

## 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

- |                                     |                                     |  |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | A description of all covariates tested   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | 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) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | 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** All the single-cell RNA-seq data sets analysed in the study were aligned to the reference genome using Cell Ranger version 6.0.1. The ATAC-seq data sets were aligned to the reference genome using BWA-MEM v0.7.17-r1188.

**Data analysis** The code to reproduce the analysis presented in this study is available at <https://github.com/cortes-ciriano-lab/SComatic>

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](#)

The raw WES and scRNA-seq data for the skin squamous cell carcinoma and matched normal samples are available at the Gene Expression Omnibus (GEO) database under the accession number GSE144240. The raw scRNA-seq data from myeloproliferative neoplasms and colorectal cancer patients are available through controlled access application via dbGaP under dbGaP Study Accession numbers phs002308.v1.p1 and phs002407.v1.p1, respectively. The cell type annotations for

the colorectal cancer data set are available at GEO database under the accession number GSE178341. Raw sequencing data and cell type annotations for 6 additional colorectal cancer patients included in this study are available at GEO database under the accession number GSE144735. The cell type annotations for the MPN data set were obtained from our previous study. The raw scRNA-seq data and cell type annotations for the human heart cell atlas were downloaded from the Human Cell Atlas Data Portal (<https://data.humancellatlas.org/>). The raw single-cell ATAC-seq data and cell type annotations are available at GEO database under the accession number GSE184462. The raw sequence data from GTEx samples are available at the Analysis Visualization and Informatics Lab-space (AnVIL; [https://anvil.terra.bio/#workspaces/anvil-datastorage/AnVIL\\_GTEEx\\_V9\\_hg38](https://anvil.terra.bio/#workspaces/anvil-datastorage/AnVIL_GTEEx_V9_hg38)) and can be downloaded through controlled data access application via dbGaP under Study Accession number: phs000424.

## Human research participants

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

Reporting on sex and gender

NA

Population characteristics

NA

Recruitment

NA

Ethics oversight

NA

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.

☒ 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](https://www.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 calculations were performed.

Data exclusions

We did not exclude any data points from the published data sets analysed.

Replication

The main findings in this study are related to the discovery of somatic mutations in single-cell data sets using a new algorithm, SComatic. To assess the performance of the algorithm to detect mutations, we analysed multiple data sets and single-cell technologies.

Randomization

No specific allocation of samples was performed.

Blinding

The investigators involved in this study were not blinded to sample allocation across groups during data analysis, which in this case refers to the genotype of samples, such as MSI vs MSS in the case of colorectal cancer data sets.

## 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
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging