

Improved analysis of *in vivo* drug combination experiments with a comprehensive statistical framework and web-tool

Received: 14 February 2025

Accepted: 10 October 2025

Published online: 19 November 2025

 Check for updates

Rafael Romero-Becerra   ^{1,2}, Zhi Zhao  ², Daniel Nebdal¹, Elisabeth Müller  ^{1,3}, Helga Bergholtz  ¹, Jens Henrik Norum  ¹ & Tero Aittokallio  ^{1,2,4,5} 

Drug combination therapy is often required to overcome the limited benefits of monotherapy in cancer treatment. While several tools exist for *in vitro* drug synergy screening and assessment, there is a lack of integrated methods for statistical analysis of *in vivo* combination experiments. To fill this gap, we present SynergyLMM, a comprehensive modeling and design framework for evaluating drug combination effects in preclinical *in vivo* studies. Unlike other methods, SynergyLMM accommodates complex experimental designs, including multi-drug combinations, and offers practical options for statistical analysis of both synergy and antagonism through longitudinal drug interaction analysis, including model diagnostics and statistical power analysis. These functionalities allow researchers to optimize study designs and determine an appropriate number of animals and follow-up time points required to achieve sufficient synergy and statistical power. SynergyLMM is implemented as an easy-to-use web-application, making it widely accessible for researchers without programming skills. We demonstrate the versatility and added value of SynergyLMM through its applications to various experimental setups and treatment experiments with chemo-, targeted- and immunotherapy. These case studies showcase its potential to improve robustness, statistical rigor and consistency of preclinical drug combination results, enabling a faster and safer transition from preclinical to clinical testing.

In advanced phase cancers, the efficacy of monotherapies is often limited by the development of drug resistance to chemo-, targeted- or immunotherapies^{1,2}. In this context, drug combinations offer a promising solution to monotherapy resistance, and various classes of combinations are currently being evaluated in numerous clinical

trials³. Cooperative or synergistic drug combinations can arise from various mechanisms, such as blocking compensatory feedback mechanisms, inhibiting complementary immunological processes, or targeting distinct subpopulations within heterogeneous tumors, among others⁴. Since the number of possible drug combinations vastly

¹Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway. ²Department of Biostatistics, Faculty of Medicine, Oslo Centre for Biostatistics and Epidemiology (OCBE), University of Oslo, Oslo, Norway. ³Division of Integrated High-Order Regulatory Systems, Center for Cancer Immunotherapy and Immunobiology, Kyoto, Japan. ⁴Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland. ⁵iCAN Digital Precision Cancer Medicine Flagship, University of Helsinki and Helsinki University Hospital, Helsinki, Finland.  e-mail: r.r.becerra@medisin.uio.no; t.a.aittokallio@medisin.uio.no

exceeds what can be tested clinically, preclinical models are the key first step toward testing interaction mechanisms and identifying potential therapeutic combinations. In vivo animal models—particularly mouse models such as patient-derived xenografts (PDXs)—are especially valuable for the identification of safe and effective drug combinations. Compared to *in vitro* cell line models, *in vivo* animal models better capture tumor heterogeneity and mimic tumor molecular features and treatment responses¹⁴. However, *in vivo* models often demonstrate substantial heterogeneity in treatment effects among animals, which, while closely reflecting clinical outcomes, poses challenges for accurate evaluation of treatment responses.

Currently, there is no established community-wide standard method to quantify synergy in preclinical cancer studies, but rather, researchers use various statistical and non-statistical approaches. The situation is similar to human clinical studies, where a number of proposed models exist to distinguish between drug interactions resulting in additivity, synergy or drug independence⁵, and to assess whether clinical benefits of approved drug combinations are due to additive or synergistic effects⁶. However, as yet, no general consensus on the definition and reporting of drug combination synergy exists⁷. Even in multi-dose combination studies in cancer cell lines *in vitro*, several drug combination synergy scoring approaches and definitions are being used, such as Bliss independence⁸, response additivity (RA)⁹, and highest single agent (HSA)¹⁰ model, each formulated under different assumptions, and which may lead to rather different interpretations of synergy¹¹. In the absence of a commonly accepted synergy definition in animal studies, the researchers may easily become biased toward the selection of a synergy scoring approach that best supports their hypotheses. Such a selection bias reduces the consistency between combination studies, leads to delays in the discovery of true synergistic drug combinations, and may negatively impact the translatability of combination discovery efforts.

In *in vivo* treatment experiments, it is critical to consider the longitudinal nature of tumor growth measurements in the statistical analyses of treatment effects¹². Still today, most animal studies use single endpoints at a prespecified time, together with simple statistical tests, such as t-test or its variants^{13,14}. However, such approaches may lead to lower statistical power, since they cannot leverage the dynamic information of the longitudinal tumor growth measurements. Narayan et al.¹³ further quantified the *in vivo* effects of drug combination using the median effect principle by Chou and Talalay¹⁵ to calculate the combination index (CI), which requires testing multiple drug dose levels to generate dose-effect curves. While this approach is widely used for *in vitro* combination discovery, it does not account for the specific aspects inherent to *in vivo* studies, such as lack of multiple doses, variability between individual mice, and the longitudinal structure of the data, among others¹². Recently, more rigorous statistical methods have been developed^{16–21}, such as publicly available tools CombPDX²⁰ and invivoSyn²¹. Even though these methods offer a clear advantage compared to the traditional analysis of *in vivo* combination experiments, they each come with important limitations (Table 1).

To fill these gaps, we developed a comprehensive statistical framework for analyzing *in vivo* drug combination experiments in animal models. The framework is based on using either exponential or Gompertz tumor growth kinetics and a (non-)linear mixed model (LMM), which effectively captures both the inter-animal heterogeneity and dynamic changes in the combination effects estimated from longitudinal tumor growth experiments. The SynergyLMM method is implemented as an R package and also deployed as an easy-to-use interactive web-tool (available at <https://synergylmm.uiocloud.no/>) for those users without programming skills. In this study, we demonstrate the unique benefits of SynergyLMM and show how it enables the identification of statistically significant and pharmaceutically meaningful combinatorial synergies and antagonistic effects, as demonstrated in selected preclinical application cases and datasets. These

Table 1 | Comparison of functionalities in methods for *in vivo* drug combination synergy analysis

	Wu J et al. ¹⁶ (2012)	Wu J ¹⁷ (2013)	Zhao W et al. ¹⁸ (2019)	Demidenko and Miller ¹⁹ (2019)	CombPDX ²⁰ (2022)	invivoSyn ²¹ (2023)	SynergyLMM (this work)
Package available	–	–	–	–	–	✓	✓
Web app available	–	–	–	–	✓	–	✓
Synergy models implemented	Bliss	Bliss	Bliss	Bliss	Bliss, HSA, RA	Bliss, HSA, RA	Bliss, HSA, RA
Growth dynamics considered	–	–	✓	✓	–	✓	✓
Uses all time points	–	–	–	✓	–	✓	✓
Statistical assessment of synergy	–	–	–	✓	✓	✓	✓
Statistical assessment of antagonism	–	–	–	–	–	–	–
Analysis of higher-order drug combinations	–	–	–	–	–	–	–
Model diagnostics	✓	–	–	–	–	–	–
Power analysis	–	–	–	–	–	✓	✓

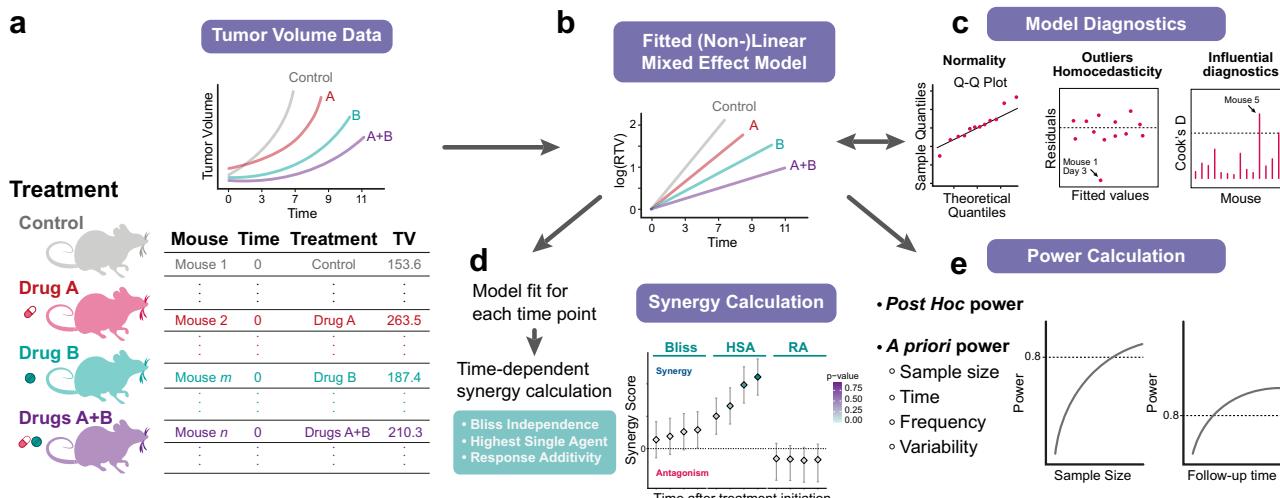


Fig. 1 | Overview of the SynergyLMM workflow. **a** The first step involves uploading the longitudinal tumor burden-related measurements for the different treatment groups. The input consists of a tabular data in long format with at least 4 columns containing information about the sample IDs, the time points for each measurement, the treatment group, and the tumor measurement. **b** The input data will then be processed to estimate the (non-)linear mixed effect model that fits the tumor growth dynamics for each subject, and estimate the growth rates for each treatment group. **c** SynergyLMM offer various functions to evaluate the model diagnostics and model performance, as well as for the identification of potential outliers and influential individuals for the model fit and the treatment groups.

