nature research

Corresponding author(s): Mark McCarthy

# Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.

For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

## Experimental design

1. Sample size

   Describe how sample size was determined.

   We aimed to bring together the largest possible sample size (N>80,000 T2D cases and >350,00 controls) to study the role of coding variants in T2D. Our sample size is adequate to recover known T2D associated regions, and identify 28 novel T2D associated regions. Also, analytical power calculation showed that our dataset has >97% power to identify variant with 20% allele frequency and 1.05 OR or variant with 1% allele frequency and OR 1.20.

2. Data exclusions

   Describe any data exclusions.

   We used established protocols to conduct rigorous data quality control for each exome-array study: variants were excluded for the following reasons: (i) not mapping to autosomes or X chromosome; (ii) multi-allelic and/or insertion-deletion; (iii) monomorphic; (iv) call rate <99%; or (v) exact p<10-4 for deviation from Hardy-Weinberg equilibrium (autosomes only) (details in Supplementary Tables 1 & 9 and Online methods pages 41-42). We made sure that the allele labels and strand were well aligned between studies. We also visually examined the allele frequencies from the sample and the reference dataset (1000 Genomes Project), and made sure that the allele frequencies are consistent.

3. Replication

   Describe the measures taken to verify the reproducibility of the experimental findings.

   Experimental replication was not attempted.

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   Not applicable.

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   Not applicable.

Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

| n/a | Confirmed |
|---|---|
| ☐ | ☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.) |
| ☐ | ☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☒ | ☐ A statement indicating how many times each experiment was replicated |
| ☐ | ☒ 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 any assumptions or corrections, such as an adjustment for multiple comparisons |
| ☐ | ☒ Test values indicating whether an effect is present<br>*Provide confidence intervals or give results of significance tests (e.g. P values) as exact values whenever appropriate and with effect sizes noted.* |
| ☐ | ☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range) |
| ☐ | ☒ Clearly defined error bars in all relevant figure captions (with explicit mention of central tendency and variation) |

*See the web collection on statistics for biologists for further resources and guidance.*

## Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

The software used has been described in Online Methods section. Softwares are: GenCall, zCall, optiCall, RAREMETALWORKER, RareMETALS, METAL, IMPUTE2, PLINK, SHAPEITv2. In addition, study-specific software, used by each study to perform analyses is listed in Supplementary Tables 1 and 9.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

## Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.

Not applicable.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

No eukaryotic cell lines were used.

b. Describe the method of cell line authentication used.

Not applicable.

c. Report whether the cell lines were tested for mycoplasma contamination.

Not applicable.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

Not applicable.

## Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study.

No animals were used in the study.

Policy information about studies involving human research participants

## 12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

We provide a description of outcome and covariates used for each study in Supplementary Tables 1 and 9. In general, association analyses were conducted with adjustment for age, sex, kinship matrix, and any other study specific covariates. Where available, analysis was also conducting after adjustment for body mass index.