nature portfolio

https://doi.org/10.1038/s41586-021-03648-3

Supplementary information

Clonal fitness inferred from time-series modelling of single-cell cancer genomes

In the format provided by the authors and unedited

Clonal fitness inferred from timeseries modeling of single cell cancer genomes: Supplementary Information

Sohrab Salehi*[1], Farhia Kabeer*[1,2], Nicholas Ceglia[3], Mirela Andronescu[1,2], Marc Williams[3], Kieran R. Campbell[6], Tehmina Masud[1], Beixi Wang[1], Justina Biele[1], Jazmine Brimhall[1], David Gee[1], Hakwoo Lee[1], Jerome Ting[1], Allen W. Zhang[1], Hoa Tran[1], Ciara O'Flanagan[1], Fatemeh Dorri[4,1], Nicole Rusk[3], Teresa Ruiz de Algara[1], So Ra Lee[1], Brian Yu Chieh Cheng[1], Peter Eirew[1], Takako Kono[1], Jenifer Pham[1], Diljot Grewal[3], Daniel Lai[1], Richard Moore[7], Andrew J. Mungall[7], Marco A. Marra[7], IMAXT Consortium[8], Andrew McPherson[3], Alexandre Bouchard-Côté[5], Samuel Aparicio[†,1,2], Sohrab P. Shah[†,3]

1. Department of Molecular Oncology, BC Cancer, Vancouver, BC, Canada
2. Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC, Canada
3. Computational Oncology, Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
4. Department of Computer Science, University of British Columbia, Vancouver, BC, Canada
5. Department of Statistics, University of British Columbia, Vancouver, BC, Canada
6. Lunenfeld-Tanenbaum Research Institute Mount Sinai Hospital Joseph & Wolf Lebovic Health Complex, Molecular Genetics, University of Toronto, Toronto, ON, Canada
7. Canada's Michael Smith Genome Sciences Centre, BC Cancer, Vancouver, BC, Canada
8. CRUK Grand Challenge IMAXT Team

* - equal contribution

† - Corresponding Authors:
saparicio@bccrc.ca
shahs3@mskcc.org

Keywords: tumour evolution, single cell sequencing, fitness, timeseries, phylogenetic reconstruction, drug resistance

Contents

1 Human mammary cell lines and serial passaging **3**

1.1 Human mammary cell lines mixing experiments 4

2 Establishment and serial passaging of patient derived xenografts **4**

2.1 Serial passaging of PDX 4

2.2 PDX tumour growth measurement curves 5

2.3 Histopathology of PDX tumours 5

2.4 TNBC PDX tumour mixing experiments 5

2.5 TNBC PDX timeseries treatment with cisplatin 5

3 Single cell whole genome sequencing and library construction with DLP+ **6**

3.1 Creation of single cell suspensions from PDX 6

3.2 Robot spotting of single cells into the nanolitre wells and library construction 6

4 DLP+ sequence analysis, copy number determination and quality control filtering **7**

4.1 Phylogenetic tree inference, clone determination and clonal abundance measurements 7

4.2 Single nucleotide variant analysis from DLP+ 8

4.3 Mutation burden analysis 8

5 Fitness modeling **8**

5.1 fitClone: a Bayesian fitness model for timeseries data 9

5.2 Wright-Fisher diffusions with selection 9

5.3 The fitClone model 9

5.4 Posterior inference under the fitClone model 10

5.5 Estimating the effective population size . . . 10

5.6 Summarising the posterior distribution . . . 11

5.7 Probability of positive selection . . . 11

5.8 Selecting the reference clone . . . 11

5.9 The deterministic logistic growth model . . . 12

5.10 Simulation benchmarking . . . 12

5.11 Software and implementation . . . 12

6 Single cell RNA sequencing (scRNAseq) 12

6.1 Processing of cell lines for scRNAseq data . . . 13

6.2 Processing of patient derived xenografts for scRNAseq data . . . 13

6.3 Quality control . . . 13

6.4 Integrative genome-transcriptome analysis with `clonealign` . . . 13

6.5 Differential expression analysis . . . 14

7 Supplementary figures 15

8 Supplementary tables 21

1 Human mammary cell lines and serial passaging

The human mammary epithelial cell line 184-hTERT wild type and isogenic 184-hTERT-P53 KO cell line, generated from 184hTERT WT-L9, were grown as previously described [1, 2]. *P53*$^{-/-}$ was knocked out by using CRISPR/Cas9 technology and one clone (99.25) was serially passaged to further subdivided at passage 10 into "branch a" and "branch b" parallel knock out branches. NM_000546(TP53):c.[156delA];[156delA], p.(GIn52Hisfs*71), and the absence of TP53 protein was confirmed with western blot (**Supplementary Fig. 1**a).

Two branches of 184-hTERT-P53$^{-/-}$ (clone 95.22) along with the counterpart wild type branch were serially passaged over $\sim$ 55-60 generations, by seeding $\sim$ 1 million cells into a new $10\,\mathrm{cm}$ tissue culture treated dish (Falcon-CABD353003) and cryopreserving every fifth passage. Mammary Epithelial Cell SingleQuot Kit Supplements (MEGM™), Growth Factors (Lonza CC-4136), with $5\,\mu\mathrm{g\,ml^{-1}}$ transferrin (Sigma) and $2.5\,\mu\mathrm{g\,ml^{-1}}$ isoproterenol (Sigma) were used as a growth media as previously described [2]. Cells were grown to around 85-90% confluence, trypsinized for 2 minutes

(Trypsin/EDTA 0.25%,VWR CA45000-664), re-suspended in cryopreservation media (10% DMSO-Sigma-D2650, 40% FBS-GE Healthcare SH30088.03, 50%media) and frozen to $-80\,^{\circ}\mathrm{C}$ at a rate of$-1\,^{\circ}\mathrm{C}\,\mathrm{min}^{-1}$. Cells were cultured continuously from passage 10 (post initial cloning as in [2]) to passage 60 for 184-hTERT WT and upto passage 57 and passage 55 for the *p53*$^{-/-}$ *branches a and b*, respectively, from initial cloning/isolation. Genome sequencing was undertaken at passages 25, 30, 51 and 60 from the wild type branch, passages 10, 15, 25, 30, 40, 50 and 57 from *p53* $^{-/-}$ *branch a* and passages 20, 30, 35, 40, 45, 50 and 55 from *p53* $^{-/-}$ *branch b*. Also, the transcriptome sequencing was carried out on passages 11 and 57 of *p53*$^{-/-}$ *branch a* and passages 15, 30 and 50 of *p53*$^{-/-}$ *branch b*.

1.1 Human mammary cell lines mixing experiments Enforced clonal competition of higher fitness clones with lower fitness counterparts should result in re-emergence or fixation of high fitness clones over time, even when re-starting from a low population prevalence. Frozen vials from 184-hTERT WT p28 (SA039) and *p53*$^{-/-}$-clone 95.22 (SA906b) passage 61 was thawed, trypsinized, counted and remixed in 3:1 ratio respectively. Prior to plating in the culture, an aliquot was subjected to DLP+ to measure the baseline clonal composition labelled as X0, and 2,701 single cell genomes were generated. At 80% confluence on the plate, cells from X1 were harvested and serially passaged upto the 20th passage. Single cell whole genome sequencing data from three time points X0, X1 and X5 with a median of 898 cells per timepoint was collected. Wildtype Clone F comprising diploid cells, had a clonal abundance of 0.72 at timepoint X0, but declined to 0.19 at X5. Clones E and D, with the highest selection coefficients in the *p53*$^{-/-}$ *branch b* cell line, were measured at clonal abundance of 0.07 and 0.18 at X0, but reached a clonal abundance of 0.37 and 0.35 at timepoint X5, respectively.

2 Establishment and serial passaging of patient derived xenografts

The Ethics Committees at the University of British Columbia approved all the experiments using human resources. Patients in Vancouver, British Columbia were recruited, and samples were collected under the tumour tissue repository (TTR-H06-00289) and transplanted in mice under the animal resource centre (ARC) bioethics protocol (A19-0298-A001) approved by the animal care committee (University of British Columbia BC Cancer Research Ethics Board H20-00170) protocols.

After informed consent, tumour fragments from patients undergoing excision or diagnostic core biopsy were collected. Tumour materials were processed as described in [3] and transplanted in mice approved by the animal care committee. Briefly, tumour fragments were chopped finely with scalpels and mechanically disaggregated for one minute using a Stomacher 80 Biomaster (Seward Limited, Worthing, UK) in $1\,\mathrm{ml}$ to $2\,\mathrm{ml}$ cold DMEM/F-12 with Glucose, L-Glutamine and HEPES (Lonza 12-719F). An aliquot of $200\,\mu\mathrm{l}$ of medium (containing cells/ clumps) from the resulting suspension was used equally for 4 transplantations in mice. Tumours were transplanted in mice as previously described [3] in accordance with SOP BCCRC 009. Briefly, female immuno-compromised, NOD/SCID/IL2r$\gamma^{-/-}$ (NSG) and NOD/Rag1$^{-/-}$Il2r$\gamma^{-/-}$ (NRG) [4] mice were bred and housed at the Animal Resource Centre (ARC) at the British Columbia (BC) Cancer Research Centre . For subcutaneous transplants, mechanically disaggregated cells and clumps of cells were re-suspended in $150\,\mu\mathrm{l}$ to $200\,\mu\mathrm{l}$ of a 1:1 v/v mixture of cold DMEM/F12: Matrigel (BD Biosciences, San Jose, CA, USA). 8-12 weeks old mice were anesthetized with isofluorane, then the mechanically disaggregated cells/clumps suspension was injected under the skin on the left flank using a $1\,\mathrm{ml}$ syringe with 21gauge needle. The animal care committee of animal welfare and ethical review committee, the University of British Columbia (UBC), approved all experimental procedures. Same method was used for maximum tolerated dose (MTD) studies later for drug experiments (**Supplementary Fig. 1**b).

2.1 Serial passaging of PDX Tumours were serially passaged as described in [3]. Briefly, for serial passaging of PDX, xenograft-bearing mice were euthanized when the size of the tumours approached $1000\,\mathrm{mm}^3$ in volume (combining together the sizes of individual tumours when more than one was present). The tumour material was excised aseptically, and processed as described for primary tumour. Briefly, the tumour was harvested and minced finely with scalpels then mechanically disaggregated for one minute using a Stomacher 80 Biomaster (Seward Limited, Worthing, UK) in $1\,\mathrm{ml}$ to $2\,\mathrm{ml}$ cold DMEM-F12 medium with Glucose, L-Glutamine and HEPES. Aliquots from the resulting suspension of cells and fragments were used for xenotransplants in the next generation of mice and cryopreserved. Serially transplanted aliquots represented approximately 0.1-0.3% of the original tumour volume. HER+SA532 and TNBC-SA609 PDX were passaged upto 10 generations and scDNAseq was carried out at each timepoint (**Supplementary Fig. 1**c). The other three untreated and treated PDX timeseries were generated in the same way for 4-5 passages.

