nature research

Corresponding author(s): Andrea Sottoriva, Trevor A. Graham

Last updated by author(s): Jun 4, 2020

Reporting Summary

Nature Research 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 Research policies, see Authors & Referees 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	Data from GDC portal was collected using the GDC Data Transfer Tool (https://gdc.cancer.gov/access-data/gdc-data-transfer-tool). Simulated tumour growth data was generated using custom code written in Julia (https://github.com/elakatos/CloneGrowthSimulation, https://julialang.org/, version 0.5.2).
Data analysis	The following open-source software was used in the bioinformatic analysis: ASCAT (https://github.com/Crick-CancerGenomics/ascat); Sequenza (https://cran.r-project.org/web/packages/sequenza/vignettes/sequenza.html, version 3.0.0); polysolver (https://software.broadinstitute.org/cancer/cga/polysolver); LOHHLA (https://bitbucket.org/mcgranahanlab/lohhla); ANNOVAR (http://annovar.openbioinformatics.org/en/latest/, version 20180416); Variant Effect Predictor (https://www.ensembl.org/info/docs/tools/vep, release 87); NeoPredPipe (https://github.com/MathOnco/NeoPredPipe). Calls derived from MANTIS (https://github.com/OSU-SRLab/MANTIS, not performed by the authors) were also used. All statistical analysis was carried out in R (https://www.r-project.org/about.html, version 3.5.0).

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/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

The datasets analysed during the current study are available from the NCI Genomics Data Commons Portal (https://portal.gdc.cancer.gov) COAD, READ, STAD and UCEC domains, and from the European Genome-Phenome Archive (https://ega-archive.org/) at accession code: EGAS00001003066.

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

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. We analysed all the suitable data (as according to exclusion criteria below) available to us, and comment on the statistical power of our analysis to resolve patterns in these data throughout the manuscript. In addition, the manuscript contains simulation-based power analysis commenting on the required data resolution.
Data exclusions	TCGA data was filtered according to pre-established criteria to ensure the quality of variant calls and VAF estimation: Only samples with available matched germline information (from blood samples) were considered. Samples with purity below 0.4 and ploidy above 3.6 were excluded before analysis. For immune escape analysis, only samples with gene expression data available in GDC, and at least one heterozygous HLA allele were considered. For comparison of escaped and non-escaped samples, all samples with less than 30 (subclonal) exonic mutations were excluded to avoid noise due to alternative biological pathways and small sample set of mutations.
Replication	No primary experiments were carried out in this study. No biological or technical replicates were considered: only a single sample was analysed for each TCGA cancer.
Randomization	Randomization was not relevant for comparison of different sample subtypes (immune escape vs no escape; MSI vs MSS) in the bioinformatic analysis. All samples meeting the preset criteria were considered.
Blinding	Blinding was not relevant due to only bioinformatic analysis carried out in the study.

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
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging