

UC Davis

UC Davis Previously Published Works

Title

Assessment of Patient-Derived Xenograft Growth and Antitumor Activity: The NCI PDXNet Consensus Recommendations.

Permalink

<https://escholarship.org/uc/item/7p99q4s2>

Journal

Molecular Cancer Therapeutics, 23(7)

Authors

Meric-Bernstam, Funda

Lloyd, Michael

Koc, Soner

et al.

Publication Date

2024-07-02

DOI

10.1158/1535-7163.MCT-23-0471

Peer reviewed



Assessment of Patient-Derived Xenograft Growth and Antitumor Activity: The NCI PDXNet Consensus Recommendations

Funda Meric-Bernstam¹, Michael W. Lloyd², Soner Koc³, Yvonne A. Evrard⁴, Lisa M. McShane⁵, Michael T. Lewis⁶, Kurt W. Evans¹, Dali Li¹, Lawrence Rubinstein⁵, Alana Welm⁷, Dennis A. Dean II³, Anuj Srivastava⁸, Jeffrey W. Grover³, Min J. Ha⁹, Huiqin Chen¹⁰, Xuelin Huang¹¹, Kaushik Varadarajan¹², Jing Wang¹⁰, Jack A. Roth¹³, Bryan Welm⁷, Ramaswamy Govinden¹⁴, Li Ding¹⁴, Salma Kaochar¹⁵, Nicholas Mitsiades¹⁶, Luis Carvajal-Carmona¹⁷, Meenhard Herylyn¹⁸, Michael A. Davies¹⁹, Geoffrey I. Shapiro²⁰, Ryan Fields²¹, Jose G. Trevino²², Joshua C. Harrell²³; NCI PDXNet Consortium, James H. Doroshow²⁴, Jeffrey H. Chuang⁸, and Jeffrey A. Moscow²⁵

ABSTRACT

Although patient-derived xenografts (PDX) are commonly used for preclinical modeling in cancer research, a standard approach to *in vivo* tumor growth analysis and assessment of antitumor activity is lacking, complicating the comparison of different studies and determination of whether a PDX experiment has produced evidence needed to consider a new therapy promising. We present consensus

recommendations for assessment of PDX growth and antitumor activity, providing public access to a suite of tools for *in vivo* growth analyses. We expect that harmonizing PDX study design and analysis and assessing a suite of analytical tools will enhance information exchange and facilitate identification of promising novel therapies and biomarkers for guiding cancer therapy.

Introduction

Interest in using patient-derived xenografts (PDX) for preclinical modeling of the antitumor activity of novel agents and combinations as well as for the discovery of novel indications and predictive biomarkers in cancer research is growing. However, study designs for experiments using PDXs have not been standardized. Furthermore, a standard approach to *in vivo* tumor growth analysis and assessment of antitumor activity of cancer therapy is lacking, complicating the comparison of studies and determination of

whether an experiment has produced the evidence needed to consider a new therapy promising.

Clinical trials in oncology have standard clinical endpoints facilitating both clinical trial design and interpretation of study results. When assessing the clinical efficacy of a drug, an effective therapeutic regimen is expected to be able to prolong overall survival (OS) more than that with established treatment regimens or, in selected circumstances, placebo administration or best supportive care. A surrogate for OS is progression-free survival (PFS), defined

¹Department of Investigational Cancer Therapeutics, The University of Texas MD Anderson Cancer Center, Houston, Texas. ²The Jackson Laboratory, Bar Harbor, Maine. ³Seven Bridges Genomics, Charlestown, Massachusetts. ⁴Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, Maryland. ⁵Biometric Research Program, DCTD, National Cancer Institute, National Institutes of Health, Bethesda, Maryland. ⁶Departments of Molecular and Cellular Biology and Radiology, Lester and Sue Smith Breast Center, Dan L. Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, Texas. ⁷Huntsman Cancer Institute, University of Utah, Salt Lake City, Utah. ⁸The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut. ⁹Department of Biostatistics, Graduate School of Public Health, Yonsei University, Seoul, Republic of Korea. ¹⁰Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas. ¹¹Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, Texas. ¹²Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas. ¹³Department of Thoracic and Cardiovascular Surgery, The University of Texas MD Anderson Cancer Center, Houston, Texas. ¹⁴Department of Medicine, Washington University School of Medicine, St. Louis, Missouri. ¹⁵Department of Medicine, Baylor College of Medicine, Houston, Texas. ¹⁶Department of Molecular Cellular Biology, Baylor College of Medicine, Houston, Texas. ¹⁷Department of Biochemistry and Molecular Medicine, University of California, Davis, California. ¹⁸The Wistar Institute, Philadelphia, Pennsylvania. ¹⁹Department of Melanoma Medical Oncology, The University of Texas MD Anderson Cancer

Center, Houston, Texas. ²⁰Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts. ²¹Department of Surgery, Washington University School of Medicine, St. Louis, Missouri. ²²Department of Surgery, Virginia Commonwealth University, Richmond, Virginia. ²³Department of Pathology, Virginia Commonwealth University, Richmond, Virginia. ²⁴Division of Cancer Treatment and Diagnosis, National Cancer Institute, National Institutes of Health, Bethesda, Maryland. ²⁵Investigational Drug Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland.

NCI PDXNet Consortium Contact: J. Moscow, Investigational Drug Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

PDXNet consortium list included in supplementary documentation.

Corresponding Author: Funda Meric-Bernstam, Department of Investigational Cancer Therapeutics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030. E-mail: fmeric@mdanderson.org

Mol Cancer Ther 2024;23:924–38

doi: 10.1158/1535-7163.MCT-23-0471

This open access article is distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license.

©2024 The Authors; Published by the American Association for Cancer Research

as survival without tumor growth beyond a certain threshold. However, because assessment of both PFS and OS require large clinical trials with long follow-up, the objective response rate (ORR) or clinical benefit rate (CBR) is frequently the primary study endpoint for early-phase clinical trials. ORR is defined as the proportion of patients with predefined tumor size reduction (for response), whereas CBR is defined as an objective response or stable disease for a predefined period. The Response Evaluation Criteria in Solid Tumors (RECIST; version 1.1) provide a standardized method of evaluating the antitumor activity in solid tumors (1). To harmonize preclinical *in vivo* studies with clinical endpoints, researchers have made some efforts to standardize *in vivo* growth metrics for preclinical modeling of antitumor activity (2–4). However, these measures have not been widely adopted.

The NCI-funded PDX Development and Trial Centers Research Network (PDXNet), as part of the Cancer Moonshot program, was established to coordinate collaborative, large-scale development, and preclinical testing of targeted therapeutic agents using patient-derived models to advance cancer precision medicine. In recent work, we showed that PDX drug responses and genomic sequencing results are reproducible across diverse experimental protocols (5). However, the collaborative effort revealed the diverse methods that each PDX Development and Trial Center employed to measure tumor drug responses and has led to this effort to establish a common basis for preclinical drug response evaluation.

We here assess the utilization of different tumor growth assessment metrics across the PDXNet as well as in the scientific literature. Consensus recommendations for metrics for assessment of PDX growth and antitumor activity of drugs and drug combinations were developed by a multidisciplinary expert panel, with input from the greater PDXNet Consortium membership.

Materials and Methods

Development of consensus recommendations

Consensus recommendations for metrics for assessment of PDX growth and antitumor activity of drugs and drug combinations were developed by a multidisciplinary expert panel, the PDXNet Volume Assessment Project Working Group, which included basic and translational researchers, clinical trialists, and statisticians. Commonly used metrics were determined through a review of the literature as well as a survey of PDXNet investigators. The expert panel met via webinar more than 2 years and contributed to the development of the consensus recommendations. The recommendations were reviewed and approved by the PDXNet investigators and steering committee. Members of the working group were responsible for reviewing the final version of the recommendations and manuscript. Funding for the administration of the project was provided by the NCI Cancer Moonshot PDXNet project.

Assessment of different metrics

For selected PDX treatment experiments, different PDX growth analyses were performed using metrics presented in the Results section.

Tumor Volume Suite

To improve the accessibility of the proposed metrics for assessment of antitumor efficacy a web-based R/Shiny application called Tumor Volume Suite was developed by our group. This tool, available at <https://tumor-volume.jax.org>, allows users to validate and then import a custom tumor volume (TV) data set. Users can

then generate plots like those shown in **Figs. 1** and **2** and compute basic statistics to assess TV response (e.g., analysis of variance, RECIST 1.1). The Tumor Volume Suite tool is based on extensively modified code originally provided with the R package DRAP (6).

Patient and animal studies and data availability

The PDXs used in the experiments shown in the figures were generated from patients with written informed consent under Institutional Review Board Approved protocols. These studies were conducted in accordance with the Declaration of Helsinki and the U.S. Common Rule. The PDX treatment experiments were performed under protocols approved by Institutional Animal Care and Use Committees at the University of Texas MD Anderson Cancer Center or the Frederick National Laboratory for Cancer Research. The PDX growth data used for the analysis shown as representative figures are available from investigators by request.

Results

PDX experimental design

PDX model generation

A variety of immunocompromised mouse models are currently being used for PDX development, and consensus has not been reached as to the optimal host. Some of the strengths of each mouse strain and PDX development strategy have been previously described (7). Although there are various ways to generate PDXs, there is agreement that minimal information on how the PDX was generated should be made available, including clinical annotation of the patient's tumor, the process of implantation and passaging, and the mouse strain for both (8).

Most groups use subcutaneous implantation of PDXs for treatment studies for solid tumors; therefore, the recommendations provided here are mainly based on the subcutaneous models in which serial tumor diameter measurements with calipers can be obtained. Some of the metrics can also be adapted for use with imaging-based assessment of tumor growth. However, further work is needed to determine the impact of orthotopic implantation versus subcutaneous implantation in different contexts and for different treatment types.

PDX research ethics

All patients should give written informed consent for the generation of PDX models and ideally, for model and data sharing. PDX studies should be conducted according to an approved institutional animal care and use committee protocol in accordance with the procedures outlined in the Guide for Care and Use of Laboratory Animals (9).

PDX model characterization

Both mouse tumors and human lymphocytic tumors may emerge upon implantation and passaging of PDXs (10, 11). Thus, screening PDXs to confirm that they are of human origin and that they match the donor patient using short tandem repeat typing, comparing PDX histology with that of the parental tumor, and testing the PDX for lymphocytic markers such as CD45 are important. Although PDXs retain many of the key genomic drivers of the parental tumor, (11, 12) tumors may demonstrate intratumoral and intertumoral heterogeneity (13–15). Therefore, molecularly characterizing PDX models after establishment with a reassessment of critical molecular features is important. Collection of blood or normal tissue from patients can facilitate analysis of whole exome

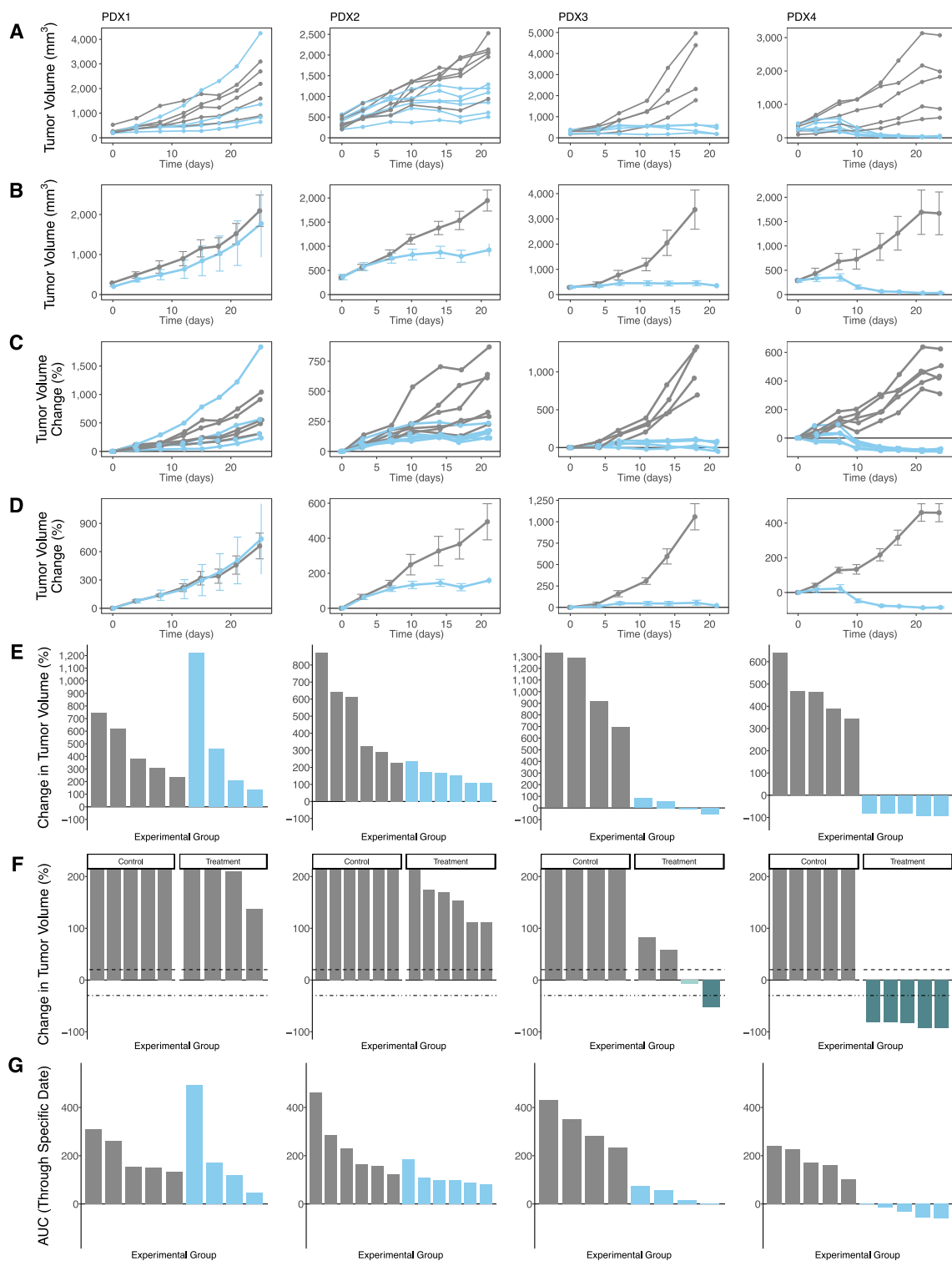


Figure 1.

Different growth metrics applied to four PDX experiments. Four PDX models were subjected to daily with a drug. The growth of treated tumors (blue) and control tumors (gray) is displayed using different metrics. **A**, Spider plots of individual tumor volumes. Calculated as follows: Tumor volume (mm³) = [tumor length × (tumor width)²]/2B). **B**, Growth curves representing average tumor volumes. **C**, Spider plots of changes in tumor volume. Calculated as follows: % tumor volume change = $\Delta V_t = \frac{V_t - V_0}{V_0} \times 100$. **D**, Growth curves representing average changes in tumor volume. **E**, Waterfall plots of changes in tumor volume on day 21. Calculated at day 21 as follows: % tumor volume change = $\Delta V_t = \frac{V_t - V_0}{V_0} \times 100$. **F**, Waterfall plots of changes in tumor volume classified by response status. Objective responses: gray, PD (≥20% growth); dark blue, PR (≤30% tumor regression); light blue, stable disease (nonPD, nonPR). Scales capped at 200% to better show response range. **G**, Area under the tumor (AUCs) for tumor volume. The area under the tumor growth curve from baseline t_0 up to time $t + t$.

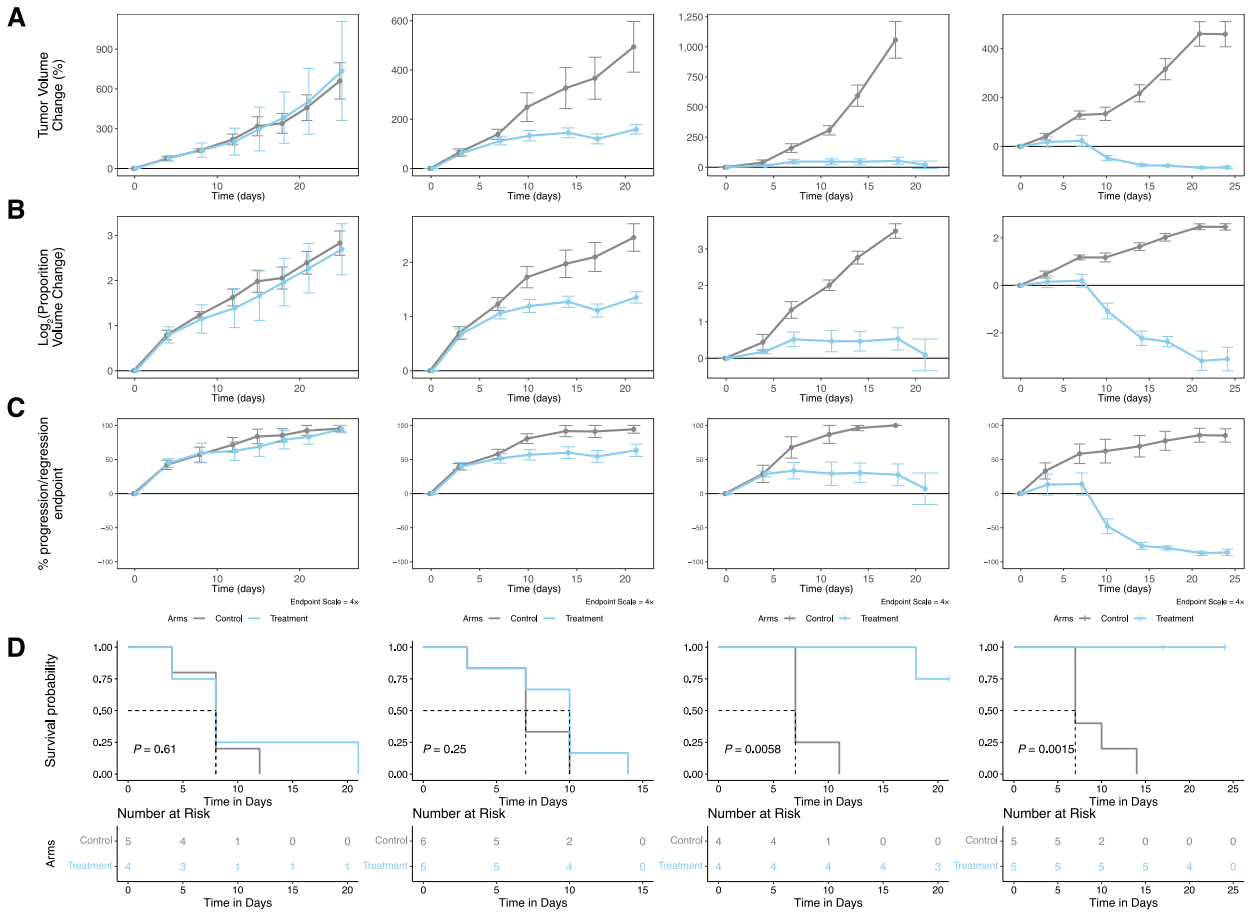


Figure 2.