2.2 PDX tumour growth measurement curves NRG mice received sub-cutaneous (SQ) inoculation of tumour cells ($150\,\mu l$) on day 0. The tumours were allowed to grow to palpable solid nodules. Around 7-9 days after they were palpable, their size were measured with calipers every 3rd day. Tumours were measured in two dimensions using a digital caliper and expressed as tumour volume in mm3; defined as: [Volume= $0.52\times$(Length)$\times$(Width)$\times$(Width)]. Record of patient derived xenografts 10 generations timeseries, HER2+SA532 and TNBC-SA609 exhibited progressively higher tumour growth rates in later passages (**Supplementary Fig. 1**d). Under drug perturbation, the treated tumours in the first two cycles of treatment showed rapid growth reduction but in the third cycle started showing non-responsive behaviour leading to total resistance in the fourth cycle. Tumour growth inhibition (TGI) percentage range was defined as: [1 - (mean volume of treated tumours)/(mean volume of control tumours) x 100%] [5].

2.3 Histopathology of PDX tumours The hormone receptor status of all tumour samples were determined by immunohistochemistry and FISH (Fluorescence in situ hybridization) copy number. Two separate tissue microarrays (TMAs) were prepared using duplicate $1\,\mathrm{mm}$ cores extracted from formalin-fixed paraffin-embedded (FFPE) blocks, containing materials from patient derived xenografts (TNBC-SA609, TNBC-535, TNBC-SA1035, and HER2+532). The slides were reviewed by a pathologist and immunohistochemistry (IHC) scores were summarized in **Supplementary Table 5**. Deparaffinized $4\,\mu\mathrm{m}$ sections of paraformaldehyde fixed tumours were processed for immunohistochemistry (IHC) using a Discovery XT automated system (Ventana Medical Systems, Tucson, AZ, USA). EGFR, INPP4B, Ki67, PR, ECAD were all performed on the Ventana Discovery XT platform using CC1 for antigen retrieval, incubating for one hour at room temperature, and using the UltraMap DAB detection kit. Primary antibodies included ERα (clone SP1, Ventana), HER2 (clone 4B5, Ventana), EGFR (clone EP22, Epitomics, Burlingamen, CA, USA) and Ki67 (clone SP6, Thermo Scientific) (**Supplementary Table 4**). Horseradish peroxidase-conjugated Discovery Universal Secondary Antibody (Ventana) was then applied, and the slides developed using 3, 3'-diaminobenzidine Map Kit (Ventana). Representative staining is shown in **Supplementary Fig. 1**e.

H&E and immunohistochemistry (IHC) of TNBC-SA609, TNBC-SA535 and TNBC-SA1035 PDX tumours showed that they all derived from a triple negative breast cancer patient while HER2+SA532 was derived from a HER2+ breast cancer patient. HER2 IHC was scored as 2+, HER2/CEP17 ratio was calculated as 6.5 (positive)[6] from FISH and focal high level amplification (average copy number state of 10) of the *ERBB2* locus (approximately Chr17:37500001-38000000) was found in the DLP+ data.

2.4 TNBC PDX tumour mixing experiments In contrast with HER2+SA532, all TNBC PDX models exhibited evidence of clonal dynamics and variation in selection coefficients consistent with positive selection and differential fitness Frozen vials from the untreated TNBC PDX passages three (X3) and eight (X8), were thawed and physically remixed in two different volumetric proportions of X3:X8 by tumour weight. The ratio of approximately 1:1 and 1:0.4, labelled as mixture branch a and branch b, respectively. From each of different dilutions, $200\,\mu l$ of aliquot was transplanted in two female 8-12 weeks old immunodeficient mice, each using the same protocol described above. Before transplantation, a small proportion of the physical mixture of cells from the 1:1 ratio, was subjected to whole genome single cell sequencing to measure the baseline clonal composition labelled as M0 and its subsequent PDX as M1 Each of the thawed X3 and X8 cell populations used for mixing were also retransplanted independently to confirm the viability of the tumour material for PDX tumour growth. The tumour cell mixture was then serially passaged over 4 generations for branch a and 5 generations for branch b, designating the transplants as M1-M4 and M1 to M5, respectively. Tumours from serial passages (X3:X8) from both mixtures branches were collected and analysed with scDNAseq (DLP+) as for other samples.

2.5 TNBC PDX timeseries treatment with cisplatin NRG mice of the same age and genotype as above were used for transplantation treatment experiments. Drug treatment with cisplatin (Platinum) was commenced when the tumour size reached approximately $300\,\mathrm{mm}^3$ to $400\,\mathrm{mm}^3$. Cisplatin (Accord DIN: 02355183) was administered intraperitoneally (IP) at $2\,\mathrm{mg\,kg}^{-1}$ every third day for 8 doses maximum (Q3Dx8). The dosage schedule was adjusted 50% less than what is mentioned in the literature [7, 8] and around one third of the maximum tolerated dose (MTD) calculated in the immunodeficient mice (**Supplementary Fig. 1**b). Low dose cisplatin pulse and tumour collection timings were optimized to achieve the experimental aims of tumour resistance. The aim was to collect the tumour at around 50% shrinkage (from the starting tumour at the time treatment started) in size when measured with a caliper. Cisplatin $1\,\mathrm{mg\,ml}^{-1}$ was diluted in 0.9% NaCl to obtain concentrations $200\,\mu l/20\,\mathrm{g}$ of mouse weight and kept in glass vials at room temperature. Quality control (QC) samples were prepared freshly on each day prior to the dosing. Mice were continually monitored for acute

signs of toxicity including pain at injection site, skin tenting, coat scruffing, sunken eyes, food consumption and behaviour for the first two hours following compound administration. For all three Tp53 mutant, TNBC PDX (**Supplementary Table 6**). , 8 mice at initial passage were transplanted in parallel for the treatment/treatment holiday study group. Half of the mice were treated with cisplatin when tumours exhibited $\sim 50\%$ shrinkage, the residual tumour was harvested as above and re-transplanted for the next passage in the group of eight mice. Again, half of the mice at X5 were kept untreated while the other half were exposed to cisplatin following the same dosing strategy. Four cycles of cisplatin treatment were generated, with a parallel drug holiday group at each passage. Cisplatin treated tumours were coded as *UT, UTT, UTTT, UTTTT* for each of the four cycles of drug respectively, while the tumours on drug holiday were labelled as *UTU, UTTU* and *UTTTU* for the three timepoints. The number of *T*s in the coded label shows the number of cycles of drug exposure. scDNAseq and scRNAseq was carried out on each tumour during the timeseries treatment with counterpart drug holiday and untreated controls.

In particular, TNBC-SA609 PDX was processed to establish 5 independent lines to explore the biological and technical replicate tumours as well as treatment series. All five lines from TNBC-SA609 were passaged identically after initial establishment. Line 1 untreated samples were seeded (X3 to X4) from a freshly dissociated tumour, whereas 4 other lines (all treated and line 2 Un-treated) were seeded (X3 to X4) from a frozen vial of tumour. Technical replicates were collected and sequenced for lines 1, 2.

3 Single cell whole genome sequencing and library construction with DLP+

All libraries, including metrics on number of cells, average number of reads per cell and quality control metrics are listed in **Supplementary Table 1**.

3.1 Creation of single cell suspensions from PDX Tumour fragments from PDX samples were incubated with collagenase/hyaluronidase, 1:10 (10X) enzyme mix (STEM CELL technologies, Catalog #07912) in $5\,\mathrm{ml}$ DMEM/F-12 with Glucose, L-Glutamine and HEPES (Lonza 12-719F) and 1%BSA (Sigma) at $37\,^{\circ}\mathrm{C}$. Intermittent gentle pipetting up and down was done every 30 min for 1 min, during the first hour with a wide bore pipette tip, and every 15-20 min for the second hour, followed by centrifugation (1100 rpm, 5 min) and supernatant removal. The tissue pellet was resuspended in $1\,\mathrm{ml}$ of 0.25 percent trypsin-EDTA (VWR CA45000-664) for 1 min, superadded by $1\,\mathrm{ml}$ of DNAse 1/dispase ($100\,\mu\mathrm{l}/900\,\mu\mathrm{l}$), (StemCell 07900,00082462) pipetted up and down 2 min, followed by neutralization with 2% FBS in Hanks' Balanced Salt Solution (HBSS) with 10 mM HEPES (STEMcells Catalog #37150). This cell suspension was then passed through a $70\,\mu\mathrm{m}$ filter to remove remaining undigested tissue and centrifuged for 5 min at 1100 rpm after topping it up to $5\,\mathrm{ml}$ with HBSS. Single cells pellet was resuspended in 0.04% BSA (Sigma) and PBS to achieve $\sim$ 1 million per ml concentration of cells for robot spotting for DLP+.

3.2 Robot spotting of single cells into the nanolitre wells and library construction scDNAseq DLP+ library construction was carried out as described in [1]. Briefly, single cell suspensions from cell lines and patient derived xenografts were fluorescently stained using CellTrace CFSE (Life Technologies) and LIVE/DEAD Fixable Red Dead Cell Stain (ThermoFisher) in a PBS solution containing 0.04% BSA (Miltenyi Biotec 130-091-376) incubated at $37\,^{\circ}\mathrm{C}$ for 20 minutes. Cells were subsequently centrifuged to remove stain, and resuspended in fresh PBS with 0.04% BSA. This single cell suspension was loaded into a contactless piezoelectric dispenser (sciFLEXARRAYER S3, Scienion) and spotted into the open nanowell arrays (SmartChip, TakaraBio) preprinted with unique dual index sequencing primer pairs. Occupancy and cell state were confirmed by fluorescent imaging and wells were selected for single cell copy number profiling using the DLP+ method [1]. Briefly, cell dispensing was followed by enzymatic and heat lysis. After cell lysis, tagmentation mix ($14.335\,\mathrm{nL}$ TD Buffer, $3.5\,\mathrm{nL}$ TDE1, and $0.165\,\mathrm{nL}$ 10% Tween-20) in PCR water were dispensed into each well followed by incubation and neutralization. Final recovery and purification of single cell libraries was done after 8 cycles of PCR. Cleaned up pooled single-cell libraries were analyzed using the Aglient Bioanalyzer 2100 HS kit. Libraries were sequenced at UBC Biomedical Research Centre (BRC) in Vancouver, British Columbia on the Illumina NextSeq 550 (mid- or high-output, paired-end 150-bp reads), or at the GSC on Illumina HiSeq2500 (paired-end 125-bp reads) and Illumina HiSeqX (paired-end 150-bp reads). The data was then processed to a quantification and statistical analysis pipeline [1].

4 DLP+ sequence analysis, copy number determination and quality control filtering

FASTQ pre-processing, sequence alignment, quality control, copy number calling and S-phase classification and filtering was performed on all libraries as detailed in [1]. Briefly, cells were assigned a QS score for data quality based on a 13 feature Random forest classifier fitted and applied as per [1]. Copy number alterations on a per cell basis were determined using a hidden Markov Model (HMM) approach using the `HMMCopy` package with parameterizations detailed in [1]. S-phase cells were identified using an automated classifier trained using cell-cycle sorted cells and where features from HMM output were used to identify cells most probably in early or late phase replication of their genomes. As these cells interfere with downstream phylogenetic analysis, they were removed from the analysis according to parameter settings and thresholds established in [1]. Doublets comprise less than 1-3% of cells after filtering. To further enable phylogenetic inference, 10-15% of cells with highest average copy number (CN) state jumps were removed. Upon inspection, these cells included early and late dividing cells that were not captured by the s-phase classifier. Host mouse infiltrating stromal cells were filtered from the PDX libraries based on proportion of reads aligning to the mouse genome. Mouse cells were filtered from xenograft libraries using fastqscreen (`https://www.ncbi.nlm.nih.gov/pubmed/30254741.2`). Fastqscreen was used to determine, for each read, whether that read originated from human, mouse and salmon DNA. Cells for which 5% or more reads mapped exclusively to either mouse or salmon were determined to be non-human or doublets and were excluded from further analysis. Attrition of cells at each stage of quality control is shown in **Supplementary Table 1**.

4.1 Phylogenetic tree inference, clone determination and clonal abundance measurements We developed a single cell Bayesian tree reconstruction method based on copy number change point binary variables called `sitka` [9] to fit phylogenetic trees to the copy number profiles. In the output of `sitka`, cells are the terminal leaf nodes of the phylogenetic topology. The inferred trees were post-processed to identify clonal populations from major clades. With clonal populations defined, their abundances were counted as a function of timeseries and these were used for fitness inference (see below). Clones were constructed by identifying connected components (each a clade or a paraphyly) in the phylogenetic tree reconstruction. The tree was 'cut' into discrete populations according to the following procedure (see Algorithm 1). Let L be a set of loci and C a set of cells. Define $\tau = (L, C, E)$ to be a rooted phylogenetic tree with E its set of directed edges. The phylogenetic tree is assumed to comprise internal nodes that are phylogenetic markers (loci) and terminal nodes that are either cells or loci. Terminal loci are considered unused phylogenetic markers and are discarded. Let $|\tau| = |C|$, that is the number of cells that belong to tree τ. By $\tau_l = (L_l, C_l, E_l)$ denote the subtree rooted at node l. Set pa(l) and child(l) be the parent node and set of immediate children of node l with desc(l) comprising all its descendants.

The inputs to the algorithm are the rooted phylogenetic tree τ and the copy number states of its cells and the minimum M_{min} and maximum M_{max} allowed clone sizes. A clone is defined as connected components (each a clade or a paraphyly) in the graph tree τ composed of cell of sufficient genomic homogeneity. The degree of homogeneity can be tuned by limiting the number of loci and the difference in copy number of sub-clades in a clone. The algorithm works by first finding the coarse structure, that is dividing the tree into major clades and then looking for fine structures within each clade by traversing the tree in a bottom up manner and merging loci that are sufficiently similar. The remaining loci constitute the roots of detected clades.

To obtain the coarse structure from the reconstructed phylogenetic tree we use a two step procedure. (i) First we identify monophyletic clades via 1. (ii) We then remove the cells comprising the clades found in step one from the tree and repeat algorithm 1. We note that these new clades (if any) could be paraphyletic.

To find the fine structures within the initial clades we use the following procedure. For each clade and its corresponding sub-tree τ_s, denote by L_s a set of loci l for which $M_{\text{min}} \leq |t_l| \leq M_{\text{max}}$. In a bottom-up traverse of the tree, for each node $l \in L_s$, remove it from L_s if $|\tau_{\text{pa}(l)}| - |\tau_l| \leq M_{\text{diff}}$, otherwise remove τ_l from τ_c. M_{diff} is a user-defined constant that intuitively results in merging clades that are too similar. At the end of the tree traversal set L_s contains new candidate roots for each initial clade. For each $l \in L_s$ define the summary copy number profile as a vector whose ith element is the median of the copy number states of the ith bin for all cells in τ_l. Compute the distance between two subclades as the mean absolute difference of their median genotypes (**Supplementary Fig. 2**). Merge subclones induced by L_s if their summary copy number profiles are too similar. We can do this by computing a t-test over the pairwise distances to exclude

outlier subclades and merge the rest. For the cell lines datasets, namely *p53 WT* and *p53-/- a* and *p53-/- b*, we opted to also split clades by the ploidy of their constituent cells, where ploidy is defined as the most recurrent CN state in the cell.

Once clones are identified, we set the abundance of each clone at a specific timepoint as the fraction of cells in that clone from that timepoint. We note that for the data from WGS bulk sequencing [3] we used the following procedure to estimate clonal fractions: (i) let ν denote the mutational cellular prevalence (rows) estimated over multiple timepoints (columns) using the multi-sample `PyClone` [10] model, (ii) define β as the genotype matrix (which mutation-cluster (rows) is present in which clones (columns)), (iii) then we set $\beta\gamma = \nu$ where $\gamma = \beta^{-1}\nu$ are the clonal fractions over time, and (iv) we solve for γ using QR-decomposition.

Algorithm 1 Heuristic (Top-down)

procedure SITKA-LUMBERJACK(τ, M_{min}, M_{max})
  locus-queue $\leftarrow$ DepthFirstSearch(τ)
  for locus l in locus-queue **do**
    if $m \leq |\tau_l| \leq M$ **then**
      $CUTS \leftarrow CUTS \cup l$
      Remove all loci below l from τ
    else
      for locus l' in descendants(l) **do**
        if $M_{\text{min}} \leq |t_{l'}| \leq M_{\text{max}}$ **then**
          $CUTS \leftarrow CUTS \cup l'$
          Remove all loci below l' from τ
      if no eligible loci were found **then**
        $CUTS \leftarrow CUTS \cup l$
        Remove τ_l from τ
  return $CUTS$

4.2 Single nucleotide variant analysis from DLP+ Single nucleotide variants (SNVs) were identified using aggregated data in which all cells were collapsed into a pseudo-bulk genome. MutationSeq and Strelka were run to identify SNVs from the pseudo-bulk sample [11, 12] as per [1]. To reduce false positives we filtered mutations with MutationSeq probability ¡ 0.9 and Strelka's Phred quality score ¡ 20 and then took the intersection of these two sets of mutations. With this set of high confidence SNVs we then genotyped each cell for every mutation, counting the number of reads supporting the reference and alternate allele. We then performed a final filtering operation, removing SNVs found in only one cell.

4.3 Mutation burden analysis To assess the differences in mutation burden within and between clones we counted the number of SNVs identified per cell as well as the number of copy number breakpoints (i.e., segments of different copy number). The number of copy number breakpoints was defined as the number of intra-chromosomal change-points identified by the hidden Markov Model copy number calling algorithm. SNVs were normalized by the coverage breadth per cell.

Clone specific CNAs could be corroborated by SVs 64% of the time in cell populations ¿ 100 cells. However, power to detect SVs is impacted by clone coverage as a function of the number of cells per clone and the total aggregate coverage (**Supplementary Fig. 3**). As such, additional granularity of clonal structure to further distinguish CN clones by SVs is limited by the combined per cell coverage and number of cells/clone. Further studies will be aimed at deciphering how depth of coverage can improve phylogenetic inference with other event types such as SNVs and SVs.

5 Fitness modeling

We describe in this section two approaches to fitness modeling. The first is a Bayesian state-space model (`fitClone`) based on the Wright-Fisher diffusion with selection (Section 5.1). The second is a deterministic logistic growth model (Section 5.9). We compare the two methods in Section 5.10. The comparison favours the Bayesian model, hence this is

the model used in our results unless we specify otherwise.

5.1 fitClone: a Bayesian fitness model for timeseries data We developed a Bayesian model and associated inference algorithm based on a diffusion approximation to K-allele Wright-Fisher model with selection. We start with timeseries clonal abundance measurements over a fixed number of clones and estimate two key unknown parameters of interest: *fitness coefficients* s_i for Clone i which represents a quantitative measure of the growth potential of a given clone; and *distributions over continuous-time trajectories*, a latent (unobserved) population structure trajectory in 'generational' time.

After briefly reviewing and setting notation for Wright-Fisher diffusions with selection (Section 5.2), we introduce the Bayesian model we used to infer quantitative fitness of clones from timeseries data (Section 5.3). We then describe our posterior inference method (Section 5.4) and ancillary methods for effective population size estimation (Section 5.5), and reference clone selection (Section 5.8).

A key difference of `fitClone` with methods that use a transformation of allele fractions to infer the existence of clones and focus on attempting to infer dynamics from bulk sequencing and single time points (e.g., the method of [13]) is that the inputs and outputs are fundamentally different, addressing fundamentally non-overlapping analytical problems. In particular, (i) `fitClone` models explicitly defined clones and their timecourse data, and (ii)`fitClone` is a generative model which allows for forecasting and prediction.

See [14] for more background on the Wright-Fisher model and [15, 16, 17, 18, 19, 20, 21, 22, 23] for previous work on inference algorithms for Wright-Fisher models.

5.2 Wright-Fisher diffusions with selection Let K denote the number of clones obtained using the tree cutting procedure described in Section 4.1, and denote by $Z_t = (Z_t^1, \ldots, Z_t^K)$ the relative abundance of each of the K clones at time t in the population. The process Z_t satisfies, for all t, the constraints $\sum_{i=1}^K Z_t^i = 1$ and $Z_t^i \geq 0$ for $i \in \{1, \ldots, K\}$. We would like to model the process Z_t using a Wright-Fisher diffusion with selection.

