Materials Design Analysis Reporting (MDAR) Checklist for Authors

The MDAR framework establishes a minimum set of requirements in transparent reporting applicable to studies in the life sciences (see Statement of Task: doi:10.31222/osf.io/9sm4x.). The MDAR checklist is a tool for authors, editors and others seeking to adopt the MDAR framework for transparent reporting in manuscripts and other outputs. Please refer to the MDAR Elaboration Document for additional context for the MDAR framework.

Completed checklist for Dietlein et al.
"Genome-wide analysis of somatic noncoding mutation patterns in cancer"

Materials

Antibodies	Yes (indicate where provided: page no/section/legend)	n/a
For commercial reagents, provide supplier name, catalogue number and RRID, if available.	**No** antibodies were used in this study.	

Cell materials	Yes (indicate where provided: page no/section/legend)	n/a
Cell lines: Provide species information, strain. Provide accession number in repository **OR** supplier name, catalog number, clone number, **OR** RRID	Information provided in Supplementary methods (section "Cell Culture and Maintenance"). All cell lines were obtained from ATCC and identity verified by STR profiling.	
Primary cultures: Provide species, strain, sex of origin, genetic modification status.	**No** primary cell cultures were used in this study	

Experimental animals	Yes (indicate where provided: page no/section/legend)	n/a
Laboratory animals: Provide species, strain, sex, age, genetic modification status. Provide accession number in repository **OR** supplier name, catalog number, clone number, **OR** RRID	**No** laboratory animals were used in this study.	
Animal observed in or captured from the field: Provide species, sex and age where possible	**No** animals were observed or captured for this study.	
Model organisms: Provide Accession number in repository (where relevant) **OR** RRID	**No** model organisms (other than cell lines, cf. above) were used in this study.	

Plants and microbes	Yes (indicate where provided: page no/section/legend)	n/a
Plants: provide species and strain, unique accession number if available, and source (including location for collected wild specimens)	**No** plants were used in this study.	
Microbes: provide species and strain, unique accession number if available, and source	**No** microbes were used in this study.	

Human research participants	Yes (indicate where provided: page no/section/legend)	n/a
Identify authority granting ethics approval (IRB or equivalent committee(s), provide reference number for approval.	**No** human research participants were involved in this study and no patient data was generated. Sequencing data published in previous studies were re-analyzed in this study. Access to these sequencing data was approved by the HMF (DR-050) and PCAWG/ICGC (DACO-1078465) consortia and these IDs are provided in the acknowledgements section.	
Provide statement confirming informed consent obtained from study participants.	**No** human research participants were involved in this study (analysis of previously published sequencing data only).	
Report on age and sex for all study participants.	**No** human research participants were involved in this study (analysis of previously published sequencing data only).	

Design

Study protocol	Yes (indicate where provided: page no/section/legend)	n/a
For clinical trials, provide the trial registration number **OR** cite DOI in manuscript.	This study is **not** a clinical trial (retrospective analysis of previously published sequencing data).	

Laboratory protocol	Yes (indicate where provided: page no/section/legend)	n/a
Provide DOI or other citation details if detailed step-by-step protocols are available.	Detailed protocols on the experimental techniques used in this study are provided in the supplementary methods.	

Experimental study design (statistics details)	Yes (indicate where provided: page no/section/legend)	n/a
State whether and how the following have been done**, or** if they were not carried out.		
Sample size determination	The full sample size of whole-genome sequencing data from the PCAWG and HMF data was used in this study. Sample sizes were not predetermined prior to this study.	
Randomisation	**No** randomization was performed.	
Blinding	Researchers were **not** blinded to the annotations of genome sequencing data (e.g. tumor type).	
Inclusion/exclusion criteria	The full sample size of whole-genome sequencing data from the PCAWG and HMF data was used in this study. Samples and mutations of low quality from these sequencing consortia were filtered/excluded. A detailed description of the exclusion criteria are provided in the supplementary methods (section "Processing and filtering of whole-genome sequencing data"). These exclusion criteria are standard in the analysis of whole-genome sequencing data and references to the literature of the exclusion criteria are provided in the supplementary methods.	

