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 wherever 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 | We used the R statistical programming environment [version 4.0], the details of the versions of all the R packages used are provided in the methods paragraphs. |
| Data analysis   | We used the R statistical programming environment [version 4.0] and the code to reproduce our results are provided here [https://github.com/S4445/Landscape-AICDA-mutations](https://github.com/S4445/Landscape-AICDA-mutations) |

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 the datasets used in this study are open access and the list and the data availability statement is provided in the supplementary material with all the datasets used in this study.
### Human research participants

Policy information about: [studies involving human research participants and Sex and Gender in Research](#).

| Reporting on sex and gender | n/a |
|-----------------------------|-----|
| Population characteristics  | n/a |
| Recruitment                 | n/a |
| Ethics oversight            | n/a |

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.

- [x] 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-faq.pdf](#).

### Life sciences study design

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

| Sample size | No sample-size calculation was performed and all the potentially available data was used as indicated in the supplementary methods. |
|-------------|-------------------------------------------------------------------------------------------------------------------|
| Data exclusions | The simulations and statistics that we performed allowed only responding whether or not, within a sample, the number of observed c-AID mutations was higher than those expected by chance, hence only allowing to further filtering by sample not by mutation within a sample. However, we removed the 3/7277 samples (within the ICGC dataset) that had more mutations by chance from further analyses. We removed those three samples from further analyses. |
| Replication | We used different datasets with different next-generation approaches (i.e. whole-exome, whole-genome or panel of genes) to validate the presence of the genomic signature related with AICDA. |
| Randomization | Not relevant for this study |
| Blinding | Blinding was not relevant to this study because there was no therapeutic treatment assessed. |

### 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 |
|-----|-----------------------|
| [x] | Antibodies           |
| [x] | Eukaryotic cell lines |
| [x] | Palaeontology and archaeology |
| [x] | Animals and other organisms |
| [x] | Clinical data        |
| [x] | Dual use research of concern |

#### Methods

| n/a | Involved in the study |
|-----|-----------------------|
| [x] | ChIP-seq              |
| [x] | Flow cytometry        |
| [x] | MRI-based neuroimaging |