nature portfolio

Corresponding author(s): McLaren PJ, Sandhu M, Fellay J

Last updated by author(s): May 15, 2023

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	1. Genome-wide association study: Variant and sample quality control was performed using plink v1.9. Genotyping data were phased and imputed using IMPUTE2. Principal components were calculated using EIGENSTRAT. Association testing (linear regression) was performed using plink v1.9. Meta-analysis was performed using meta v1.7. Additional analyses of variants were performed using R v3.4.3. Functional consequences of the variants were assessed using the Variant Effect Predictor. 2. RNA and eQTL analyses: RNA sequencing reads were inspected using the FastQC tool for quality control. Reads were trimmed using Cutadapt for polyA and adaptors prior to mapping. Reads were mapped to the human reference sequence v38 using the STAR alignment tool in two pass mode. Quantification of gene expression was performed with htseq-count v0.9.1 using genes as features. Normalization was performed using the DESeq2 tool. eQTLs were identified by linear regression where normalized gene expression was regressed on variant dosage correcting for covariates using Matrix eQTL. Data were combined across studies using a random effects model implemented by METASOFT.

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 GWAS summary statistics are deposited in the NHGRI-EBI Catalog of human genome-wide association studies (https://www.ebi.ac.uk/gwas/home) under accession number GCP000635. RNA sequencing data are available at https://www.ebi.ac.uk/ena/browser/view/PRJEB18581 and the eQTL results are available at https://github.com/smontgomlab/AFGR.

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	We used genetically-inferred sex as covariate in the genome-wide association study No sex-stratified analysis was performed. We did not collect any information on gender.
Reporting on race, ethnicity, or other socially relevant groupings	We used the biological construct of genetic ancestry as an inclusion criterion. We inferred ancestry using principal components analysis of the genome-wide genotype data, comparing each participant to the 1000 Genomes phase 3 reference sample. Individuals clustering with the 1000 Genomes African populations (LWK, GWD, MSL, ACB, ASW, YRI, ESN) were deemed "of African ancestries" and included in the analyses. We did not use the sociological constructs of race/ethnicity at any point in the study.
Population characteristics	All study participants were HIV-1 infected adults (>18 years of age) of African ancestries recruited in one of nine different studies or cohorts, who provided informed consent for human genetic research. No other inclusion or exclusion criteria were applied.
Recruitment	The study participants were recruited as part of the original studies or cohorts in Africa, Europe and the USA, which are described in the Supplementary Materials.
Ethics oversight	All participants gave written informed consent for genetic testing and ethical approval was obtained from institutional review boards for each of the respective contributing centers (table S1).

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 performed. The International Collaboration for the Genomics of HIV sought to collect the largest possible number of genotyped samples from individual of African ancestries.
Data exclusions	During quality control of the genotyping data, the following exclusion criteria were applied: - Variants were removed that displayed high missingness (>2%), low minor allele frequency (<1%) or deviation from Hardy-Weinberg equilibrium (P<5x10-6) - Samples were excluded if they exhibited high missingness (>2%), high heterozygosity (Inbreeding coefficient >0.1 or <-0.1) or shared greater than 12.5% similarity in an identity by descent analysis
Replication	After initial discovery in a set of 2,682 individuals, we replicated the CHD1L association result in four additional, independent groups of HIV-1 infected individuals (N total =1,197 individuals). Association testing in these independent studies showed consistent replication of the CHD1L signal.
Randomization	This is not relevant to our study, because we performed an association study without any intervention on the participants.
Blinding	Blinding is not relevant for genome-wide association studies.

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)	The hiPSC line used in this study, KOLF2, was generated by the Human Induced Pluripotent Stem Cells Initiative (HipSci) consortium at the Sanger Institute using fibroblast from a healthy adult male with the CytoTune 1 Sendai method. Details on the generation of the line and the Certificate of Analysis are available on the HipSci project website (www.hipsci.org). Other cell lines.....
Authentication	The cell lines were not authenticated.
Mycoplasma contamination	The cell lines were not tested for Mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	We did not use any commonly misidentified cell line.