Sample definition and in-laboratory replication	Yes (indicate where provided: page no/section/legend)	n/a
State number of times the experiment was replicated in laboratory	The text and figure legends provide for each experiment the number of times an experiment was replicated: CRISPRi screen - 2 replicates shown in Figure 5. Validation by qPCR - 3 replicates shown as error bars in Figure S43. Sanger sequencing validation of WGS data- no replicates (Figure S41 reports unsuccessful sequencing reactions for 2 samples, no re-attempt). Luciferase reporter experiments - 3 experiments with triplicates in each experiment (individual datapoints shown in Figure S45A).	
Define whether data describe technical or biological replicates	CRISPRi screen replicates are biological replicates (described in the main text). qPCR validation are technical replicates (qPCRs from the same RNA sample. described in the figure legend). Luciferase reporter experiments were performed in biological replicates (described in the figure legend).	

Ethics	Yes (indicate where provided: page no/section/legend)	n/a
Studies involving human participants: State details of authority granting ethics approval (IRB or equivalent committee(s), provide reference number for	**No** human research participants were involved in this study and no patient data was generated.	
Studies involving experimental animals: State details of authority granting ethics approval (IRB or equivalent committee(s), provide reference number for approval.	**No** experimental animals were involved in this study.	

Studies involving specimen and field samples: State if relevant permits obtained, provide details of authority approving study; if none were required, explain why.	**No** specimens or field samples were involved in this study.	

Dual Use Research of Concern (DURC)	Yes (indicate where provided: page no/section/legend)	n/a
If study is subject to dual use research of concern, state the authority granting approval and reference number for the regulatory approval	This study (statistics, cancer genomics, CRISPR screening, luciferase reporter) is not subject to dual use research of concern.	

Analysis

Attrition	Yes (indicate where provided: page no/section/legend)	n/a
State if sample or data point from the analysis is excluded, and whether the criteria for exclusion were determined and specified in advance.	No experimental data or data points were excluded. Samples and mutations of low quality from these HMF and PCAWG sequencing consortia were filtered/excluded. A detailed description of the exclusion criteria are provided in the supplementary methods (section "Processing and filtering of whole-genome sequencing data"). These exclusion criteria are standard in the analysis of whole-genome sequencing data and references to the literature of the exclusion criteria are provided in the supplementary methods.	

Statistics	Yes (indicate where provided: page no/section/legend)	n/a
Describe statistical tests used and justify choice of tests.	Standard statistical tests are described in the figure legends. A detailed description of the new statistics developed in this study (genome-wide analysis of significance for somatic mutations) is provided in the supplementary methods (section "Identification of significantly mutated regions in whole cancer genomes") as well as the main text (section: "Genome-wide detection of significantly mutated regions in somatic whole cancer genomes") and "Methods"). The choice of the new statistics is justified in the main text and differences to existing statistical models are described.	

Data Availability	Yes (indicate where provided: page no/section/legend)	n/a
State whether newly created datasets are available, including protocols for access or restriction on access.	**No** new datasets were generated in this study (analysis of previously published sequencing data).	
If data are publicly available, provide accession number in repository or DOI or URL.	**No** new datasets were generated in this study analysis of previously published sequencing data).	
If publicly available data are reused, provide accession number in repository or DOI or URL, where possible.	Sequencing data published in previous studies were re-analyzed in this study. Access to these sequencing data was approved by the HMF (DR-050) and PCAWG/ICGC (DACO-1078465) consortia and these IDs are provided in the acknowledgements section. All other data are publicly available without restrictions or data access requests. Sources and download links are provided in the supplementary methods.	

Code Availability	Yes (indicate where provided: page no/section/legend)	n/a
For all newly generated code and software essential for replicating the main findings of the study:		
State whether the code or software is available.	Yes, the source code of our genome-wide analysis software is publicly available. Furthermore, we provide all files needed to run this source code (annotation files of the human genome) as well as test files to run this software.	
If code is publicly available, provide accession number in repository, or DOI or URL.	The Supplementary User Manual provides all download links for source code, annotation files and test files to run the software developed in this study. Furthermore, the user manual provides detailed step by step on how to run the source code. Based on our user manual, scientists not involved in this study, were able to run the source code independently and reproduce significant findings of this study.	

Reporting

Adherence to community standards	Yes (indicate where provided: page no/section/legend)	n/a
MDAR framework recommends adoption of discipline-specific guidelines, established and endorsed through community initiatives. Journals have their own policy about requiring specific guidelines and recommendations to complement MDAR.		
State if relevant guidelines (eg., ICMJE, MIBBI, ARRIVE) have been followed, and whether a checklist (eg., CONSORT, PRISMA, ARRIVE) is provided with the manuscript.	Discipline-specific guidelines in cancer genomics and experimental cancer biology have been followed (reproducibility, data and source code availability, data access requests, sufficient number of replicates, positive and negative controls, inclusion of all experimental data, verification of cell line identity). No additional checklists are provided with the manuscript.	