Cancer Chemotherapy and Pharmacology (2019) 84:1167–1178
https://doi.org/10.1007/s00280-019-03942-y

ORIGINAL ARTICLE

A reliable method to determine which candidate chemotherapeutic drugs effectively inhibit tumor growth in patient-derived xenografts (PDX) in single mouse trials

Derek Gordon[1] · **David E. Axelrod**[2]

Received: 2 May 2019 / Accepted: 24 August 2019 / Published online: 12 September 2019

Abstract

Purpose We report on a statistical method for grouping anti-cancer drugs (GRAD) in single mouse trials (SMT). The method assigns candidate drugs into groups that inhibit or do not inhibit tumor growth in patient-derived xenografts (PDX). It determines the statistical significance of the group assignments without replicate trials of each drug.

Methods The GRAD method applies a longitudinal finite mixture model, implemented in the statistical package PROC TRAJ, to analyze a mixture of tumor growth curves for portions of the same tumor in different mice, each single mouse exposed to a different drug. Each drug is classified into an inhibitory or non-inhibitory group. There are several advantages to the GRAD method for SMT. It determines that probability that the grouping is correct, uses the entire longitudinal tumor growth curve data for each drug treatment, can fit different shape growth curves, accounts for missing growth curve data, and accommodates growth curves of different time periods.

Results We analyzed data for 22 drugs for 18 human colorectal tumors provided by researchers in a previous publication. The GRAD method identified 18 drugs that were inhibitory against at least one tumor, and 10 tumors for which there was at least one inhibitory drug. Analysis of simulated data indicated that the GRAD method has a sensitivity of 84% and a specificity of 98%.

Conclusion A statistical method, GRAD, can group anti-cancer drugs into those that are inhibitory and those that are non-inhibitory in single mouse trials and provide probabilities that the grouping is correct.

Keywords Drug screening · Chemotherapeutic drugs · Statistical classification · Finite mixture model · Antitumor

Introduction

Several model systems have been utilized for the pre-clinical evaluation of anti-cancer drugs. These include cell lines derived from multiple tissue types [1, 2], single tissue types grown in 2D or 3D cultures [3, 4], stem cell-derived organoids [5–7], genetically engineered mice [8, 9], cell-line derived mouse xenografts [10], and patient-derived tumor xenografts (PDX) [11, 12].

Although each model system has advantages and limitations, PDX models promise to offer several advantages. Among these are that the tumor cells grow in a mammalian host, and that the tumor response can be tested after a minimal number of passages to avoid tumor evolution [13]. The effect of candidate drugs on the inhibition of tumor growth can be monitored by measuring tumor volume over time. Also, possible adverse side effects on the health of the host can be monitored in live animals by measuring physiological parameters such as body weight.

PDX models have been developed to test the inhibitory effect of drugs against tumors from a specific tissue. Examples include tumors from lung [14], colon [15–18], liver [19], blood [20], breast [21], ovary, pancreas, and stomach.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00280-019-03942-y) contains supplementary material, which is available to authorized users.

✉ David E. Axelrod
axelrod@biology.rutgers.edu

[1] Department of Genetics and Human Genetics Institute, Rutgers, The State University of New Jersey, 145 Bevier Road, Piscataway, NJ 08854-8082, USA

[2] Department of Genetics and Rutgers Cancer Institute of New Jersey, Rutgers, The State University of New Jersey, 604 Allison Road, Piscataway, NJ 08854-8082, USA

Other studies have considered tumors from multiple different tissues [22, 23].

Commercial companies have developed xenograft models to test their own drugs that are in development, such as Novartis' Institute for Biomedical Research PDX Encyclopedia [23]. Other companies have developed xenograft models to test drugs on a contract basis, such as Charles River Laboratories [24], Jackson Labs [25], CrownBio [26], Champions Oncology [27], and XenTech [28].

Non-commercial PDX repositories include the following: the EuroPDX Consortium; IMODI Cancer, a consortium to develop new experimental models; PRoXe Public Repository of Xenografts; the National Cancer Institute's repository of patient-derived models (PDMR); PDX Net Consortium; and PPTC, the Pediatric Preclinical Testing Consortium of the Children's Oncology Group (COG) cell culture and xenograft repository. Continued activity in pre-clinical drug screening using PDX models is expected in this area since the National Institutes of Health announced funding for new centers to develop PDX models for the pre-clinical testing of drugs and for the coordination of the large-scale data analysis.

Inhibition of tumor growth by a drug in PDXs is typically determined in the following experimental design: a biopsy is engrafted in a mouse, and after it takes, portions of the tumor are implanted in 5–10 genetically identical mice treated with the same drug, and a similar number of mice not treated with the drug. Tumor volume measurements are made every few days for three or more weeks. Then the tumor growth inhibition is determined by comparing the drug-treated and non-drug-treated tumor volumes at one time point, or longitudinally at multiple time points.

Several formulae for calculating a difference, or ratio, at one time point have been proposed. These include: a tumor growth index [29–32]; percent tumor growth inhibition and percent tumor regression [33]; time to tumor failure [34]; and the Response Evaluation Criteria in Solid Tumors, RECIST [30, 35]. For experiments that record data at a single time point, a two-sample *t* test or an Analysis of Variance (ANOVA) (for more than two groups) is applied to the data to determine if there are differences in the sample means.

Also, tumor inhibition has been evaluated using longitudinal data rather than data recorded at just one time. Some published methods include: using data at two time points to calculate a tumor control index [36], comparing the areas under the growth curves [37], fitting non-parametric splines [38], fitting exponential curves [39–42], computing the BestAvgResponse [23], or applying classical mathematical models to describe tumor growth [42].

An alternative to the experimental design that uses multiple mice is the single mouse trial (SMT) (Charles River Oncotest [24]) (Fig. 1). This type of trial also goes by the names "Clinical Mouse Trial" (CrownBio) [26] and "1 × 1 × 1" (one mouse, one drug, one tumor type) (Novartis)