Additional metrics used to display tumor volume and antitumor activity. **A**, comparison of average changes in tumor volume. **B**, Log₂ proportional changes in tumor volume. **C**, Scaled average changes in tumor volume. The tumor volumes range from –100% (complete regression) to 100% (endpoint). Tumor volume four times the baseline volume was chosen as the maximum growth endpoint. **D**, Event-free Survival. EFS is defined as the time until the tumor volume increases by a multiple of δ or reaches a certain volume cutoff. Tumor doubling (EFS2) was selected as an event. For example, $\min(\tau : \Delta v_t \geq \delta \times 100)$; $\delta = 1$ corresponds to time until tumor doubles in size (EFS2).

sequencing. If feasible, collecting and profiling a matched patient tumor sample to establish the parental tumor profile would be of interest.

When testing the antitumor activity, PDX models with clinical and molecular characteristics relevant to the intent-to-treat population should be selected. Sharing how a PDX was derived, (16) and when possible, the deidentified patient and treatment histories will facilitate optimal model selection and additional secondary use of the models. Publicly accessible central repositories, such as the NCI Patient-Derived Models Repository, are essential for sharing models to expedite the discovery of novel targets, novel biomarkers, novel therapeutic indications, and novel combinations. Furthermore, central repositories ensure that the scientific community has access to models to ensure reproducibility.

PDX treatment

In general, tumors at a preselected diameter/volume are treated, and tumor-bearing mice are randomized to treatment and control groups. Treatments must be initiated when all tumors to be enrolled have demonstrated growth and should continue long enough to

assess effects on tumor growth more than a period in which the control tumors go through at least two volume doublings to assess the effects of the drug in the treatment group(s). Treatments may be started simultaneously for all models within an acceptable size window (such as 200–300 mm³) or independently for each mouse when a certain tumor size is reached (e.g., at a predetermined size such as 200 mm³) as long as mice are enrolled in all groups in a parallel fashion. The latter generally gives fewer variable results but may not accurately reflect the variation captured in clinical scenarios. It is often optimal to implant more mice than needed for the planned sample size (based on previous growth information) to ensure sufficient mice with active tumors in the size window are available to meet the statistical plan.

Dose, schedule, and toxicity monitoring

Drug doses and schedules that approximate clinical exposure and clinically achievable drug concentrations should be used in pre-clinical models when possible. Mouse models may not recapitulate the toxicity seen in patients owing to differences in target specificity and metabolism. Furthermore, toxicity may vary based on mouse

strain. In PDX experiments, the overall health and body weights of the mice should be monitored. Treatment interruption, dose reduction, or change in treatment schedule should be considered if at least 15% weight loss is observed in an individual mouse or as stated in the approved experimental plan. If antitumor activity is observed only at a dose associated with weight loss, this information should be shared, and repeating experiments with different doses/schedules should be considered to demonstrate that a therapeutic index is achievable. Selected toxic effects can be monitored with additional testing (e.g., laboratory testing).

Study controls

Preclinical modeling has the advantage that untreated PDXs can be directly compared with treated PDXs. Different PDX models may grow at different rates, so having a control arm (untreated or vehicle-treated, if appropriate) can differentiate a generally slow PDX growth rate from the treatment effect.

In combination therapy studies, treatment arms must include single-agent administration of each agent to assess the contributions of each agent. In an emerging trend, some investigators elect to do single mouse or small cohort (two or three mice) screening for rational combination strategies as a signal-seeking experiment, with subsequent larger confirmatory studies powered to capture the treatment effect of interest. Comparison of the antitumor activity of new agents with that of established agents may be considered as an additional control in certain scenarios.

Assessing PDX growth and single-agent drug activity

Several metrics have been described to assess PDX growth and antitumor activity. We describe below some of the metrics commonly used to analyze tumor growth and antitumor efficacy, with a discussion of special considerations for each metric. Each metric has advantages and disadvantages as shown in **Table 1**. **Figs 1** and **2** provide short descriptions of different metrics and demonstrate their application to an experimental set of four PDX models with differential sensitivity to a targeted therapy. For many metrics, tumor growth is only compared at one timepoint. Pre-specification of a timepoint for analysis may overcome potential biases that can occur if the timepoint for analysis is chosen to give the “best” results.

Tumor growth curves and changes in TV

Approach

PDX volumes are often displayed in tumor growth curves (**Fig. 1A** and **B**). In subcutaneous models, tumor diameter/volume should be assessed two or three times a week using caliper measurements. The following formula approximates the volume of the ellipsoid corresponding to the tumor: $TV \text{ (mm}^3\text{)} = [\text{tumor length} \times (\text{tumor width})^2]/2$. Tumor growth curves show median or average TVs, and the standard error (SE) or standard deviation is usually added to demonstrate variability.

Changes in TV (**Fig. 1C** and **D**) are assessed by comparing the volume at a defined timepoint with the baseline volume, thus adjusting for baseline TV. For an individual mouse, the tumor response according to the percent change in TV is calculated from baseline to the time of assessment as follows: % TV change $\Delta V_t = \frac{V_t - V_0}{V_0} \times 100$, where V_t is the TV at time t and V_0 is the TV at baseline.

Tumor treatment and observation are usually continued for a predetermined amount of time. Animals are euthanized when

tumors reach the endpoint in the preapproved institutional protocol, which can be defined by as a certain duration of treatment and maximum tumor size.

In certain scenarios, when tumor regression or durable disease control is observed, treatment can be continued to determine the durability of response on-treatment or develop models with acquired resistance. Alternatively, treatment can be withheld after a certain period to assess the durability of the treatment effect and compare the times to tumor regrowth [see event-free survival (EFS) section below].

Considerations

When demonstrating tumor growth and antitumor activity, using tumor growth curves rather than comparative TVs at one point in time only is very helpful.

Many *in vivo* model/treatment pairs demonstrate only slowing of tumor growth with treatment. Importantly, changes in TV that achieve statistical significance (compared with controls) may occur after a tumor has at least doubled in size, so the observed differences do not necessarily represent clinically meaningful tumor responses. If PDX criteria that truly predict clinical significance are desired, one may have to raise the bar by trying to identify antitumor agents or combinations that will cause prolonged stable disease or, even better, tumor regression.

Tumor growth curves can yield several important insights:

- 1) Growth inhibition compared with controls. PDXs grow at different rates. Growth curves allow for comparison with untreated or vehicle-treated controls.
- 2) Growth curves allow for easy visualization of whether a tumor regresses from the baseline, and the dynamics of antitumor activity (e.g., how fast a tumor regresses). Rapid tumor shrinkage followed by regrowth between cycles may suggest the need to change the treatment schedule.
- 3) Changes in TV curves should look similar to the TV curves themselves (**Fig. 1A**). However, demonstrating changes in TV rather than TVs controls for differences in baseline volumes between groups. Furthermore, displaying changes in TV can readily reveal whether the PDX model has tumor regression (negative TV changes).
- 4) Large SEs in growth curves may suggest heterogeneity of treatment response. This can be better elucidated with spider plots, showing PDX volumes in individual mice that can be heterogeneous in tumor response or growth rate.
- 5) Growth curves depicting TVs during prolonged treatment can give insight into the dynamics of acquired resistance.
- 6) Continued monitoring of tumor growth after treatment termination can allow for assessment of the durability of response.

Mice may be censored because of unexpected one-off events (e.g., losses owing to technical errors in gavage or intravenous injection or biologic problems prevalent in laboratory animals, such as rectal prolapse). If the cause of animal loss is not clear, repeat experiments can help better elucidate the safety of treatment. A bigger problem occurs if mice must be killed early owing to tumor burden. In this case, if TV averages are shown, the volume curves show a sudden drop at the timepoint in which mice with larger tumors were euthanized. Showing individual mouse plots would be preferable in such cases. If averages are to be shown, some imputation of TV at timepoints beyond this (even if it is a simple last observation carried forward or linear extrapolation) is preferable to ignoring the missing values.

Table 1. Advantages and disadvantages of different PDX tumor growth metrics.

| Tumor growth metrics | Advantages | Disadvantages |
|---|--|---|
| Comparison of tumor volume/size/weight at a timepoint | <ul style="list-style-type: none"> • Easy-to-use metric for statistical analysis | <ul style="list-style-type: none"> • Captures differences at one point in time • Relies on similar starting volumes • For tumors with slow growth, differences in volume change will be blunted • Analysis impacted if some mice do not reach the desired timepoint • Does not demonstrate disease control compared with baseline • Potential for bias unless the timepoint is predefined |
| % change in tumor volume | <ul style="list-style-type: none"> • Captures tumor regression and progression • Does not require a comparator if interested only in activity | <ul style="list-style-type: none"> • Often used at a single timepoint • Does not take into account differences in baseline PDX growth |
| Tumor volume T/C ratio | <ul style="list-style-type: none"> • Easy to calculate • Continuous metric • Accounts for control group growth | <ul style="list-style-type: none"> • Does not account for starting volumes • Does not clearly demonstrate when tumor regression is present, but a lower ratio is better |
| T/C tumor growth | <ul style="list-style-type: none"> • Easy to calculate • Continuous metric • Accounts for starting volumes • Accounts for control group growth | <ul style="list-style-type: none"> • Does not clearly demonstrate when tumor regression is present, but lower growth is better |
| Objective response | <ul style="list-style-type: none"> • Classification in response groups • Does not require a comparator if interested only in activity | <ul style="list-style-type: none"> • Tumor regression in PDXs is relatively rare • Not a continuous metric |
| AUC | <ul style="list-style-type: none"> • Integrates antitumor activity over time | <ul style="list-style-type: none"> • Complicated to analyze |
| Scaled average change in tumor volume | <ul style="list-style-type: none"> • Easy to recognize treatments that lead to tumor regression • Focused on clinically relevant endpoints of tumor stabilization and regression | <ul style="list-style-type: none"> • Hard to calculate • Suppresses antitumor activity signals of agents that have modest growth-inhibitory effects |
| EFS | <ul style="list-style-type: none"> • The event can be predefined but allows for flexibility in the definition of the event (e.g., tumor doubling or quadrupling) | <ul style="list-style-type: none"> • Long study • Requires many resources • Susceptible to technical variation in tumor measurements |
| OS | <ul style="list-style-type: none"> • Clinically relevant endpoint | <ul style="list-style-type: none"> • Statistical analysis is complicated • Long study • Requires more resources • Treatment until death usually is not allowed in vertebrate animal experiments • Surrogate death measure • Study is often terminated owing to tumor size/ulceration or failure of mice to thrive |

For animals not subjected to TV measurement at time t but subjected to flanking volume measurements at times t_0 and t_1 such that $t_0 < t < t_1$, linear interpolation can be used to compute the measurement. That is, $\Delta V_t = \Delta V_0 + \beta(t - t_0)$ is computed, where $\beta = (\Delta V_{t_1} - \Delta V_{t_0}) / (t_1 - t_0)$. Alternatively, using log-linear interpolation with ΔV_t replaced by $\log \Delta V_t$ in the above formula may be advantageous.

Finally, when PDX experimental differences are observed but the treated group still has continued to substantially grow, the clinical relevance of the growth inhibition must be questioned.

Comparison of TVs at a predetermined timepoint

Approach

TVs in two or more treatment groups at a specific timepoint can be displayed as average volumes for each group (\pm SE) using bar graphs or by representing each tumor in box plots, violin plots, or similar visualization approaches. Similar approaches also can be used to represent differences in tumor weight or diameter.

The mean TVs on a certain day in treatment and control groups can be compared using two-sample t tests. Similarly, combination treatments can be compared with their component single-agent treatments. However, t tests rely on the assumption of normally distributed data, which would be much more plausible with log-transformed volumes. If sample sizes are large (rare in PDX studies), then a t test is well approximated by a Z -test. Alternatively, the Wilcoxon rank sum test could be used, but if sample sizes are very small, the power would be low.

Considerations

TV comparisons at a single timepoint have the advantage of being relatively straightforward. However, differences in baseline TV may impact comparisons. Concerns also emerge if some mice have already died prior to the date chosen for comparison. Moreover, TV comparisons with those in control groups do not give insight into whether the treatment was able to stabilize growth or cause regression. Unlike clinical studies, in PDX experiments the timepoint for TV comparisons is not always prespecified, partly

because tumor growth rates between PDXs may differ. Bias may be introduced into the data via these comparisons if the timepoint is intentionally chosen to show the largest difference in TVs.

Small amounts of growth inhibition may be statistically significant with large cohorts. However, if tumor growth stabilization or regression is not achieved, these differences may not be clinically meaningful. Increasingly PDX experiments are being designed with smaller cohorts but more models.

Presenting TV changes as waterfall plots

Approach

In clinical oncology, waterfall plots are used to show the percent change in the sum of tumor diameters (based on RECIST; ref. 1) with radiographic tumor assessments usually occurring every two or three cycles.

Presentation of the percent TV change at a certain timepoint with waterfall plots (Fig. 1E) has the following advantages: i) it allows for readily identifying treatments that lead to tumor regression; ii) it allows for easy visual comparison of outcomes among multiple treatment groups; iii) it allows for depiction of responses in individual mice, facilitating assessment of heterogeneity; and iv) it allows for visualization of data on individual mice, facilitating molecular subtype/phenotype correlation.

Considerations

With fast-growing PDXs, if one or more mice must be euthanized, the TV at the last assessment can be used for the waterfall plot, or a linear extrapolation may be used (with sharing of this extrapolation in the figure legend). If the entire control cohort must be killed early, options include using the last TV at euthanasia for the control group, the volume at the planned assessment timepoint for the treatment group, or the volumes in both groups at the time of control group euthanasia, although the best response in the treatment group may be missed under this scenario. If any measurement used in a statistical comparison or plot is taken at a timepoint other than that indicated (e.g., if the value had to be imputed using the last observation carried forward), it should be clearly acknowledged.

Another consideration is whether the best change in TV should be displayed in a waterfall plot rather than the volume change at a certain time. Transient responses with subsequent tumor regrowth within the first few weeks of treatment are rare, whereas increasing growth inhibition with treatment beyond the first cycle is more common.

Objective response assessment

Approach

Currently, data supporting the use of specific cutoffs for tumor response and progression in PDX models are lacking. The most commonly used metrics are the following:

- 1) Progressive disease (PD): increase in TV by 20% or more above baseline at planned assessment.
- 2) Stable disease (SD): change in TV between 20% above and 30% below baseline.
- 3) Partial response (PR): decrease in TV of 30% or more from baseline. If desired, the study can be designed to see if this response is maintained more than two or more TV measurements (confirmed response).

- 4) Complete response: complete disappearance of palpable tumor (95% reduction) from baseline maintained more than a pre-defined period (with or without treatment).

Considerations

Objective response assessment (Fig. 1F) mimics the most commonly used antitumor activity metric in early-phase clinical trials, with the difference that the clinically used RECIST metric assesses the change in the sum of the diameters of multiple target lesions, whereas PDX measurements are usually based on a volumetric change in a single tumor, often using subcutaneously implanted models.

Several different cutoffs for response have been used in the literature. For example, in a highly cited report by Gao and colleagues (2), the best response was described as the minimum value of ΔVol_t for $t \geq 10$ days. For each time t , the average of ΔVol_t from $t = 0$ to t was also calculated. The best average response was defined as the minimum value of this average for $t \geq 10$ days. The criteria for response were defined as follows: modified complete response, BestResponse $< -95\%$ and BestAvgResponse $< -40\%$; modified PR, BestResponse $< -50\%$ and BestAvgResponse $< -20\%$; modified stable disease, BestResponse $< 35\%$ and BestAvgResponse $< 30\%$; and modified PD, not otherwise categorized. Other approaches used include reaching a prespecified cutoff, or defining progression at the group level as a statistically greater average TV than at baseline, or less than 50% regression from the initial volume during the study period and greater than a 25% increase in the initial volume at the end of the study period as defined by the Pediatric Preclinical Testing Program (4). When assessing response, some groups use the greatest TV decrease at any point within a defined time period or the best response that was confirmed in at least two consecutive measurements, whereas others reported the best response at a specific timepoint (often on day 21 or 28).

In larger cohort studies, the ORR may be reported as the number of PDX-bearing mice that had a response (e.g., 5 of 8 mice = 63%). The Fisher exact test may be used to compare response rates in different treatment arms. In signal-seeking small cohort studies, objective responses are often the basis for go/no-go decisions in looking for treatments that lead to objective responses in selected models.

Area under the tumor growth curve

Approach

To calculate the area under the curve (AUC; Fig. 1G), the area under the tumor growth curve from baseline t_0 up to time t is normalized by dividing by t . The definition of this measure is the average tumor size from baseline to time of interest.

Statistical analysis of the AUC

To summarize TV dynamics, the AUC value for TV) for each animal is normalized by dividing by time t to account for different study durations across animals, which is called the adjusted AUC. Variability in the measurement at different timepoints across animals is handled by imputing missing TVs. If a tumor measurement is absent at time t but flanking measurements at times t_0 and t_1 are present such that $t_0 < t < t_1$, then the linear interpolation can be used to estimate the TV at t . If the TV value is missing at the final time t , linear extrapolation could be considered.

The $aAUC_{max}$ for each animal is defined as the adjusted AUC from baseline up to the last measurement. The mean AUCs of

treatment groups can be compared using two-sample *t* tests. Log transformation or a nonparametric test also can be considered.

Considerations

AUC has the advantage of dynamically comparing tumor growth rather than limiting the comparison to one timepoint. However, the AUC curves may be driven by the mice that experience progression without adequately highlighting the mice that experience regression.

Log-based change and scaled average change in TV

When PDX growth curves are drawn to scale, the extent of tumor progression dwarfs the extent of regression. For example, a PDX that starts with a volume of 100 mm³ can only shrink to 0 but could increase to 2000 mm³ before the experiment stops. The area above 100 mm³ in a growth curve representing progression in this example is 19-fold greater than the area representing regression. This representation of TV can therefore exaggerate small degrees of growth inhibition and hide the extent of tumor regression (Fig. 2A). To overcome this problem, one option would be to display the log change in TV (Fig. 2B; Supplementary Fig. S1).

As a novel approach, we propose displaying the scaled average change in TV (Fig. 2C). With this approach, the PDX growth curve is generated by scaling from -100% (complete regression) to +100% (the defined growth endpoint limit). The maximum growth endpoint may be a certain percent increase in TV from baseline or an increase to a predetermined TV.

Considerations

The relative tumor regression in responding PDX models is much easier to appreciate when the scaled average change in TV is displayed compared to traditional growth curves. Fig. 2C shows the endpoint scaling limit presented as four times the baseline TV in which the disease stabilization and regression in models are much more apparent than with standard growth curves (Fig. 2A). This approach highlights treatments that cause regression rather than merely slowing growth in the context of continued progression, which may be considered a strength if the goal of the metric is to compare the potential clinical effectiveness of therapies. Metrics such as the TV T/C ratio (described below) may help supplement this information.

EFS

Approach

EFS (Fig. 2D) is defined as the time from treatment initiation until the time TV increases by a certain amount or ratio: $(\Delta(V\tau) = (V(t) - V_0)/V_0 \times 100)$. For example, $\min(\tau : \Delta V_{\tau} \geq \delta \times 100)$ (e.g., $\delta = 3$) corresponds to the time until a tumor quadruples in size (EFS4). EFS is censored at the last tumor measurement for animals whose tumors have never increased by that predetermined amount. The EFS T/C ratio is defined as the ratio of the median time to event in the treatment group to that of the respective control group. If EFS is a planned metric, the study groups should be followed until all cohorts achieve tumor doubling (or quadrupling, depending on the metric).

Considerations

Observing tumor growth three to four times longer than the duration required for controls to reach the target endpoint may better elucidate growth inhibition. For PDXs that have durable responses, the treated groups may not reach the target endpoint (e.g., tumor doubling or quadrupling); then, the EFS T/C ratio is defined as

greater than the EFS T/C ratio based on the last day of the study for the treatment group divided by the median time to event in the control group. Some investigators elect to stop treatment in all groups after a certain duration (e.g., 21–48 days) and observe the growth in all groups up to the endpoint. This may facilitate EFS analysis, but it may underestimate the durability of control if treatment is continued as it would be done in the clinic. Importantly, the endpoints of tumor doubling or quadrupling are much greater than the 20% increase in tumor size that is considered progressive in the clinic.

OS

Approach

OS is technically defined as the time from initiation of treatment to the date of death of PDX-bearing mice. Killing mice when they have a significant tumor burden (e.g., maximum tumor diameter of 1.5 cm) or signs of compromised health, such as lethargy and hunching, is usually considered best practice. Timepoint when mice were euthanized due to health concerns or tumor burden is usually substituted for the date of death when OS is used as an endpoint.

Considerations

OS is a more difficult metric than EFS and is essentially the time from treatment initiation to a predefined event endpoint, which is defined by institutional parameters. Although OS is the gold standard for assessing antitumor activity in clinical trials, this is a difficult endpoint in PDX experiments because of ethical concerns. Assessment of OS also leads to long experiments with resource implications. We thus recommend assessing EFS rather than OS in PDX experiments.

Assessing combination therapy activity

In clinical trials, when patients receive combination therapy with two agents (A and B) and a response is observed, whether the antitumor activity is attributable to agent A, agent B, or the combination is unclear. PDXs represent a unique opportunity to dissect the contribution of agents to the antitumor efficacy of a combination therapy. Therefore, when the intention is to determine if a combination therapy improves antitumor activity, experiments should include monotherapy treatment with each agent. Then, the combination therapy arm should be compared with each monotherapy arm, as well as the vehicle or untreated control arm; furthermore, each monotherapy treatment arm should be compared with the controls to determine if each agent has antitumor activity. Effective combination treatments should enhance antitumor activity compared with not only no treatment or vehicle controls but also each of the single agents.

Fig. 3 shows examples of combination therapy experiments. In Fig. 3A, two targeted agents had limited activity as monotherapy but achieved disease stabilization in combination therapy. In Fig. 3B, both agents had antitumor activity, and the combination led to tumor regression. In cases where one or both agents have substantial activity, prolonged treatment (Fig. 3C) or prolonged treatment followed by observation after treatment cessation (Fig. 3D) may help delineate differences in antitumor activity.

In contrast, Fig. 3E and F show examples of experiments without significant antitumor activity attributable to combination therapy. Figure 3E shows that whereas the combination therapy had significant antitumor activity, the activity was primarily driven by the cytotoxic agent without potentiation by the combination. This highlights the value of having single-agent cohorts in an experiment.

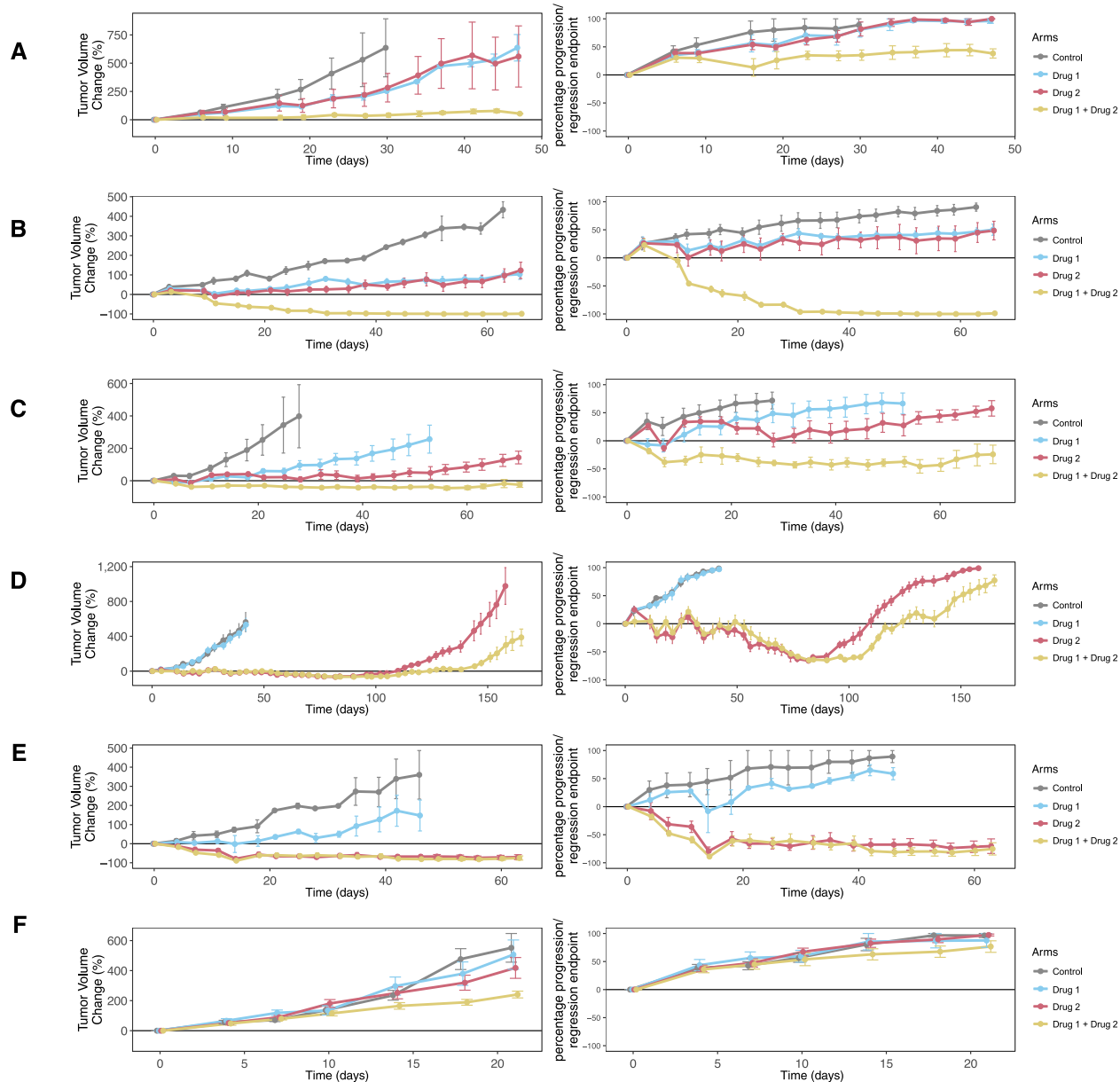


Figure 3.

Combination therapy testing. The panels show the effects of treatment of PDXs with single agents (blue and green) and combination therapies (red). The black lines represent untreated controls. The panels on the left show changes in tumor volume and panels on the right show scaled average changes in tumor volume. **A**, PDX treated with two therapeutic agents which achieved limited activity as monotherapy but achieved disease stabilization in combination therapy. **B**, Both drugs had monotherapy antitumor activity, but the combination led to tumor regression. **C**, Prolonged treatment with two drugs, demonstrating greater antitumor activity with the combination. **D**, Prolonged treatment with two agents alone and in combination, followed by treatment cessation, helping to delineate differences in durability of antitumor activity. **E**, A PDX experiment in which the combination therapy had significant antitumor activity was primarily driven by one agent without potentiation of the antitumor activity by the combination. **F**, PDX experiment here shows limited single-agent activity for each therapy with slightly more growth inhibition with the combination but still with continued tumor growth.

Figure 3F shows limited single-agent activity for each agent, with slightly more growth inhibition with the combination but still with continued tumor growth with the combination.

Figure 3 also shows the scaled average change in TV in these experiments. This approach can better demonstrate the separation of the growth curves and better highlight the tumor

regression seen with combination therapy than a traditional growth curve.

Drug dose and schedule choice are important considerations in planning combination therapy studies. It is optimal to test biologically and pharmacologically relevant doses and schedules for both drugs. It is possible that the highest achievable dose for both drugs

may not be achievable when given in combination, or a lower dose of one agent or both agents may be advantageous from a toxicity perspective. However, if combinations are studied with lower doses than biologically achievable monotherapy doses, one should consider evaluating whether a higher dose of monotherapy will be equally or more effective.

In some scenarios, combination therapy experiments may be done without monotherapy cohorts. One such scenario is when signal-seeking small cohort testing is performed to identify active combinations with antitumor activity. In this scenario, testing is often followed up with confirmatory studies to determine the contribution of each agent to the antitumor efficacy observed. A second scenario is when a combination is already of interest and is being tested using different models to look at the relative activity of the combination therapy in a larger cohort or models with different histologies or different molecular backgrounds. In this scenario, follow-up studies with monotherapy controls can help dissect whether differential tumor sensitivity is attributable to a single agent or a combination.

Comparing drug activity across PDX models

PDXs are often used to determine the activity of drugs against different tumor types to identify indications for drug development as well as to compare molecular features of PDX models with differential antitumor activity for the identification of biomarkers of response. Therefore, standardized approaches to comparing this activity across models are needed.

Waterfall plots of average change in TV

In clinical trials, a common approach to comparing drug activity is to display the activity in waterfall plots depicting changes in the sum of tumor diameters. A similar approach could be used to compare PDX models by presenting the average change in TV after treatment by model (Fig. 4A). The waterfall plot figures could be merged with selected molecular data (e.g., genomics and transcriptomics) to demonstrate molecular subtype–phenotype associations.

However, waterfall plots have an important disadvantage: When they are used to depict treatment effects across multiple models, they do not account for differences in baseline PDX growth in the control arms. Thus, if only the treatment arm is presented, it may lead to a perception of a lack of activity in fast-growing models or an overestimation of the drug effect in slow-growing models.

TV T/C ratio

Approach

Tumor growth may be assessed by comparing the TVs in the treatment group with those in the control group at time t , V_t . This approach has been referred to as tumor growth inhibition in the literature. However, because the baseline TVs may not be perfectly matched in PDX models, adjusting the volumes for the baseline TV is preferable, thus focusing on the on-treatment change in TV. This is calculated as follows:

TV T/C ratio at time t : TVs from baseline to time t , $R_t = V_t/V_0$, with E denoting mean.

$$\gamma = \frac{E(R_t|Treatment)}{E(R_t|Control)}$$

For example, if the mean TV in both the treatment group and control group started at 200 mm³ and that in the treatment group

increased to 250 mm³, whereas that in the control group increased to 600 mm³ by day 21, the TV T/C₂₁ would be 0.42 ($\gamma = 250/200 \div 600/200$).

The time when this comparison is performed is also an important consideration. Usually, the timepoint is preselected and may be 21 or 28 days. However, if the control group has early animal losses owing to tumor burden, the TV T/C ratio can be calculated based on TVs measured at an earlier date when all animals were alive and evaluable for response. An alternative approach is to select the day of calculation of antitumor activity based on the day that the median control TV growth reaches a predetermined level. For example, the T/C ratio could be calculated on the day that the median TV in the control arm has grown 300%.

Considerations

With this approach, the TV T/C ratio is assigned a value from 1 to 0, with 1 indicating a complete lack of growth inhibition and 0 indicating complete regression at the selected timepoint. In the rare scenario that treatment leads to enhanced tumor growth, it will result in a value greater than 1. Although the TV T/C ratio has been used as an antitumor activity measure in several studies, the proposed cutoffs for antitumor activity have varied from study to study. Some studies considered a TV T/C ratio of up to 0.6 as representing meaningful antitumor activity, whereas others considered a T/C ratio of 0.15 or lower as representing high antitumor activity, greater than 0.45 as low activity, and from greater than 0.15 to 0.45 as intermediate activity.(4, 17) Further study is needed to benchmark the TV T/C ratio that is predictive of clinical benefit in pre-clinical models.

The TV T/C ratio has the disadvantage of not being able to designate a standard cutoff that can clearly distinguish between tumor regression and growth inhibition. The numeric cutoff for the T/C ratio to represent regression as opposed to growth inhibition varies across experiments, although it can be identified for example, if an untreated tumor tripled in TV from baseline, a T/C ratio of 1/3 is the threshold between regression and growth inhibition. The T/C ratio has the important advantage of being a continuous metric, with smaller numbers depicting greater treatment effects.

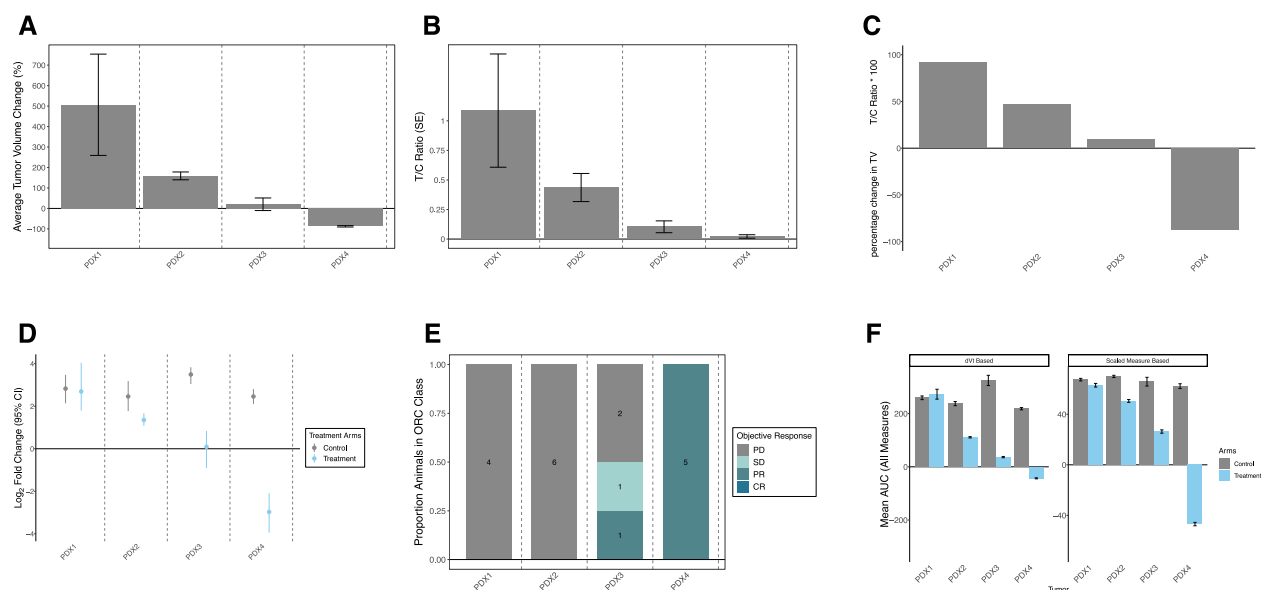
An alternative way to visualize both tumor growth inhibition and regression is to generate hybrid graphs displaying the percent regression in models that have tumor shrinkage and the TV T/C ratio in models without regression to demonstrate whether growth was inhibited compared with that in controls. Figures 4B and C show TV T/C ratio and hybrid tumor regression/growth inhibition graphs, respectively.

Logarithmic tumor growth in treatment versus control groups

Another way to display differences in tumor growth between treatment and control groups is to display the log₂ growth in both groups (Fig. 4D). This approach allows for the display of the relative tumor growth as well as regression compared with that in controls.

ORRs

In clinical trials, comparison of ORRs between treatment arms is a common clinical endpoint. Although cohort sizes in PDX studies are usually relatively small, a comparison of response outcomes may give insight into differences in antitumor activity between treatment groups. Figure 4E shows that in a small cohort of PDXs tested, objective responses differed across models.

**Figure 4.**

Comparison of the antitumor activity in four PDX models (**A**) waterfall plot of the average tumor volume changes at day 21. **B**, Graph of tumor volume T/C ratios. **C**, Hybrid tumor regression/growth inhibition plot of T/C ratios and tumor regression. The average percent decrease in tumor volume (TV) is shown for PDXs with tumor regression, whereas the T/C ratios are shown for PDXs with tumor growth above baseline volume (**D**) Log₂ fold change in tumor volumes. **E**, Objective response classification in different PDX models. Gray, PD ($\geq 20\%$ tumor growth); dark blue, PR ($\leq 30\%$ tumor regression); light blue, stable disease (SD); nonPD, nonPR. **F**, AUCs based on average tumor volumes (left) and scaled (right).

AUCs for scaled average change in TV

AUCs have the benefit of integrating changes in tumor growth over time. Thus, another approach to compare antitumor activity across models is to compare the AUCs (dVt; % change in different models; **Fig. 4F**, left panel). Moreover, the AUC can be calculated based on the scaled change in TV (**Fig. 4F**, right panel). In scaling dVt to a maximum of 100, by definition, the y-axis range of the curve is suppressed. As a result, the $AUC_{(\text{scaled measure})}$ is less than the $AUC_{(\text{dVt})}$ except when the dVt for an individual is negative throughout the entire study. The benefit of presenting the AUC dVt is that tumor growth inhibition in the context of continued progression is easier to visualize, whereas the $AUC_{(\text{scaled measure})}$ is better able to demonstrate tumor regression. Because the goal of cancer therapy is to produce durable tumor regressions and not to merely slow tumor growth, the $AUC_{(\text{scaled measure})}$ metric provides an improved method for displaying results to emphasize clinically relevant endpoints.

Special considerations in PDX metrics

The use of more than one PDX metric may better capture tumor growth and treatment effect than the use of only one metric. In certain scenarios (described below), some metrics may have specific advantages/disadvantages.

Tumor growth characteristics

Fast-growing tumors: In fast-growing PDX models, the control animals may be euthanized before the planned assessment time. This usually does not affect the ability to perform analyses such as EFS. However, early euthanasia of controls can interfere with the ability to compare TVs at a preplanned date. For TV comparisons, performing analysis at an earlier date than the preplanned date would be preferred. Because continued treatment may improve the best response, if a change in TV is displayed (often in a waterfall

plot) along with the changes in the control volumes, the volumes in the treatment group on the day of euthanasia of the control group can be displayed, albeit with specification that the controls were killed earlier than the treated mice. Transparently showing how the analysis was performed is important.

Slow-growing tumors: PDX models with slow-growing tumors may require longer follow-up to determine whether a treatment effect occurs. Calculation of the ΔV_t at a certain timepoint may suggest growth inhibition even if it does not exist, if the change in TV in controls is low.

Tumor ulceration and cystic growth: Some subcutaneous PDX models develop ulceration. This often not only leads to difficulty for tumor measurements but usually is considered an indication for sacrifice of the mouse bearing the ulcerated PDX. Furthermore, some PDXs develop a cystic growth pattern, making it more difficult to accurately measure growth or antitumor activity.

PDX models with initial tumor regression followed by regrowth

Some treatments may lead to a rapid tumor response followed by rapid regrowth in a PDX model. This is often seen with treatments that are intermittent, frequently with agents with short half-lives in mice, although it may also be related to rapid adaptive changes in the tumor. With such models, the use of growth curves and analysis with dynamic metrics such as the AUC may better capture transient tumor regression than metrics such as ΔV_t if t is after tumor regrowth, which may miss regression.

Models with initial growth and subsequent regression

Some models and treatments may have a delayed onset of treatment effect, with initial tumor growth followed by tumor

stabilization or regression. For metrics that assess growth at a specific timepoint, the time for assessment use should be based on the mechanism of action of the drug and the expected timeline for antitumor activity. Moreover, metrics such as EFS2 (time to tumor doubling) and assessment of progression with relatively low cutoffs, such as 20% growth, may be adversely affected by delayed growth inhibition.

On-treatment and off-treatment observation periods

Optimal cancer treatment effects remain durable after treatment has stopped. PDX data are confounded by varying practices of treatment administration. Some experiments are performed with relatively short treatment periods followed by posttreatment observation, and others are performed with constant drug administration throughout the experiment. The duration of treatment should be clearly indicated in graphical representations of tumor growth and time to progression. Off-treatment observations in PDX experiments are as valuable as those in the clinic in assessing the durability of therapeutic interventions.

Other important considerations

Limitations

Although PDXs have many strengths, they also have several limitations. First, PDX implantation may have a patient selection bias. For example, many PDX collections are enriched for surgical samples representing early-stage disease or tumors that have undergone limited treatment. When preclinical modeling is done with the intent of planning clinical trials, consideration should be given to expected prior treatment exposures of the tumor and to seeking model development in that setting (e.g., BRAF V600E mutant melanoma after treatment with RAF/MEK inhibitors). Second, PDXs that grow *in vivo* may be biologically different. For example, they may represent more aggressive tumors compared to tumors that have been implanted but that did not grow as a PDX (11). PDXs that engraft may also arise from tumors with lower levels of immune infiltration (18). Third, PDXs grow faster than human tumors do in the clinic. How much antitumor activity in PDX experiments translates into successful agents or combinations in the clinic is not known. Fourth, mouse experiments often do not capture the drug toxicity seen in humans; thus, doses required to have target engagement and antitumor activity in mice may not be achievable in the clinic.

Below are a few more points for consideration in assessing PDX experiments.

Site of implantation

Many investigators perform subcutaneous implantation of PDXs. This is more convenient than orthotopic implantation, as it allows for close monitoring of tumor size without the use of imaging procedures; however, the impact of the implantation site on tumor growth and treatment response is not well understood. For tumors implanted orthotopically, usually, TVs can only be assessed at a selected timepoint, often through imaging.

Stroma, immune environment, and humanized models

PDXs are proposed to better capture the heterogeneity of tumors than cell line-derived xenografts. Although PDXs often retain some of the human stroma and other components of the immune environment in the first passage, the stroma is replaced with mouse stroma over the next few passages (11).

Another limitation of PDXs is that experiments are performed in immunodeficient mice. Therefore, PDX experiments may miss the immune effects of therapies and thus will have limited utility in testing immunotherapies. Several different approaches to “humanizing” the PDX immune environment are now being developed, such as the use of human peripheral blood mononuclear cells and implantation of cord blood CD34⁺ hematopoietic stem cells (19, 20–23).

Mouse sex

In humans, sex is known to be an important risk factor for the development of certain tumors (e.g., breast cancer), and more recently has been shown to impact the efficacy of certain therapeutics. (24, 25) The impact of mouse sex on xenograft growth is relatively unknown. While some studies have demonstrated an effect of sex on therapeutic response, others have not (26). At this time, for endocrine-related cancers sex-matching is expected, and for others, systematic studies are needed to better address these questions experimentally.

Other endpoints

In this report, we focused on the most common PDX growth assays. However, several other endpoints can be used depending on the clinical scenario being modeled, such as recurrence-free survival (after surgical resection), time to distant metastasis, tumor response in tumor dimensions/volume according to imaging (such as computerized tomography), or functional imaging (such as positron emission tomography).

Biomarker assessment

Predictive biomarkers: The ability to test multiple models with different molecular backgrounds makes PDXs important tools for the predictive biomarker discovery and validation of predictive markers.

Pharmacodynamic biomarker assessment: An important strength of PDXs is the ability to assess biomarker engagement and adaptive responses by comparing treated models with untreated and/or vehicle-treated controls or by comparing models treated with single agents with those treated with combinations. Pharmacodynamic markers may include those indicative of target inhibition, for example, by assessing downstream cell signaling, DNA damage, cell proliferation, or apoptosis, depending on drug type. Biomarkers assessed may include comprehensive or targeted RNA or proteomic profiling or assessing selected biomarkers evaluated via immunohistochemistry.

Comprehensive approaches may also give insight into adaptive responses, such as alternate survival pathways activated due to treatment. Importantly, PDX models provide an experimental platform to guide the subsequent timing of the acquisition of clinical specimens. Planning the timing of biomarker assessment is important, taking into consideration both the number of days from the start of the treatment and collection of PDX samples, as well as the number of hours or days from the last treatment to PDX collection. PDX studies can incorporate additional specific biomarker cohorts for PDX sampling at timepoints earlier than the efficacy endpoints for comprehensive pharmacodynamic analysis, with the use of separate cohorts for conducting studies with longer treatment durations to enable assessment of the durability of treatment effects and EFS endpoints.

Discovery of mechanisms of acquired resistance: Prolonged continuous treatment of PDXs or treatment for a duration followed by a

Table 2. PDXNet consensus recommendations for assessment of PDX growth and antitumor activity.

Experimental design

- PDX studies should be conducted according to approved institutional animal care and use committee protocols. Patients should have provided informed consent for PDX model development and, ideally, for model and data sharing.
- PDX experiments should be performed using clinically relevant doses and schedules, if known.
- Antitumor activity should be demonstrated in at least two clinically relevant models.
- Tolerability should be monitored. Treatments that lead to growth inhibition only at doses that have toxicity, may not have an adequate therapeutic window.

Assessment of antitumor activity

- Antitumor activity is best assessed using a combination of two or more metrics. At least one metric (usually % tumor growth) should be used to compare tumor volume on-treatment with that at baseline, to demonstrate whether tumor regression or stabilization of growth occurs. A second metric, such as comparison of tumor volume change or comparison of event-free survival, can be used to compare the growth in the treatment cohort with that in controls.
- Percent tumor volume change or objective response classification allows for assessment of whether tumor regression is observed. Agents that lead to tumor regression or prolonged growth inhibition (stable disease) are considered to have more promising therapies
- Comparison of tumor volume or change in tumor volume at a planned timepoint, area under the curve C, or event-free survival are approaches to allow for assessment of statistical differences in the growth of treated vs. control group. Active treatments are expected to lead to statistically significant differences in these metrics. In contrast, treatments leading to statistically significant differences in growth but in the setting of continued growth in the treatment group, are not as compelling.
- Depicting tumor volume changes using scaled average change or log change in tumor volume may better elucidate tumor regression vs. progression than standard tumor volume growth curves.
- Treatment control ratio for tumor volume (or TV change) can serve as a screening tool for signal of antitumor activity.

Treatment duration

- Continuing studies until the control arm demonstrates substantial tumor growth (e.g., quadrupling in volume) may better elucidate the extent of antitumor activity than stopping treatments early, at a predetermined timepoint.
- For EFS analysis, observation should not be stopped for any treatment groups until they have reached the prespecified endpoint or the observation period has exceeded two or three times the time to event for the untreated/vehicle-treated controls.
- With treatments/models for which tumor regression is observed, continued monitoring of tumor volume beyond treatment cessation can provide insight into the durability of disease control.

Combination treatment

- Combination treatment experiments should include treatment groups that are given treatment with each of the components of the combinations alone in addition to untreated or vehicle-treated controls.
- Animals given combination treatments should have enhanced antitumor activity when compared with not only untreated or vehicle-treated controls but also those given each of the combination treatment components alone.
- If lower doses of agents are used in combination experiments, consideration should be given to assessing whether either agent at a higher dose has equivalent or greater efficacy, or lower toxicity.

Biomarker analysis

- Having cohorts where PDXs are treated shorter term with the intent of doing biomarker analysis in treated vs. control PDXs, can provide valuable insight into target modulation and adaptive responses. The timing of sample collection can be chosen to capture the optimal pharmacodynamic effect.
- Long-term treatment studies may provide novel insights into mechanisms of acquired resistance and strategies for overcoming resistance.

period of observation can be used to generate PDX models with acquired resistance. Alternately, PDXs can be generated from tumor biopsies (or effusions) from patients progressing after initial clinical benefit (objective response or prolonged stable disease). These models may be invaluable to identify mechanisms of acquired resistance and to test combinations that can overcome resistance as well as next-generation agents.

PDXNet consensus recommendations for the design and analysis of PDX growth experiments

PDXs are growing in importance as models that may better recapitulate the heterogeneity of human cancers. They are often used for preclinical drug screens, to identify novel combination therapies, in order to facilitate the identification of potential biomarkers of drug response and resistance. However, there is significant variability in approaches used in assessment of *in vivo* growth and antitumor efficacy. The context of each potential PDX experiment, including specific objectives, models, and agents may differ. Thus, there is value in having several well-defined tools that can be used to

universally communicate tumor growth and antitumor activity. Therefore, we here reviewed many of the commonly used tools and created a publicly accessible suite of analytic tools. The graphic representations of PDX experiments should place equal emphasis on tumor regression vs tumor growth so that growth inhibition is not given more emphasis than the achievement of tumor regression. One method of presenting PDX data that equalizes the representation of tumor growth vs regression is the metric $AUC_{(scaled\ measure)}$. **Table 2** summarizes the NCI PDXNet Consensus Recommendations, representing best practice guidelines for PDX experiments in solid tumors. Further programmatic studies are needed to improve PDX science as well as better determine the extent of antitumor activity needed in different experimental designs to translate PDX findings into successful therapeutic strategies in the clinic.

