

## 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

*Provide a description of all commercial, open source and custom code used to collect the data in this study, specifying the version used OR state that no software was used.*

Data analysis

plink 1.9 (<https://www.cog-genomics.org/plink/1.9/>) , MTAG (<https://github.com/JonJala/mtag>) PRSice-2, R Language 3.46

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 relevant data are included in the manuscript or as supplementary materials of this paper

# 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://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	6,407 rural Ugandans
Data exclusions	NA
Replication	2,598 South African Zulu
Randomization	NA
Blinding	NA

## 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 type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern		

## Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	The Uganda General Population Cohort (GPC) is a population-based cohort study founded in late 1980, and it has over 22,000 participants from 25 neighbouring villages in Kyamolibwa in rural Uganda. This open cohort study was established to investigate the trends of HIV infection in Uganda. However, the cohort's focus now is to examine the role of host genetic variants associated with communicable and non-communicable diseases in rural Ugandans.
Recruitment	From 2010-2011, the research questions have included the epidemiology and genetics of communicable and non-communicable diseases (NCDs) to address the limited data on the burden and risk factors for NCDs in sub-Saharan Africa. The cohort comprises all residents (52% aged $\geq 13$ years, men and women in equal proportions) within one-half of a rural sub-county, residing in scattered houses, and largely farmers of three major ethnic groups.
Ethics oversight	Data collection was conducted from 2009 to 2013 for the DCC and from 2013 to 2014 for the DDS. The survey questionnaire included socioeconomic factors, health information, lifestyle factors, anthropometric measurements (including height, weight, systolic blood pressure, diastolic blood pressure, and hip and waist circumferences), biomarkers for communicable and non-communicable diseases, and genetic data.

Note that full information on the approval of the study protocol must also be provided in the manuscript.