/.
Raw transcriptomic files are available at GEO under the accession number GSE130991. Me-tabolite abundances file is available at BioStudies under the accession number S-BSST1479. Cluster annotations of transcriptomic and metabolomic samples are available at https://gitlab.com/bililleˆ/2024-raverdy_et_al-masld_clusters/-/tree/main/Data

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	The sex of participants was reported and was determined based on self-reporting
Population characteristics	Six variables associated with MASLD physiopathology and increased risk of MASH were selected for clustering in ABOS, namely age, body mass index (BMI), HbA1c, ALT, LDL cholesterol, and circulating triglycerides. Cluster analysis and identification of MASLD subtypes were performed on 1,389 ABOS participants, after the exclusion of 54 patients for self-declaration alcohol consumption above 50/60 g per day for women and men respectively at the first visit, to avoid any risk of inclusion of patients with alcohol-related liver disease (ALD); 58 participants for a BMI≤30 kg/m2; 27 participants for missing values in clustering traits (i.e., age, BMI, HbA1c, ALT, LDL cholesterol, and circulating triglycerides) and 17 participants having absolute standardized values of 5 or higher in at least one of the clustering traits (Supplementary Figure 1). The analysis was performed using the partitioning around medoids (pam) method in R (package 'cluster', version 2.1.4) 56, which is a more robust version of k-means clustering. Distances were computed as Euclidean distances using standardized variables scaled to a mean of 0 and a standard deviation of 1. Data from the UZA, MAFALDA and Helsinki were normalized using ABOS values for centering and scaling. Then, participants were allocated to the cluster they were most similar to after the exclusion of participants having absolute standardized values of 5 or higher in at least one of the clustering traits, calculated as their Euclidean distance from the nearest cluster medoid derived from ABOS coordinates. Data from the UK biobank cohorts were normalized using ABOS values for centering and scaling. Participants were allocated to the cluster they were most similar to after the exclusion of those with self-reported history or medical diagnosis of other causes of liver disease, with a medical diagnosis of the target longitudinal outcome at baseline, or having absolute standardized values of 5 or higher in at least one of the clustering traits, calculated as their Euclidean distance from the nearest cluster medoid derived from ABOS coordinates.
Recruitment	ABOS is a prospective study (NCT01129297) aiming to identify the key factors influencing the outcomes of bariatric surgery. A total of 1,545 participants enrolled between 2006 and 2021 at the Lille University Hospital, Lille, France, were included in the present analysis.UK Biobank cohort. The UK Biobank is a large prospective cohort study recruiting approximately 500,000 participants (age 40-69 years) between 2006-2010 throughout the UK . Helsinki Cohort. The Helsinki Cohort enrolled 343 consecutive individuals with morbid obesity eligible for bariatric surgery and 42 consecutive individuals with a body mass index (BMI) ≥25 kg/m2 undergoing liver biopsy for suspected MASH, all recruited between 2006 and 2018 at the Helsinki University Hospital, Helsinki, Finland. Molecular Architecture of FAtty Liver Disease in individuals with obesity undergoing bAriatric surgery (MAFALDA) cohort. A total of 264 participants with liver biopsy data from the MAFALDA cohort were included in the analyses. Universitair Ziekenhuis Antwerpen (UZA). The UZA cohort included 467 patients referred to the Obesity Clinic at Antwerp University Hospital, Edegem, Belgium, for suspected MASLD based on imaging and biochemistry data. The collection of clinical, anthropometric, and histological data has been previously described .

Ethics oversight	ABOS cohort : Ethical approval for the ABOS study was granted by the Comité de Protection des Personnes Nord Ouest VI (Lille, France). Universitair Ziekenhuis Antwerpen (UZA). Written informed consent was obtained from all patients in both cohorts and the studies were conducted in conformity with the Helsinki Declaration. Molecular Architecture of FAtty Liver Disease in individuals with obesity undergoing bAriatric surgery (MAFALDA) cohort. The MAFALDA study has been approved by the Local Research Ethics Committee (no. 16/20) and it was conducted in accordance with the principles of the Declaration of Helsinki. All participants gave written informed consent to the study.Helsinki Cohort. The study was approved by the Local Research Ethics Committee at Helsinki University Hospital. All participants gave written informed consent to the study.UK Biobank cohort. The UK Biobank study has been approved by the NorthWest Multicenter Research Ethics Committee (no. 21/NW/0157). All participants gave written informed consent to the study. Data used in this study were obtained under application number 37142.

