nature portfolio

Corresponding author(s): Johan Van Weyenbergh

Last updated by author(s): Apr 1, 2025

# 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<br>*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<br>*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 custom or commercial code was used for data collection. |
| Data analysis | Custom code used in this study is available at https://github.com/isaacracine/ham_tsp_multi_omics.git. Software used included GraphPad Prism 10.0, XL-STAT, WEKA 4.0, nSolver 4.0, TRACTOR, Regenie, Prsice-2, lassosum v1. |

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 anonymized data are available from the authors upon reasonable request. GWAS summary statistics are available at https://doi.org/10.48804/ITSA6K

## 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 | For this study, only biological sex at birth, as informed by the patients and controls, was used in conjunction with the genetic data from genomic analysis, in agreement with the current data on HAM/TSP being associated to female sex. We did not ask patients or controls about their gender nor was that a topic of the current research, as this is unrelated to HAM pathogenesis, as far as our current understanding goes. |
| Reporting on race, ethnicity, or other socially relevant groupings | No social constructs were used and "race" was not used as a variable, only ethnicities were used as strictly defined by genomic analyses. Moreover, in the Brazilian cohort most individuals were admixed so the EUR, AFR, AMR and ASN ethnicities were applied to genomic subsets/stretched, as identified by ADMIXTURE and TRACTOR software. |
| Population characteristics | Samples were collected through a Brazilian nation-wide effort, detailed in Suppl. Fig. S1. Each sample's HTLV-1 seropositivity was confirmed via Western Blot along with nested PCR. HAM diagnosis (cases) was made according to Castro-Costa crtiteria (reference 6), those without HAM were considered neurologically asymptomatic controls. |
| Recruitment | Samples were collected through a nation-wide effort, detailed in Suppl. Fig. S1, local clinicians were responsible for diagnosis of cases and controls, as described above. |
| Ethics oversight | The Ethical Board of "Instituto de Infectologia Emilio Ribas", Sao Paulo-Brazil, approved the protocol (Number 07688818.2.1001.0061). Signed informed consent was obtained from all participants prior to study inclusion. |

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 | Sample size for the Brazilian GWAS study was estimated for 500 cases and 500 controls according to Hong & Park (Genomics & Informatics 2012;10(2):117-122): for an OR of 1.5 this number achieves 60% statistical power but for an OR of 2.0 achieves >95% statistical power, under the assumption of 5% prevalence, 5% MAF, complete LD, and 5% α level. |
| Data exclusions | The original genotyping returned 902,527 variants for 1,043 samples. A first round of quality control resulted in 49 samples being removed due to failing quality control. Forty more samples were removed due to having a call rate error greater than 3%. Lastly, 41,374 variants with a call rate error also greater than 3% were removed. Thus, the received data consisted of 954 samples and 861,153 genotyped polymorphisms. Of the 954 samples, three were plate controls. The remaining 951 samples consisted of 416 HAM cases and 535 asymptomatic individuals (Fig. S1). |
| Replication | Published Japanese GWAS data was replicated in our Brazilian cohort using polygenic scores. Transcriptomic data were replicated using four different cohorts, as indicated in the Methods and in each figure. Monocyte counts were replicated in Brazilian and Peruvian cohorts as indicated in Figure 2. All other data were from single cohorts (all Brazilian except SOD2 in Peruvian cohort). |
| Randomization | Participants were not randomized but classified as either HAM (cases) or asymptomatic (controls), according to Castro-Costa criteria (reference 6). |
| Blinding | Investigators were not blinded to study outcomes (HAM cases or asymptomatic controls). |

# 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 |

# 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.*