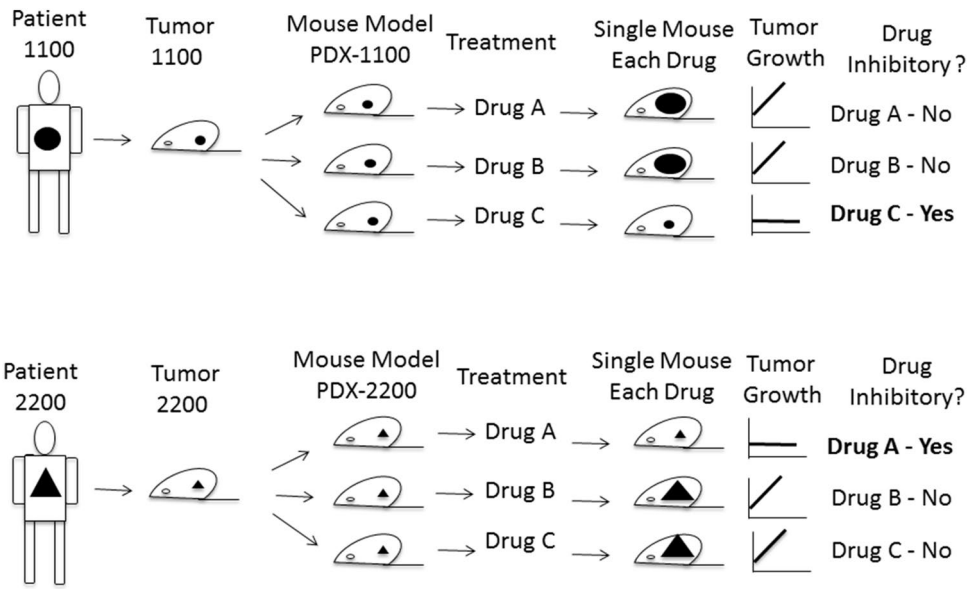

Fig. 1 Drug screening using patient-derived tumor xenografts (PDX) in single mouse trials. For each patient (patient 1100 or patient 2200), a tumor (represented by a black circle or black triangle) is removed and engrafted in a mouse. After the tumor has grown in the mouse, portions are implanted in several different mice. Mice with pieces of tumor from the same patient are collectively referred to as a PDX model for the same human tumor (e.g., Mouse Model PDX-1100). Each single mouse with a piece of the same tumor has a different drug treatment. Tumor volumes are measured at different time points to determine tumor growth. In each mouse, the tumor volume may increase, remain the same, or decrease. Statistical analyses described in "Methods" section are used to determine if a drug inhibits tumor growth

[23]. The SMT has a cost advantage for drug screening because data about many drugs can be obtained with fewer mice.

For the single mouse trials there is an issue that needs to be addressed. Specifically, how representative are the tumor growth data that are being collected and analyzed? With a single mouse trial for each drug, the variance of the tumor growth trajectory cannot be estimated, since $n = 1$.

Other issues with SMTs include that the observed tumor growth curves over time may be non-linear and/or non-monotonic, making analysis based on any one time less than fully informative (Fig. 2). Also, there may be missing data in any one growth curve, different ranges of times for different mice, and different single times chosen for comparison. Also, with an $n = 1$, it is challenging to estimate the confidence of any conclusion.

Here we propose a new method to evaluate a panel of drugs for their inhibition of tumor growth and for their possible adverse side effects on the health of the host animal. The method is appropriate to evaluate the kind of experimental data that are collected when screening drugs in PDX models in SMTs. We call this method GRAD, for GRouping Anti-cancer Drugs. This work comes under the general heading of finite mixture model analysis [43–45]. In this work, we apply the PROC TRAJ subroutine in the statistical application SAS [46–48]. The goal of this analysis is classification of each drug treatment as inhibitory or non-inhibitory for a given tumor model.

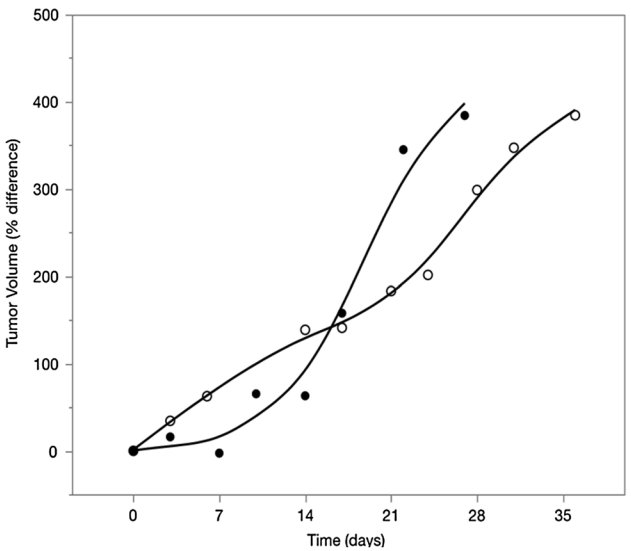

Fig. 2 Assessing tumor growth curves at a single time may be insufficient. Tumor growth curves of model X-0933 with the symbols: untreated (open circle) and treated (closed circle). At 7 and 14 days, the treated tumor volume is less than the untreated tumor volume. However, at 21 and 28 days the treated tumor volume is greater than the untreated tumor volume

As a demonstration, we use the GRAD method to analyze longitudinal data for the growth of 18 different human colorectal tumors in mice, each tumor exposed to 22 different drug treatments. The GRAD method assigns each drug into a group that is effective, or a group that is not effective, in inhibiting tumor growth in single mouse trials. And it provides a probability that the assignment is correct.

Methods

We give a brief description of the experimental method to obtain data in a SMT and how it differs from a conventional multiple mouse trial. This is followed by a step-by-step description of the GRAD method to analyze the data. Definitions and further detailed statistical and computational methods are provided in a Supplement.

In a single mouse trial, a patient-derived tumor xenograft (PDX) is divided into several different mice. Each mouse is exposed to a different drug and the corresponding tumor growth curves are measured over time (Fig. 1). The goal is to classify each drug into a group of drugs that inhibit tumor growth or a group that does not. This is achieved by comparing several tumor growth curves with each other. The GRAD method of Grouping Anti-cancer Drugs determines the most parsimonious number of groups by applying a longitudinal finite mixture model. This requires several steps listed below. Each of these groups has a trajectory that is described by a polynomial equation. The coefficients of the equation indicate if the trajectory of the group increases or does not increase with time. If a drug is in a group that does not increase with time, the drug is considered to be inhibitory for tumor growth.

Conventional multiple mice trials differ both in experimental design and in analysis of data. Typically, in conventional mouse trials of a drug, 5–10 tumor-bearing mice are treated with the same drug and another 5–10 mice bearing the same tumor are treated with a vehicle as a control. The tumor growth curves are measured over time. The tumor volume averages and standard deviations of the treated and control groups at a single time are compared. A statistical test, such as a *t*-test, is used to determine if the tumor growth curves of the drug-treated and untreated mice are significantly different. If so, then the drug may be considered to be inhibitory.

In single mouse trials there is no "average" and no "standard deviation" of the single tumor growth curve in the single mouse treated with a single drug. Therefore, a different statistical method is required to determine if a drug is inhibitory. Instead of comparing multiple drug-treated growth curves with multiple untreated growth curves at a single time, a longitudinal finite mixture model is used to compare all drug-treated growth curves. The drug treatments

are then classified into inhibitory and non-inhibitory groups. The finite mixture model we employ is implemented with the PROC TRAJ package in the SAS statistical application.

The steps of the GRAD method for Grouping Anti-cancer Drugs are illustrated in Fig. 3 and described below.

1. Determine drug treatments that are non-toxic by mouse weight and tumor size.

Retain non-toxic drug treatments.

Remove from further consideration drug treatments where tumor size indicates clinical signs of morbidity, including tumor weight loss greater than 10% of body weight, and any tumor dimension exceeding 20 mm [49]. Such drug treatments are referred to as vehicle-like.

2. Determine the optimal number of groups of growth curves.

Tumor growth curves may be classified into one group, two groups, or three groups, hereafter referred to as the number of groups. To determine the optimal number of groups, the output of the statistical package PROC TRAJ for each of the possible number of groups are compared. The output for each number will include values for the following parameters: the percent of growth curves in each group, the p-value for the percentage of growth curves in each group, the coefficients of a polynomial equation for the trajectory of each group, the Bayesian Posterior Probability (BPP) for each growth curve assigned to each group, and the Bayesian Information Criterion (BIC) for the specific number of groups.

