

RESEARCH ARTICLE

Open Access



The design, analysis and application of mouse clinical trials in oncology drug development

Sheng Guo^{1*} , Xiaoqian Jiang¹, Binchen Mao¹ and Qi-Xiang Li^{2,3*}

Abstract

Background: Mouse clinical trials (MCTs) are becoming wildly used in pre-clinical oncology drug development, but a statistical framework is yet to be developed. In this study, we establish such as framework and provide general guidelines on the design, analysis and application of MCTs.

Methods: We systematically analyzed tumor growth data from a large collection of PDX, CDX and syngeneic mouse tumor models to evaluate multiple efficacy end points, and to introduce statistical methods for modeling MCTs.

Results: We established empirical quantitative relationships between mouse number and measurement accuracy for categorical and continuous efficacy endpoints, and showed that more mice are needed to achieve given accuracy for syngeneic models than for PDXs and CDXs. There is considerable disagreement between methods on calling drug responses as objective response. We then introduced linear mixed models (LMMs) to describe MCTs as clustered longitudinal studies, which explicitly model growth and drug response heterogeneities across mouse models and among mice within a mouse model. Case studies were used to demonstrate the advantages of LMMs in discovering biomarkers and exploring drug's mechanisms of action. We introduced additive frailty models to perform survival analysis on MCTs, which more accurately estimate hazard ratios by modeling the clustered mouse population. We performed computational simulations for LMMs and frailty models to generate statistical power curves, and showed that power is close for designs with similar total number of mice. Finally, we showed that MCTs can explain discrepant results in clinical trials.

Conclusions: Methods proposed in this study can make the design and analysis of MCTs more rational, flexible and powerful, make MCTs a better tool in oncology research and drug development.

Keywords: PDX, CDX, Syngeneic model, Mouse clinical trials, Linear mixed models, Survival analysis, Statistical power, Biomarker

Background

Cancer is a heterogeneous disease with intra- and inter-tumor genomic diversity that determines cancer initiation, progression and treatment. The understandings of cancer biology and the development of therapeutics have been aided greatly by a variety of mouse tumor models,

including cell line-derived xenografts (CDXs), patient derived-xenografts (PDXs), genetically engineered mouse models (GEMMs), cell line- or primary tumor-derived homografts in syngeneic mice and so on (reviewed by [1–4]). These models differ in their generation, host and tumor genomics and biology, availability, and research utilizations. For example, immunotherapies are tested in immunocompetent models such as GEMMs and syngeneic models.

Past decades witnessed the accelerated creation, distribution, profiling and characterization of mouse tumor

* Correspondence: guosheng@crownbio.com; henryli@crownbio.com

¹Crown Bioscience Inc., Suzhou Industrial Park, 218 Xinghu Street, Jiangsu 215028, China

²Crown Bioscience, Inc, 3375 Scott Blvd, Suite 108, Santa Clara, CA 95054, USA

Full list of author information is available at the end of the article



models [5–10]. The abundant collections made it possible to conduct the so-called “mouse clinical trials (MCTs)”, in which a panel of mouse models, dozens to hundreds, are used to evaluate therapeutic efficacy, discover/validate biomarkers, study tumor biology and so on. MCTs demonstrated faithful clinical predictions in multiple studies [6, 11–15]. While most reported MCTs used PDXs, MCTs using other mouse models, such as syngeneic models, are now widely performed as well.

Because of their resemblance to clinical trials, MCTs are often analyzed by methods for clinical trials. For example, overall survival (OS) and progression-free survival (PFS) are estimated by tumor volume increase, Cox proportional hazards models are used for survival analysis, response categories are defined by tumor volume change and objective response rate (ORR) is calculated [6, 13, 16]. However, MCTs differ from clinical trials in many ways. (1) In an oncology clinical trial, a patient is enrolled in only one arm, while in a MCT, multiple mice bearing tumor from the same mouse model are made so that mice can be placed in all arms. Mice from the same mouse model capture intra-tumor heterogeneity for tumor growth and drug response, and mice from different mouse models capture inter-tumor heterogeneity. Measurement error can be quantified when multiple mice are used in each arm. Furthermore, since there are mice of same mouse models in both arms, they themselves can serve as control across arms for better measurement of drug efficacy. (2) tumor volumes are routinely measured every few days; (3) mouse models are usually characterized with genomic/pharmacology/histopathology annotations; (4) MCTs are done in labs that reduces/removes various noise and inconvenience encountered in clinical trials, such as dropouts, long trial time and concomitant medication.

In this study, we combine empirical data analysis, statistical modeling and computational simulations to address some key issues for MCTs, including the determination of animal numbers (number of mouse models and number of mice per mouse model), statistical power calculation, quantification of efficacy difference between mice/mouse models/drugs, survival analysis, biomarker discovery/validation with and beyond simple efficacy readouts, handling of mouse dropouts, missing data and difference in tumor growth rates, study of mechanisms of action (MoA) for drugs. We will also show MCTs can explain discrepant clinical trial results.

Methods

Mouse models, studies and transcriptomic profiling

The establishment of mouse models and the conduct of mouse efficacy studies were described previously [17–19]. Briefly, for PDX models, freshly resected

patient tumors were sliced into roughly $3 \times 3 \times 3 \text{ mm}^3$ chunks and engrafted subcutaneously on the flanks of immunocompromised mice (BALB/c, NOD/SCID, NOG, etc.). Tumor growth was monitored by a caliper twice a week to establish the first passage of a PDX model. Tumor was harvested for next round of engraftment when it reached $500\text{--}700 \text{ mm}^3$ ($1/2 \text{ length} \times \text{width}^2$). A series of engraftment produced subsequent passages of the model. For CDX and syngeneic models, cell suspension ($0.1\text{--}5 \times 10^6$ cell/mouse) was injected into immunocompromised mice and immunocompetent mice (C57BL/6, BALB/c, etc.), respectively, to induce tumor. Pharmacological dosing started when a tumor was normally $100\text{--}300 \text{ mm}^3$, tumor volume was measured twice a week until the tumor was reaching 3000 mm^3 , by then the mouse was euthanized. All animal studies were conducted at Crown Bioscience SPF facility under sterile conditions and were in strict accordance with the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. Protocols of all studies were approved by the Committee on the Ethics of Animal Experiments of Crown Bioscience, Inc. (Crown Bioscience IACUC Committee). Mouse models and cell lines were profiled by RNA-seq on Illumina HiSeq series platforms by certified service providers, as previously described [7].

Categorical efficacy endpoints in mouse studies

Four categorical endpoint methods were evaluated, including the Response Evaluation Criteria In Solid Tumors (RECIST) criteria [20], a 3-category or 3-cat method [13], the 4-response mRECIST criterion [6], and a 5-category or 5-cat method [16]. Briefly, the RECIST-based criterion categorizes drug responses into complete response (CR), partial response (PR), stable disease (SD) and progressive disease (PD) based on relative tumor volume, or RTV, at a later day relative to treatment initiation day (CR: $\text{RTV} = 0$, PR: $0 < \text{RTV} \leq 0.657$, SD: $0.657 < \text{RTV} \leq 1.728$, PD: $\text{RTV} > 1.728$). Metastasis is not considered because it rarely occurs in subcutaneous implantation. The 3-cat method classifies response into PD, SD and objective response (OR) based RTV as well (OR: $\text{RTV} \leq 0.65$, PD: $\text{RTV} \geq 1.35$, SD: $0.65 < \text{RTV} < 1.35$). The mRECIST method considers tumor growth kinetics 10 days after treatment initiation and classifies responses into CR, PR, SD and PD using two RTV-based quantities: best response and best average response. The 5-cat method classifies responses into maintained CR (MCR), CR, PR, SD and PD based on RTV (PD: $\text{RTV} > 0.50$ during the study period and $\text{RTV} > 1.25$ at end of study, SD: $\text{RTV} > 0.50$ during the study period and $\text{RTV} \leq 1.25$ at end of study, PR: $0 < \text{RTV} \leq 0.50$ for at least one time point, CR: $\text{RTV} = 0$ for at least one time point, MCR:

RTV = 0 at end of study). In the definitions of MCR and CR, we also use RTV = 0 to designate disappearance of measurable tumor mass to replace the convention ($TV < 0.10 \text{ cm}^3$) used in Houghton et al., 2007. For all 4 methods, the admissible initial tumor volume is 50~300mm³. Objective response is defined as OR, CR + PR, MCR + CR + PR in the 3-cat, RECIST/mRECIST and 5-cat methods, respectively.

Continuous efficacy endpoints in mouse studies

We briefly describe 4 continuous endpoints here. (a) Progression-free survival (PFS) is defined as tumor volume doubling time and obtained by linear intrapolation on tumor growth data. Specifically, if the PFS is between day d_1 and day d_2 , then it is $d_1 + (d_2 - d_1)(2TV_0 - TV_1)/(TV_2 - TV_1)$ where TV_1 , TV_2 and TV_0 are tumor volumes at d_1 , d_2 and treatment initiation day. (b) RTV ratio is the ratio of RTV between drug group and vehicle group at a specific day d and equals RTV_t / RTV_c , where RTV_t is the relative tumor volume between day d and treatment initiation day for the drug treatment group, and RTV_c is accordingly defined for the vehicle group. (c) Tumor growth inhibition (TGI) has several definitions, it can be defined as $1 - RTV_t / RTV_c$, or as $1 - \Delta T / \Delta C$ where ΔT and ΔC are tumor volume changes relative to initial volume for drug group and vehicle group, respectively, at a specific day. (d) The ratio of growth rates between drug group and vehicle group is defined as k_t / k_c where k_t and k_c are the growth rates obtained by modeling tumor growth data for the two groups by Eq. 1. More general, we can introduce a new endpoint called AUC ratio, which reduces to ratio of growth rates when tumor grows under exponential kinetics (Fig. S5). Unique treatment models with at least 10 mice were used to calculate continuous endpoints, including 621 unique treated PDXs, 739 CDXs and 438 syngeneic models.

Modeling tumor growth

Tumor growth under exponential kinetics is modeled by

$$TV_d = TV_0 e^{kd} \tag{1}$$

Where TV_0 is the initial tumor volume, TV_d is the tumor volume at day d , and k is the tumor growth rate. A logarithmic transformation gives

$$\ln TV_d = \ln(TV_0) + kd \tag{2}$$

Linear mixed models for the cisplatin dataset

A general model can be specified for tumor volume, in log scale, at day t for mouse i within PDX j as follows:

$$\begin{aligned} \log TV_{ij} = & \beta_0 + \beta_1 \times Day_t + \beta_2 \times Day_t \\ & \times CancerTypeGA_j + \beta_3 \times Day_t \\ & \times CancerTypeLU_j + \beta_4 \times Day_t \\ & \times Treatment_{ij} + \beta_5 \times Day_t \\ & \times CancerTypeGA_j \times Treatment_{ij} \\ & + \beta_6 \times Day_t \times CancerTypeLU_j \\ & \times Treatment_{ij} + u_{0j} + u_{1j} \times Day_t \\ & + u_{(0|j)} + u_{(1|j)} \times Day_t + \epsilon_{ij} \end{aligned} \tag{3}$$

LU is lung cancer, GA is gastric cancer and ES is esophageal cancer. The model uses vehicle in ES as the reference. There are 6 fixed effects: β_0 for the intercept, β_1 for the time slope, β_2 and β_3 quantify the growth rate difference of GA and LU with respect to ES, β_4 measures cisplatin effect, β_5 and β_6 measures if GA and LU respond differently to cisplatin. The model also has 5 random effects, including the residual ϵ_{ij} . In a MCT, we view the cohort of PDXs as random samples from a PDX or patient population, therefore, they have different growth rates, which is modeled by random effect u_{1j} associated with the time slope. Similarly, we model growth difference for mice within a PDX by the random effect $u_{1i|j}$. Mice and PDX may have different starting tumor volumes, modeled by the two random effects on intercept u_{0j} and $u_{0i|j}$.

Power calculation based on computational simulation

Power calculation was based on parameters (e.g., variance and covariance of random effects) estimated from fitting the cisplatin dataset by a LMM:

$$\begin{aligned} \log TV_{ij} = & \beta_0 + \beta_1 \times Day_t + \beta_2 \times Day_t \\ & \times Treatment_{ij} + u_{0j} + u_{1j} \times Day_t \\ & + u_{(0|j)} + u_{(1|j)} \times Day_t + \epsilon_{ij} \end{aligned} \tag{4}$$

At significance level $\alpha = 0.05$, we obtained power curves by simulations for $\beta_2/\beta_1 = -0.1$ to -0.9 , that is, drug treatment reduces tumor growth rate by 10 to 90%.

Additive frailty models for survival analysis

In the additive frailty model, the hazard function for the j -th mouse of the i -th mouse model is given by

$$h_{ij}(t) = h_0(t) \exp(u_i + (w + v_i)T_{ij} + \beta^T X_i) \tag{5}$$

where $h_0(t)$ is the baseline hazard function. Parameter u_i is the random effect (the first frailty term) associated with the i -th mouse model that captures its characteristic growth, thus survival behavior, without drug treatment. Parameter v_i is the random effect (the second frailty term) associated with the i -th mouse model that depicts its drug response. Parameter w measures the drug treatment effect on all mouse models. T_{ij} is the treatment variable and equals 0 for the vehicle treatment and 1 for the drug treatment; X_i is a vector for the mouse model's covariates, e.g., cancer type and genomic

features; β^T is the parameter vector quantifying the fixed effects of the covariates. The two random effects u_i and v_i assume a bivariate normal distribution with zero means, variance σ^2 and τ^2 , and covariance $\rho\sigma\tau$. If the two random effects u_i and v_i are removed, the model reduces to the Cox proportional hazards model. Model fitting was done by the R package frailtypack (version 2.12.6), assuming Weibull distribution for the hazard function [21].

Linear mixed models for the biomarker discovery

The following LMM is used for single-gene biomarker discovery by fitting efficacy data from a MCT:

$$\begin{aligned} \log TV_{tij} = & \beta_0 + \beta_1 \times Day_t + \beta_2 \times Day_t \\ & \times Gene_j + \beta_3 \times Day_t \times Treatment_{ij} \\ & + \beta_4 \times Day_t \times Gene_j \times Treatment_{ij} \\ & + u_{0j} + u_{1j} \times Day_t + u_{(0|j)} + u_{(1|j)} \\ & \times Day_t + \varepsilon_{tij} \end{aligned} \quad (6)$$

In this model, *Gene* is a covariate for the genomic status (expression, mutation, copy number variation, etc.) of a gene.

Gene list enrichment analysis

A list of top ranked genes were used as input to the Enrichr web server (<http://amp.pharm.mssm.edu/Enrichr/>) for their enrichment in the “Reactome 2016” pathway database and in the “GO Biological Process 2018” database [22]. Adjusted *p*-values were used to rank enriched pathways and biological processes.

Protein-protein interaction network analysis

A list of top ranked genes were analyzed for protein-protein interactions in the STRING database (version 10.5 at <https://string-db.org>) [23]. Default settings were used except the value for “minimum required interaction score” changed from “medium confidence (0.400)” to “high confidence (0.700)”.

Results

Determining number of mice for categorical responses

We collected tumor volume data under drug treatment for 26127 mice from 2883 unique treatment PDXs, 11139 mice from 1219 unique treatment CDXs, and 5945 mice from 637 unique treatment syngeneic models. A unique treatment model is a mouse model treated by a drug in a study. Every unique treatment has at least 8 mice. Categorical drug response was determined by 4 methods (see Materials and Methods), and we illustrate the results using the mRECIST criteria, which classifies drug response into 4 categories: complete response (CR), partial response (PR), stable disease (SD), and progressive disease (PD). For each unique treatment model, its

response is the majority response of all mice. We observed that individual mouse responses matched the majority response most often for PD: 90% for PDXs, 95% for CDXs and syngeneic models (Fig. 1a-c). The other 3 response categories exhibit lower concordance, particularly so for syngeneic models. Of the 10 unique treatment syngeneic models classified as CR, only half of the mice had complete response as well, while 17% of mice were PD and resistant to treatment. Such polarized response pattern is observed in the other 3 methods, too (Additional file 1: Figure S1-S3). Large variance exists for all 4 response categories. For example, only about 70% of individual responses matched the majority response for a third of the 107 unique treatment PDX models categorized as CR, although the average is 83%.

Measurement accuracy increases with number of mice. We randomly sampled n ($n = 1, 3, 5, 7$) mice from all the mice in a treatment and obtained a majority response, which was then compared with the actual majority response. The procedure was repeated for 1,000 times to generate statistical results (Fig. 1d-f). Accuracy increases with mouse number for all 4 categories, and their unweighted average is highest in CDXs, which is slightly higher than PDXs, while syngeneic models have much lower accuracy (Fig. 1g). Therefore, more mice are needed for syngeneic models to achieve similar accuracy as PDXs/CDXs. For example, accuracy is comparable between syngeneic studies with 5 mice per model and PDX/CDX studies with 1 mouse per model. Similar patterns are also seen in the other 3 methods (Additional file 1: Figure S1-S3).

All the 4 methods categorize responses based on relative tumor volume (RTV) at a later day to treatment initiation day, but differ in specific thresholds. As such, a unique treatment model can be categorized differently. We found that there is a good overlapping for unique treatment models classified as objective response between the 4 methods (Fig. 1h-j), and their objective response rates (ORR) are similar. (Additional file 1: Table S1). Nevertheless, there are many models only unique to some methods as OR, cautioning method-specific bias and applicability. For example, the mRECIST considers averaging tumor reduction for a period of time, therefore, a unique treatment model can be classified as PD even though tumor completely disappears at end of study (Additional file 1: Figure S4).

Determining number of mice for continuous responses

Drug efficacy can be measured by continuous responses, some are direct adaptation of clinical endpoints (e.g., PFS and OS), others are unique to mouse studies that use data from both vehicle and drug treatment groups (e.g., RTV ratio between drug and vehicle groups). We calculated the estimation errors of PFS and RTV ratio

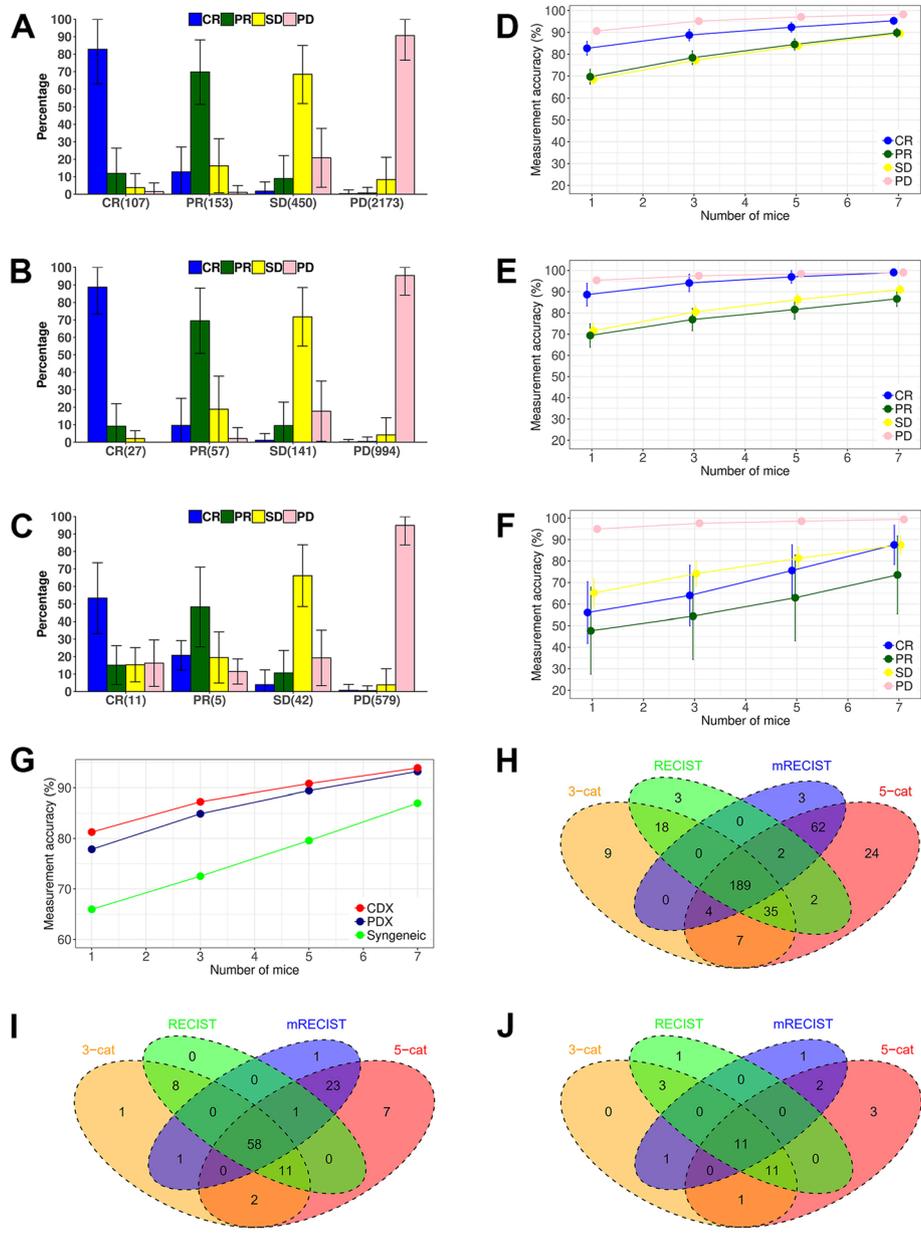


Fig. 1 Mouse number and measurement accuracy of categorical responses defined by the mRECIST criteria. (a-c): individual mouse response and majority response in PDX (a), CDX (b) and syngeneic models (c), x axis is the number of majority response from 4 response categories (CR: complete response, PR: partial response, SD: stable disease, PD: progressive disease), y axis is the percentage of individual mouse response relative to the majority (average \pm s.d.). There are 26,127 mice in 2,883 unique treatment PDX models, 11,139 mice in 1,219 unique treatment CDX models, and 5,945 mice in 637 unique treatment syngeneic models. Each unique treatment model had at least 8 mice. (d-g): measurement accuracy increases with number of mice for PDX (d), CDX (e) and syngeneic models (f). For each unique treatment model, the majority response of n ($n = 1, 3, 5, 7$ in x axis) randomly sampled mice was obtained to see if it agreed with the actual majority response. The procedure was repeated 1,000 times to obtain the accuracy—percentage of times (average \pm s.d.) that they agreed—for the 4 response categories, whose unweighted average is shown in (g). (h-j): Venn diagram showing the overlap of unique treatment PDX models classified as objective response by 4 categorical methods in PDX (h), CDX (i), and syngeneic models (j). Objective response is OR in the 3-cat method, CR + PR in the mRECIST and RECIST methods, MCR + CR + PR in the 5-cat method

computed from n ($n = 1$ to 9) mice randomly sampled from the ≥ 10 mice in a study, and obtained the quantitative relationship between estimation errors and mouse numbers (Fig. 2). For each n , we obtained the empirical

cumulative density function (ECDF) with respect to percentage error of PFS estimate for PDX, CDX and syngeneic models (Fig. 2a-c), and with respect to the absolute error of RTV ratio estimate for the three types

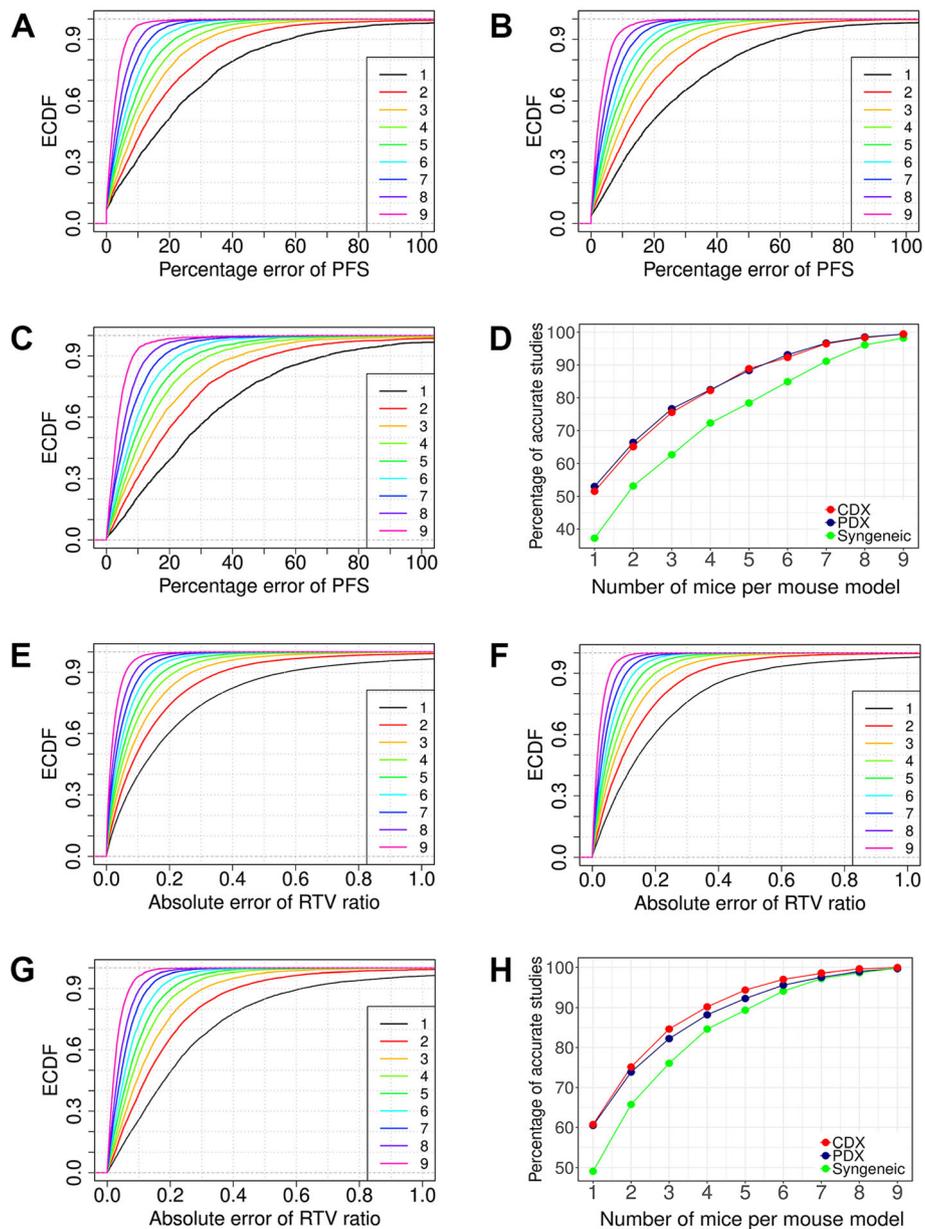


Fig. 2 Determining mouse numbers for continuous responses. (a-c): Progression-free survival, or PFS, calculated from n mice ($n = 1$ to 9) randomly sampled from a unique treatment model with at least 10 mice shows relative deviation to the PFS calculated from all mice in PDX (a), CDX (b), and syngeneic models (c), x axis is the percent error of PFS, and y axis is the empirical cumulative density function (ECDF) estimated from the random samplings for each n . Percent error of PFS decreases with increased number of mice, and the error is larger for syngeneic models than PDXs/CDXs. (d): Percentages of unique treatment models with percent error less than 20% in the 3 types of mouse models. (e-g): RTV ratio between drug and vehicle groups, calculated from n mice ($n = 1$ to 9) randomly sample from a study with at least 10 mice in both drug and vehicle groups, shows deviation to the RTV ratio calculated from all mice in both groups in PDX (e), CDX (f), and syngeneic models (g), x axis is the absolute error, and y axis is the empirical cumulative density function (ECDF) estimated from the random samplings for each n . Absolute error of RTV ratio decreases with increased number of mice, and the error is larger for syngeneic models than PDXs/CDXs. (h): Percentages of studies with absolute error less than 0.2 in the 3 types of mouse models

of models (Fig. 2e-g). Large estimation errors are inherent to small sample sizes, particularly so for syngeneic models. For example, percent error of PFS is greater than 20% for 63% syngeneic mice and for about half of

PDX/CDX mice (Fig. 2d). Estimation errors are reduced sharply by addition of more mice when n is small. For RTV ratio, 3 mice in both drug and vehicle group already lift mice with absolute error < 0.2 from 60% to

above 80% for PDXs/CDXs (Fig. 2h). Similar results hold for other continuous endpoints as well (Additional file 1: Figure S5).

Modeling MCTs as clustered longitudinal studies

It is convenient to measure drug efficacy by a categorical or continuous endpoint, but those approaches also suffer from loss of information and other drawbacks. For example, it is somewhat arbitrary to choose a day to calculate RTV ratio and TGI; it adds logistic burden to match mice with comparable tumor volume at treatment initiation day [24]; it is difficult to deal with mouse dropouts. These shortcomings can be overcome by modeling MCTs as clustered longitudinal studies, in which a cluster is consisted of all mice of a mouse model so they share genomic profile and have more similar drug response. Each mouse is in a longitudinal study. It can be shown that tumor growth in majority of mice follows exponential kinetics (Additional file 1: Figure S6). Therefore, we can model the clustered longitudinal studies by a 3-level linear mixed model (LMM) on the log-transformed tumor volumes (logTV) and day (Fig. 3a). There are covariates associated with mouse models such as cancer type and genomic features, which can be used for examining efficacy difference on cancers and for discovering predictive biomarkers.

We use one example to demonstrate the modeling of MCTs by LMMs for efficacy evaluation and comparison. In this MCT, cisplatin—a chemotherapy drug—was administered to 42 PDXs (4 mg/kg, weekly dosing for 3 weeks), including 13 esophageal cancers (ES), 21 gastric cancers (GA) and 8 lung cancers (LU), each PDX with 5 to 9 mice (Additional file 1: Figure S7). We fit the efficacy data by a LMM (Eq. 3 in Materials and Methods), which explicitly models tumor growth rate heterogeneity and drug response heterogeneity at both PDX level and mouse level. Model fitting is satisfactorily (Table 1, Additional file 1: Figure S8). We conclude that (1) under vehicle treatment, tumor in GA grows slightly faster than ES, while tumor growth is much faster in LU; (2) cisplatin has comparable efficacy on the 3 cancers (p -values for β_5 and β_6 are > 0.05). The results can be readily visualized from the mean growth curves for the 3 cancers under (Fig. 3b).

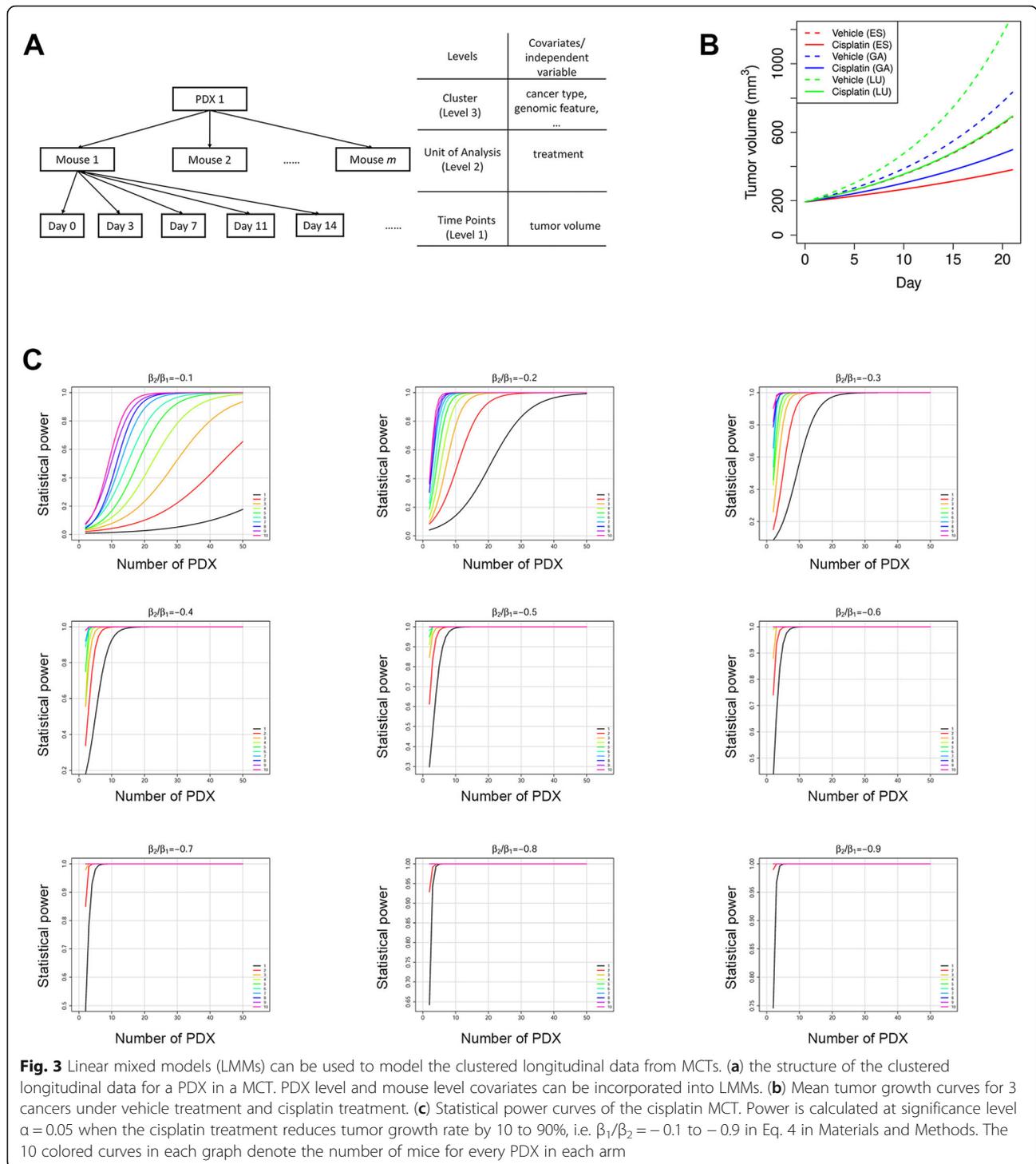
Statistical power and sample size determination in MCTs

Much like clinical trials, rational design of MCTs requires statistical power calculation and sample size determination—number of mouse models and number of mice per mouse model. We demonstrate this under the LMM framework with the following assumptions (1) a balanced $n:n$ design in which there are n (≥ 1) mice in both drug and vehicle groups, and (2) a 21-day trial with tumor volume measured at treatment initiation and then

twice every week to produce 8 data points for every mouse. Drug efficacy is measured by how much drug treatment slows down tumor growth (β_2/β_1 in Eq. 4). Power curves were obtained by computational simulations based on parameters obtained from fitting the cisplatin dataset by Eq. 4 (Fig. 3c).

We observed that if the number of PDXs is the same, more mice per PDX confer better statistical power. For example, to achieve 80% power, we need about 28 PDXs for the 1:1 design (1 mouse each in the vehicle and drug treatment groups), and 11 PDXs for the 3:3 design (3 mice each in the vehicle and drug treatment groups). More importantly, statistical power is comparable for designs with similar number of total mice. For example, when the drug efficacy is 20%, that is, the drug reduces tumor growth rate by 20%, the following designs all achieve 90% power at 0.05 significance level: 36 PDX with 1:1 design, 19 PDXs with 2:2 design, 13 PDXs with 3:3 design, 10 PDXs with 4:4 design, and so on. However, it is important to note that such designs with similar statistical power and total number of mice have different biological implications. A design with a larger number of PDX but fewer mice or even one mouse per PDX can give better representation and measurement of inter-tumor heterogeneity, while a design with a smaller number of PDX but more mice per PDX sacrifices such inter-tumor heterogeneity to give more accurate measurement of drug efficacy for each PDX. It depends on study aims to choose a design. For example, we likely prefer a design with more PDX each with fewer mice for biomarker discovery because it would give us a broader representation of inter-tumor heterogeneity and more genomic datasets to work with. In the extreme case, we can use the 1:1 design if there are many PDXs at disposal—the 1x1x1 approach [6], in which Gao et al. showed that the 1:1 design is effective in biomarker assessment and efficacy evaluation. But for biomarker validation, we may use a design with a limited number of selected PDX models that are predicted to be responsive or resistant, and each PDX should have a relatively high number of mice so that the efficacy measurement is accurate enough to gauge the effectiveness of the biomarker. The design also are constrained by available resource, for example, when there is only a limited number of suitable PDXs, e.g., PDXs carrying a particular mutation or PDXs of a specific subtype, we can increase the number of mice per PDX to boost statistical power.

We also observed that fewer PDXs are needed for a more potent drug to reach same statistical power. For example, to achieve 80% statistical power at 0.05 significance level by the 3:3 design, we need about 40, 11, and 5 PDXs for drugs with 10, 20, and 30% efficacy, respectively. When a drug is potent enough, all $n:n$ designs achieve high power with very small number of PDXs. In



such cases, we use a good number of PDXs not for statistical power but for better representation of tumor heterogeneity.

Survival analysis in MCTs

In clinical trials, patient survival is usually assumed to be independent of each other. In MCTs, this assumption

no longer holds because mice are now clustered within PDXs, and mice of same PDX tend to have more similar survival time, while their survival time between treatments is highly correlated (Fig. 4a). Further, PDXs can vary greatly in growth rate (or hazard) and drug response (Additional file 1: Figure S9). Therefore, we use an additive frailty model to model the heterogeneity on

Table 1 Parameters estimated for the LMM (Eq. 3) of the cisplatin dataset

Fixed-Effect Parameters	Estimate*	p-value
β_0 (Intercept)	5.2641 (0.0257)	0
β_1 (Day)	0.0605 (0.0043)	1.5E-43
β_2 (Day \times CancerTypeGA)	0.0091 (0.0055)	0.098
β_3 (Day \times CancerTypeLU)	0.0297 (0.0071)	2.8E-5
β_4 (Day \times Treatment)	-0.0282 (0.0031)	1.2E-19
β_5 (Day \times CancerTypeGA \times Treatment)	0.0037 (0.0039)	0.35
β_6 (Day \times CancerTypeLU \times Treatment)	-0.0011 (0.0052)	0.84

*parameters estimated by the REML method in the R nlme package

hazard and drug efficacy under the clustered population structure of MCTs (see Eq. 5 in Materials and Methods). The additive frailty model is an extension of the Cox proportional hazards model widely used in clinical trials. It has two frailty terms, the first one u_i quantifies PDX growth rate heterogeneity and the second one v_i measures drug response heterogeneity.

We use the cisplatin MCT to illustrate the utilization of the additive frailty model. Overall survival (OS) is defined as tumor volume tripling time. We fit the cisplatin MCT dataset by Eq. 5, and observed that both frailty terms are significant larger than 0 (Wald test p -value < 0.05), proving that the PDXs grow at different rate and had different responses to cisplatin. In fact, the first frailty term u_i is negatively correlated with tumor growth rate in the vehicle group, as expected ($R^2 = 0.85$, Fig. 4b).

Drug efficacy can be estimated more accurately by excluding the influence of tumor growth heterogeneity and considering drug response heterogeneity, which is measured by the second frailty term v_i . Indeed, the hazard ratio (HR) is estimated to be 0.21 (95% CI: 0.15–0.31), much smaller than that obtained from the Cox proportional hazards model, which gives HR = 0.36 (95% CI, 0.28–0.46) (Fig. 4c). These results show that without considering PDX heterogeneity, drug effect can be severely misestimated.

We performed statistical power analysis for the survival analysis by assuming the n:n designs and using parameters estimated from the cisplatin MCT with Weibull hazard functions (Fig. 4d). Like in LMMs, statistical power is similar for designs with similar total number of mice.

Biomarker discovery in MCTs

Genomic correlation to cetuximab efficacy in solid tumors has been well documented [13], and we previously reported a MCT for a cohort of 20 gastric cancer PDXs, each with 3–10 mice in the vehicle and cetuximab treatment arms. We found that EGFR expression to be a predictive biomarker for cetuximab

on gastric cancer [19]. The cohort is now expanded to 27 PDXs (Additional file 1: Figure S10). We observed a strong correlation between EGFR expression and drug efficacy measured by tumor growth inhibition or TGI (Fig. 5a). When all 18586 genes were ranked from high to low by the absolute value of correlation coefficient between their expression and TGI, EGFR is ranked 157 out of all these genes, demonstrating that such simple methods in biomarker discovery can yield many false positives with seemingly better predictivity than the true biomarker.

We used a LMM that explicitly models a gene's effect on tumor growth to fit the efficacy data (Eq. 6 in Materials and Methods). EGFR stands out as the most significant gene and its p -value, being 1.5×10^{-23} , is at least five orders of magnitude smaller than all other genes (Fig. 5b). EGFR as a predictive biomarker for cetuximab on gastric cancer is supported by a phase 2 clinical trial [25] and a phase 3 clinical trial with data re-interpretation (Additional file 1: Figure S11) [26]. This study shows that simple analysis can produce many false positive hits to hamper biomarker discovery, especially when a drug target is unknown or there are off-target effects, while the more sophisticated LMM method can be superior in biomarker discovery.

Mechanism of action study in MCTs

MCTs are used for drug efficacy evaluation and biomarker discovery, the latter can be facilitated by a better understanding of a drug's mechanism of action (MoA), which helps identify relevant genes, pathways and gene sets, and remove false positive genes that could have higher statistical significance, i.e. lower p -values, in some analysis. Biomarkers constructed from genes selected this way have explicit biological relevance and oftentimes are preferred.

With the readily available genomic and efficacy data from a MCT, MoA studies can be readily performed. Like in biomarker discovery, simple categorical and continuous endpoints, as a gross summary of efficacy, have various drawbacks. For example, the 4 categorical methods only measure efficacy in drug treatment group, ignoring the relative drug-to-vehicle efficacy. RTV ratio and TGI are dependent on calculation day and tumor growth rate (Additional file 1: Figure S12). Again, we can use LMM for a better study of MoA, as shown by the example below.

Irinotecan is a DNA topoisomerase I inhibitor that interrupts cell cycle in the S-phase by irreversibly arresting the replication fork, therefore causing cell death [27]. We conducted a MCT for 16 PDXs (Additional file 1: Figure S13), each PDX with 3 to 10 mice. We modeled the effect of gene expression on drug efficacy by a LMM

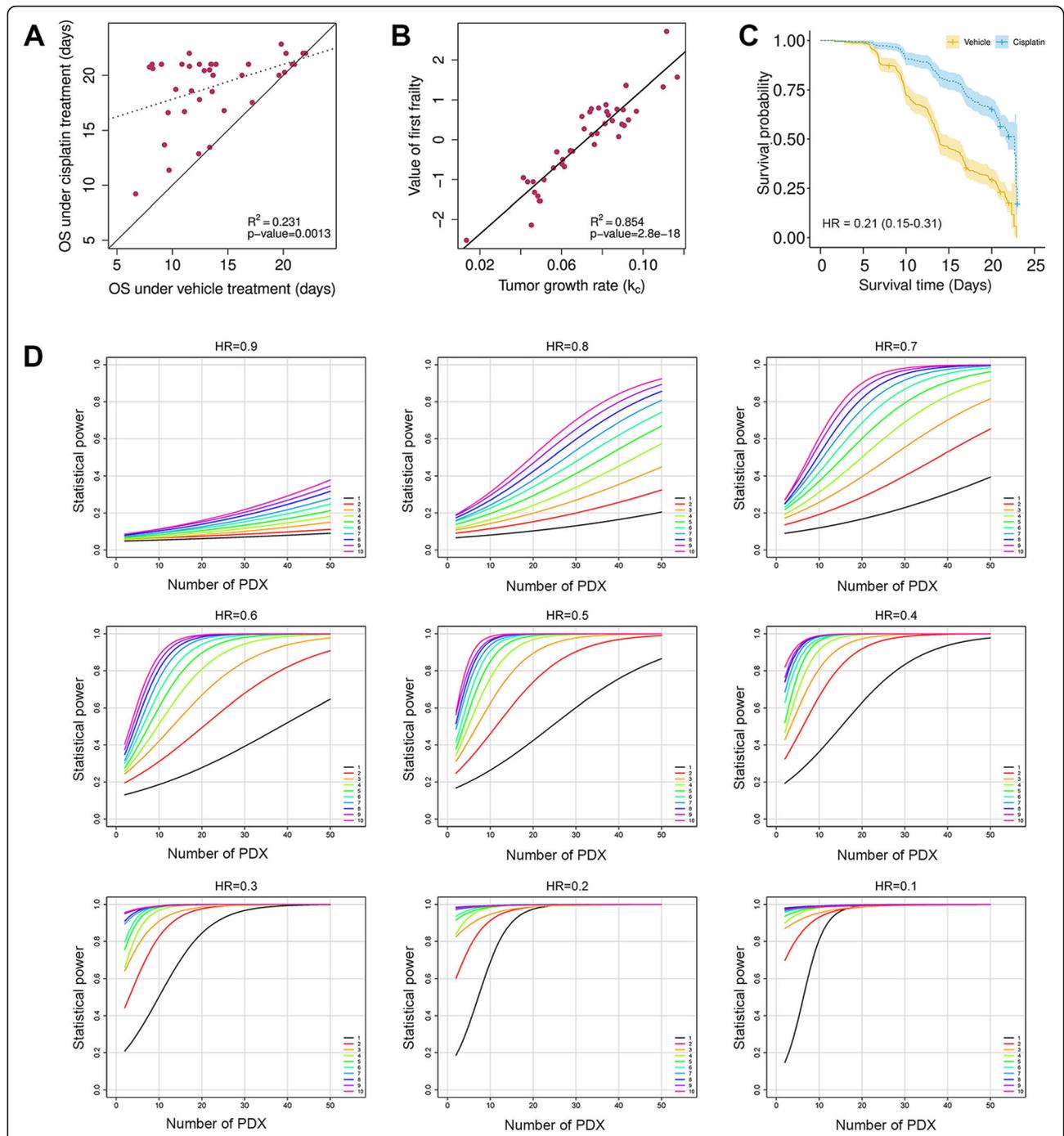


Fig. 4 Survival analysis in a cisplatin MCT. (a) The median progression free survival (PFS) times of PDXs under cisplatin and vehicle treatment are highly correlated. The dotted line is the linear regression lines, and the solid line is a line with unit slope. (b) The first frailty term u_i in Eq. 5 is positively correlated with the tumor growth rate k_c . (c) Survival curves under cisplatin and vehicle treatments. Additive frailty model gives more accurate hazard ratio (HR) than the Cox proportional hazards model whose estimation is 0.36 (95% CI: 0.28–0.46). (d) Statistical power curves at significance level $\alpha = 0.05$ when the hazard ratio is 0.9 to 0.1 for the survival analysis. The 10 colored curves in each graph denote the number of mice per PDX per arm

(Eq. 6). Top ranked genes were highly enriched for the cell cycle pathway R-HSA-160170 in the Reactome 2016 database (Fig. 5c), and for DNA replication initiation

(Gene Ontology annotation GO: 0006270) (Fig. 5d), which perfectly reveals the MoA for irinotecan. A highly connected protein-protein interaction network for cell

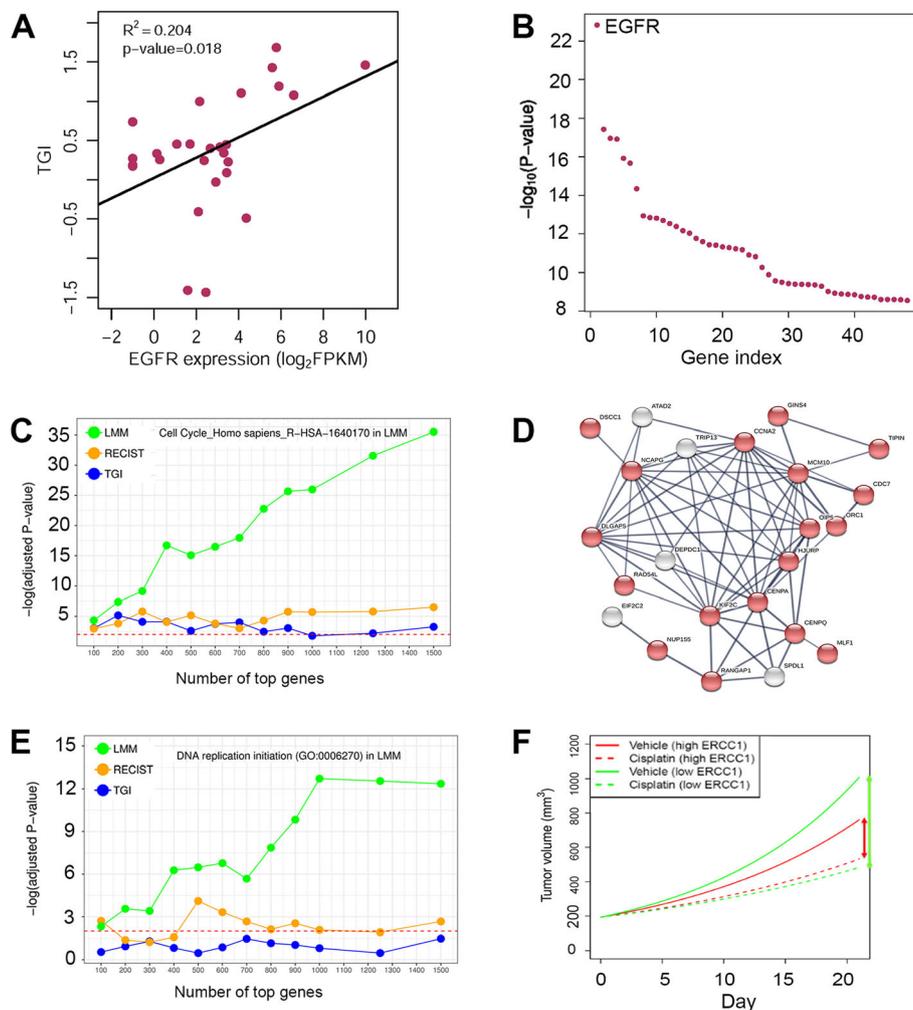


Fig. 5 Biomarker discovery and MoA study in MCTs. (a-b): A MCT of 27 gastric cancer PDXs treated with cetuximab. EGFR is ranked 157th among all genes based on Spearman rank correlation between EGFR expression and TGI (a), but is the top gene in predicting cetuximab efficacy based on a linear mixed model (LMM) (b). (c-e): A MCT of 16 PDXs treated with intraperitoneal injection of Irinotecan (100 mg/kg, once per week for 2 to 3 weeks). (c): R-HAS-160170, the cell cycle pathway in Reactome2016 database, is consistently ranked as the most enriched pathway with 100 to 2000 top genes selected by a LMM, superior to top genes selected by methods based on categorical endpoints (e.g., RECIST, Table S2) and continuous endpoints (e.g., TGI). (d): DNA replication initiation (GO: 0006270) is the most enriched GO term based on top genes selected by the LMM. (e): A highly enriched protein-protein interaction network (p -value < 10^{-16}) consisted of 23 genes in the top 100 genes selected by the LMM. Red-colored nodes are ones involved in cell cycle (GO: 0007049). Dashed horizontal lines in (c-d) denotes p -value = 0.01. (f): Mean tumor growth curves for PDXs with highest and lowest ERCC1 mRNA expression in a MCT of 21 gastric PDXs treated by cisplatin

cycle is also identified from the 100 top ranked genes (Fig. 5e). In contrast, endpoint based methods are far less insightful (Fig. 5c-d, Additional file 1: Table S2-S4).

MCTs can explain paradoxical clinical trial results

Conflicting clinical trial reports exist regarding the role of ERCC1 expression in predicting cisplatin treatment on gastric cancer: some claimed that patients benefit more from low ERCC1 expression [28–32], some stated the opposite [33–35], while still others found no connection at all [36].

In a previous section, we described a cisplatin MCT which included 21 gastric cancer PDXs. We fit the tumor volume data by Eq. 6. Parameter β_2 quantifies how ERCC1 expression affects tumor growth when there is no drug intervention, as seen from the vehicle growth curves (Fig. 5f). Parameter β_4 evaluates how ERCC1 expression impacts cisplatin’s efficacy on tumor growth, as seen by comparing the cisplatin growth curves with corresponding vehicle growth curves. These two parameters are at comparable magnitude but with opposite signs ($\beta_2 = -0.0155$ and $\beta_4 = 0.0136$). Therefore, when ERCC1 expression gets higher, tumor grows

slower, but the benefit of cisplatin treatment is smaller as well (Fig. 5f).

In a clinical trial, patients with low/negative ERCC1 expression would have worse prognosis if they were not treated, and they could benefit more from cisplatin treatment. With treatment, their prognosis is improved, but whether it is better than the prognosis of ERCC1 high/positive patients is undetermined and depends on the trial population, hence we saw conflicting study conclusions.

Discussion

MCTs are population-based efficacy trials mimicking human trials. Multiple mice are usually used per mouse model per arm to improve accuracy of efficacy measurement. For example, Bertotti et al. used 6 mice per PDX per arm in a two-arm MCT with 85 colorectal cancer PDXs to identify HER2 as a therapeutic target in Cetuximab-resistant colorectal cancers [11]. It may also be feasible to use one mouse per model per arm when there is a large number of mouse models, which compensate the loss of measurement accuracy on individual mice [6, 8, 37, 38]. Caution must be exercised to use this approach though, when the number of mouse models is small, or high measurement accuracy of individual mouse models is mandated, or response varies greatly among mice of same mouse models, as commonly observed for immunotherapeutic agents on syngeneic models. Syngeneic models, unlike PDX or CDX that are immunodeficient mouse tumor models, have intact immune system, which likely is the source for large variation of drug response among mice within a syngeneic model, because individual mice can vary greatly in tumor immunity including the levels of T-cell infiltration, Th1 cytokine expression, and immunogenicity [39].

Our study established theoretic foundations for the design and analysis of MCTs. We first investigated tumor growth kinetics. Many complex mathematical models were used to describe tumor growth [40], but might not be particularly advantageous at the expense of more parameters and the need of more data points for model fitting. The exponential growth model is simple, interpretable and linear after a logarithmic transformation, and was shown to be adequate in most cases. Consequently, LMMs can describe nearly all MCTs, using quadratic terms of time if necessary.

We introduced additive frailty models to perform survival analysis for MCTs. The definition of PFS/OS can vary. For example, OS can be defined same as in human trials for leukemia PDXs [8]. For both LMMs and frailty models, we performed power simulations that give concrete recommendations on trial design. In particular, we answered the frequently asked questions, from a statistical perspective, on how many mouse models and how

many mice per model to use, with flexible combination of the two numbers. We emphasize that it is equally important to consider the purpose of MCTs, e.g., biomarker discovery versus biomarker validation, in the study design, and designs with more PDX but fewer mice per PDX (e.g. 1:1 design) have better representation of inter-tumor heterogeneity than ones with fewer PDX but more mice per PDX (e.g., 3:3 design), but the latter gives more accurate measurement of drug efficacy. MCTs can be asymmetric, i.e., unequal numbers of mice in arms. LMMs and frailty models are flexible for covariates, for example, a fixed effect for site can be incorporated if a MCT is conducted at multiple sites.

Conclusions

In conclusion, methods proposed in this study make the design and analysis of MCTs more rational, flexible and powerful when mouse tumor models are used in oncology research and drug development.

Additional file

Additional file 1: Figure S1. Mouse number and measurement accuracy of categorical responses defined by the RECIST criteria. **Figure S2.** Mouse number and measurement accuracy of categorical responses defined by the 3-cat criterion. **Figure S3.** Mouse number and measurement accuracy of categorical responses defined by the 5-cat criterion. **Figure S4.** A unique treatment model classified as PD by mRECIST method, though tumor completely disappeared at end of study. **Figure S5.** AUC ratio as a continuous metric for MCTs. **Figure S6.** (a) Distribution of coefficient of determination between log-transformed tumor volume and day for PDX mice under vehicle treatment. **Figure S7.** Growth curves of 42 PDXs under vehicle treatment (a) and cisplatin treatment (b). **Figure S8.** Fitting diagnostics of the linear mixed model in Eq. 3 for the cisplatin MCT dataset (cf. Fig. S7). **Figure S9.** Tumor volume doubling time in PDXs for 10 cancers. **Figure S10.** Growth curves of 27 PDXs under (a) vehicle treatment and (b) cetuximab treatment (1 mg/mouse, intraperitoneal injection, once per week). **Figure S11.** In the EXPAND phase III trial (1), for patients with IHC score greater than ~200, the 7 patients receiving cetuximab in addition to had significantly longer (a) PFS and (b) OS than the 19 patients receiving only chemotherapies. **Figure S12.** TGI is a growth rate biased and time-dependent efficacy metric. **Figure S13.** Growth curves of 16 PDXs under (a) vehicle treatment and (b) Irinotecan treatment (100 mg/kg, intraperitoneal injection, once per week for 2–3 weeks. **Table S1.** Objective response rate (ORR) in 4 categorizing methods. **Table S2.** Irinotecan response of 16 PDX models by 4 categorical endpoint methods. **Table S3.** Most enrichment pathways in Reactome 2016 database for the Irinotecan MCT. **Table S4.** Most enrichment terms in GO Biological Processes for the Irinotecan MCT. (PDF 2790 kb)

Abbreviations

CR: Complete response; CDX: Cell line-derived xenograft; EGFR: Epidermal growth factor receptor; ERCC: Excision repair 1 (ERCC1); ES: Esophageal cancer; GA: Gastric cancer; GEMM: Genetically engineered mouse model; HER2: Human epidermal growth factor receptor 2; HR: Hazard ratio; LMM: Linear mixed model; LU: Lung cancer; MCT: Mouse clinical trial; MoA: Mechanism of action; OR: Objective response; OS: Overall survival; PD: Progressive disease; PDX: Patient-derived xenograft; PFS: Progression-free survival; PR: Partial response; RCT: Randomized controlled trial; RECIST: Response Evaluation Criteria in Solid Tumors; RTV: Relative tumor volume; SD: Stable disease; TGI: Tumor growth inhibition

Acknowledgements

The authors would like to express their gratitude to the *in vivo* team members at the Translational Oncology Division of Crown Bioscience, Inc. for contributing all the *in vivo* efficacy data.

Authors' contributions

SG and QL designed the study and wrote the manuscript, SG, XJ, BM analyzed the data. All authors have read and approved the manuscript.

Funding

Not applicable.

Availability of data and materials

Datasets used in the current study are available from the corresponding authors on reasonable request.

Ethics approval and consent to participate

All animal studies were conducted at Crown Bioscience SPF facility under sterile conditions and were in strict accordance with the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. Protocols of all studies were approved by the Committee on the Ethics of Animal Experiments of Crown Bioscience, Inc. (Crown Bioscience IACUC Committee).

Consent for publication

Not applicable.

Competing interests

This research was funded by Crown Bioscience Inc. and all authors were employees thereof at the time the study was performed. The authors declare no other competing interests.

Author details

¹Crown Bioscience Inc., Suzhou Industrial Park, 218 Xinghu Street, Jiangsu 215028, China. ²Crown Bioscience, Inc, 3375 Scott Blvd, Suite 108, Santa Clara, CA 95054, USA. ³State Key Laboratory of Natural and Biomimetic Drugs, Peking University, Beijing 100191, China.

Received: 8 December 2018 Accepted: 5 July 2019

Published online: 22 July 2019

References

- Day CP, Merlino G, Van Dyke T. Preclinical mouse cancer models: a maze of opportunities and challenges. *Cell*. 2015;163(1):39–53.
- Khaled WT, Liu P. Cancer mouse models: past, present and future. *Semin Cell Dev Biol*. 2014;27:54–60.
- Li QX, Feuer G, Ouyang X, An X. Experimental animal modeling for immuno-oncology. *Pharmacol Ther*. 2017;173:34–46.
- Walrath JC, Hawes JJ, Van Dyke T, Reilly KM. Genetically engineered mouse models in cancer research. *Adv Cancer Res*. 2010;106:113–64.
- Byrne AT, Alferez DG, Amant F, Annibaldi D, Arribas J, Biankin AV, Bruna A, Budinska E, Caldas C, Chang DK, et al. Interrogating open issues in cancer precision medicine with patient-derived xenografts. *Nat Rev Cancer*. 2017;17(4):254–68.
- Gao H, Korn JM, Ferretti S, Monahan JE, Wang Y, Singh M, Zhang C, Schnell C, Yang G, Zhang Y, et al. High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat Med*. 2015;21(11):1318–25.
- Guo S, Qian W, Cai J, Zhang L, Wery JP, Li QX. Molecular pathology of patient tumors, patient-derived xenografts, and Cancer cell lines. *Cancer Res*. 2016;76(16):4619–26.
- Townsend EC, Murakami MA, Christodoulou A, Christie AL, Koster J, DeSouza TA, Morgan EA, Kallgren SP, Liu H, Wu SC, et al. The public repository of xenografts enables discovery and randomized phase II-like trials in mice. *Cancer Cell*. 2016;30(1):183.
- Krupke DM, Begley DA, Sundberg JP, Richardson JE, Neuhauser SB, Bult CJ. The mouse tumor biology database: a comprehensive resource for mouse models of human Cancer. *Cancer Res*. 2017;77(21):e67–70.
- Stewart E, Federico SM, Chen X, Shelat AA, Bradley C, Gordon B, Karlstrom A, Twarog NR, Clay MR, Bahrami A, et al. Orthotopic patient-derived xenografts of paediatric solid tumours. *Nature*. 2017;549(7670):96–100.
- Bertotti A, Migliardi G, Galimi F, Sassi F, Torti D, Isella C, Cora D, Di Nicolantonio F, Buscarino M, Petti C, et al. A molecularly annotated platform of patient-derived xenografts ("xenopatient") identifies HER2 as an effective therapeutic target in cetuximab-resistant colorectal cancer. *Cancer Discov*. 2011;1(6):508–23.
- Migliardi G, Sassi F, Torti D, Galimi F, Zanella ER, Buscarino M, Ribero D, Muratore A, Massucco P, Pisacane A, et al. Inhibition of MEK and PI3K/mTOR suppresses tumor growth but does not cause tumor regression in patient-derived xenografts of RAS-mutant colorectal carcinomas. *Clin Cancer Res*. 2012;18(9):2515–25.
- Bertotti A, Papp E, Jones S, Adleff V, Anagnostou V, Lupo B, Sausen M, Phallen J, Hruban CA, Tokheim C, et al. The genomic landscape of response to EGFR blockade in colorectal cancer. *Nature*. 2015;526(7572):263–7.
- Bardelli A, Corso S, Bertotti A, Hobor S, Valtorta E, Siravegna G, Sartore-Bianchi A, Scala E, Cassingena A, Zecchin D, et al. Amplification of the MET receptor drives resistance to anti-EGFR therapies in colorectal cancer. *Cancer Discov*. 2013;3(6):658–73.
- Yao YM, Donoho GP, Iversen PW, Zhang Y, Van Horn RD, Forest A, Novosiadly RD, Webster YW, Ebert P, Bray S, et al. Mouse PDX trial suggests synergy of concurrent inhibition of RAF and EGFR in colorectal Cancer with BRAF or KRAS mutations. *Clin Cancer Res*. 2017;23(18):5547–60.
- Houghton PJ, Morton CL, Tucker C, Payne D, Favours E, Cole C, Gorlick R, Kolb EA, Zhang W, Lock R, et al. The pediatric preclinical testing program: description of models and early testing results. *Pediatr Blood Cancer*. 2007;49(7):928–40.
- Yang M, Shan B, Li Q, Song X, Cai J, Deng J, Zhang L, Du Z, Lu J, Chen T, et al. Overcoming erlotinib resistance with tailored treatment regimen in patient-derived xenografts from naive Asian NSCLC patients. *Int J Cancer*. 2013;132(2):E74–84.
- Yang M, Xu X, Cai J, Ning J, Wery JP, Li QX. NSCLC harboring EGFR exon-20 insertions after the regulatory C-helix of kinase domain responds poorly to known EGFR inhibitors. *Int J Cancer*. 2016;139(1):171–6.
- Zhang L, Yang J, Cai J, Song X, Deng J, Huang X, Chen D, Yang M, Wery JP, Li S, et al. A subset of gastric cancers with EGFR amplification and overexpression respond to cetuximab therapy. *Sci Rep*. 2013;3:2992.
- Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancy J, Arbuck S, Gwyther S, Mooney M, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009;45(2):228–47.
- Rondeau V, Gonzalez JR. frailtypack: a computer program for the analysis of correlated failure time data using penalized likelihood estimation. *Comput Methods Prog Biomed* 2005, 80(2):154–164.
- Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013;14:128.
- Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015;43(Database issue):D447–52.
- Laajala TD, Jumppanen M, Huhtaniemi R, Fey V, Kaur A, Knuutila M, Aho E, Oksala R, Westermarck J, Makela S, et al. Optimized design and analysis of preclinical intervention studies *in vivo*. *Sci Rep*. 2016;6:30723.
- Zhang X, Xu J, Liu H, Yang L, Liang J, Xu N, Bai Y, Wang J, Shen L. Predictive biomarkers for the efficacy of cetuximab combined with cisplatin and capecitabine in advanced gastric or esophagogastric junction adenocarcinoma: a prospective multicenter phase 2 trial. *Med Oncol*. 2014;31(10):226.
- Lordick F, Kang YK, Chung HC, Salzman P, Oh SC, Bodoky G, Kurteva G, Volovat C, Moiseyenko VM, Gorbunova V, et al. Capecitabine and cisplatin with or without cetuximab for patients with previously untreated advanced gastric cancer (EXPAND): a randomised, open-label phase 3 trial. *Lancet Oncol*. 2013;14(6):490–9.
- Xu Y, Villalona-Calero MA. Irinotecan: mechanisms of tumor resistance and novel strategies for modulating its activity. *Ann Oncol*. 2002;13(12):1841–51.
- De Dosso S, Zanellato E, Nucifora M, Boldorini R, Sonzogni A, Biffi R, Fazio N, Bucci E, Beretta O, Crippa S, et al. ERCC1 predicts outcome in patients with gastric cancer treated with adjuvant cisplatin-based chemotherapy. *Cancer Chemother Pharmacol*. 2013;72(1):159–65.
- Hirakawa M, Sato Y, Ohnuma H, Takayama T, Sagawa T, Nobuoka T, Harada K, Miyamoto H, Sato Y, Takahashi Y, et al. A phase II study of neoadjuvant combination chemotherapy with docetaxel, cisplatin, and S-1 for locally advanced resectable gastric cancer: nucleotide excision repair (NER) as potential chemoresistance marker. *Cancer Chemother Pharmacol*. 2013;71(3):789–97.

30. Kwon HC, Roh MS, Oh SY, Kim SH, Kim MC, Kim JS, Kim HJ. Prognostic value of expression of ERCC1, thymidylate synthase, and glutathione S-transferase P1 for 5-fluorouracil/oxaliplatin chemotherapy in advanced gastric cancer. *Ann Oncol.* 2007;18(3):504–9.
31. Metzger R, Leichman CG, Danenberg KD, Danenberg PV, Lenz HJ, Hayashi K, Groshen S, Salonga D, Cohen H, Laine L, et al. ERCC1 mRNA levels complement thymidylate synthase mRNA levels in predicting response and survival for gastric cancer patients receiving combination cisplatin and fluorouracil chemotherapy. *J Clin Oncol.* 1998;16(1):309–16.
32. Miura JT, Xiu J, Thomas J, George B, Carron BR, Tsai S, Johnston FM, Turaga KK, Gamblin TC. Tumor profiling of gastric and esophageal carcinoma reveal different treatment options. *Cancer Biol Ther.* 2015;16(5):764–9.
33. Baek SK, Kim SY, Lee JJ, Kim YW, Yoon HJ, Cho KS. Increased ERCC expression correlates with improved outcome of patients treated with cisplatin as an adjuvant therapy for curatively resected gastric cancer. *Cancer Res Treat.* 2006;38(1):19–24.
34. Bamias A, Karina M, Papakostas P, Kostopoulos I, Bobos M, Vourli G, Samantas E, Christodoulou C, Pentheroudakis G, Pectasides D, et al. A randomized phase III study of adjuvant platinum/docetaxel chemotherapy with or without radiation therapy in patients with gastric cancer. *Cancer Chemother Pharmacol.* 2010;65(6):1009–21.
35. Kim KH, Kwon HC, Oh SY, Kim SH, Lee S, Kwon KA, Jang JS, Kim MC, Kim SJ, Kim HJ. Clinicopathologic significance of ERCC1, thymidylate synthase and glutathione S-transferase P1 expression for advanced gastric cancer patients receiving adjuvant 5-FU and cisplatin chemotherapy. *Biomarkers.* 2011;16(1):74–82.
36. Sonnenblick A, Rottenberg Y, Kadouri L, Wygoda M, Rivkind A, Vainer GW, Peretz T, Hubert A. Long-term outcome of continuous 5-fluorouracil/cisplatin-based chemotherapy followed by chemoradiation in patients with resected gastric cancer. *Med Oncol.* 2012;29(5):3035–8.
37. Williams JA. Using PDX for preclinical Cancer drug discovery: the evolving field. *J Clin Med.* 2018;7(3):41.
38. Murphy B, Yin H, Maris JM, Kolb EA, Gorlick R, Reynolds CP, Kang MH, Keir ST, Kurmasheva RT, Dvorchik I, et al. Evaluation of alternative in vivo drug screening methodology: a single mouse analysis. *Cancer Res.* 2016;76(19):5798–809.
39. Mosely SI, Prime JE, Sainson RC, Koopmann JO, Wang DY, Greenawalt DM, Ahdesmaki MJ, Leyland R, Mullins S, Pacelli L, et al. Rational selection of syngeneic preclinical tumor models for immunotherapeutic drug discovery. *Cancer Immunol Res.* 2017;5(1):29–41.
40. Benzekry S, Lamont C, Beheshti A, Tracz A, Ebos JM, Hlatky L, Hahnfeldt P. Classical mathematical models for description and prediction of experimental tumor growth. *PLoS Comput Biol.* 2014;10(8):e1003800.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

