

Corresponding author(s): Yaoling Yang
Daniel John Lawson

Last updated by author(s): 13/1/25

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

- | | | |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | 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) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | 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

Data analysis

ChromoPainter
PBWT
FLARE
RFMix
MOSAIC
Neural-ADMIXTURE
gprofiler2
sparsesvd
nnls
bcftools
QCTools

The C++ code for SparsePainter is available on GitHub at <https://github.com/YaolingYang/SparsePainter>, and the website for SparsePainter is at <https://sparsepainter.github.io/>. PBWTPaint is available on GitHub at <https://github.com/richarddurbin/pbwt>.

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 phased 1000 Genomes Project data build GRCh37/hg19 are available at https://bochet.gcc.biostat.washington.edu/beagle/1000_Genomes_phase3_v5a/b37.vcf/. The genetic map build GRCh37/hg19 is available from https://bochet.gcc.biostat.washington.edu/beagle/genetic_maps. The UK Biobank data can be accessed by approved researchers through <https://www.ukbiobank.ac.uk>. We used the UK Biobank data under project 81499. The UK map data are available at <https://gadm.org>.

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	Sex-based analyses were not performed.
Reporting on race, ethnicity, or other socially relevant groupings	Our study included 7 self-reported ethnic groups from the UK Biobank: British, Irish, Indian, Pakistani, African, Caribbean, Chinese. We investigated the genes with LDAS and AAS signals which are shared between all these 7 self-reported ethnic groups.
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	We used the UK Biobank data under project 81499.

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](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We present PBWTPaint and SparsePainter which are used for efficient local ancestry inference. We show haplotypes better represent ancestries than principal components, whilst linkage-disequilibrium of ancestry identifies signals of recent changes to population-specific selection for many genomic regions associated with immune responses.
Research sample	Our samples were taken from the UK Biobank and 1000 Genomes Project. All the samples are included in our study.
Sampling strategy	We randomly sampled 10000 individuals from self-reported British ethnic background in the UK Biobank, and used all the individuals from the other 6 ethnicities we investigate. The sampling strategy ensured the sample size in each ethnicities do not differ substantially.
Data collection	Data were collected by the UK Biobank and 1000 Genomes Project.
Timing and spatial scale	These details are available from the official UK Biobank and 1000 Genomes Project websites.
Data exclusions	Low coverage samples were excluded.
Reproducibility	N/A
Randomization	UK Biobank target samples are allocated by their self-reported ethnic groups. 1000 Genome Project samples are allocated by their populations and continents.
Blinding	Blinding not possible.

Did the study involve field work? ☐ Yes ☒ No

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		Methods	
n/a	Involved in the study	n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies	<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines	<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology	<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants		

Plants

Seed stocks	<div>N/A</div>
Novel plant genotypes	<div>N/A</div>
Authentication	<div>N/A</div>