The optimal number of groups will have the following characteristics: all parameters can be estimated, group percentages will be greater than 5% with all percentage p-values less than 0.10, at least one drug treatment can be placed into a group (with BPP > 0.9), and it will have the largest BIC value of one, two, or three groups.

The result will be a specified optimal number of groups of growth curves.

3. Determine the trajectory of each group of growth curves.

The trajectory of each group of growth curves is described by the coefficients of a polynomial equation determined by PROC TRAJ. The trajectory of each group may increase over time, remain constant, or decrease.

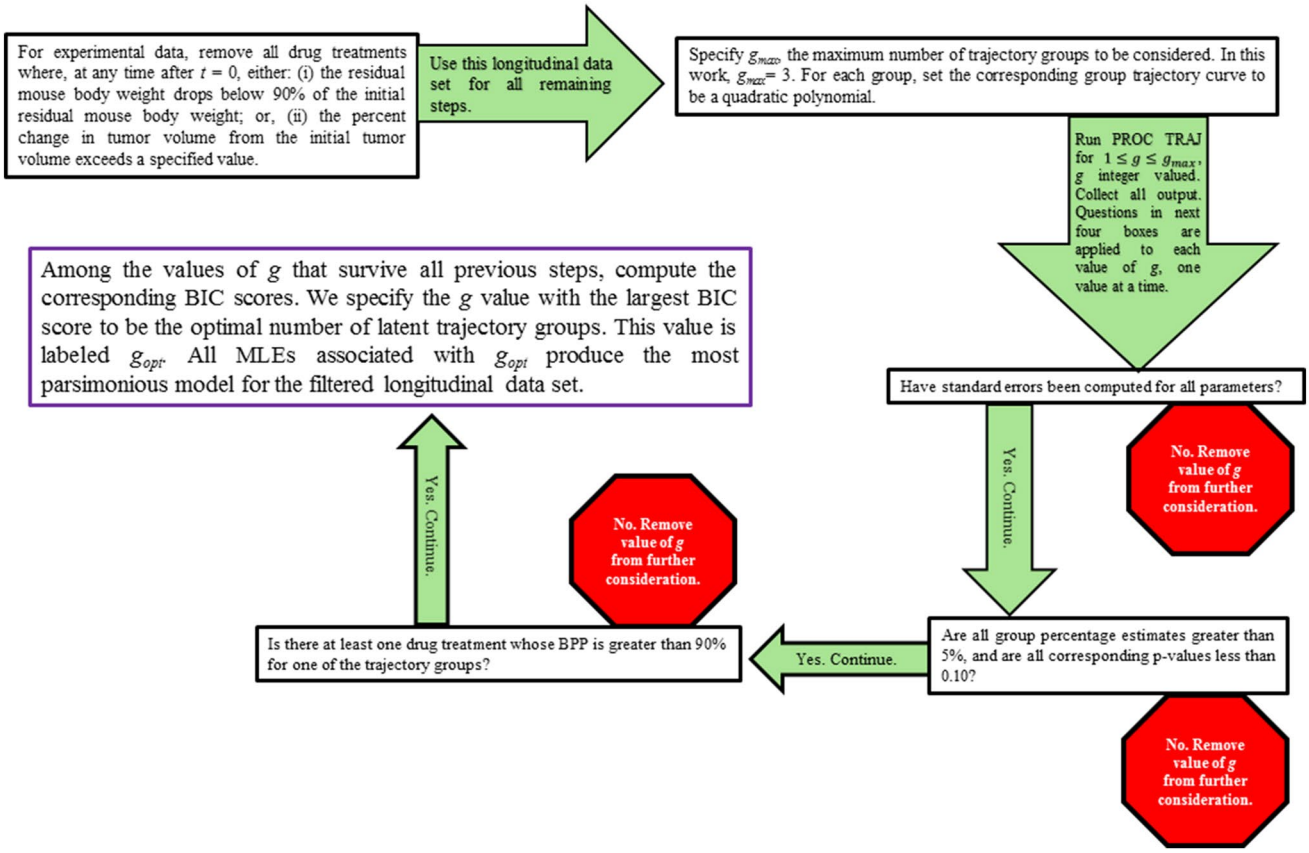

Fig. 3 Flow chart illustrating the GRAD method

4. Determine drug treatments that are inhibitory or non-inhibitory.

Drug treatments whose growth curves are in a group whose trajectory is constant or decreases over time are considered inhibitory. Drug treatments whose growth curves are in a group whose trajectory increases are considered non-inhibitory.

5. The result is a list of drugs, grouped into those that are effective and those that are non-effective.

Effective drugs are defined as those that are both inhibitory for tumor growth and are non-toxic in single mouse trials.

Results

Example tumor model X-0933

We analyzed tumor growth curves for 18 human colorectal tumors treated with 22 drugs in single mouse trials, using data provided by Gao et al. [23]. As an example we describe the results of our analysis using the drug treatments for tumor model X-0933. Details of the computational and statistical methods are provided in "Supplement".

All of the drug-treated tumor growth curves are shown in Fig. 4a, where the Tumor Volume (% difference) is plotted as a function of time in days. Inspection is insufficient to objectively classify all curves as either inhibitory or non-inhibitory.

Tumor growth curves for drug treatments that were non-toxic (retained more than 90% of their initial weight at all

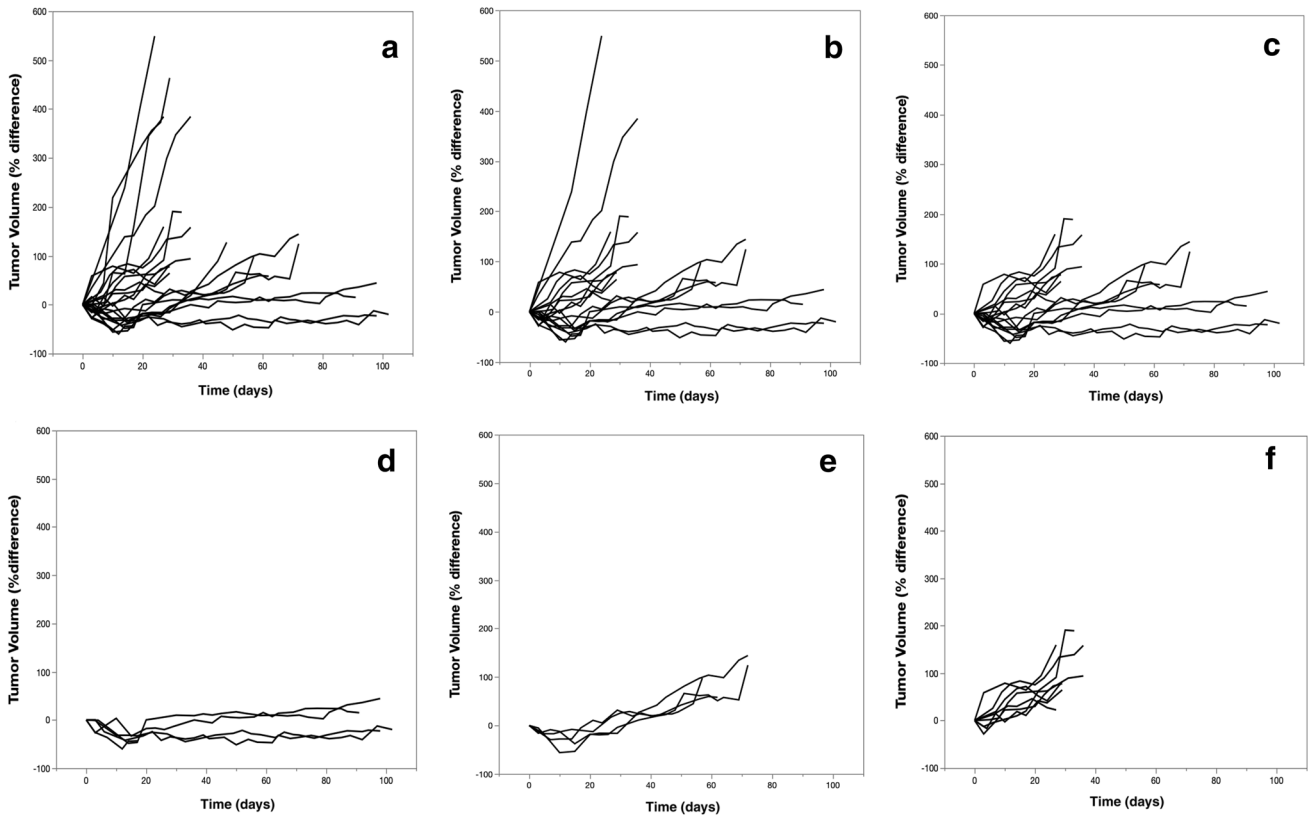

Fig. 4 Protocol for identifying effective drug treatments. Drug treatments are considered to be effective if they are non-toxic to mice and they inhibit tumor growth. Drugs are evaluated in three successive steps. In this example, 22 mice each with the PDX tumor model X-0933 were exposed to 22 different drug treatments. Tumor volumes (percent difference from initial volume) were plotted as a function of time in days. **a** Tumor growth curves for mice exposed to all 22 drug treatments, including 21 candidate drugs and one vehicle as an untreated control. **b** Tumor growth curves for drugs that were non-toxic to mice. **c** Tumor growth curves for drugs that were non-toxic, and in addition, that were substantially different than the tumor growth curves of the vehicle exposed mice. The tumor growth curves in panel **c** were analyzed by the PROC TRAJ statistical procedure and separated into three groups, panels **d**, **e** and **f**. Drug treatments resulting in growth curves in panel **d** (Group 1) were classified as inhibitory because of the estimated constant trajectory and the sufficiently high BPP values. Drug treatments resulting in growth curves in panels **e** and **f** (Groups 2 and 3, respectively) were classified as non-inhibitory because of the estimated positive trajectories and the sufficiently high BPP values

times) are shown in Fig. 4b. Those curves for drug treatments that were both non-toxic and not vehicle-like are shown in Fig. 4c.

The retained drug-treated tumor growth curves were classified into three trajectory groups using the procedure PROC TRAC in the statistical package SAS. The number three was determined since the Bayesian Information Criterion (BIC) for three groups (− 1188.39) was greater than the BIC for two groups (− 1252.33) and for one group (− 1308.06).

The coefficient estimates for the polynomials describing the group-specific trajectory curves are shown in Table 1. The equations for Groups 1, 2, and 3 are the following:

Tumor Volume (% difference) ≈ − 18.3291,

Tumor Volume (% difference) ≈ − 16.8454 + 0.0294 t,

Tumor Volume (% difference) ≈ − 1.6899t + 0.0599 t^2,

where t is time units measured in days. In the supplement, we use the notation TumorVol%Δ(t). Coefficients with corresponding p-values greater than 0.10 are omitted.

The equations for the trajectory curves for Groups 2 and 3 have positive coefficients, indicating that the tumor volumes increase over time. Therefore, drugs in these groups are not inhibitory. The equation for trajectory Group 1 has only one significant coefficient (for the intercept), indicating that the tumor volumes do not increase over time.

Using the procedure described in the "Methods" section, we determined that the optimal number of groups for the X-0933 tumor model is three.

In Table 2, we provide BPPs that a given drug treatment is in Group 1, 2, or 3. We see that 17/22 ≈ 77% of the drug treatments were retained after Step 1 (Methods). Of the 17 drug treatments, 15 had BPPs that were greater than 0.90 for at least one of the three groups, and hence by GRAD criteria, are classified as being either inhibitory or non-inhibitory.

Tumor growth curves for specific drug-treatment subsets of the X-0933 tumor model are presented in panels Fig. 4a–f. Briefly, Fig. 4a is for all drug treatments, Fig. 4b is for those that are non-toxic, Fig. 4c are those in Fig. 4b that also are not vehicle-like. Figure 4 (panels d, e, and f) display those growth curves from Fig. 4c for Groups 1, 2, and 3, respectively, as determined by the GRAD method.

For Group 1, that contains inhibitory treatments according to Table 1, the inhibitory drug treatments are CLR457, CKX620, BYL719 + cetuximab + encorafenib, and BYL719 + binimetinib. From the results of Table 2, we see that all four of these drug treatments have BPP equal to 1.00 for being in Group 1. Note that LFW527 + binimetinib is not classified as inhibitory by GRAD, since the BPP for it being

Table 1 Summary of trajectory curve information produced by PROC TRAJ

The SAS System
Maximum likelihood estimates
Model: censored normal (CNORM)

Group	Parameter	Estimate	Standard error	t Statistic for H_0: Parameter = 0	Prob > |t|
1	Intercept	− 18.3291	6.36754	− 2.879	0.0043
	Linear	− 0.13735	0.30067	− 0.457	0.6482
	Quadratic	0.00339	0.00302	1.122	0.2628
2	Intercept	− 16.84541	8.08897	− 2.083	0.0383
	Linear	− 0.20336	0.50699	− 0.401	0.6887
	Quadratic	0.02937	0.00717	4.098	0.0001
3	Intercept	1.62395	7.04741	0.230	0.8179
	Linear	1.68991	0.98082	1.723	0.0861
	Quadratic	0.05988	0.02888	2.073	0.0392
	Sigma	23.81645	1.09708	21.709	0.0000
Group membership					
1	(%)	27.99185	11.54991	2.424	0.0161
2	(%)	25.43935	11.31270	2.249	0.0254
3	(%)	46.5688	12.53946	3.714	0.0003
BIC = − 1188.39 (N = 248)		BIC = − 1172.31 (N = 17)		AIC = − 1167.31	L = − 1155.31

Here, we present PROC TRAJ output when fitting tumor trajectory curves of model X-0933 data with quadratic equation coefficients. We specify g = 3 groups. The trajectory curve for Group 1 has only one significant coefficient (for Intercept), indicating that tumor volumes are constant over time. Therefore, drug treatments that are in this group are considered to be inhibitory. The trajectory curves for Groups 2 and 3 have significant positive quadratic coefficients, indicating that the tumor volumes continued to increase, and, therefore, the drug treatments in these groups are considered to be non-inhibitory. Groups 1, 2, and 3 have 28, 25, and 47% of the drug treatments, respectively, based on Group membership values. The Bayesian Information Content (BIC) for g = 3 groups is − 1188.39 (N = 248)

Table 2 Bayesian Posterior Probabilities (BPPs) that a given drug treatment is in a particular tumor trajectory group for tumor model X-0933, as determined by PROC TRAJ

Drug treatment index #	Drug treatment name	BPP–Grp01	BPP–Grp02	BPP–Grp03	GROUP
2	LKA136	0	0	1	3
4	LFW527 + binimetinib	0.758	0.242	0	1
5	LEE011	0	0	1	3
7	Encorafenib	0	0.077	0.923	3
8	CLR457	1	0	0	1
9	CKX620	1	0	0	1
10	CGM097	0	0	1	3
11	Cetuximab + encorafenib	0	1	0	2
12	Cetuximab	0	0	1	3
14	BYL719 + encorafenib	0	1	0	2
15	BYL719 + cetuximab + encorafenib	1	0	0	1
16	BYL719 + cetuximab	0	0	1	3
17	BYL719 + binimetinib	1	0	0	1
18	BYL719	0	1	0	2
19	BKM120 + LJC049	0	0	1	3
20	BKM120	0	1	0	2
21	Binimetinib	0	0	1	3

Using information from Table 1, we report PROC TRAJ output when computing the BPPs for all model X-0933 drug treatment data with $g = 3$ groups. The column headings are defined as follows: BPP-Grp0x = Bayesian Posterior Probability of the given Drug Treatment Name for trajectory group x ($x = 1$, 2, or 3). Group = Group assignment, computed as the maximum of the three BPP values. Note: Ninety-four percent (16/17) drug treatments have BPPs greater than 0.90 for one of the three groups

in Group 1 is 0.758, which is less than 0.90. We observe tumor growth curves for these drug treatment in Fig. 4d.

Group 2 contains non-inhibitory treatments according to Table 1. The treatments are cetuximab + encorafenib, BYL719 + encorafenib, BYL719, and BKM120. Using Table 2, we see that, as with the inhibitory drug treatments in Group 1, all four of these drug treatments have BPP equal to 1.00 for being in Group 2. Growth curves for these drug treatments are presented in Fig. 4e.

The drug treatments in Group 3 are non-inhibitory treatments according to Table 1. The treatments are LKA136, LEE011, encorafenib, CGM097, cetuximab, BYL719 + cetuximab, BKM120 + LJC049, and binimetinib. All of these drug treatments have BPP equal to 1.00 for being in Group 3, with the exception of encorafenib. This drug treatment has a BPP equal to 0.92 for being in Group 3. The growth curves are presented in Fig. 4f.

All tumor models

Eighteen tumor models and the drugs that were effective against each are shown in Table 3. A drug was considered effective if the tumor had non-positive growth and it was non-toxic. Fifty-six percent (10/18) of the tumor models were effectively treated by more than one drug, and eight of the models were not effectively treated by any of the drugs tested. Three of the models were effectively treated by the least number of drugs, two, and two models were effectively treated by the most number of drugs, seven. This result suggests that application of our GRAD method can discover more than one effective drug treatment for some tumors.

Over all tumor models for which there was at least one effective drug treatment (sample size of ten), the number of effective drug treatments ranged from two to seven. The mean, median, and mode numbers were 4.43, 4, and 2. Examples include model X-2145, for which two drug treatments each were effective as single treatments, namely CLR457 and binimetinib. Model X-1119 was not effectively treated by drug BYL719 alone, but was effectively treated by BYL719 in combination with LJM716, or BYL719 in combination with encorafenib. Model X-1167 had four drug treatments that were effective, including CKX620 that was effective as a single drug; BYL719 + binimetinib that was effective as a combination of two drugs; and cetuximab + BYL719 + encorafenib that was effective as a combination of three drugs.

All drug treatments

There were 22 drug treatments tested, some singly and others in combination. Eighty-one percent (18/22) of the drug treatments were effective for at least one tumor model, and

Table 3 Effective drug treatments for each tumor model

Drug Treatment Index #	Drug Treatment Name	Tumor Model (X-*)																		Row Total
		0933	1055	1119	1167	1290	1329	1443	1500	2145	2182	2483	2573	2659	2861	3792	5254	5438	5495	
1	untreated																	■		1
2	LKA136			■		■		■						■						4
3	LJC049																			0
4	LFW527 + binimetinib							■						■						2
5	LEE011							■						■						2
6	HDM201																			0
7	encorafenib																			0
8	CLR457	■		■				■												3
9	CKX620	■		■	■			■			■			■						6
10	CGM097			■														■		2
11	cetuximab + encorafenib									■	■							■		3
12	cetuximab					■														1
13	BYL719 + LJM716			■		■														2
14	BYL719 + encorafenib			■		■														2
15	BYL719 + cetuximab + encorafenib	■			■									■				■		4
16	BYL719 + cetuximab					■														1
17	BYL719 + binimetinib	■			■								■	■						4
18	BYL719													■						1
19	BKM120 + LJC049																			0
20	BKM120				■															1
21	binimetinib			■						■			■							3
22	5FU					■														1
Column Total		4	0	7	4	6	0	5	0	2	2	0	2	7	0	0	0	4	0	43

Drug treatments that were effective for given tumor models are indicated with gray cells. We have removed the "X-" from each tumor model due to space consideration. For example, the second tumor model, labeled with column heading "1055" corresponds to the tumor model X-1055

four were not effective for any model. Fifty-five percent (12/22) were effective for more than one tumor model. Twenty-seven percent (6/22) were effective for exactly one tumor model. One drug treatment, CKX620, was effective for the most number of tumor models, six.

It might be expected that combinations of drug treatments would be more effective than single drug treatments. This was observed, for example, for encorafenib and cetuximab. Each was not effective singly against model X-5438, but they were effective in combination. Another example occurs for the drugs encorafenib, cetuximab, and BYL719. They were not effective singly, or in any double combination against model X-1167, but were effective in the triple combination (Table 3).

The drug 5FU was effective for only one of the 22 models, X-1290. The drug treatment 5FU, or its precursor Capecitabine, has been used clinically for decades. It is recommended for colon cancer patients [50]. However, 5FU was not classified as effective in 21 treated colorectal tumor models because it did not inhibit tumor growth or it was toxic to mice, or both.

One exception was noted. Without treatment with a drug (intended as an "untreated" control, vehicle), the tumor model X-5438 did not grow and, therefore, was classified as if the "treatment" was inhibitory. This one exception indicates the need for replication [51]. Specifically, it would be strongly advised that researchers confirm conclusions from screening by single mouse trials with additional observations of multiple mice.

Sensitivity and specificity

In this context, sensitivity is the proportion of true effective treatments that are classified as being effective by the GRAD method. Similarly, specificity is the proportion of true non-effective treatments that are classified as being non-effective by the GRAD method.

We estimate, through simulation, how frequently the GRAD method can identify effective drug treatments when they exist (sensitivity), and how frequently the method can identify non-effective treatments when there are no effective treatments (specificity).

