nature research

Corresponding author(s): L. Vermeulen

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.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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
☐	☒	An indication of 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 statistics 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
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Leica LAS-X v3.1.5.16308, BD FACSDIVA
Data analysis	IMARIS (version 8.3.0); Extreme Limiting Dilution Analysis (ELDA) 'limdil' function (http://bioinf.wehi.edu.au/software/elda/); Graphpad Prism 7; MATLAB; FlowJo v10; ImageJ Programming code has been deposited in GitHub: (https://github.com/dmmiedema/Tumor-Growth-Model)

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 upon request. 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

Source data for Figures 1, 2, 6 and 7 and Supplementary Figures 3 and 6 have been provided in Supplementary Table 1. RNA sequencing data have been deposited

in the Gene Expression Omnibus (GEO) under accession number GSE95499. Programming code has been deposited in GitHub: (https://github.com/dmmiedema/Tumor-Growth-Model). Additional theoretical information regarding the modelling can be found in Supplementary Note 1. All other data supporting the findings of this study are available from the corresponding author on reasonable request.

Field-specific reporting

Please select 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/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

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

Sample size	Animal sample sizes were estimated on the basis of previous work (Vermeulen, L. et al. Science 342, 995-998 (2013) in which we performed a similar study in mouse intestine. Pilot experiments were used to estimate the amount of clones per tumour.
Data exclusions	Animals were only excluded from analyses when no tumours appeared.
Replication	All experiments were repeated at least three times. Clonal data was analysed in three independent xenograft models, with similar results. All attempts at replication were succesful.
Randomization	Animals were randomly assigned to experimental groups.
Blinding	No blinding was performed during in vivo experiments, since our study is descriptive and data was analysed using automated image analysis, blinding is not relevant.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☐	☒ Unique biological materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants

Methods

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

Unique biological materials

Policy information about availability of materials

Obtaining unique materials	All unique materials used are readily available from the authors.

Antibodies

Antibodies used	Rabbit monoclonal anti-Ki67, SAB5500134, Sigma, 1:200, lot#170407LV2D Mouse monoclonal anti-Ki67 (MIB-1) Agilent M7240, 1:200, lot#20052254 Rabbit polyclonal anti-cleaved caspase3 (Asp175) Cell Signaling 9661, 1:600 , lot#45 Rabbit polyclonal anti-alpha-SMA, Abcam 5694, 1:100 Mouse monoclonal, anti-alpha-SMA, A-2547, Sigma, 1:100, lot#076M4784V Rabbit anti-osteopontin ab91655, Abcam, 1:100, lot#GR64316-31 Mouse anti-AC133-PE Miltenyi 130-080-801, 1:1000, lot#5170809431 Mouse anti-CD133/1-Biotin (AC133) Miltenyi 130-090-664, 1:20, lot#5060531037 Goat-anti-rabbit-Alexa488 Invitrogen A11008, 1:500, lot#1853312 Goat-anti-rabbit-Alexa546 Invitrogen A11035, 1:500, lot#1810820 Donkey-anti-mouse-IRdye680 Li-Cor Biosciences 926-32222, 1:500, lot#C70419-08 Rabbit polyclonal anti-CD31 Abcam ab28364, 1:20, lot#GR150486-9

Mouse anti-Hif1-alpha (clone 54/HIF1a) BD Transduction Laboratories, 610959, 1:100, lot#4073775
Mouse anti-Mucin2 Santa Cruz CCP58, 1:100, lot#A1818
Mouse anti-cytokeratin20 Genetex GTX15205, 1:100, lot#31258
Rabbit anti-Intestine alkaline phosphatase Genetex GTX27322, 1:100, lot#10563
Rabbit polyclonal anti-LysozymeEC 3.2.1.17, DAKO A0099, 1:100 , lot#00061477
Mouse anti-alpha-defensin5 Abcam ab90802, 1:100, lot#GR232-3
Goat anti-mouse/rabbit/rat Power vision Poly-HRP Immunologic dpvp110HRP
Mouse IgG1-biotin eBioscience P3 13-4714-85, 1:500, lot#015567
Streptavidin-APC BD pharmingen 554067, 1:500, lot#34340

Validation

All antibodies were used as validated by the manufacturer for their specific assay according to their datasheet. Rabbit antiKi67, mouse anti-ki67, anti-cleaved caspase3, anti-SMA, anti-osteopontin, anti-AC133, anti-CD133, anti-CD31, anti-HIF1a, anti-cytokeratin20, anti-intestine alkaline phosphatase, anti-lysozymeEC, anti-alpha-defensin5 were all published before. anti-Mucin2 was validated by the manufacturer for immunofluorescence.
We performed primary and secondary antibody controls for all immunofluorescent stainings.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

Human primary colon cancer cultures were established in our lab. HCT-15 (ATCC) and HT29 (Sanger) cells were purchased.

Authentication

Cell lines have been authenticated by STR profiling and mutation analysis

Mycoplasma contamination

All cell lines were tested negative for mycoplasm.

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified cell lines were used

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals

Mouse: Hsd:Athymic Nude-Foxn1nu, female, 6-12 weeks, obtained from Envigo.
Mouse: NOD-scid IL2rγnull mice (NSG; NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ), female, 6-12 weeks, were bred in our facility.

Wild animals

The study did not involve wild animals.

Field-collected samples

The study did not involve field collected samples.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Dissociated tumour cells were washed in FACS buffer (PBS + 1% FCS (Life Technologies)). Cells were incubated for 1 hour at 4°C with either mouse anti-AC133-biotin (130-090-664, Miltenyi, 1:20), or isotype controls (mouse IgG1-biotin (eBioscience, 1:500). As secondary antibody streptavidin-APC (BD pharmingen, 1:500) was used. Dead cells were excluded by 7-AAD staining (BD Biosciences).

Instrument

Cells were analysed using FACSCanto (BD Biosystems)

Software

Flowjo v10

Cell population abundance

mStrawberry+ cell population abundance was dependent on tamoxifen induction.

Gating strategy

Cells were selected in FSC/SSC dot plot to remove debris, single cells were gated using the FSC-H/FSC-W dot plot. GFP+, mStrawberry+ or PE+ cells were gated and compared with a control sample with no detectable fluorochrome expression.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.