

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work we publish. This form is published with all life science papers and is intended to promote consistency and transparency in reporting. All life sciences submissions use this form; while some list items might not apply to an individual manuscript, 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](#).

► Experimental design

1. Sample size

Describe how sample size was determined.

We recruited all available population studies with adequate DNA samples and eligible to deposit data into a public database. Our new sample series genotyped with OncoArray includes over 73,000 samples.

2. Data exclusions

Describe any data exclusions.

We excluded all samples that performed poorly with respect to genotyping. We also limited our analyses to individuals of European ancestry to avoid the potential for spurious associations.

3. Replication

Describe whether the experimental findings were reliably reproduced.

We meta-analyzed our new genotyped samples with previous studies. We set a stringent statistical threshold to minimize false positive results.

4. Randomization

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

We recruited individuals based on presence or absence of prostate cancer. Covariate information was collected and controlled in the regression models (i.e. age, recruitment center, country, etc.).

5. Blinding

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

Blinding was not necessary in our study. This was a population-based observation study.

Note: all studies involving animals and/or human research participants must disclose 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 the Methods section if additional space is needed).

n/a Confirmed

- | | | |
|-------------------------------------|-------------------------------------|--|
| ☐ | ☒ | The <u>exact</u> 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 (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section) |
| ☐ | ☒ | A description of any assumptions or corrections, such as an adjustment for multiple comparisons |
| ☐ | ☒ | The test results (e.g. p values) given as exact values whenever possible and with confidence intervals noted |
| ☐ | ☒ | A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range) |
| ☒ | ☐ | Clearly defined error bars |

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.

All statistical packages used are publicly available and referenced in the manuscript. These include IMPUTE2, SHAPEIT, SNPTEST, and R.

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The *Nature Methods* [guidance for providing algorithms and software for publication](#) may be useful for any submission.

► 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 for-profit company.

No unique materials were used.

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.

No eukaryotic cell lines were used.

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

No eukaryotic cell lines were used.

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

No commonly misidentified cell lines were used.

► 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 details on animals and/or animal-derived materials used in the study.

No animals were used.

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.

Our study was limited to males (due to disease). Cases were individuals diagnosed with prostate cancer. Controls were prostate cancer free at time of recruitment. All baseline information such as age, country of residence were collected. Average age and clinical features (where applicable) are reported in the manuscript by epidemiological study. Genotype summaries are reported.

The OncoArray genotype data and relevant covariate information (i.e. ethnicity, country, principal components, etc.) generated during this study are currently being deposited into dbGAP for access (Accession #: phs001391.v1.p1; the submission is still underway). In total 47 of the 52 OncoArray studies, encompassing nearly 90% of the individual samples, will be available (Supplementary Table 19). The previous meta-analysis summary results and genotype data currently are available in dbGAP (Accession #: phs001081.v1.p1). The complete meta-analysis summary associations statistics is publicly available at the PRACTICAL website (<http://practical.icr.ac.uk/blog/>).