A Wright-Fisher diffusion can be written in stochastic calculus notation as

$$dZ_t = \mu^{s,N_e}(Z_t)\mathrm{d}t + \sigma(Z_t)\mathrm{d}W_t, \tag{1}$$

where $\{W_t\}$ is a K-dimensional Brownian motion, and the functions μ and σ, defined below, respectively control the deterministic and stochastic aspects of the dynamics. For $z = (z^1, z^2, \ldots, z^K)$, the vector-valued function $\mu^{s,N_e} : \mathbb{R}^K \to \mathbb{R}^K$ is defined as

$$\begin{aligned}\mu^{s,N_e}(z) &= (\mu_1^{s,N_e}(z), \ldots, \mu_K^{s,N_e}(z)) \\ \mu_i^{s,N_e}(z) &= N_e z^i (s_i - \langle s, z\rangle),\end{aligned}$$

where $\langle x, y\rangle$ is the inner product of vectors x and y, N_e, the *effective population size*, discussed in more details in Section 5.5, and the parameters $s = (s_1, s_2, \ldots, s_K)$ are called *fitness coefficients*. The interpretation of the fitness parameters is that if $s_i > s_j$, then subpopulation i has higher growth potential compared to subpopulation j. The matrix-valued function $\sigma : \mathbb{R}^K \to \mathbb{R}^{K^2}$ is defined as

$$\begin{aligned}\sigma^2(z) &= [\sigma_{i,j}^2(z)]_{i,j\in\{1,\ldots,K\}} \\ \sigma_{i,j}^2(z) &= z^i(\delta_{i,j} - z^j),\end{aligned}$$

where $\delta_{i,j}$ is the Kronecker delta. Given an initial value z, we denote the marginal distribution of the process at time t by $Z_t \sim \mathrm{WF}(s, N_e, t, z)$.

5.3 The fitClone model Given as input timeseries data measuring the relative abundances of K populations at a finite number of timepoints, the output of the `fitClone` model is a posterior distribution over the unknown parameters of

interest: the fitness parameters s described in the previous section, and the continuous-time trajectories interpolating and extrapolating the discrete set of observations.

To do this, `fitClone` places a prior on the fitness parameters s, and uses a state space model in which the latent Markov chain is distributed according to a Wright-Fisher diffusion, and the observation model encodes noisy sampling from the population at a discrete set of timepoints.

Each component of the fitness parameter, now a random variable S_i, is endowed with a uniform prior over a prior range I,

$$S_k \sim \text{Uniform}(I), k > 1,$$

where we set $S_1 = 0$ to make the model identifiable (see Section 5.8 for details). We used $I = (-10, 10)$ in our experiments. Note that the posterior is contained far from the boundaries of this prior range in all experiments.

The initial distribution, i.e. the distribution of the value of the process at time zero, is endowed a Dirichlet distribution with hyper-parameter $(1, 1, \ldots, 1)$,

$$Z_0 \sim \text{Dirichlet}(1, 1, \ldots, 1).$$

This can equivalently be seen as a uniform distribution over the K-simplex.

Let $t_1 < t_2 < \cdots < t_{T-1} < t_T$ denote a set of process times at which measurements are available. Ideally, we would like the latent transition kernels to be given by the marginal transitions of the Wright-Fisher diffusion from last section,

$$Z_{t_m} | Z_{t_{m-1}}, S \sim \text{WF}(S, N_\text{e}, t_m - t_{m-1}, Z_{t_{m-1}}), \tag{2}$$

where N_e is estimated as a pre-processing step (Section 5.5). In practise we resort to approximating the distribution in Equation (2) via a Euler-Maruyama scheme (Section 5.4).

Finally, for each $t \in \{t_1, t_2, \ldots, t_T\}$, let $Y_t = (Y_t^1, \ldots Y_t^K)$ denote a noisy observation of the population prevalences at process time t. In the single-cell context, this is obtained by counting, for each clone defined in Section 4.1, the number of cells coming from each passage, and normalizing by the number of cells sequenced in that passage. In the bulk sequencing context, see Section 4.1. For simplicity, in both cases we use a normal observation model, i.e., $Y_t^i \mid Z_t^i \sim \mathcal{N}(Z_t^i, \sigma^2_{\text{obs}})$, where $\sigma^2_{\text{obs}} = np_i\,(1 - p_i)$ and $n = \sum_j Y_t^j$ and $p_i = Y_t^i/n$..

5.4 Posterior inference under the fitClone model Since the marginal distributions of the Wright-Fisher diffusion do not admit closed form expressions, and previous work on exact simulation does not scale to high values of K, we resort to discretization using a Euler-Maruyama scheme [24]. For simplicity, we used the same number of grids between all observed time steps. We used a particle Markov chain Monte Carlo (pMCMC) method called Particle Gibbs with Ancestor Sampling and particle rejuvenation [25, 26] to sample from Z_{t_m} and the intermediate Euler-Maruyama steps, and a Metropolis within Gibbs sampler to sample the selection parameters S.

5.5 Estimating the effective population size Following [27] we use $F_s^{'}$ an unbiased moment-based estimator of the N_e where $N_\text{e} = \frac{1}{F_S^{'}}$; and t is the number of generations between each passage.

$$F_s^{'} = (1/t)\frac{F_s(1 - 1/(2\tilde{n})) - 1/\tilde{n}}{(1 + F_S/4)(1 - 1/n_y)}, \tag{3}$$

where $F_s = \frac{(x-y)^2}{z(1-z)}$ and $z = (x+y)/2$ and $\tilde{n} = \frac{2n_y n_x}{n_y+n_x}$, the harmonic mean of the sample size (initial population size at the passage) n_x and n_y at the two timepoints. x and y are the minor allele frequencies at the two timepoints.

In the multi-allelic case, we have:

$$F_s = \frac{1}{K}\sum_{i=1}^{K}\frac{(x_i - y_i)^2}{z_i(1-z_i)}.$$

This is equivalent to plan 2 in [27], sampling before reproduction and without replacement.

We used the sum of clone sizes as the approximate initial population size at each timepoint/passage. **Supplementary Table 3** lists the resulting N_e estimates. Since `fitClone` is robust to the choice of N_e in this range (**Supplementary Fig. 4**c), we set $N_e = 500.0$ for all datasets analysed in this paper. We note that in our model we assume that the effective population size remains constant over all timepoints. This does not take in account the potential changing population growth rate or the bottleneck effect due to passaging. These phenomena may scale the diffusion time and bias our estimates of evolutionary events including fixation or extinction times. This stretching and compressing of time could be accounted for by adding random effects to the number of generations in the model, for example by taking $N_{e,t}$ as a piece-wise constant random variable that can vary between passages. This is subject to our ongoing research.

Supplementary Table 3 shows the other parameters used in the inference over the real datasets.

5.6 Summarising the posterior distribution In all cases 10,000 particles and a burn-in equal to 10% of the MCMC samples were used. For the trajectories, we reported $\hat{z}_{1:T}$ where $\hat{z}_t = (\hat{z}_t^1, \ldots, \hat{z}^K)$ and $\hat{z}_t^k$ encodes the posterior clonal fraction of Clone k at time t.

The posterior of the selection coefficient vector was summarised by $\hat{s} = (\hat{s}_2, \ldots, \hat{s}_K)$ where $\hat{s}_k$ denotes the posterior mean of the selection coefficient of Clone k. To compare the selection coefficients of two clones we used a posterior ordering matrix $P_{(K-1)\times(K-1)}$ (**Supplementary Fig. 4**b,c). $P_{i,j} = P(s_i \leq s_j)$ shows the posterior probability that Clone i has higher selection coefficient than Clone j, with the stronger purple hues (close to 1.0) representing a higher confidence that clone i dominates Clone j, and conversely the stronger grey hues (close to 0.0) denote that Clone j dominates Clone i. Colours closer to white (0.5) represent no dominance. Note that for the lower diagonal elements $P(s_j \leq s_i) = 1 - P(s_i \leq s_j)$ and are omitted for clarity. The diagonal entries are to guide the eyes only.

5.7 Probability of positive selection Distribution of the probability of positive selection over pairs of clones was computed as $\max(P(s_i > s_j), 1 - P(s_i > s_j))$ for all pairs of clones i, j such that $i > i$. Let $s_{1:M} = (s_1, s_2, \ldots, s_M)$ be the M post burn-in MCMC samples for the selective coefficients where $s_m = \{s_{m,1}, s_{m,2}, \ldots, s_{m,K-1}\}$ are the sampled selective coefficients of clones 1 to $K-1$ at iteration m. Define $P(s_i > s_j) = \Sigma_{m=1}^{M}\mathbb{I}(s_{m,i} > s_{m,j})$ for $i, j \in \{1, \ldots, K-1\}$ be the posterior probability of Clone i having a larger coefficient than Clone j. We computed the effect size as the absolute value of the expected difference between the selection coefficients of clones i, j, that is $\big|E(s_i - s_j)\big| = \big|\Sigma_{m=1}^{M}(s_{m,i} - s_{m,j})\big|$.

5.8 Selecting the reference clone In our formulation of the Wright-Fisher diffusion one reference clone with selection coefficient of zero has to be chosen. The selection coefficient of the other clones are reported relative to this value. For instance, if the fittest clone is chosen as reference, the other clones will have negative selection coefficients. We chose to set the reference to a clone with an approximately monotonically decreasing trajectory (clonal abundance over time). This choice was motivated by a desire to infer a non-negative value for the fittest clones. **Supplementary Fig. 4**a,b shows that the model is robust to the choice of the reference clone. We run the inference procedure over the same dataset multiple times, each time changing the reference. The posterior ordering of clones over different choices of clones remained mostly identical.

5.9 The deterministic logistic growth model We developed a closely related population genetics model which incorporates selection via deterministic differential equations, but has closed form solutions. In this model [28], the solution of a deterministic DE, the frequency of each population c out of possible K populations, at time t, with selection coefficient s_c is proportional to its starting prevalence $p_0(c)$ multiplied by a power of the relative fitness coefficient $w_c = (s_c + 1)$,

$$f(c, t, s) = \frac{(s_c + 1)^t p_0(c)}{\sum_{k=1}^{K} (s_k + 1)^t p_0(k)}. \tag{4}$$

To estimate fitness coefficients $w_{1:T}$ from observed clonal fractions $Y_{1:T}$, we solve the optimization problem in Equation (5) using a limited memory Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimization procedure with box constraints [29]. Note that $f_{t,c}(w)$ is the clonal fraction estimates from Equation (4) and $Y_{t,c}$ is the observed clonal fraction for Clone c at time t.

$$\min_{w \in \mathcal{W}} (\Sigma_{t=1}^{T} d_t(w)), \tag{5}$$

where $d(w) = \big(d_1(w), d_2(w), \ldots, d_T(w)\big)$ is a vector valued function whose elements are $d_t(w) = \Big(\sum_{c=1}^{K} (f_{t,c}(w) - Y_{t,c})\Big)^{\frac{1}{2}}$ and $\mathcal{W} = \big(\mathcal{R} \cup \{0\}\big)^K$ and $w_1 = 1.0$.

5.10 Simulation benchmarking We forward simulated $L = 40$ datasets from the joint distribution of the Wright-Fisher model, with K = 4 (5 clones) and $L = 40$ datasets with K = 10 (11 clones). For each simulated dataset l, we sampled initial clonal abundance vector $Z_{1,l} \sim$ Dirichlet$(\alpha_{1:K})$ where $\alpha_i = 1$ and selective coefficients from a normal distribution with $s_{i,l} \sim$ Normal$(0.0, 0.3)$ truncated at $(-0.5, 1)$ for $i \in \{2, \ldots, K\}$ assuming the index of the reference clone is $i = 1$ and $s_1 = 0$. Discretisation constant (step size) $\Delta\tau$ = 0.001 and the standard deviation of the emission model was set to $\sigma_{obs,\text{simul}}$ = 0.001. At simulation and inference, we set $N_\text{e} = 500$. The simulation was continued to diffusion time of 0.1 after which 10 equidistant samples were recorded as observed values for the process. In all models except the Logistic growth, we put a uniform prior on each component of the s vector, that is, $s_i \sim$ Uniform$(-5.0, 5.0)$ and a Dirichlet prior on the initial clonal distributions. We set step size $\Delta\tau$ = 0.001, $\sigma_{obs,\text{infer}}$ = 0.01 and use 10,000 particles for 10,000 MCMC iterations. We run 5 different models on the simulated dataset as follows (**Supplementary Fig. 4**): (i) WFda that is the Wright-Fisher model with diffusion approximation. (ii) Logistic growth is the deterministic differential equation Wright-Fisher model. (iii) One-step is the identical to the standard model but applies no discretisation for the time between observations. (iv) Single Clone is the standard model with $K = 1$. See section 5.9 for how logistic model was fit. For the other models, we used the mean absolute error (MAE) of the the post burn-in mean posterior of marginal selection coefficients for each method and the selection coefficient used to generate the simulated data. For each dataset l, the Single Clone model is run K times, once for each Clone k, where its input consists in the observed clonal fraction of only one clone. For this model we report the averaged MAE across clones per dataset. The single clone model performs worst which suggests that treating each clone independently is sub-optimal.

We carried out 2 types of withholding analysis: leave one out (where single timepoint measures were withheld), and right censoring (where we progressively removed timepoints from the end of the timeseries). We applied these methods and quantified the trajectory predictions of the `fitClone` method over 3 cell lines and 6 PDX timeseries to estimate prediction error (**Supplementary Fig. 5**. As expected, the prediction error increases when the timeseries is more right-truncated at inference. We also performed interpolation leave-one-out prediction experiments and observed a prediction error limited to ¡ 0.1 in 85% and ¡0.05 in 53% of subsamplings (**Supplementary Fig. 5**a, b). In 69.62% of cases the removal of a timepoint increases the average CI width of inferred selection coefficients (**Supplementary Fig. 5**c).

5.11 Software and implementation The software implementation of `fitClone` is available at: [`https://github.com/UBC-Stat-ML/fitclone`]

6 Single cell RNA sequencing (scRNAseq)

All libraries generated using 10x scRNAseq are listed in **Supplementary Table 7**.

6.1 Processing of cell lines for scRNAseq data Suspensions of 184-hTERT *p53 WT* and KO cells were fixed with 100% ice-cold methanol prior to preparation for scRNAseq. Single cell suspensions were loaded onto the 10x Genomics single cell controller and libraries prepared according to the Chromium Single Cell 3' Reagent Chemistry kit standard protocol. Libraries were then sequenced on an Illumina Nextseq500/550 with 42bp paired-end reads, or a HiSeq2500 v4 with 125bp paired-end reads. 10x Genomics CellRanger, V3.0.2 (V3 chemistry), was used to perform demultiplexing, alignment and counting.

6.2 Processing of patient derived xenografts for scRNAseq data PDX tumours were harvested and mechanically disaggregated into small fragments to viably freeze them. One of the viable frozen tumour vial was thawed and after washing out the freezing media, the tumour clumps and fragments were incubated with digestion enzymes as with DLP+ single cells preparation (as above) and the cells were resuspended in 0.04% BSA in PBS. Dead cells were removed using the Miltenyi MACS Dead Cell Removal kit and cells were processed as previously described [30]. To avoid processing artifacts and dissociation methods, the times were tightly controlled for treatment and treatment holiday pairs. Library construction of the samples at the same time point was performed on the same chips. Library construction sample batch groupings are listed in **Supplementary Table 7**.

We next profile the impact of clone specific gene expression changes as a higher order representation of phenotypic properties. We tested if the genotypes of high fitness clones exhibited changes in their transcriptional program, with scRNAseq performed on matched aliquots of samples sequenced using DLP+ (**Supplementary Table 7**, Supplementary Information). We applied a statistical model, clonealign[31], to map scRNAseq transcriptional profiles to their clone of origin, and to investigate gene dosage effects of copy number alterations on transcription. DLP+ and `clonealign` clone abundance measures were positively correlated across all libraries (Pearson correlation coefficient geq 0.76, p¡0.001, **Supplementary Fig. 6** a), consistent with clone specific gene dosage effects in the majority of libraries. Differentially expressed genes in regions of clone specific CNAs were strongly correlated with clone abundance (Pearson correlation coefficient was $r > 0.42$, $p < 0.023$, **Supplementary Fig. 6** b).

Transcriptional profiles from late timepoints (when high fitness clones had grown in abundance) indicated copy number driven gene expression in each of the *p53-/-a, p53-/-b* and TNBC series (**Supplementary Fig. 6** c). Examples from *in vitro* and PDX series showed between 1 and 52% of all differentially expressed genes have clone specific copy number differences (**Supplementary Fig. 6** c, FDR¡0.01). Together these data indicate that clonal genotypes driving high fitness trajectories are accompanied by changes in gene expression at both chromosomal and focal level copy number alterations.

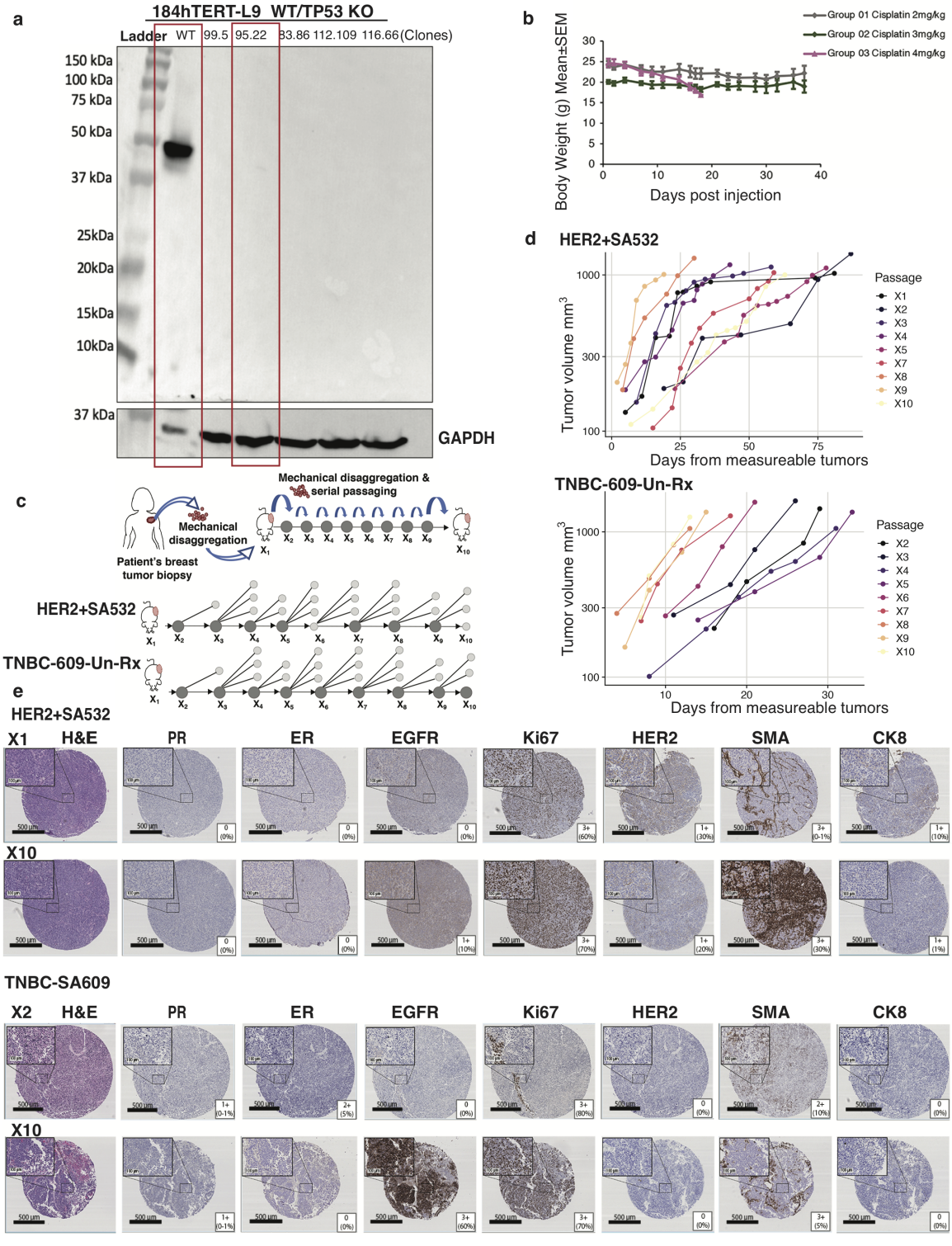
6.3 Quality control Count matrices were generated using CellRanger version 3.0.2 (V3 chemistry). Cells were considered to have passed a quality control filter (QC-filter) and retained for subsequent analysis if they met the following criteria: (i) at least 1000 genes detected, (ii) less than 20% of counts (UMIs) mapping to genes from the mitochondrial genome ("mitochondrial genes"), (iii) less than 60% of counts (UMIs) mapping to ribosomal genes, and (iv) the total counts (UMIs) per cell was at most 3 median absolute deviations lower than the overall median. Cells not matching all criteria were filtered using the `calculateQCMetrics` and `isOutlier` functions in the `scater` package [32]. Then, all mouse cells were eliminated. A cell is classified as a mouse cell if the total number of counts in a mouse alignment of the 10x sample was greater than the total number of counts in a human alignment. Finally, we eliminated doublets using package `scrublet` [33] (**Supplementary Table 7**).

6.4 Integrative genome-transcriptome analysis with `clonealign` `clonealign` version 1.99.2 was used to align scRNAseq cells to the DLP+ clones obtained, with the parameters described in Supplementary Table **8**. Manual tuning was necessary for some of the parameters due to the high heterogeneity of our samples, as follows: (1) Purity of a gene in a clone, defined as the percentage of cells that have the modal copy number for that gene and clone, was set to 0.4-0.6, depending on the observed variability of copy number values across cells in the DLP+ data (0.4 for larger variability and 0.6 for lower variability). (2) `initial_shrink`, defined as the strength with which the variational parameters for clone assignments are initially shrunk towards the most likely assignments, was set to 0 or 10. (3) `data_init_mu` was set to TRUE if the mu parameters need to be initialized using the data, or FALSE otherwise. Other parameters that had the same values for all libraries: `n_repeats = 3`, `mc_samples = 1`, `learning_rate = 0.07`, `max_iter =`

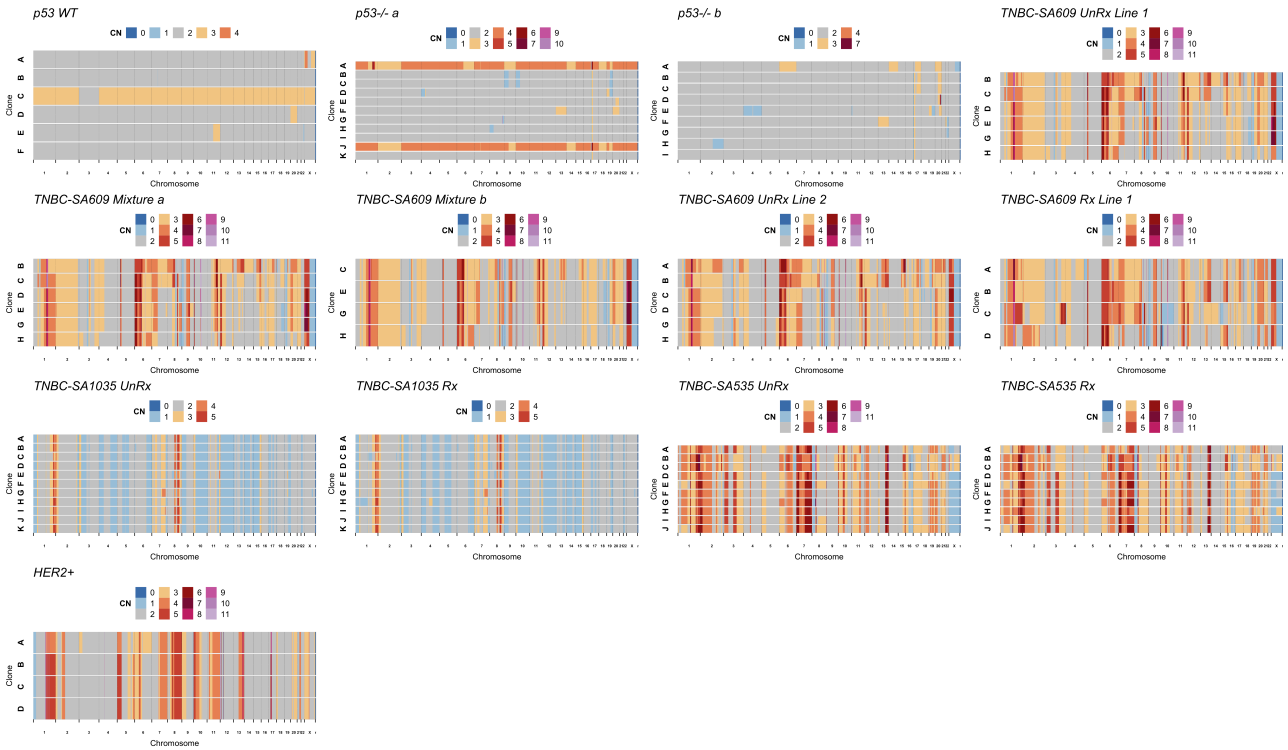
500, saturation_threshold = 6 and clone_call_probability = 0.9.

6.5 Differential expression analysis Differential expression quantifiers including $\log 2$ fold change and FDR were computed using the R 3.6.0 Bioconductor package edgeR_3.26.0 that implements scRNAseq differential expression analysis methodology based on the Negative Binomial distribution. Given the raw counts for the cells in two clones, we first call the estimateDisp() function to estimate the dispersion by fitting a generalized linear model that accounts for all systematic sources of variation. Next, we use the edgeR functions glmQLFit() and glmQLFTest() to perform a quasi-likelihood dispersion estimation and hypothesis testing that assigns false discovery rate values to each gene. In the track scRNAseq plots, a positive $\log 2$ fold change value for Clone X relative to Clone Y signifies that the gene is significantly more upregulated (at a given FDR threshold) in X than in Y while taking into consideration all the expression values for all the genes in both clones. Similarly, a gene with negative $\log 2$ fold change is significantly more downregulated in X than in Y.

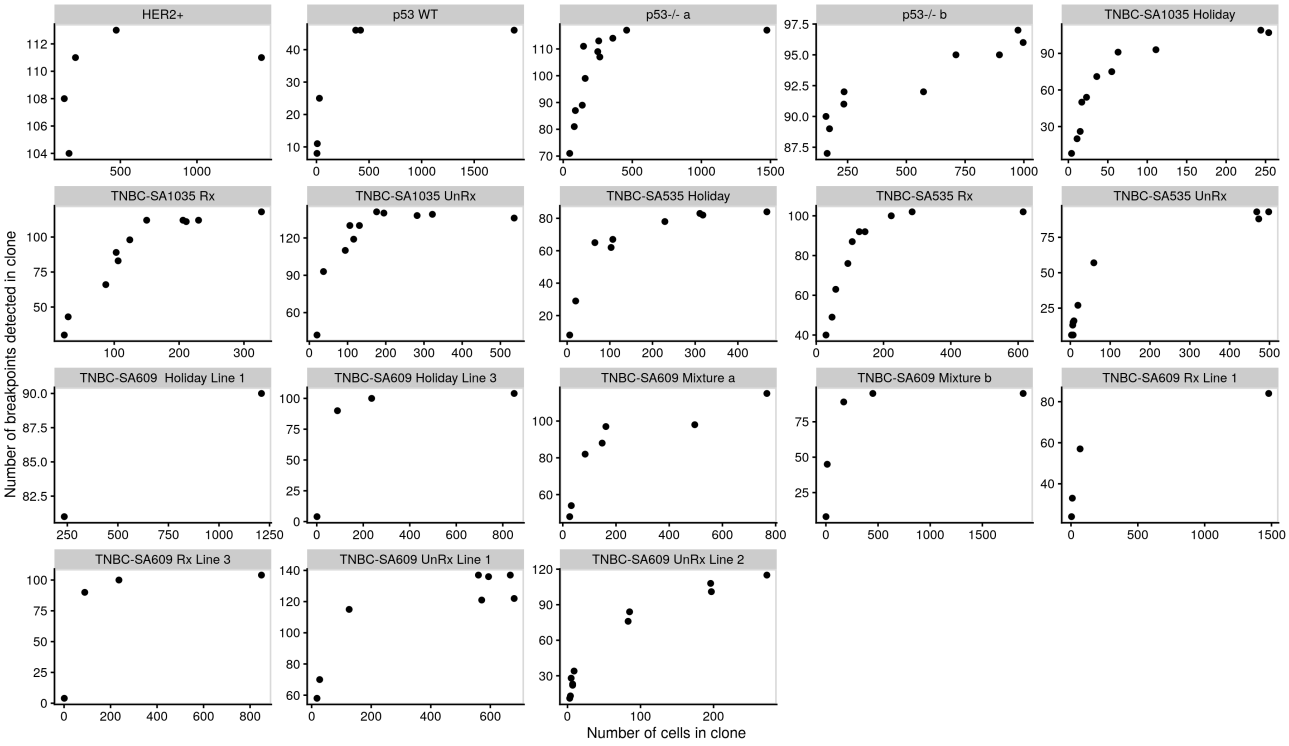
7 Supplementary figures



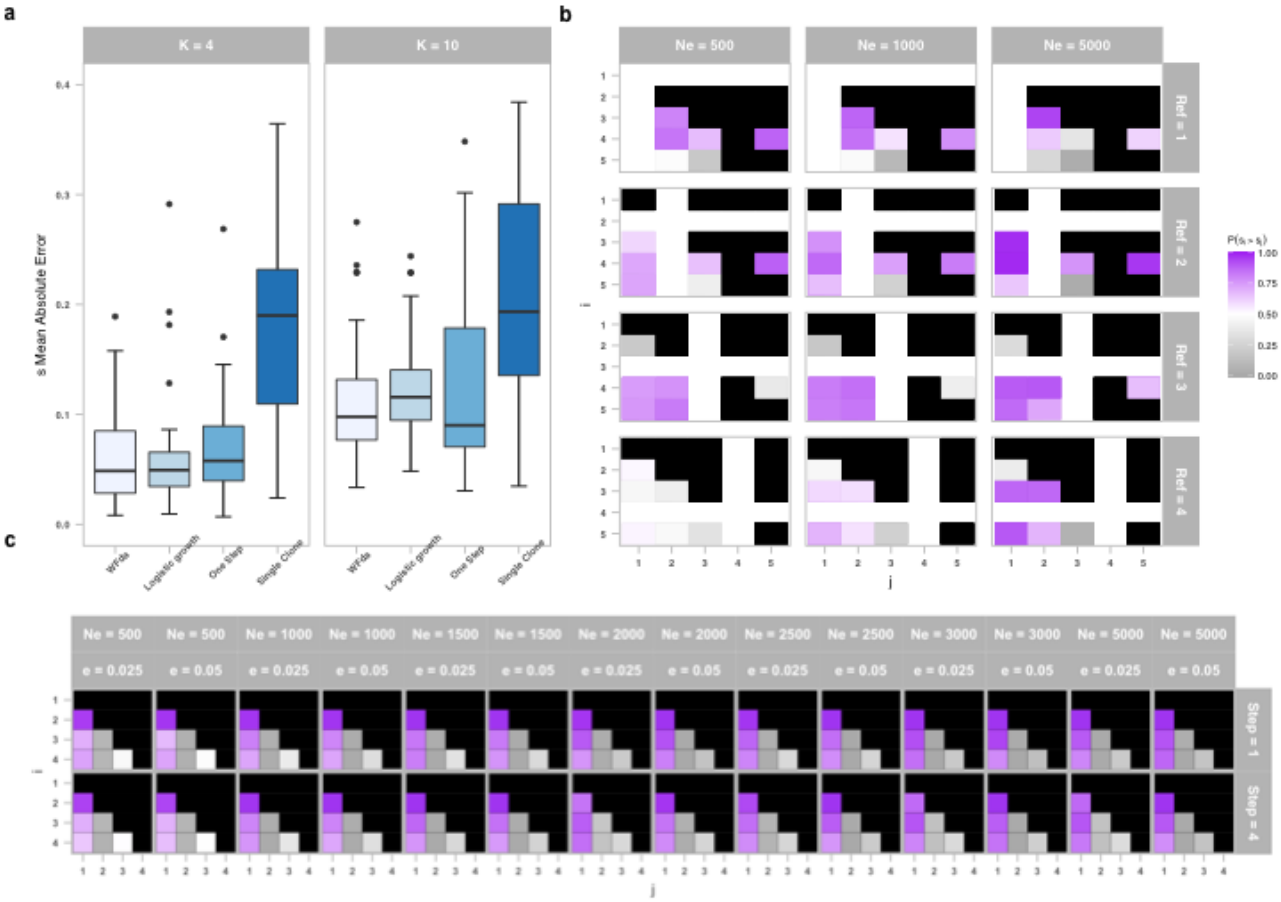
Supplementary Figure 1 *(previous page)* Overview of substrates and PDX growth rates curves. **a)** Western blot confirming absence of TP53 from 184-hTERT wildtype cell line (n=3). Clones shown along the top. **b)** Mouse body weight graph recorded during maximum tolerated dose (MTD) evaluation of cisplatin in NRG mice (n=3 in each study cohort). Error bars represent one standard deviation from the mean. **c)** (top) Schematic for PDX timeseries; (bottom) serial sampling of HER2+SA532 and TNBC-SA609 PDX tumours; dark grey circles represent each sampled mouse for scDNAseq. The light grey circles: technical replicates at the same timepoint. **d)** Representative individual tumour growth from each passage of TNBC-SA609 (X2-X10) and HER2+SA532 PDXs (X1 to X10). The vertical axis: the tumour volume, horizontal axis: days. **e)** IHC (n=3) of HER2+SA532 and TNBC-SA609 tumours at early and late passages, 4x and 20x (insets). Scale bars 500 µm and 100 µm (insets). Antibodies and TMA scores (**Supplementary Tables 4** and **5**).



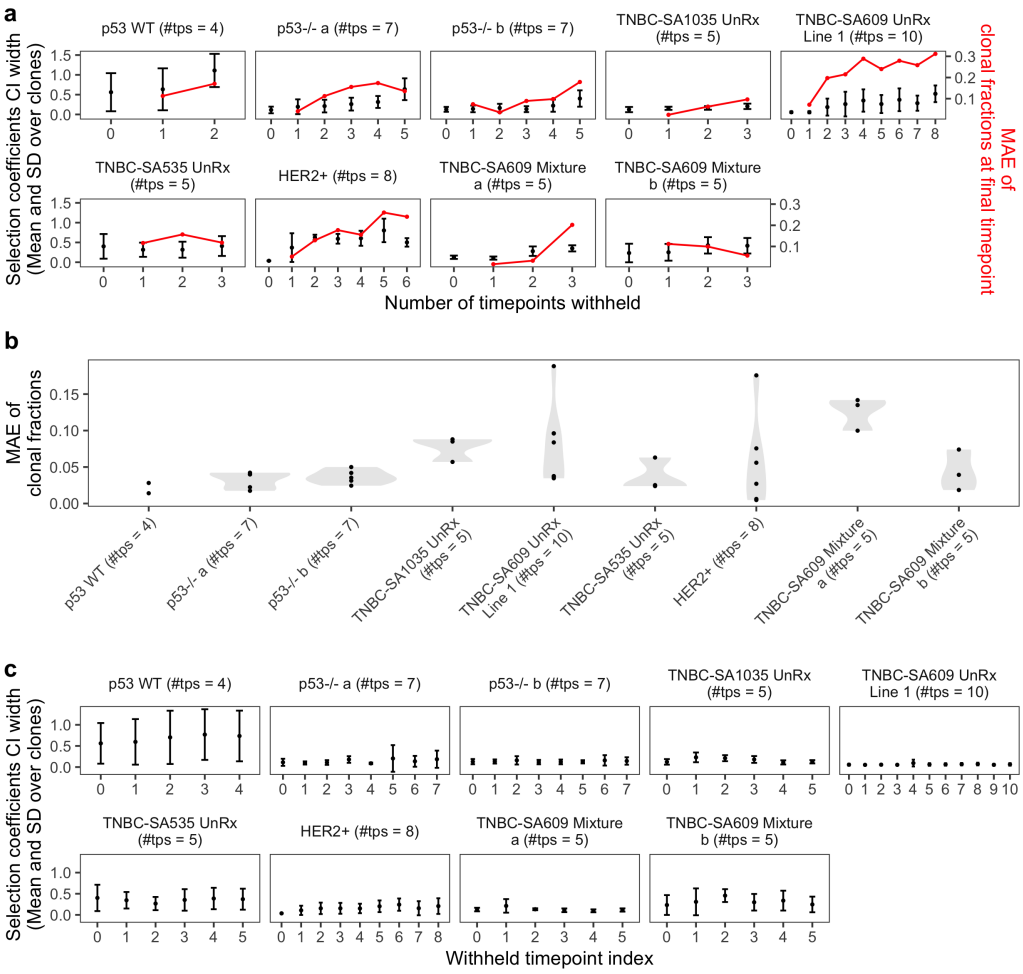
Supplementary Figure 2 The heatmap representation of per clone median genotypes for the 184hTERT cell lines *p53 wildtype (WT)* and *p53-/- a* and *p53-/- b* and the TNBC PDX series TNBC-SA609 UnRx Line 1, TNBC-SA609 mixtures a and b, SA609 UnRx Line 2, SA609 Rx line 1, TNBC-SA1035 and TNBC-SA535 untreated and treated (Un and Rx), as well as the HER2+SA532 PDX.



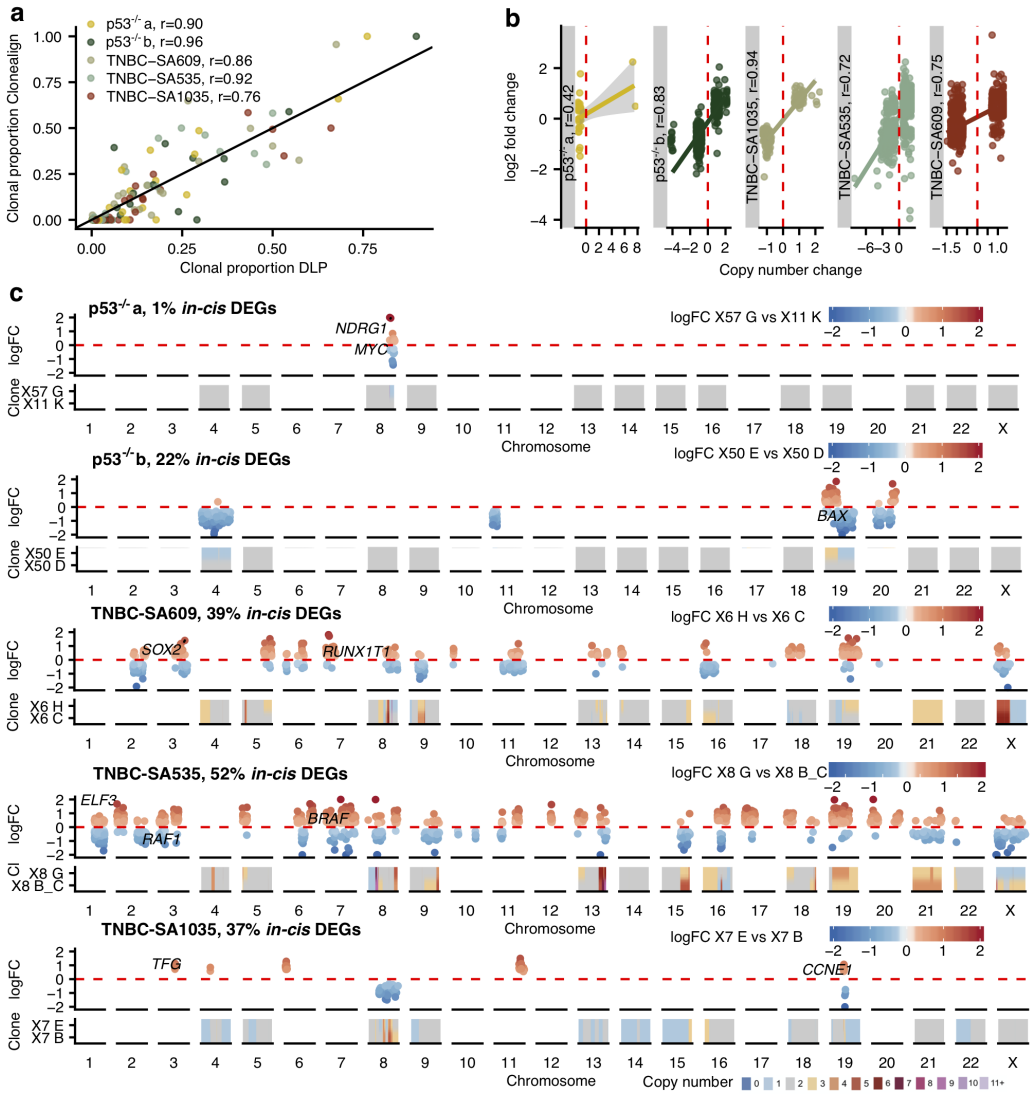
Supplementary Figure 3 Relationship between the number of cells present in a clone against the number of breakpoints detected in that clone for each dataset.