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	No sample size calculation was used, but we included all available records with a full set of the clustering variables in ABOS, MAFALDA, Helsinki, UZA and UK Biobank cohorts.
Data exclusions	Cluster analysis and identification of MASLD subtypes were performed on 1,389 ABOS participants (Supplementary Figure 1), after the exclusion of 27 participants for missing values in clustering traits (i.e., age, BMI, HbA1c, ALT, LDL cholesterol, and circulating triglycerides); 58 participants for a BMI≤30 kg/m2; 54 patients for self-declaration alcohol consumption above 50/60 g per day for women and men respectively at the first visit, to avoid any risk of inclusion of patients with alcohol-related liver disease (ALD); and 17 participants having absolute standardized values of 5 or higher in at least one of the clustering traits
Replication	We validated the distribution of MASLD phenotypes in MAFALDA, Helsinki, UZA and UK Biobank cohorts.
Randomization	No intervention was planned for this study and thus no randomisation was needed and study utilized observational data.
Blinding	Study participants were not allocated to any groups for interventions thus blinding was not required 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

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	ABOS cohort study NCT01129297
Study protocol	cohort study

Data collection	ABOS cohort: Demographic characteristics, anthropomorphic measurements, medical history, concomitant medication, and laboratory tests were collected before surgery. A 75-g oral glucose tolerance test (OGTT) was performed after overnight fasting at baseline and one year after surgery. Diabetes status was defined at baseline based on a previous history of diabetes, use of anti-diabetic medications, fasting plasma glucose ≥126 mg/dL (7.0 mmol/L) and/or 2-hour plasma glucose ≥ 200 mg/dL (11.1 mmol/L) during OGTT, and/or HbA1c ≥ 6.5% (48 mmol/L). UK Biobank: Clinical and laboratory data were collected using highly standardized procedures. Medical diagnoses were obtained through linkage of hospital admissions, death and cancer registers from the National Health Service records (data-fields 41270, 40001, 40002, 40006).
Outcomes	ABOS, UZA, MAFALDA and Helsinki : Liver histology was obtained at baseline through a percutaneous liver needle biopsy performed during surgery. UK Biobank: Hepatic steatosis was defined by PDFF >5.5% 46, MASH by PDFF >5.5% and cT1 >800 msec. We analyzed the risk of developing hepatic and extrahepatic outcomes and overall mortality in the UK Biobank cohort.

Magnetic resonance imaging

Experimental design

Design type	cohort study
Design specifications	cohort study
Behavioral performance measures	Non applicable

Acquisition

Imaging type(s)	LiverMultiScan©
Field strength	Siemens 1.5T MAGNETOM Aera
Sequence & imaging parameters	Hepatic steatosis was defined by PDFF >5.5% , MASH by PDFF >5.5% and cT1 >800 msec
Area of acquisition	A shortened modified look locker inversion (ShMOLLI) was used to quantify liver T1, and a multi-echo-spoiled gradient-echo was used to quantify liver iron and fat.

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software	LiverMultiScan© Discover 4.0 software
Normalization	*If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*
Normalization template	*Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*
Noise and artifact removal	*Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*
Volume censoring	*Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.*

Statistical modeling & inference

Model type and settings	MRI data were reported as median (interquartile range) for continuous variables and frequencies (percentages) for categorical variables. MRI data across Clusters were compared using the Kruskal-Wallis test, Chi-squared test, or Fisher's exact test, as appropriate.
Effect(s) tested	*Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.*

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016)	*Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.*
Correction	*Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).*

Models & analysis

n/a	Involved in the study
☒	☐ Functional and/or effective connectivity
☒	☐ Graph analysis
☐	☒ Multivariate modeling or predictive analysis

Multivariate modeling and predictive analysis

We confirmed the association of the three clusters with liver phenotype in a subset of the UK Biobank participants (N=6,792) who underwent liver magnetic resonance imaging (MRI).