Authors' Disclosures

F. Meric-Bernstam reports personal fees from AbbVie, Aduro BioTech Inc., Alkermes, AstraZeneca, Daiichi Sankyo Co. Ltd., Calibr (a division of Scripps

Research), DebioPharm, Ecor1 Capital, eFFECTOR Therapeutics, F. Hoffman-La Roche Ltd., GT Apeiron, Genentech Inc., Harbinger Health, IBM Watson, Incyte, Infinity Pharmaceuticals, Jackson Laboratory, Kolon Life Science, LegoChem Bio, Lengo Therapeutics, Menarini Group, OrigiMed, PACT Pharma, Parexel International, Pfizer Inc., Protai Bio Ltd, Samsung Bioepis, Seattle Genetics Inc., Tallac Therapeutics, Tyra Biosciences, Xencor, Zymeworks, and Black Diamond, Biovica, Eisai, FogPharma, Immunomedics, Inflection Biosciences, Karyopharm Therapeutics, Loxo Oncology, Mersana Therapeutics, OnCusp Therapeutics, Puma Biotechnology Inc., Seattle Genetics, Sanofi, Silverback Therapeutics, Spectrum Pharmaceuticals, Theratechnologies, Zentalis, grants from Aileron Therapeutics, Inc. AstraZeneca, Bayer Healthcare Pharmaceutical, Calithera Biosciences Inc., Curis Inc., CytomX Therapeutics Inc., Daiichi Sankyo Co. Ltd., Debiopharm International, eFFECTOR Therapeutics, Genentech Inc., Guardant Health Inc., Klus Pharma, Takeda Pharmaceutical, Novartis, Puma Biotechnology Inc., and Taiho Pharmaceutical Co., and personal fees from Dava Oncology and other support from European Organisation for Research and Treatment of Cancer (EORTC), European Society for Medical Oncology (ESMO), Cholangiocarcinoma Foundation, Dava Oncology outside the submitted work. M.W. Lloyd reports grants from the NCI [HHS–NIH] during the conduct of the study. M.T. Lewis reports grants from NIH/NCI during the conduct of the study and being a founder of and limited partner in StemMed Ltd. and manager in StemMed Holdings LLC, its general partner, and being a founder of and equity stakeholder in Tvardi Therapeutics Inc. A.L. Welm reports grants from Aslan Pharmaceuticals during the conduct of the study and other support from Thunder Biotechnology and Modulus Therapeutics outside the submitted work, and the University of Utah may license the models described herein to for-profit companies, which may result in tangible property royalties to the university and members of the Welm labs who developed the models. D.A. Dean reports grants from Velsera (Seven Bridges) during the conduct of the study. X. Huang reports grants from the National Institutes of Health during the conduct of the study. B.E. Welm reports grants from NCI during the conduct of the study and has received royalties from licenses of previously developed PDX models issued by the University of Utah. The University of Utah may issue new licenses in the future at its discretion, which may result in additional royalties to the authors. N. Mitsiades reports grants from NCI during the conduct of the study. M.A. Davies reports grants from NCI during the conduct of the

study; personal fees from Roche/Genentech, Pfizer, Novartis, BMS, Iovance, and Eisai; grants and personal fees from AMB Therapeutics; and grants from LEAD Pharm outside the submitted work. G.I. Shapiro reports grants and personal fees from Merck KGaA/EMD-Serono; grants from Tango Therapeutics, Bristo Myers Squibb, Pfizer, and Eli Lilly; personal fees from Bicycle Therapeutics, Boehringer Ingelheim, Concarlo Holdings, Zentalis, Kymera Therapeutics, Janssen, Xinthera, Syros, ImmunoMet, and Blueprint Medicines outside the submitted work; additionally, G.I. Shapiro has a patent for Dosage regimen for sapacitabine and seliciclib issued to G.I. Shapiro and Cyclacel Pharmaceuticals and a patent for Compositions and methods for predicting response and resistance to CDK4/6 inhibition issued to G.I. Shapiro and Liam Cornell. No disclosures were reported by the other authors.

Acknowledgments

We thank Lori Henderson and Tiffany Wallace for their support of PDXNet in their roles as Program Directors at the NCI, Susanna Brisendine for her administrative assistance, and Donald R. Norwood from The University of Texas MD Anderson Cancer Center Editing Services, Research Medical Library, for their editorial assistance. We especially thank our patients who have kindly agreed to participate in protocols generating patient-derived models. This work was supported in part by NCI PDX Development and Trial Centers grants U54CA224065, U54CA224083, U54CA224076, U54CA224070, CPRIT RP170691, and 4UM1CA186688; NIH Clinical Translational Science Award U01TR003167 and the NIH/NCI under award numbers P30CA016672 and P30CA125123.

Note

Supplementary data for this article are available at Molecular Cancer Therapeutics Online (<http://mct.aacrjournals.org/>).

Received July 21, 2023; revised December 8, 2023; accepted March 29, 2024; published first April 19, 2024.

References

- Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;45:228–47.
- Gao H, Korn JM, Ferretti S, Monahan JE, Wang Y, Singh M, et al. High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat Med* 2015;21:1318–25.
- Ortmann J, Rampásek L, Tai E, Mer AS, Shi R, Stewart EL, et al. Assessing therapy response in patient-derived xenografts. *Sci Transl Med* 2021;13:eabf4969.
- Houghton PJ, Morton CL, Tucker C, Payne D, Favours E, Cole C, et al. The pediatric preclinical testing program: description of models and early testing results. *Pediatr Blood Cancer* 2007;49:928–40.
- Evrard YA, Srivastava A, Randjelovic J, Doroshow JH, Dean DA 2nd, Morris JS, et al. Systematic establishment of robustness and standards in patient-derived xenograft experiments and analysis. *Cancer Res* 2020;80:2286–97.
- Li Q, Dai W, Liu J, Li Y-X, Li Y-Y. DRAP: a toolbox for drug response analysis and visualization tailored for preclinical drug testing on patient-derived xenograft models. *J Transl Med* 2019;17:39.
- Dobrolecki LE, Airhart SD, Alferrez DG, Aparicio S, Behbod F, Bentires-Alj M, et al. Patient-derived xenograft (PDX) models in basic and translational breast cancer research. *Cancer Metastasis Rev* 2016;35:547–73.
- Meehan TF, Conte N, Goldstein T, Inghirami G, Murakami MA, Brabetz S, et al. PDX-MI: minimal information for patient-derived tumor xenograft models. *Cancer Res* 2017;77:e62–6.
- National Research Council. National science education standards. Council NR, editor. Washington DC: The National Academies Press; 1996.
- Bondarenko G, Ugolov A, Rohan S, Kulesza P, Dubrovskiy O, Gursel D, et al. Patient-derived tumor xenografts are susceptible to formation of human lymphocytic tumors. *Neoplasia* 2015;17:735–41.
- McAuliffe PF, Evans KW, Akcakanat A, Chen K, Zheng X, Zhao H, et al. Ability to generate patient-derived breast cancer xenografts is enhanced in chemoresistant disease and predicts poor patient outcomes. *PLoS One* 2015;10:e0136851.
- Woo XY, Giordano J, Srivastava A, Zhao Z-M, Lloyd MW, de Bruijn R, et al. Conservation of copy number profiles during engraftment and passaging of patient-derived cancer xenografts. *Nat Genet* 2021;53:86–99.
- Zhang J, Fujimoto J, Zhang J, Wedge DC, Song X, Zhang J, et al. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* 2014;346:256–9.
- Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* 2012;366:883–92.
- Dentro SC, Leshchiner I, Haase K, Tarabichi M, Wintersinger J, Deshwar AG, et al. Characterizing genetic intra-tumor heterogeneity across 2,658 human cancer genomes. *Cell* 2021;184:2239–54.e39.
- Meehan TF, Conte N, West DB, Jacobsen JO, Mason J, Warren J, et al. Disease model discovery from 3,328 gene knockouts by the international mouse phenotyping consortium. *Nat Genet* 2017;49:1231–8.
- Corbett TH, White K, Polin L, Kushner J, Paluch J, Shih C, et al. Discovery and preclinical antitumor efficacy evaluations of LY32262 and LY33169. *Invest New Drugs* 2003;21:33–45.
- Petrosyan V, Dobrolecki LE, LaPlante EL, Srinivasan RR, Bailey MH, Welm AL, et al. Immunologically cold triple negative breast cancers engraft at a higher rate in patient derived xenografts. *NPJ Breast Cancer* 2022;8:104.
- Verma B, Wesa A. Establishment of humanized mice from peripheral blood mononuclear cells or cord blood CD34+ hematopoietic stem cells for immunology studies evaluating new therapeutic agents. *Curr Protoc Pharmacol* 2020;89:e77.
- Meraz IM, Majidi M, Meng F, Shao R, Ha MJ, Neri S, et al. An improved patient-derived xenograft humanized mouse model for evaluation of lung cancer immune responses. *Cancer Immunol Res* 2019;7:1267–79.

21. Ito M, Hiramatsu H, Kobayashi K, Suzue K, Kawahata M, Hioki K, et al. NOD/SCID/gamma(c)(null) mouse: an excellent recipient mouse model for engraftment of human cells. *Blood* 2002;100:3175–82.
22. Shultz LD, Lyons BL, Burzenski LM, Gott B, Chen X, Chaleff S, et al. Human lymphoid and myeloid cell development in NOD/LtSz-scid IL2R gamma null mice engrafted with mobilized human hemopoietic stem cells. *J Immunol* 2005;174:6477–89.
23. Okada S, Harada H, Ito T, Saito T, Suzu S. Early development of human hematopoietic and acquired immune systems in new born NOD/Scid/Jak3null mice intrahepatic engrafted with cord blood-derived CD34 + cells. *Int J Hematol* 2008;88:476–82.
24. Vellano CP, White MG, Andrews MC, Chelvanambi M, Witt RG, Daniele JR, et al. Androgen receptor blockade promotes response to BRAF/MEK-targeted therapy. *Nature* 2022;606:797–803.
25. McQuade JL, Daniel CR, Hess KR, Mak C, Wang DY, Rai RR, et al. Association of body-mass index and outcomes in patients with metastatic melanoma treated with targeted therapy, immunotherapy, or chemotherapy: a retrospective, multicohort analysis. *Lancet Oncol* 2018;19:310–22.
26. Qi L, Kogiso M, Du Y, Zhang H, Braun FK, Huang Y, et al. Impact of SCID mouse gender on tumorigenicity, xenograft growth and drug-response in a large panel of orthotopic PDX models of pediatric brain tumors. *Cancer Lett* 2020;493:197–206.