Supplementary Figure 4 Simulation studies for the `fitClone` model. **a)** Comparison to baseline methods for K = 5 clones (left) and K = 11 clones (right). **b)** Posterior ordering of clones based on their inferred posterior selection coefficients across three values of effective population size (columns) and different choice of the reference clone (rows). **c)** Posterior ordering of clones across different hyper parameters in the `fitClone` model. Effective population size in the range of 500 to 5,000 (top column), and observation error (bottom column). Minimum number of interpolations between two observations (rows). In the boxplots, the lower and upper whiskers, and the lower and upper hinges represent the -1.5×IQR and 1.5×IQR, 25th and 75th percentiles respectively, while the white line indicates the median of the distribution, and IQR is the interquartile range.



Supplementary Figure 5 Robustness analysis in fitClone using cell lines and untreated timeseries. **a)** Credible Interval width (the dots and bars show the mean and standard deviation of the CI width respectively; n=#clones used to compute the statistic is indicated on the top of each subpanel) of inferred selection coefficients (left axis - in black) and the prediction error at the last observed timepoint over increasingly shorter (right-censored) timeseries (right-axis in red). **b)** Interpolation leave-one-out error over possible timepoint indices. **c)** Interpolation leave-one-out Credible Interval width for selection coefficients averaged over clones. The dots and bars show the mean and standard deviation CI width respectively (n=#clones used to compute the statistic is indicated on the top of each subpanel). Timepoint index zero represents the full (uncensored) timeseries. CI: credible interval; MAE: mean absolute error.



Supplementary Figure 6 Transcriptome data analysis. **a)** Clonal proportions of scRNAseq derived from `clonealign` estimates against CNA derived clones: two-sided Pearson's product-moment correlation coefficient for each series was $r = 0.77$ or more, $p = 1e-5$ or less, 19 libraries, 46 clones in total. **b)** The $\log 2$ fold change of differentially expressed genes in panel **c)** in regions of clone specific CNAs: two-sided Pearson's product-moment correlation coefficient was $r = 0.72$ or more, $p = 1e-15$ or less for each series except $p53^{-/-}a$, for which $r = 0.42$, $p = 0.023$. **c)** Differential gene expression between the most fit clone and another abundant clone in the same sample, in the areas of CN change (for $p53^{-/-}a$, we used different samples for consistence of genome ploidy). In each panel, the top plot shows $\log 2$ fold change, with red showing higher log fold change in the top clone and blue showing higher log fold change in the bottom clone. The bottom plot shows the median CN of each gene from the DLP+ data, in the rank order of appearance in the genome. The percentage of *in cis* differentially expressed genes (DEGs) is calculated as the percentage of DEGs that have a corresponding change in CN, out of all the DEGs considered.

8 Supplementary tables

Supplementary Table 1 DLP+ libraries generated for $p53^{wt}$, $p53^{-/-}a$, $p53^{-/-}b$, HER2+, TNBC-SA609, TNBC-1035, and TNBC-535 timeseries indicating timepoint, number of cells (initially and after different QC levels), median reads per cell and quality metrics.

Supplementary Table 2 Clones and their respective $1+s$ coefficients inferred for WGS-bulk-TNBC, $p53^{wt}$, $p53^{-/-}a$, $p53^{-/-}b$, HER2+, TNBC-SA609, TNBC-1035, and TNBC-535 timeseries indicating timepoint, clonal abundance (frac), number of cells (if applicable), and mean, median and standard deviation of selective coefficients.

Supplementary Table 3 Real data parameters, estimated effective population size and number of SNVs. rej_rate stands for rejection rate.

Supplementary Table 4 List of antibodies and experimental conditions. Summary of antibodies clones and their suppliers used for staining TMAs for IHC and performing western blots for $p53^{wt}$and TP53 null. **(*RTU: Ready to use; N/A: not applicable)**

.

Supplementary Table 5 IHC Scores of un-treated and treated a time series tissue microarray (TMA). HER2+ and TNBC PDX serial passages and cisplatin treated serial passages on TMA with scores of various stains

Supplementary Table 6 The TP53 mutation and its variant allele frequency in the four PDX series, TNBC-SA1035, TNBC-SA535, TNBC-SA609, and HER2+

Supplementary Table 7 10x scRNAseq libraries indicating dataset, sample name, time point, initial number of cells, number of cells after three quality control measures and mean, standard deviation and median of counts in the entire library (entries with counts < 100 were eliminated from the mean, standard deviation and median metrics).

Supplementary Table 8 Parameters used as input in our `clonealign` analyses, including default parameters for most of the libraries and specific parameters for the remaining libraries.

References

1. Laks, E. *et al.* Clonal decomposition and DNA replication states defined by scaled Single-Cell genome sequencing. *Cell* **179**, 1207–1221.e22 (2019).

2. Burleigh, A. *et al.* A co-culture genome-wide RNAi screen with mammary epithelial cells reveals transmembrane signals required for growth and differentiation. *Breast Cancer Res.* **17**, 4 (2015).

3. Eirew, P. *et al.* Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature* **518**, 422–426 (2015).

4. Pearson, T. *et al.* Non-obese diabetic–recombination activating gene-1 (nod–rag 1 null) interleukin (il)-2 receptor common gamma chain (il 2 rγnull) null mice: a radioresistant model for human lymphohaematopoietic engraftment. *Clinical & Experimental Immunology* **154**, 270–284 (2008).

5. Hather, G. *et al.* Growth rate analysis and efficient experimental design for tumor xenograft studies: Supplementary issue: Array platform modeling and analysis (a). *Cancer Informatics* **13**, CIN–S13974 (2014).

6. Ahn, S., Woo, J. W., Lee, K. & Park, S. Y. Her2 status in breast cancer: changes in guidelines and complicating factors for interpretation. *Journal of pathology and translational medicine* **54**, 34 (2020).

7. Li, D. *et al.* Enhanced tumor suppression by adenoviral pten gene therapy combined with cisplatin chemotherapy in small-cell lung cancer. *Cancer gene therapy* **20**, 251–259 (2013).

8. Wang, Y. *et al.* Klotho sensitizes human lung cancer cell line to cisplatin via pi3k/akt pathway. *PloS one* **8** (2013).

9. Dorri, F. *et al.* Efficient bayesian inference of phylogenetic trees from large scale, low-depth genome-wide single-cell data. *bioRxiv* (2020).

10. Roth, A. *et al.* Pyclone: statistical inference of clonal population structure in cancer. *Nature methods* **11**, 396–398 (2014).

11. Ding, J. *et al.* Feature-based classifiers for somatic mutation detection in tumour-normal paired sequencing data. *Bioinformatics* **28**, 167–175 (2012).

12. Saunders, C. T. *et al.* Strelka: Accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811–1817 (2012).

13. Williams, M. J. *et al.* Quantification of subclonal selection in cancer from bulk sequencing data. *Nat. Genet.* **50**, 895–903 (2018).

14. Tataru, P., Simonsen, M., Bataillon, T. & Hobolth, A. Statistical inference in the wright–fisher model using allele frequency data. *Systematic biology* **66**, e30–e46 (2017).

15. Beaumont, M. A., Cornuet, J.-M., Marin, J.-M. & Robert, C. P. Adaptive approximate bayesian computation. *Biometrika* **96**, 983–990 (2009).

16. Foll, M. *et al.* Influenza virus drug resistance: a time-sampled population genetics perspective. *PLoS Genet* **10**, e1004185 (2014).

17. Bollback, J. P., York, T. L. & Nielsen, R. Estimation of 2nes from temporal allele frequency data. *Genetics* **179**, 497–502 (2008).

18. Malaspinas, A.-S., Malaspinas, O., Evans, S. N. & Slatkin, M. Estimating allele age and selection coefficient from time-serial data. *Genetics* **192**, 599–607 (2012).

19. Ferrer-Admetlla, A., Leuenberger, C., Jensen, J. D. & Wegmann, D. An approximate markov model for the wright–fisher diffusion and its application to time series data. *Genetics* **203**, 831–846 (2016).

20. Beskos, A., Roberts, G. O. *et al.* Exact simulation of diffusions. *The Annals of Applied Probability* **15**, 2422–2444 (2005).

21. Pollock, M., Johansen, A. M., Roberts, G. O. *et al.* On the exact and ε-strong simulation of (jump) diffusions. *Bernoulli* **22**, 794–856 (2016).

22. Jenkins, P. A., Spano, D. *et al.* Exact simulation of the wright–fisher diffusion. *The Annals of Applied Probability* **27**, 1478–1509 (2017).

23. Blanchet, J. Exact simulation for multivariate itô diffusions (2017).

24. Malham, S. J. & Wiese, A. An introduction to sde simulation. *arXiv preprint arXiv:1004.0646* (2010).

25. Lindsten, F., Jordan, M. I. & Schön, T. B. Particle gibbs with ancestor sampling. *The Journal of Machine Learning Research* **15**, 2145–2184 (2014).

26. Lindsten, F., Bunch, P., Singh, S. S. & Schön, T. B. Particle ancestor sampling for near-degenerate or intractable state transition models. *arXiv preprint arXiv:1505.06356* (2015).

27. Jorde, P., Palm, S. & Ryman, N. Estimating genetic drift and effective population size from temporal shifts in dominant gene marker frequencies. *Molecular Ecology* **8**, 1171–1178 (1999).

28. Otto, S. P. & Day, T. *A biologist's guide to mathematical modeling in ecology and evolution* (Princeton University Press, 2011).

29. Nocedal, J. & Wright, S. *Numerical optimization* (Springer Science & Business Media, 2006).

30. O'Flanagan, C. H. *et al.* Dissociation of solid tumor tissues with cold active protease for single-cell rna-seq minimizes conserved collagenase-associated stress responses. *Genome biology* **20**, 1–13 (2019).

31. Campbell, K. R. *et al.* clonealign: statistical integration of independent single-cell RNA and DNA sequencing data from human cancers. *Genome Biol.* **20**, 54 (2019).

32. McCarthy, D. J., Campbell, K. R., Lun, A. T. & Wills, Q. F. Scater: pre-processing, quality control, normalization and visualization of single-cell rna-seq data in r. *Bioinformatics* **33**, 1179–1186 (2017).

33. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* **8**, 281—291 (2019).