d Once a proper model with satisfactory diagnostics has been obtained, the statistical assessment of combination effects is performed, with a time dependent estimation of the synergy score and combination index, along with their confidence intervals and statistical significance. The method allows for testing synergy using three reference models: Bliss independence, highest single agent (HSA), and response additivity (RA). **e** SynergyLMM implements versatile functions to calculate the post hoc power of the experiment and the a priori power by modifying the values of key experimental variables, such as sample size, follow-up time, or frequency of the measurements.

case studies cover several experimental settings and treatment classes of chemo-, targeted-, and immunotherapy, and show that careful interpretation of the identified synergy and antagonism patterns at various time points is essential for avoiding false positive and false negative findings. The SynergyLMM approach enables drug combination discovery with higher confidence, reproducibility, and interpretability, and we expect it will support many exciting applications of combinatorial discovery and clinical translation.

Results

Overview of the SynergyLMM framework

The SynergyLMM workflow for the assessment of combination effects in vivo drug combination experiments involves five main steps (Fig. 1). The input data consists of longitudinal tumor burden-related measurements in various treatment groups and control (treatment-naïve) animals (Fig. 1a). In most applications, this would be tumor volume measurements at different time points, but the input data can be any measurement that reflects the tumor burden or the number of cancerous cells, such as luminescence signal. For each animal, the tumor measurements are normalized against the treatment initiation time point to adjust for the variability in initial tumor burden across animals. These data are then used to fit a linear mixed effect model in the case of the exponential model, or a non-linear mixed effect model in the case of the Gompertz model, for describing the tumor growth dynamics and to estimate the growth rate parameters for each treatment group (Fig. 1b). Once the model is estimated, the SynergyLMM framework provides several tools for its statistical diagnosis to check how well the model fits the data, along with several functions to identify potential outlier observations and highly influential subjects in the observed data (Fig. 1c).

Based on the well-estimated mixed effect model, time-resolved synergy scores (SSs) and CI for the drug combination effect can be estimated. The SynergyLMM method supports the use of various synergy scoring models, such as Bliss independence, HSA, and RA (see “Methods” for details), accompanied by uncertainty quantification and

statistical assessment of synergy and antagonism (*p* values) (Fig. 1d). Finally, the SynergyLMM method also provides guidance for the experimental design and calculation of the statistical power of the synergy analysis, including post-hoc power analysis of completed experiments, and a priori power analysis for a new experiment, which enables investigation of the effect of various experimental factors, such as sample size and number and frequency of measurements (Fig. 1e). To assist end-users, a detailed guideline document—covering topics such as reference model selection, model diagnostics, and results interpretation—is provided as Supplementary Information S1. The following sections demonstrate the unique benefits of the SynergyLMM method in various experimental setups and study questions.

Time-resolved evaluation of synergy and antagonism with SynergyLMM

To test and validate the SynergyLMM method for time-dependent in vivo synergy evaluation, we reanalyzed the drug combination experiments from Narayan et al.¹³ (Fig. 2). We first reanalyzed the results from the U87-MG Fluc-Mcherry (FM) orthotopic glioblastoma model and Docetaxel + GNE-317 combination (Fig. 2a). The SynergyLMM analysis showed a significant synergy under the HSA model, but not enough statistical evidence to reject additivity under the Bliss model (Fig. 2b and Source Data S1). This also contrasted with the results obtained by the authors using the median-effect principle, which pointed to a synergistic effect. This demonstrates the importance of choosing the appropriate synergy reference model that is compatible with the drug combination under investigation.

Interestingly, the SynergyLMM results for the Imatinib + Dasatinib combination in the BV-173-Gluc leukemia model (Fig. 2c) showed opposite results with the Bliss and HSA models (Fig. 2d and Source Data S1). While the HSA analysis indicated a significant synergy at each time point, the Bliss synergy results indicated an antagonistic effect for most time points (Fig. 2d). This is due to the strong monotherapy effect that each individual drug alone had on tumor growth. Therefore, the expected drug combination additive effect by the Bliss model is

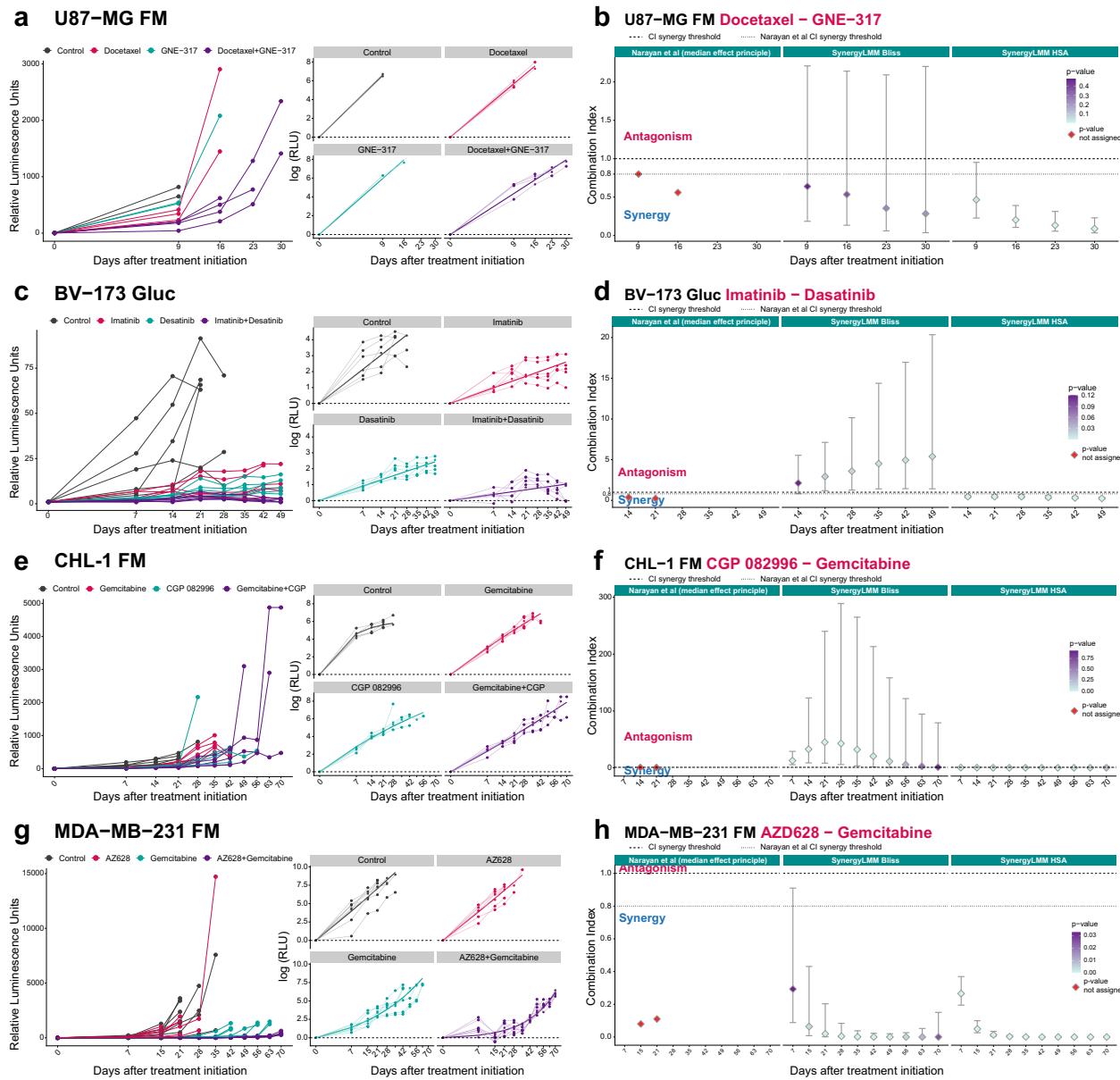


Fig. 2 | Time-dependent assessment of synergy using SynergyLMM.

a, b Evaluation of combination Docetaxel + GNE-317 in U87-MG FM orthotopic glioblastoma model. Each curve corresponds to an individual mouse: control $n = 4$, Docetaxel $n = 4$, GNE-317 $n = 4$, and Docetaxel + GNE-317 $n = 4$. **c, d** Evaluation of combination Imatinib + Dasatinib in BV-173-Gluc leukemia model. Each curve corresponds to an individual mouse: control $n = 6$, Imatinib $n = 6$, Dasatinib $n = 6$, and Imatinib + Dasatinib $n = 6$. **e, f** Evaluation of combination Gemcitabine + CGP-082996 in CHL-1 FM melanoma model. Each curve corresponds to an individual mouse: control $n = 6$, Gemcitabine $n = 6$, CGP-082996 $n = 6$, and Gemcitabine + CGP-082996 $n = 6$. **g, h** Evaluation of combination AZD628 + Gemcitabine in the MDA-MB-231 FM triple-negative breast cancer model. Each curve corresponds to an individual mouse: control $n = 8$, AZD628 $n = 7$, Gemcitabine $n = 7$, and AZD628 +

Gemcitabine $n = 8$. **a, c, e, g** display tumor growth data for the different drug combinations across the distinct cancer models. The bold lines in the right plots represent the fitted regression lines for the fixed effects of each treatment estimated by the (non-)linear mixed effects models. The exponential growth model was used in **(a, c)**, while the Gompertz growth model was used in **(e, g)**. The x-axis of the plots represents the original time in square-root scale. **b, d, f, h** show the synergy evaluations reported in the original publication and reanalyzed using SynergyLMM. The error bars indicate the 95% confidence intervals for the combination index. The dot in the center indicates the estimated combination index value. The dot colors indicate the two-sided (non-adjusted) p value from SynergyLMM. CI combination index. RLU relative luminescence units. HSA highest single agent. Source data are provided as a Source data file.

higher than the observed combination effect, resulting in an antagonistic effect for the drug combination. Similar results were obtained in the melanoma model CHL1 FM and CGP-082996 + Gemcitabine combination (Fig. 2e, f). The HSA synergy results indicated a significant synergy at several of the early time points, which was lost at the final time points, where more variability in the tumor measurements was observed, and combination additivity could not be rejected (Fig. 2f and Source Data S1).

Finally, the SynergyLMM analysis of the drug combination AZD628 + Gemcitabine in the triple-negative breast cancer cell model MDA-MB-231 FM (Fig. 2g) supported the conclusions of the authors by showing a significant synergistic effect at multiple time points when using both the Bliss and HSA synergy models (Fig. 2h and Source Data S1). However, the SynergyLMM results obtained with the Bliss model revealed differences in the synergistic effects across time points, and more accurately reflected the tumor growth pattern in the

combination group, with a stronger effect at the initial time points that is gradually reduced at the later time points (Fig. 2a).

The model diagnostics indicated that the model assumptions were largely satisfied across all the experiments (Supplementary Fig. 1). For U87-MG FM and BV-173 Gluc data, the exponential growth model offered the best diagnostics and performance metrics. In contrast, the Gompertz growth model was selected for the CHL-1 FM and MDA-MB-231 FM data. For the CHL-1 FM experiment, the Gompertz model offered the best model performance metrics, while the exponential growth did not properly explain the tumor growth and showed evident violations of the model assumptions in the diagnostic plots in the case of the MDA-MB-231 FM dataset. A degree of small deviations from the normality of the random effects were observed in the MDA-MB-231 FM and U87-MG FM experiments, likely due to potential outlier measurements that might deserve further analysis. Additionally, plots of observed versus predicted values supported the adequacy of the models in explaining the longitudinal data (Supplementary Fig. 2).

As mentioned in the user guidelines document, differences in the tumor volumes between groups at the moment of treatment initiation might affect the final results (Supplementary Information S1). We evaluated whether there were differences in baseline tumor volumes that could have influenced the outcomes (Supplementary Fig. 2e). We found significant differences in the baseline tumor volume between groups in the melanoma model CHL1 FM. Hence, the results obtained from the analysis of this experiment should be interpreted with caution, since the differences in baseline tumor volume could have had an influence on the treatment response. To avoid such potential problems, we recommend a careful selection of tumor size windows at treatment initiation to ensure comparability across groups (Supplementary Information S1).

Taken together, these results support the value of SynergyLMM for time-dependent drug combination evaluation, allowing not only the statistical identification of synergy but also antagonism, something that is missing in the existing synergy analysis methods. These results also demonstrate the importance of selecting an appropriate tumor growth and synergy model based on the combination experiment under evaluation.

SynergyLMM enables the identification of influential subjects

In addition to providing a variety of tools for model diagnostics that enable testing of model assumptions, SynergyLMM also offers several options to identify highly influential subjects (animals) in the experiment. These options include detecting subjects with a substantial impact on the fitted values, as well as identifying those subjects with a significant influence on the estimation of control and treatment group fixed effects. Both options are based on the calculation of Cook's distances, as described in Supplementary Methods.

As an example, Supplementary Fig. 3 illustrates the application of these diagnostics options in the MDA-MB-231 FM AZD628-Gemcitabine combination experiment from Narayan et al.¹³. The Cook's distances based on the change of the fitted values reveal a subject in the control group that has a greater influence on the fitted values than the other subjects (Supplementary Fig. 3a). Similarly, the analysis of Cook's distances based on the change in the fixed effects indicates that the same individual strongly affected the estimation of the control group's growth rate coefficient (Supplementary Fig. 3b). Supplementary Fig. 3c confirms that this subject exhibits a tumor growth curve that is notably distinct from the other animals.

These diagnostic analyses provide a deeper understanding of the experiment and the modeling results by identifying potential outliers, measurement errors, or subjects with unique tumor growth patterns. Such insights improve the reliability of the conclusions by highlighting individuals or measurements that may warrant further investigation to reveal the reasons behind their unusual growth behavior, and

potentially exclude these before reanalysis after careful reporting and justification.

Evaluation of tumor-related pathways using SynergyLMM

SynergyLMM provides a robust statistical framework for evaluating synergy and antagonism, making it a versatile tool for assessing drug combination effects *in vivo*. Beyond mere validation of hypothesized synergy between treatments, SynergyLMM also supports the exploration of drug interaction effects on molecular pathways, hence offering opportunities to uncover insights into tumor biology.

To demonstrate its wide utility, we reanalyzed the data from the study of Gilfillan et al.²², which hypothesized that enhanced HER2/HER3 signaling could drive tumor growth independently of estrogen receptor (ER) activity in hormone therapy-resistant tumors. To corroborate this hypothesis, the authors tested the effect of Heregulin (HER3 agonist), EGF (EGFR agonist), Fulvestrant (ER antagonist), Trastuzumab (HER2 inhibitor), and pairwise combinations of these, in a luminal-like breast cancer patient-derived PDX mouse model (MAS98.06), in which tumor growth is dependent on estrogen²³ (Fig. 3a, b). Unlike typical drug combination studies, this experiment specifically tested the interactions of growth factors with antitumor agents.

We first confirmed that there were no significant differences in the initial tumor volume between the treatment groups (Supplementary Fig. 4a). After fitting the models separately for each drug combination experiment, we evaluated the adequacy of the models for capturing the observed values (Supplementary Fig. 4b–e), and the model diagnostics to corroborate the validity of the model assumptions (Supplementary Fig. 5). The model diagnostics indicated that the model assumptions were largely satisfied across all experiments and supported the validity of the models for fitting the longitudinal data, although some extreme observations were also detected that might deserve further scrutiny.

We analyzed the effects of the different combinations at the final experimental time point (28 days) using the HSA model (Fig. 3c and Source Data S2). Consistent with the authors' proposed mechanism, the combination of Heregulin and Fulvestrant demonstrated an antagonistic effect, leading to a greater tumor growth compared to Fulvestrant monotherapy. Notably, SynergyLMM analysis also revealed a synergistic effect between EGF and Fulvestrant, resulting in reduced tumor growth compared to Fulvestrant monotherapy. Finally, as expected, the results from the combinations of Fulvestrant + Trastuzumab and Heregulin + Trastuzumab showed no significant differences compared to their respective monotherapy groups (Fig. 3c).

These results exemplify the utility of SynergyLMM for the study of interaction effects of targeted agents on tumor biology. The results did not only support the hypothesized mechanism, but also identified an interesting synergistic interaction between EGF and Fulvestrant, which deserves further investigation.

SynergyLMM enables analysis of triple drug combination effects

The combination treatments for cancer therapy often go beyond the standard two-drug combinations, and analysis of higher-order (3 or more drugs) combinations is routinely done *in vitro* combination experiments²⁴. Although the *in vivo* assessment of higher-order combinations is not commonly performed for practical reasons, we extended the application of SynergyLMM for the analysis of three-drug combination experiments to offer a solution for statistical synergy evaluation required in these experiments (see "Methods" and Supplementary Methods).

In addition to the two-drug combination responses, Gilfillan et al.²² also investigated the effect of Heregulin + Fulvestrant + Trastuzumab combination in the MAS98.06 PDX model (Fig. 4a). We re-analyzed these data using SynergyLMM with the Bliss, HSA, and RA models (Fig. 4b and Source Data S3). The model diagnostics analysis indicated the adequacy of the model (Supplementary Fig. 6a, b). We also confirmed that there were no differences in the initial tumor volume

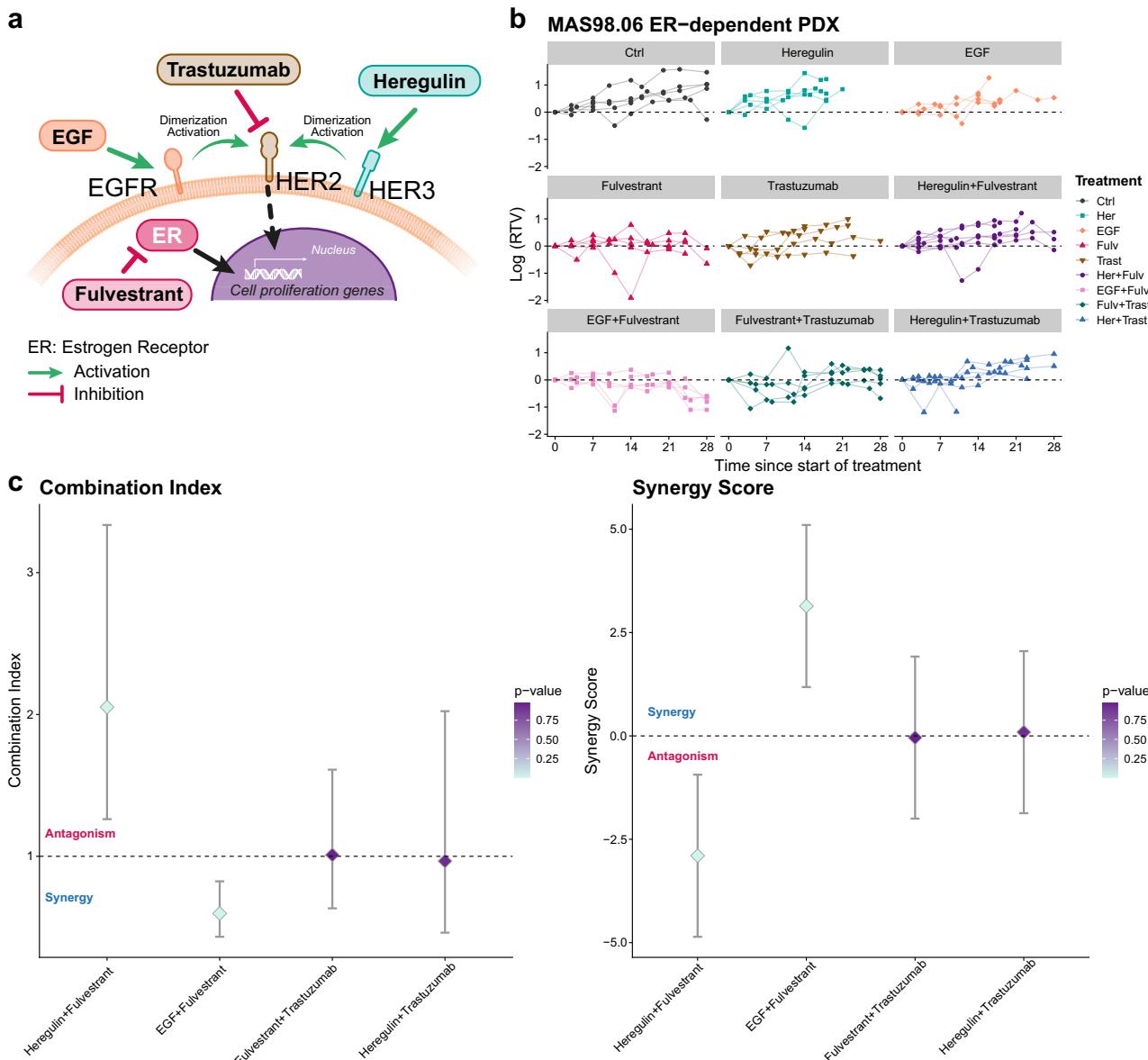


Fig. 3 | Analysis of drug combination effects on tumor growth in a luminal-like breast cancer PDX model using SynergyLMM. **a** Schematic representation of the signaling pathway investigated by Gilfillan et al. MAS98.06 is a luminal-like breast cancer PDX model in which tumor growth is promoted by the activity of the estrogen receptor (ER). Fulvestrant acts as an ER antagonist, inhibiting tumor growth. The activation of HER3 by Heregulin, or EGFR by EGF, promotes their dimerization with HER2, inducing tumor growth in an ER-independent manner. Trastuzumab is a monoclonal antibody that targets HER2, inhibiting this pathway. **b** Tumor growth in Heregulin, EGF, Fulvestrant, and Trastuzumab combination experiments in the breast cancer PDX MAS98.06 model. Each curve corresponds to an individual mouse: control $n = 5$, Heregulin $n = 5$, EGF $n = 4$, Fulvestrant $n = 5$,

Trastuzumab $n = 5$, Heregulin + Fulvestrant $n = 8$, EGF + Fulvestrant $n = 5$, Fulvestrant + Trastuzumab $n = 5$, Heregulin + Trastuzumab $n = 6$. **c** Results of the highest single agent (HSA) synergy evaluation using SynergyLMM, at day 28, based on the experiments shown in **(b)**. The horizontal dashed lines indicate the synergy thresholds of 1 for the combination index and 0 for the synergy score. The error bars indicate the 95% confidence intervals for the combination index and synergy score. The dot in the center indicates the estimated combination index and synergy score values. The dot colors indicate the two-sided (non-adjusted) p value from SynergyLMM. Ctrl Control, Her Heregulin, Fulv Fulvestrant, Trast Trastuzumab. Source data are provided as a Source data file.

between the groups that could affect the results (Supplementary Fig. 6c). The results for the triple drug combination were similar to those obtained for the Fulvestrant+Trastuzumab double drug combination (Source Data S2). No evidence of synergy was found with the Bliss or RA models, and only the HSA model suggested a statistically significant synergy at day 8 after treatment initiation, which was lost at later time points (Fig. 4b). These results support the mechanism proposed by the authors, by which the antagonistic effect of Heregulin would be lost when HER2 is inhibited by Trastuzumab (Fig. 3a).

