nature portfolio

Corresponding author(s): Xiao Chang

Last updated by author(s): Oct 14, 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 | The GWAS datasets used in this study were obtained from publicly available sources, and no custom software or code was employed during the original data collection phase. |
| Data analysis | The analysis was performed using our novel InsightGWAS framework, a Transformer-based deep learning model implemented in Python 3.11.7 with PyTorch 2.3.1 (CUDA 11.8). Key dependencies include NumPy 2.2.6, pandas 2.3.0 and scikit-learn 1.7.0. The complete codebase has been deposited in a GitHub repository (https://github.com/ziangmeng/MA-MDD-Transformer-based-model-) with detailed documentation. |

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 datasets used in this study are publicly available. The GWAS summary statistics for migraine can be obtained upon application from the International Headache

Genetics Consortium (IHGC). Summary statistics for mental disorders are available for direct download from the Psychiatric Genomics Consortium (PGC) website at: https://pgc.unc.edu/for-researchers/download-results/.

## 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 | The migraine GWAS dataset analyzed in this study were obtained from publicly available datasets including the International Headache Genetics Consortium, UK Biobank, GeneRISK, and the Nord-Trøndelag Health Study. Biological sex was recorded as male/female through either self-reporting or genotype-derived chromosomal analysis in the original studies. As the source GWAS datasets did not provide sex-disaggregated summary statistics, we were unable to conduct sex-stratified analyses in our secondary analysis. |
| Reporting on race, ethnicity, or other socially relevant groupings | The analysis was restricted to individuals of European ancestry to minimize population stratification confounding, as determined by genotype-based principal component analysis (PCA) and self-reported ancestry in the original studies. These ancestry classifications were used exclusively for genetic control purposes in our analyses, not as proxies for socioeconomic or cultural factors. The source GWAS incorporated ancestry-informative principal components as covariates. |
| Population characteristics | All participant characteristics were derived from the existing GWAS summary statistics originally reported by Hautakangas et al. (Nature Genetics 2022). The cohorts consisted of European-ancestry adults with age distributions varying across studies. Detailed cohort descriptions are available in the Supplementary Note of the source publication and through the UK Biobank documentation (https:// www.ukbiobank.ac.uk). |
| Recruitment | Participants were recruited through multiple avenues including population-based cohorts, case-control studies, biobanking initiatives, direct-to-consumer studies, and clinical settings. While most cases were self-reported, a subset underwent phenotyping in specialized headache centers. Full recruitment details are documented in the Supplementary Note of the source GWAS publication (Hautakangas H. et al., Nat Genet, 2022). |
| Ethics oversight | All original studies obtained approval from their respective local research ethics committees and collected written informed consent from participants, as described in the Supplementary Note of the source GWAS (Hautakangas et al., Nature Genetics 2022). |

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 | Our analysis utilized existing GWAS summary statistics comprising 102,084 migraine cases and 771,257 controls from the International Headache Genetics Consortium (IHGC) meta-analysis. |
| Data exclusions | Standard GWAS quality control procedures were applied in the source studies to exclude individuals and variants. Further details are described in the Method section and Supplementary Note of the source GWAS (Hautakangas et al., Nature Genetics 2022). |
| Replication | we assessed the effectiveness of InsightGWAS by comparing its prioritized loci with those identified in larger, more powered GWAS, as well as testing replication in an independent cohort (MVP). These analyses demonstrate that InsightGWAS can successfully recover loci that are missed in smaller GWAS but later confirmed in larger studies or independent datasets, further details are described in the Results section. |
| Randomization | Randomization is not applicable in our secondary analysis of summary statistics. |
| Blinding | Blinding is not applicable in our secondary analysis of summary statistics. |

# 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.*