More specifically, we randomly generated a set of longitudinal data replicates where a portion of the drug treatments were effective. Separately, we generated a set of replicates where all of the drug treatments were non-effective. Every replicate in the first set is labeled a true positive, and every replicate in the second set is labeled a true negative.

Parameters for the first set of simulations were drawn from tumor model X-5438. We concluded from the results of Table 3 that a portion of the 22 drug treatments were effective for this tumor model. Parameters for the second set were drawn from tumor model X-5254. By GRAD, no drug treatment was effective for this tumor model.

For each model, we simulated 500 replicates based on these parameter settings. We then applied GRAD to each replicate. For the sensitivity analyses, we scored a replicate as being a "positive" (effective) if GRAD identified at least one effective drug treatment. For the specificity analyses, we scored a replicate as being a "negative" (non-effective) if GRAD identified all drug treatments as being non-effective.

Sensitivity was computed as the proportion of the 500 replicates that were labeled positive by GRAD. Specificity was computed as the proportion of the 500 replicates that were labeled negative by GRAD.

The GRAD method sensitivity was 84% (80.5%, 87.1%) and its specificity was 98% (96.4%, 99.0%). The 95% confidence interval values, presented in parentheses, were computed using the method implemented in the BINOM program [52].

Discussion

We have developed a statistical method for analyzing data obtained in pre-clinical screening of candidate chemotherapeutic drugs. The method, Grouping Anti-cancer Drugs (GRAD), evaluates drugs based on their ability to inhibit human tumors growing in single mouse trials (SMTs) of patient-derived tumor xenografts (PDXs). In this experimental design, a series of mice carrying portions of the same human tumor are each exposed to a different drug treatment, one drug treatment per mouse, without replicates. The data consist of measurements of tumor size and mouse body weight obtained at multiple time points.

The experimental design of SMTs is intended as a screening procedure with minimal costs compared to experimental designs that have replicate mice exposed to the same drug. The purpose of each experimental design is to distinguish drugs that may inhibit human tumors, from those drugs that may not inhibit human tumors. However, the measurements obtained in SMTs present several challenges for data analysis. One challenge is not having replicate observations of multiple mice exposed to the same drug. Also, the tumor growth curves may not be monotonic, or there may be measurements missing at some times. In addition, the overall time periods may be different for tumor growth curves in different mice.

Nevertheless, the growth curves of the same human tumor exposed to a number of different drugs in different mice offer the opportunity to apply a statistical method that can objectively group drugs into those that inhibit tumor growth and those that do not. We determine drugs as being inhibitory if and only if their corresponding tumor trajectory curve is non-positive (constant, linear with a negative slope, or quadratic concave down). Likewise, we determine drugs as being non-inhibitory if and only if their corresponding fitted tumor trajectory curve is positive (linear with positive slope, or quadratic concave up).

To determine groups of drugs that are either inhibitory or not, the GRAD method uses finite mixture modeling [43–45]. The group-based modeling of longitudinal data complements previously used hierarchical modeling and latent curve analysis. Finite mixture models can handle missing data, irregularly spaced measurements, tumor growth curves of different lengths, and multiple different non-monotonic growth curves. These models have proved informative for analysis with various kinds of longitudinal data [53–56].

A key feature of the finite mixture model is that the set of tumor growth curves is composed of distinct groups. Each group of similar growth curves is described by a polynomial whose coefficients are estimated by maximum likelihood using the EM algorithm [57]. The probability that each tumor growth curve belongs to a group is determined by its Bayesian Posterior Probability. We have demonstrated the utility of finite mixture modeling for analysis of SMT data through application of the PROC TRAC procedure implemented in the SAS software. An introduction to PROC TRAC and the description of further developments have appeared. Other implementations of finite mixture models include Mplus [58], and packages in R [59, 60].

To classify a drug as "effective" we use a decision rule with two criteria: (i) the drug inhibits tumor growth, and (ii) it is non-toxic to the host. Drugs are considered to be inhibitory if their estimated tumor trajectories do not increase over time. Drugs are considered non-toxic if the mice continue to be healthy. A mouse is considered to be healthy if it loses no more than 10% of its initial residual weight at any time. Residual mouse weight is calculated as the difference of the measured weight of the mouse with the tumor and the weight of the tumor. The weight of the tumor is calculated from the measured volume of the

tumor, as described in the Supplement. The 10% threshold of weight loss is consistent with welfare guidelines for the use of animals in research [61] and the report that human cancer patients who lose more than 10% of their weight have a worse prognosis [62].

The effect of various drugs on tumor growth has previously been determined in different ways, either by comparing the tumor size measured at a single time, or by comparing entire growth curves obtained by measuring tumor size at multiple time points [63, 64]. Using tumor size at a single time discards most of the measurements and may lose the information about the entire growth curve. Also, single time comparisons may lose the ability to predict future tumor growth. By contrast, growth curves can be compared using statistical methods that use data on multiple time points. For instance, assuming exponential growth, logarithmic equations can be used to compare growth curves. Without assuming a parametric fit, growth curves can be compared by considering the area under each curve.

The GRAD method of grouping candidate anti-cancer drugs differs from methods that compare drug-treated tumor growth and control non-drug-treated tumor growth. GRAD does not depend on comparison of various drugs with a control. Rather it uses all of the tumor growth curves and groups those that are inhibitory versus those that are not inhibitory. In this way it overcomes one of the limitations of the SMT, in having only one control growth curve to which all of the other single drug growth curves are compared.

When using finite mixture models that employ the EM algorithm, one should be aware that there are situations where the algorithm may not converge, that is, one cannot obtain maximum likelihood estimates of the coefficients for the fitted trajectory polynomials. This situation tends to occur when one attempts to fit more parameters.

Based upon analysis of the experimental data generously provided by Gao et al. [23], we make several suggestions for future single mouse trials. One suggestion is that the interval and the length of time that each mouse is observed should be consistent, within the limitations of the mouse health. Second, once an experimental design is established for one drug, it should be followed for the other drugs to which it is being compared. Third, although the finite mixture model and GRAD procedure do not compare growth curves from drug-treated mice with growth curves from untreated mice, as in conventional mouse trials, it is necessary to establish the limit of tumor size that an untreated mouse can tolerate.

In conclusion, we describe a statistical method, GRAD, for GRouping Anti-cancer Drugs. This method identifies groups of drugs in patient-derived tumor xenograft single mouse trials that are effective, and those that are not, with potentially high sensitivity and specificity. By effective, we mean that the drugs inhibit tumor growth and are non-toxic to the mice. The GRAD method uses a finite mixture model implemented with the PROC TRAJ procedure in the statistical application SAS.

Acknowledgements We thank Gao et al. for making their extensive PDX data publically available, and Payal Patel and Lisheng Zhou for programming assistance.

Authors' contributions DEA and DG equally conceived of all study designs, methods, and statistical analyses. Both authors wrote and approved of the manuscript.

Funding D.E.A. was supported by the Human Genetics Institute of New Jersey, the New Jersey Breast Cancer Research Fund, and Rutgers Cancer Institute of New Jersey (PA30CA072720).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with animals performed by any of the authors.

References

1. Shoemaker RH (2006) The NCI60 human tumour cell line anticancer drug screen. Nat Rev Cancer 6(10):813–823. https://doi.org/10.1038/nrc1951
2. Sharma SV, Haber DA, Settleman J (2010) Cell line-based platforms to evaluate the therapeutic efficacy of candidate anticancer agents. Nat Rev Cancer 10(4):241–253. https://doi.org/10.1038/nrc2820
3. Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, Clark L et al (2006) A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell 10(6):515–527. https://doi.org/10.1016/j.ccr.2006.10.008
4. Weiswald LB, Bellet D, Dangles-Marie V (2015) Spherical cancer models in tumor biology. Neoplasia 17(1):1–15. https://doi.org/10.1016/j.neo.2014.12.004
5. Lancaster MA, Knoblich JA (2014) Organogenesis in a dish: modeling development and disease using organoid technologies. Science 345(6194):1247125. https://doi.org/10.1126/science.1247125
6. Clevers H (2016) Modeling development and disease with organoids. Cell 165(7):1586–1597. https://doi.org/10.1016/j.cell.2016.05.082
7. Crespo M, Vilar E, Tsai SY, Chang K, Amin S, Srinivasan T, Zhang T et al (2017) Colonic organoids derived from human induced pluripotent stem cells for modeling colorectal cancer and drug testing. Nat Med 23(7):878–884. https://doi.org/10.1038/nm.4355
8. Gould SE, Junttila MR, de Sauvage FJ (2015) Translational value of mouse models in oncology drug development. Nat Med 21(5):431–439. https://doi.org/10.1038/nm.3853
9. Day CP, Merlino G, Van Dyke T (2015) Preclinical mouse cancer models: a maze of opportunities and challenges. Cell 163(1):39–53. https://doi.org/10.1016/j.cell.2015.08.068
10. Rubio-Viqueira B, Hidalgo M (2009) Direct in vivo xenograft tumor model for predicting chemotherapeutic drug response in cancer patients. Clin Pharmacol Ther 85(2):217–221. https://doi.org/10.1038/clpt.2008.200

11. Aparicio S, Hidalgo M, Kung AL (2015) Examining the utility of patient-derived xenograft mouse models. Nat Rev Cancer 15(5):311–316. https://doi.org/10.1038/nrc3944
12. Li G (2015) Patient-derived xenograft models for oncology drug discovery. J Cancer Metastasis Treat 1:8–15
13. Ben-David U, Ha G, Tseng YY, Greenwald NF, Oh C, Shih J, McFarland JM et al (2017) Patient-derived xenografts undergo mouse-specific tumor evolution. Nat Genet 49(11):1567–1575. https://doi.org/10.1038/ng.3967
14. Morgan KM, Riedlinger GM, Rosenfeld J, Ganesan S, Pine SR (2017) Patient-derived xenograft models of non-small cell lung cancer and their potential utility in personalized medicine. Front Oncol 7:2. https://doi.org/10.3389/fonc.2017.00002
15. Nunes M, Vrignaud P, Vacher S, Richon S, Lievre A, Cacheux W, Weiswald LB et al (2015) Evaluating patient-derived colorectal cancer xenografts as preclinical models by comparison with patient clinical data. Cancer Res 75(8):1560–1566. https://doi.org/10.1158/0008-5472.can-14-1590
16. Seol HS, Kang HJ, Lee SI, Kim NE, Kim TI, Chun SM, Kim TW et al (2014) Development and characterization of a colon pdx model that reproduces drug responsiveness and the mutation profiles of its original tumor. Cancer Lett 345(1):56–64. https://doi.org/10.1016/j.canlet.2013.11.010
17. Vlachogiannis G, Hedayat S, Vatsiou A, Jamin Y, Fernandez-Mateos J, Khan K, Lampis A et al (2018) Patient-derived organoids model treatment response of metastatic gastrointestinal cancers. Science 359(6378):920–926. https://doi.org/10.1126/science.aao2774
18. Maekawa H, Miyoshi H, Yamaura T, Itatani Y, Kawada K, Sakai Y, Taketo MM (2018) A chemosensitivity study of colorectal cancer using xenografts of patient-derived tumor-initiating cells. Mol Cancer Ther 17(10):2187–2196. https://doi.org/10.1158/1535-7163.Mct-18-0128
19. Gu Q, Zhang B, Sun H, Xu Q, Tan Y, Wang G, Luo Q et al (2015) Genomic characterization of a large panel of patient-derived hepatocellular carcinoma xenograft tumor models for preclinical development. Oncotarget 6(24):20160–20176. https://doi.org/10.18632/oncotarget.3969
20. Townsend EC, Murakami MA, Christodoulou A, Christie AL, Koster J, DeSouza TA, Morgan EA et al (2016) The public repository of xenografts enables discovery and randomized phase II-like trials in mice. Cancer Cell 29(4):574–586. https://doi.org/10.1016/j.ccell.2016.03.008
21. Bruna A, Rueda OM, Greenwood W, Batra AS, Callari M, Batra RN, Pogrebniak K et al (2016) A biobank of breast cancer explants with preserved intra-tumor heterogeneity to screen anticancer compounds. Cell 167(1):260–274. https://doi.org/10.1016/j.cell.2016.08.041
22. Pauli C, Hopkins BD, Prandi D, Shaw R, Fedrizzi T, Sboner A, Sailer V et al (2017) Personalized in vitro and in vivo cancer models to guide precision medicine. Cancer Discov 7(5):462–477. https://doi.org/10.1158/2159-8290.cd-16-1154
23. Gao H, Korn JM, Ferretti S, Monahan JE, Wang Y, Singh M, Zhang C et al (2015) High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. Nat Med 21(11):1318–1325. https://doi.org/10.1038/nm.3954
24. Charles River Laboratories. (Accessed April 17 2019). https://www.criver.com/products-services/discovery-services/vivo-pharmacology/oncology-pharmacology-models/patient-derived-xenografts-pdx?region=3601
25. Mouse tumor biology database. (Accessed 17 April 2019). http://tumor.informatics.jax.org/mtbwi/index.do
26. CrownBio. (Accessed 17 April 2019). https://www.crownbio.com/oncology/invivo-services/patient-derived-xenograft-pdx-tumor-models
27. Champions Oncology. (Accessed 17 April 2019). https://championsoncology.com
28. XenTech. (Accessed 17 April 2019). http://www.xentech.com
29. Wu J, Houghton PJ (2010) Interval approach to assessing antitumor activity for tumor xenograft studies. Pharm Stat 9(1):46–54. https://doi.org/10.1002/pst.369
30. Bertotti A, Migliardi G, Galimi F, Sassi F, Torti D, Isella C, Cora D et al (2011) A molecularly annotated platform of patient-derived xenografts ("xenopatients") identifies HER2 as an effective therapeutic target in cetuximab-resistant colorectal cancer. Cancer Discov 1(6):508–523. https://doi.org/10.1158/2159-8290.CD-11-0109
31. Yang M, Shan B, Li Q, Song X, Cai J, Deng J, Zhang L et al (2013) Overcoming erlotinib resistance with tailored treatment regimen in patient-derived xenografts from naive Asian NSCLC patients. Int J Cancer 132(2):E74–E84. https://doi.org/10.1002/ijc.27813
32. Corcoran RB, Atreya CE, Falchook GS, Kwak EL, Ryan DP, Bendell JC, Hamid O et al (2015) Combined BRAF and MEK inhibition with dabrafenib and trametinib in BRAF V600-mutant colorectal cancer. J Clin Oncol 33(34):4023–4031. https://doi.org/10.1200/JCO.2015.63.2471
33. Stebbing J, Paz K, Schwartz GK, Wexler LH, Maki R, Pollock RE, Morris R et al (2014) Patient-derived xenografts for individualized care in advanced sarcoma. Cancer 120(13):2006–2015. https://doi.org/10.1002/cncr.28696
34. Liang H (2007) Comparison of antitumor activities in tumor xenograft treatment. Contemp Clin Trials 28(2):115–119. https://doi.org/10.1016/j.cct.2006.05.001
35. Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, Verweij J et al (2000) New guidelines to evaluate the response to treatment in solid tumors. J Natl Cancer Inst 92(3):205–216
36. Corwin WL, Ebrahimi-Nik H, Floyd SM, Tavousi P, Mandoiu II, Srivastava PK (2017) Tumor control index as a new tool to assess tumor growth in experimental animals. J Immunol Methods 445:71–76. https://doi.org/10.1016/j.jim.2017.03.013
37. Duan F, Simeone S, Wu R, Grady J, Mandoiu I, Srivastava PK (2012) Area under the curve as a tool to measure kinetics of tumor growth in experimental animals. J Immunol Methods 382(1–2):224–228. https://doi.org/10.1016/j.jim.2012.06.005
38. Liang H (2005) Modeling antitumor activity in xenograft tumor treatment. Biom J 47(3):358–368
39. Wu J (2011) Assessment of antitumor activity for tumor xenograft studies using exponential growth models. J Biopharm Stat 21(3):472–483. https://doi.org/10.1080/10543406.2010.481802
40. Hather G, Liu R, Bandi S, Mettetal J, Manfredi M, Shyu WC, Donelan J et al (2014) Growth rate analysis and efficient experimental design for tumor xenograft studies. Cancer Inform 13(Suppl 4):65–72. https://doi.org/10.4137/cin.s13974
41. Yao YM, Donoho GP, Iversen PW, Zhang Y, Van Horn RD, Forest A, Novosiadly RD et al (2017) Mouse PDX trial suggests synergy of concurrent inhibition of RAF and EGFR in colorectal cancer with braf or kras mutations. Clin Cancer Res 23(18):5547–5560. https://doi.org/10.1158/1078-0432.ccr-16-3250
42. Benzekry S, Lamont C, Beheshti A, Tracz A, Ebos JM, Hlatky L, Hahnfeldt P (2014) Classical mathematical models for description and prediction of experimental tumor growth. PLoS Comput Biol 10(8):e1003800. https://doi.org/10.1371/journal.pcbi.1003800
43. Titterington DM, Smith AFM, Makov UE (1985) Statistical analysis of finite mixture distributions. J. Wiley and Sons, Hoboken
44. McLachlan G, Peel D (2000) Finite mixture models. Wiley series in probability and statistics. J. Wiley and Sons, New York

45. Vermunt JK (2010) Longitudinal research using mixture models. In: Monfort K, Oudh JHL, Satorra A (eds) Longitudinal research with latent variables. Springer, Berlin
46. Jones BL, Nagin DS (2007) Advances in group-based trajectory modeling and an SAS procedure for estimating them. Sociol Method Res 35(4):542–571. https://doi.org/10.1177/0049124106292364
47. Jones BL, Nagin DS, Roeder K (2001) A SAS procedure based on mixture models for estimating developmental trajectories. Sociol Method Res 29(3):374–393. https://doi.org/10.1177/0049124101029003005
48. SAS. (Accessed 9 April 2019). https://www.sas.com/en_us/home.html
49. Burkholder T, Foltz C, Karlsson E, Linton CG, Smith JM (2012) Health evaluation of experimental laboratory mice. Curr Protoc Mouse Biol 2:145–165. https://doi.org/10.1002/9780470942390.mo110217
50. Benson AB, Venook AP, Cederquist L, Chan E, Chen YJ, Cooper HS, Deming D et al (2017) Colon cancer, version 1.2017, NCCN clinical practice guidelines in oncology. J Natl Compr Cancer Netw 15(3):370–398
51. Igl BW, Konig IR, Ziegler A (2009) What do we mean by 'replication' and 'validation' in genome-wide association studies? Hum Hered 67(1):66–68. https://doi.org/10.1159/000164400
52. Ott J (1999) Analysis of human genetic linkage, 3rd edn. The John Hopkins University Press, Baltimore
53. Wu R, Lin M (2006) Functional mapping—how to map and study the genetic architecture of dynamic complex traits. Nat Rev Genet 7(3):229–237. https://doi.org/10.1038/nrg1804
54. Chang SW, Choi SH, Li K, Fleur RS, Huang C, Shen T, Ahn K et al (2009) Growth mixture modeling as an exploratory analysis tool in longitudinal quantitative trait loci analysis. BMC Proc 3(Suppl 7):S112
55. Musolf A, Nato AQ, Londono D, Zhou L, Matise TC, Gordon D (2014) Mapping genes with longitudinal phenotypes via bayesian posterior probabilities. BMC Proc 8(Suppl 1):S81. https://doi.org/10.1186/1753-6561-8-s1-s81
56. Londono D, Chen KM, Musolf A, Wang R, Shen T, Brandon J, Herring JA et al (2013) A novel method for analyzing genetic association with longitudinal phenotypes. Stat Appl Genet Mol Biol 12(2):241–261. https://doi.org/10.1515/sagmb-2012-0070
57. Muthen B, Shedden K (1999) Finite mixture modeling with mixture outcomes using the EM algorithm. Biometrics 55(2):463–469
58. Muthen LK, Muthen BO (1998–2010) Mplus user's guide, 5th edn. Muthen & Muthen, Los Angeles
59. Leisch F (2004) FlexMix: A general framework for finite mixture models and latent class regression in r. J Stat Softw 11(8):21374
60. R Development Core Team (2012) R: A language and environment for statistical computing. R Foundation for Statistical Computing Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/
61. Ullman-Cullere MH, Foltz CJ (1999) Body condition scoring: a rapid and accurate method for assessing health status in mice. Lab Anim Sci 49(3):319–323
62. Vergidis J, Gresham G, Lim HJ, Renouf DJ, Kennecke HF, Ruan JY, Chang JT et al (2016) Impact of weight changes after the diagnosis of stage III colon cancer on survival outcomes. Clin Colorectal Cancer 15(1):16–23. https://doi.org/10.1016/j.clcc.2015.07.002
63. Heitjan DF, Manni A, Santen RJ (1993) Statistical analysis of in vivo tumor growth experiments. Cancer Res 53(24):6042–6050
64. Li Q, Dai W, Liu J, Li Y-X, Li Y-Y (2019) DRAP: a toolbox for drug response analysis and visualization tailored for preclinical drug testing on patient-derived xenograft models. J Transl Med 17:39–47. https://doi.org/10.1186/s12967-019-1785-7