Narayan et al.¹³ also tested the combination of three drugs (Osimertinib, AZD2014, and Docetaxel) in the U87-MG FM orthotopic

glioblastoma model (Fig. 4c). Even though not well-suited for fixed-dose treatments experiments, the authors used the median effect principle by Chou and Talalay¹⁵ to report the synergistic drug combination effect. Our reanalysis of these data using SynergyLMM could not corroborate this conclusion, and even showed some tendency of an antagonistic effect of the three-drug combination at the early time points with the Bliss and RA models (Fig. 4d and Source Data S3). SynergyLMM analysis at the remaining time points indicated that the additive effect of the drug combination could not be statistically rejected. The model diagnostics confirmed that the model assumptions were largely met (Supplementary Fig. 7a, b), supporting the

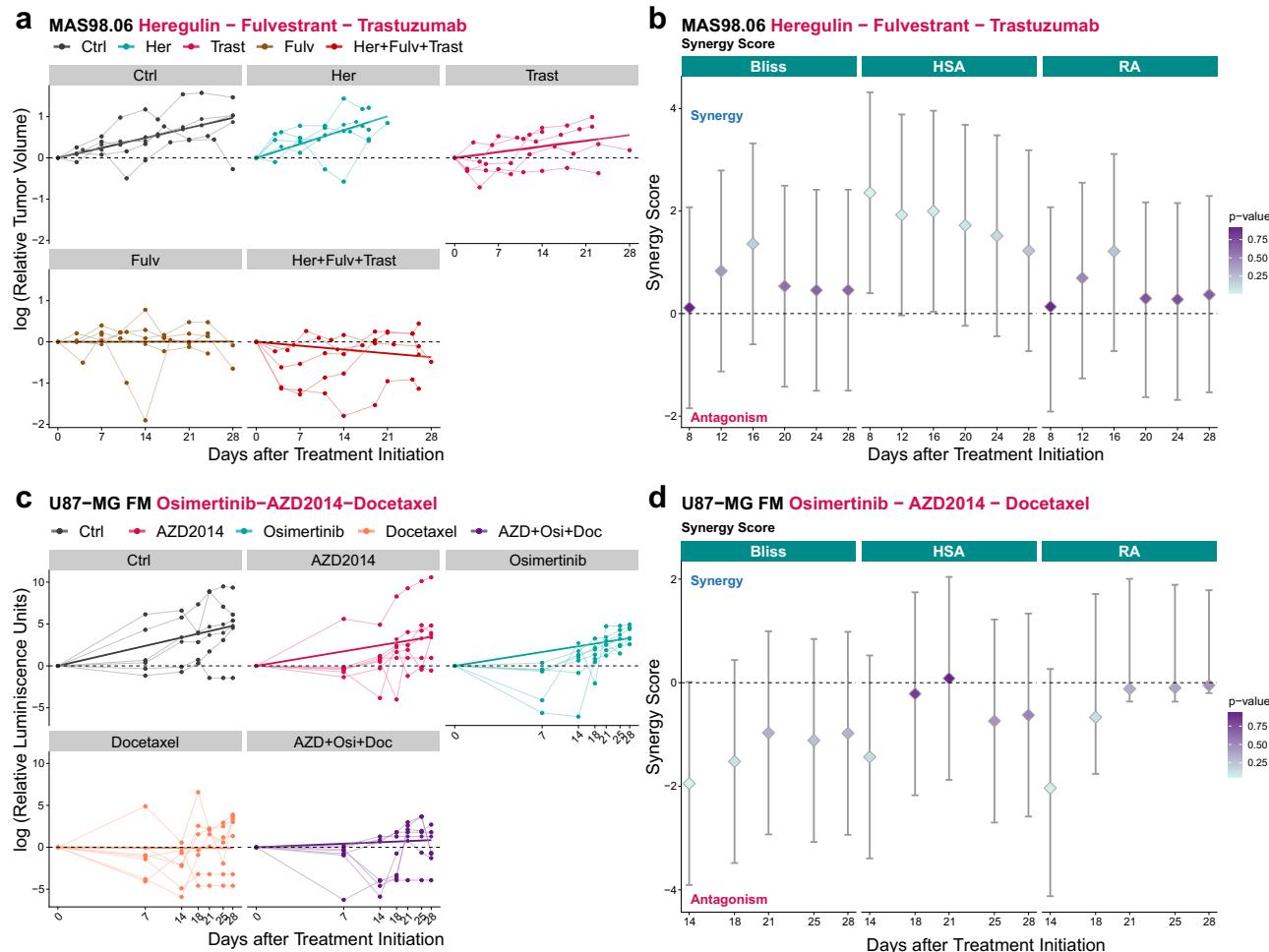


Fig. 4 | Assessment of triple drug combinations using SynergyLMM.

a, b Heregulin + Fulvestrant + Trastuzumab triple drug combination effects on tumor growth in the MAS98.06 luminal-like breast cancer PDX model. The measurements for Control and Heregulin, Fulvestrant, and Trastuzumab single treatments correspond to the same data shown in Fig. 4b. Each curve corresponds to an individual mouse: control $n = 5$, Heregulin $n = 5$, Fulvestrant $n = 5$, Trastuzumab $n = 5$, Heregulin + Fulvestrant + Trastuzumab $n = 5$. **c, d** In vivo evaluation of Osimertinib + AZD2014 + Docetaxel triple drug combination effect on tumor growth in U87-MG FM orthotopic glioblastoma model. Each curve corresponds to an individual mouse: control $n = 6$, AZD2014 $n = 7$, Osimertinib $n = 7$, Docetaxel $n = 7$, AZD2014 + Osimertinib + Docetaxel $n = 7$. **a, c** display tumor growth data for the different drug combinations across the distinct cancer models. The bold lines in the

plots represent the fitted regression lines for the fixed effects of each treatment estimated by the linear mixed-effects models. The x-axis of the plot in **c** represents the original time in square-root scale. **b, d** show the synergy evaluation using SynergyLMM for MAS98.06 (**b**) and U87-MG FM (**d**) tumor models for Bliss, highest single agent (HSA), and response additivity (RA) synergy models. The horizontal dashed lines indicate the synergy score threshold of 0 to assess synergy or antagonism. The error bars indicate the 95% confidence intervals for the synergy score. The dot in the centers indicate the estimated synergy score value. The dot colors indicate the two-sided (non-adjusted) p value from SynergyLMM. Ctrl Control, Her Heregulin, Fulv Fulvestrant, Trast Trastuzumab, AZD AZD2014, Osi Osimertinib, Doc Docetaxel. Source data are provided as a Source data file.

validity of the combination testing results. We did not find any significant differences in the initial tumor volume between the groups, although we found a potential outlier individual with a larger initial tumor volume than the rest of animals, which might deserve further exploration (Supplementary Fig. 7c).

These two experiments demonstrate the capability of SynergyLMM to evaluate combination effects in more complex, triple-drug combination experiments, with rigorous and time-dependent statistical modeling. Therefore, SynergyLMM offers a unique tool for in vivo studies that aim to assess the therapeutic potential of higher-order drug combinations, something that is not implemented in the other in vivo combination evaluation methods (Table 1).

Comparison of SynergyLMM with other statistical methods

One common statistical approach to assess drug combination effects is to compare treatment groups at a single endpoint using t-test or its variants^{21,25–27}. However, these approaches pose several limitations,

including low statistical power due to the loss of dynamic information from the longitudinal tumor growth measurements. To exemplify the advantages of using SynergyLMM over these simpler approaches, we compared the analysis of the Heregulin–Fulvestrant combination in MAS98.06 PDX model shown in Fig. 3 using SynergyLMM and the standard t-test at different time points (Supplementary Fig. 8). This dataset offers a practical example, because the measurements per animal are not evenly distributed across the treatment groups (Supplementary Fig. 8a). Direct statistical comparison between the groups using t-test cannot deal with incomplete data, and only the time points in which there are measurements for both groups being compared can be analyzed. In contrast, SynergyLMM estimates the growth rates of the treatment groups based on all the available data, and it uses the estimated coefficient to make the comparison (Supplementary Fig. 8b). Another limitation is that the standard t-test does not account for intra-subject correlations, and it treats each time point independently, losing the longitudinal information provided by the history of

the tumor growth. Besides making impossible the comparison at several time points, t-test failed to detect significant differences at the last time points (Supplementary Fig. 8b). We note that such statistical comparison of the combination therapy vs the individual drug treatments at a single time-point is effectively equal to using the HSA model for synergy quantification. In the statistical assessment of synergy based on the other synergy models, such as Bliss independence or RA, all the four (or more) treatment groups are considered simultaneously for the longitudinal comparison, similarly as SynergyLMM does.

More comprehensive statistical methods have been developed to overcome the limitations of the simple statistical approaches. However, only a few existing methods developed for *in vivo* synergy analysis currently provide open-access source codes or easy-to-use web-based tools for their wide application (Table 1). One such method is invivoSyn, which has been developed to identify synergy from *in vivo* drug combinations experiments²¹. To further evaluate the relative power of SynergyLMM to detect time-dependent synergy effects, we reanalyzed the data for the combination effect of Cetuximab + Palbociclib in the colon PDX model CR1197²¹. This model has been shown to exhibit a delayed response to the drug combination in some individuals²¹ (Fig. 5a). The SynergyLMM analysis corroborated that a significant Bliss synergy was only identified at the last time points of the experiment (Fig. 5b). Although the analysis with invivoSyn also showed that the CI values varied with time, the statistical *p* values remained constant (Fig. 5b and Source Data S4). This is due to the default use of all data points for the statistical assessment of synergy in invivoSyn, hence leading to a *p* value that is calculated independently of the values of the CI or SS at each time point²¹. Therefore, the SynergyLMM results provided a more precise description of the time-dependent drug combination effects and their statistical significance.

One of the unique features of SynergyLMM is the statistical assessment of antagonism in addition to synergism. As shown in Fig. 2d, f, SynergyLMM correctly reported antagonism in those cases in which the drug combination effect was lower than the expected additive effect according to the Bliss model. To further validate the identification of antagonism using SynergyLMM, we evaluated the effect of GABA + anti-PD1 treatment in the breast cancer cell model 4T1 (Fig. 5c). GABA is a metabolite and neurotransmitter, which has been shown to mediate immune suppression and antagonize immune checkpoint inhibition therapy, such as anti-PD1 treatment^{28,29}. As expected, and in agreement with previous results²⁸, the GABA treatment alone did not have any effect on the tumor growth (Fig. 5c). Therefore, we evaluated the drug combination effect using the HSA model, and compared the results obtained with CombPDX, invivoSyn, and SynergyLMM (Fig. 5d). The three methods showed negative SSs and confidence intervals, pointing to the expected antagonistic effect. However, only SynergyLMM assigned a significant *p* value for the antagonistic effect, while CombPDX and invivoSyn reported highly insignificant *p* values (Fig. 5d and Source Data S4). This is due to the one-tailed tests used in the statistical assessment of synergy in CombPDX and invivoSyn methods, which do not allow for testing of antagonistic effects, and which can artificially inflate the statistical significance of synergy.

Finally, we also compared the results obtained with CombPDX, invivoSyn, and SynergyLMM in the assessment of Bliss synergy for the combination of Rabusertib and Irinotecan in various colon cancer models¹⁴. The three methods pointed to a synergistic effect for the drug combination in the SW837 model (Fig. 5e, f). Notably, invivoSyn was the only method that assigned a significant synergistic effect, while SynergyLMM showed borderline significance only at the last time point (Fig. 5f and Source Data S4). This is because invivoSyn performs a one-tailed test in the assessment of synergy, whereas a two-tailed test would have provided results similar to those of SynergyLMM. CombPDX showed a positive SS and confidence interval for the last time point, indicating synergy. However, this was inconsistent with the

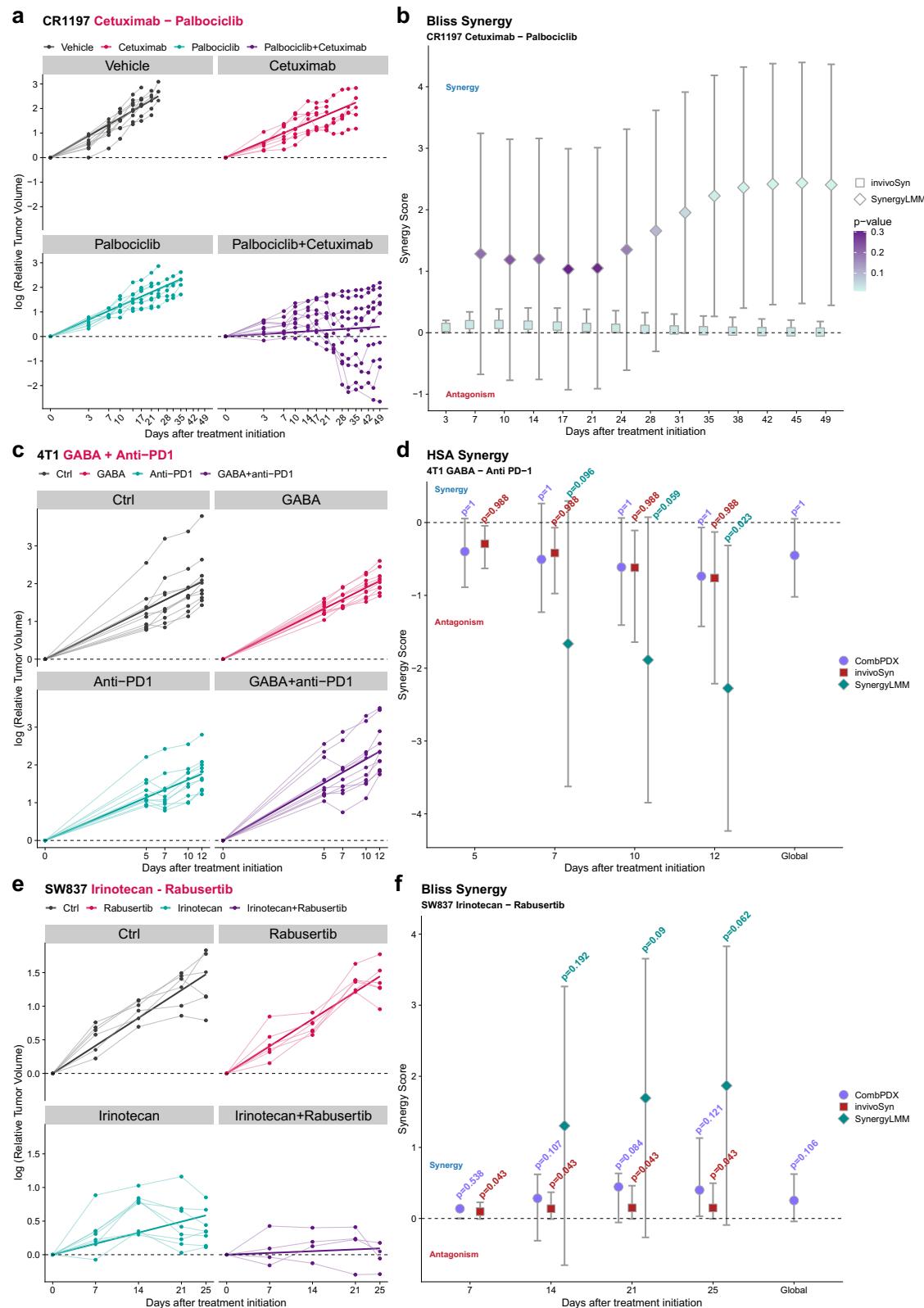
reported *p* value, which indicated a non-significant effect (Fig. 5f and Source Data S4). In the case of LS-1034 and SNU-81 models, SynergyLMM and CombPDX showed similar results at the initial time points (Supplementary Fig. 9 and Source Data S4). However, CombPDX could not analyze the last time points in these experiments, since tumor measurements from all the groups were not available. SynergyLMM was the only method that assigned a significant synergistic effect at the later time points in the LS-1034 model, while invivoSyn suggested a tendency of significant synergy through all time points (Supplementary Fig. 9b). In contrast, SynergyLMM indicated a tendency, although not statistically significant, of a synergistic effect on the SNU-81 model at the last time point, while invivoSyn reported a significant synergy through the whole experiment time period (Supplementary Fig. 9d).

The evaluation of the model diagnostics confirmed that the model assumptions were met for all the datasets, only showing some deviation from the normality of the random effects for the 4T1 model experiment (Supplementary Fig. 10). However, we found significant differences in the baseline tumor volumes in the 4T1 dataset (Supplementary Fig. 10d), although the absolute range of initial volumes (2.1–37.7 mm³) is within what is typically acceptable for PDX models, and may therefore exert a limited influence on treatment response interpretation. The plotting of observed versus fitted values also supported the validity of the models for fitting the longitudinal data (Supplementary Fig. 11).

Taken together, these results suggest a synergistic effect of the drug combination across the three colon cancer models. However, CombPDX was not able to provide information beyond the initial time points for the LS-1034 and SNU-81 models, due to insufficient data across all treatment groups. Both SynergyLMM and invivoSyn yielded similar results, indicating significant or nearly significant Bliss synergy for the three colon cancer models. One of the key factors affecting the magnitude of the *p* values obtained by SynergyLMM was the sample size. LS-1034 model experiment had a larger sample size (41 mice) compared to SW837 (24 mice) and SNU-81 (26 mice) models. SynergyLMM accounts for sample size in its statistical assessment of synergy, resulting in increased statistical power for experiments with larger sample sizes, such as the LS-1034 experiment. SynergyLMM also accounts for variability in the tumor measurements at each time point. For instance, in the SNU-81 experiment, SynergyLMM assigned a smaller *p* value for Bliss synergy at day 8 post-treatment initiation, where the variability in relative tumor volume among animals in each group—particularly in the drug combination group—was lower compared to later time points (Supplementary Fig. 9c). These findings highlight the sensitivity and specificity of SynergyLMM in detecting synergy and/or antagonism through rigorous statistical evaluation.

Simulations demonstrate consistency and robustness of SynergyLMM results

Reproducibility of results from preclinical cancer models remains a critical challenge, requiring improvements in consistency^{30,31}. To evaluate the sensitivity, specificity, and consistency of SynergyLMM results, we conducted a series of simulation studies using the SynergyLMM functions. First, we simulated 1000 typical drug combination experiments with four groups (control, drug A, drug B, and combination), assigning a synergistic tumor growth rate for the combination group according to the Bliss independence model and using realistic model parameters (see “Methods”). As anticipated, SynergyLMM consistently detected highly significant synergy across multiple time points in all these model sensitivity simulations (Supplementary Fig. 12a and Source Data S5). In a separate set of simulation studies, we modified the growth rates of the combination group such that the additive effects of the two drugs could not be rejected according to the Bliss model. As expected, SynergyLMM consistently reported non-significant results at all time points across all these model specificity simulations (Supplementary Fig. 12b and Source Data S5).



Biological variability, and particularly outlier observations, poses a significant challenge in *in vivo* drug combination experiments, where sample sizes are typically relatively small. Outliers can lead to both false positive and false negative findings, undermining the reliability of conclusions. To assess the robustness of SynergyLMM, we introduced an artificial outlier subject with a higher growth rate into one of the simulated synergistic experiments to investigate the robustness of the

method to maintain its sensitivity to detect synergy, despite the outlier subject. We then compared the performance of the analysis with SynergyLMM, CombPDX, and *invivoSyn* under the same conditions. Without the outlier, all the three methods detected significant Bliss synergy at multiple time points (Supplementary Fig. 12c, d and Source Data S5). However, when the outlier was introduced, CombPDX failed to detect synergy at any time point (Supplementary Fig. 12e, f and

Fig. 5 | Drug combination evaluation using CombPDX, invivoSyn, and SynergyLMM. **a** Tumor growth in Cetuximab + Palbociclib combination experiment in the colon PDX CRII97 model. Each curve corresponds to an individual mouse: vehicle $n=10$, Cetuximab $n=8$, Palbociclib $n=9$, Palbociclib + Cetuximab $n=10$. The bold lines in the plots represent the fitted regression lines for the fixed effects of each treatment estimated by the linear mixed-effects model. The x-axis of the plots represents the original time in square-root scale. **b** Comparison of Bliss synergy evaluation using invivoSyn and SynergyLMM. The horizontal dashed line indicates the combination index threshold of 1 to assess synergy and antagonism. The vertical lines indicate the 95% confidence intervals for the combination index. **c, d** Evaluation of GABA + anti-PD1 combination effect on tumor growth in 4T1 breast cancer model. Each curve corresponds to an individual mouse: control $n=10$, GABA $n=10$, anti-PD1 $n=10$, GABA + anti-PD1 $n=11$. **e, f** Evaluation of combination Rabusertib + Irinotecan in SW837 colon cancer model. Each curve corresponds to an individual mouse: control $n=6$, Rabusertib $n=6$, Irinotecan $n=8$,

Rabusertib + Irinotecan $n=4$. **c, e** display tumor growth data for the different drug combinations across the distinct cancer models. The bold lines in the right plots represent the fitted regression lines for the fixed effects of each treatment estimated by the linear mixed-effects models. The x-axis of the plot in **(c)** represents the original time in square-root scale. The synergy evaluation using CombPDX, invivoSyn, and SynergyLMM for 4T1 **(d)** and SW837 **(f)** tumor models. The horizontal dashed lines indicate the synergy score threshold of 0 to assess synergy and antagonism. The error bars indicate the 95% confidence intervals for the synergy score. The dot in the center indicates the estimated synergy score value. The dot colors in **(b)** indicate the non-adjusted p values for the one-sided tests from invivoSyn and the two-sided tests from SynergyLMM. The non-adjusted p values reported in **(d)** and **(f)** correspond to the one-sided tests from CombPDX and invivoSyn, and two-sided tests from SynergyLMM. HSA highest single agent. Source data are provided as a Source data file.

Source Data S5). In contrast, both SynergyLMM and invivoSyn continued to report significant Bliss synergy, albeit with higher p values (Supplementary Fig. 12e, f, and Source Data S5).

We also tested the robustness of SynergyLMM in identifying false positives (i.e., tested the model specificity). Using the simulated non-synergistic experiment, we introduced an outlier subject with a lower growth rate into the combination group. A robust method should not report a synergistic drug combination effect under these conditions. Without the outlier, both SynergyLMM and invivoSyn correctly reported non-significant results at all time points, while CombPDX erroneously reported significant p values at certain time points, despite confidence intervals indicating that additivity could not be rejected (Supplementary Fig. 12g, h, and Source Data S5). The analysis of the data with the outlier subject showed that CombPDX again reported non-significant results for most of the time points, yet it still assigned significant synergy at some time points (Supplementary Fig. 12i, j, and Source Data S5). The invivoSyn results were markedly affected by the outlier, and it reported significant Bliss synergy in this case. In contrast, SynergyLMM reported non-significant results across all time points, demonstrating its robustness and specificity (Supplementary Fig. 12i, j and Source Data S5).

These simulation studies demonstrate that SynergyLMM consistently delivers reproducible results, maintains robustness against outliers, and effectively minimizes the risk of false negatives and false positives, highlighting its reliability for in vivo synergy analysis.

Robustness of SynergyLMM against violations of modelling assumptions

Finally, we evaluated the robustness of SynergyLMM against model misspecification and violations of the distributional assumptions by fitting an exponential model and a Gompertz model to simulated data following a logistic growth model (Supplementary Fig. 13a, b). In this case, when the exponential model is used without any transformation of the time unit or specifying a heteroscedastic structure of the residuals, certain model assumptions are violated, such as the normality and homoscedasticity of the residuals (Supplementary Fig. 13c). The Gompertz model provided a better model describing the tumor growth in this case, offering an alternative model with satisfactory diagnostics (Supplementary Fig. 13d). Next, we simulated 1000 typical drug combination experiments with the four treatment groups (control, drug A, drug B, and combination), assigning tumor growth rates compatible with a synergistic drug combination effect. The synergy results obtained both from the exponential and the Gompertz model showed, as expected, significant Bliss synergy across multiple time points in all the simulations (Supplementary Fig. 13e, f and Source Data S6). However, the exponential model resulted in higher significance levels from day 15 to the last time point, compared to the Gompertz model, coincident with those time points in which the experimental measurements

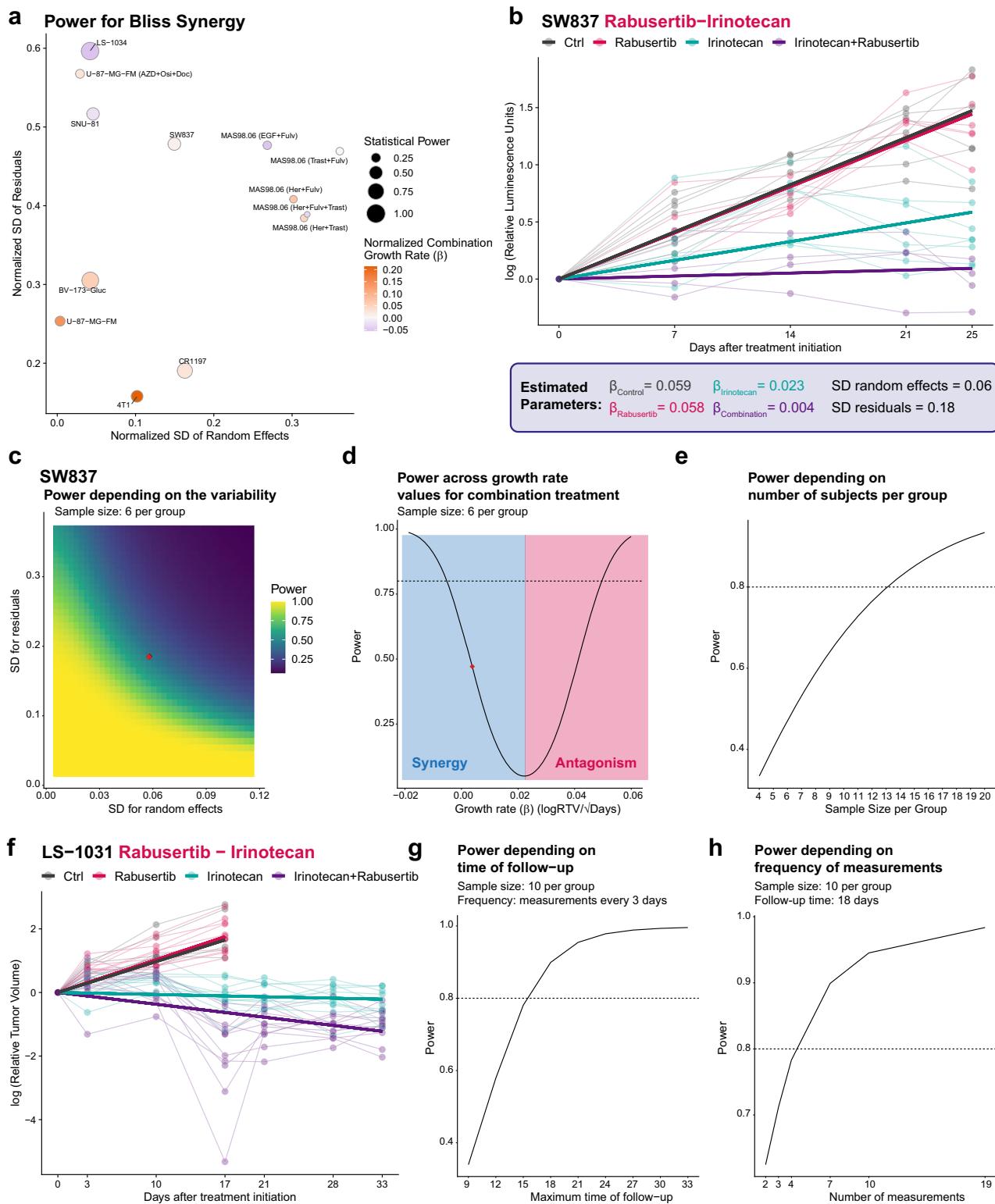
clearly deviated from the exponential model (Supplementary Fig. 13b and Source Data S6).

In a separate set of simulation studies, we modified the growth rates of the combination group such that the additive effects of the two drugs could not be rejected according to the Bliss model. Again, we evaluated the synergy results by fitting the exponential model and the Gompertz model to the logistic growth data (Supplementary Fig. 14a, b). As previously, the exponential model showed violation of the assumptions, including the normality and homoscedasticity of the residuals (Supplementary Fig. 14c), while the Gompertz model offered a better description of the tumor growth and no deviations from the model assumptions (Supplementary Fig. 14d). The synergy assessment again showed comparable results, where both models correctly did not reject the additive effect of the drug combination at multiple time points (Supplementary Fig. 14e, f and Source Data S6). The Gompertz model assigned a significant antagonism at the early time points, which was actually expected by the defined logistic growth pattern. However, this was not detected by the exponential model. Additionally, the exponential model resulted again in higher significance levels at the last time points, compared to the Gompertz model, but the p values from both models were non-significant ($p > 0.05$), in line with the simulation setup.

These simulation results indicate that the exponential model provided consistent results, even when the model was misspecified and the distributional assumptions were violated. However, the sensitivity and specificity of the detections were compromised as the discrepancy between the model assumptions and the observed data increased. This highlights the importance of using the appropriate model to obtain reliable results.

SynergyLMM implements statistical power analysis to improve experimental designs

A key determinant of statistical power is the effect size of the drug combination effect, given by the magnitude of the estimated coefficient for the growth rate after combination treatment, relative to that observed in the monotherapy and control groups. Treatments with larger effect sizes are easier to detect, even with smaller sample sizes, resulting in higher statistical power. Another critical factor is the variance components of the model, including residual variance (within-subject variance) and random effects variance (between-subject variance). High estimated values of the coefficients for these variances reduce precision, thereby diminishing statistical power. To illustrate the influence of these factors, we evaluated the retrospective (post hoc) power of the synergy analyses across the datasets described in the previous sections using the built-in functions in SynergyLMM (see “Methods” and Supplementary Methods). To simplify the power analyses, we used the synergy results calculated at the last time point, although a similar analysis could be performed for each time point with SynergyLMM. As shown in Fig. 6a, the models with high residuals



and/or random effects variance and high growth rates in the combination group (indicating lower effect size) exhibited lower statistical power. Conversely, the models with low effect size resulted in higher power only when the variance components were small. Similarly, the models with high variance components demonstrated higher power only when the effect sizes were large, with the LS-1034 model being a prominent example.

SynergyLMM also provides functionalities for a priori power analysis (see "Methods" and Supplementary Methods). This enables

the prospective evaluation of the statistical power before the study is conducted by varying the values of combination group growth rates or variance components, while keeping the other model parameters constant. As an example, we analyzed the Rabusertib and Irinotecan combination experiment in the SW837 colon cancer model from Jaaks et al.¹⁴. Using the estimated parameters from the model at day 25 (Fig. 6b), we evaluated the power for Bliss synergy across a range of residuals and random effects variances (Fig. 6c), and combination group growth rates (Fig. 6d). The results demonstrate that the power

Fig. 6 | Statistical power analysis using SynergyLMM. **a** Bubble plot indicating the statistical power for Bliss synergy, combination group growth rate, and residuals and random effects residuals across the different *in vivo* drug combination experiments. For each experiment, we used the information from the statistical analysis at the last time point. The values of the standard deviation (SD) and combination treatment group growth rate were normalized using L1 norm to make the values between different experiments on the same scale. Low values for the combination group growth rate indicate a higher drug combination effect, and vice versa. The information about all model estimates and statistical power for each experiment can be found in Source Data S7. **b** Tumor growth curves and estimated model parameters from the evaluation of drug combination Rabusertib and Irinotecan effect in the SW837 colon cancer model. The bold lines in the plots represent the fitted regression lines for the fixed effects of each treatment estimated by the linear mixed-effects model. The x-axis of the plots represents the original time in square-root scale. **c** Colormap of the statistical power for Bliss synergy across different residuals and random effects variances for an experiment with 6 animals per group and the same growth rates and duration than in the experiment shown in **(b)**. The red dot indicates the estimated power for the residuals and random effects SD indicated in **(b)**. **d** Statistical power for Bliss synergy

across different values of combination group growth rate for an experiment with 6 animals per group and the same duration and estimated parameters as in the experiment shown in **(b)**. The red dot indicates the estimated power for the combination group growth rate indicated in **(b)**. **e** Statistical power for Bliss synergy across different sample sizes per group for an experiment with the same duration and estimated parameters as the experiment shown in **(b)**. **f** Tumor growth curves for the evaluation of drug combination Rabusertib and Irinotecan effect in the LS-1034 colon cancer model. The bold lines in the plots represent the fitted regression lines for the fixed effects of each treatment estimated by the linear mixed-effects model. **g** Statistical power for Bliss synergy across different values of maximum follow-up time per group for an experiment with the same sample size and estimated parameters than in the original experiment shown in **(f)**, and measurements performed every 3 days. **h** Statistical power for Bliss synergy across different frequencies and total number of measurements for an experiment with the same sample size and estimated parameters than in the experiment shown in **(f)**, and with 18 days of follow-up. Note: 2 measurements indicate an experiment in which the tumor volume was measured only at day 0 (initial measurement) and at day 18, while 19 measurements correspond to an experiment measuring the tumor volume every single day. Source data are provided as a Source data file. Source Data S7.

increases as the residual and random effects variances decrease (Fig. 6c). Furthermore, Fig. 6d reveals a U-shaped power profile as a function of the combination group growth rate. This indicates that with lower growth rates (i.e., higher effect size), the power is increased for detecting synergy, but it decreases as growth rates approach the expected additive effect under the Bliss independence model until it reaches its minimum. As growth rates exceed the expected additive effect, the power starts to increase again, reflecting the model's ability to detect negative deviations from additivity (indicating antagonism).

An important experimental factor influencing statistical power and the study design and costs is sample size (i.e., the number of animals in each treatment group). SynergyLMM includes functions for a priori power analysis to evaluate how changes in the sample size affect power to detect synergy, if present, while keeping the other model parameters fixed. Using the parameters estimated from the data and model in Fig. 6a, we analyzed the a priori power for Bliss synergy in the SW837 Rabusertib + Irinotecan combination experiment (Fig. 6e). In the original experiment, the average sample size was six animals per group, resulting in a statistical power of 0.46 for Bliss synergy and a tendency, but not statistically significant results (Fig. 5e). This resulted in unnecessary and futile suffering of the animals to end up with non-conclusive results. Our analysis revealed that it would be necessary to increase the sample size to 14 animals per group in order to achieve a statistical power of 0.8 for detecting Bliss synergy for this combination at the evaluated time point. The SynergyLMM R package and web-tool enables the users to choose the most appropriate synergy model (Bliss or HSA) for their experiment when deciding the sample size for sufficient statistical power.

Finally, the number of measurement points per subject is another experimental factor that can influence the experimental design and statistical power. This depends on the duration of tumor growth follow-up and the frequency of the measurements during that period, both of which SynergyLMM includes functions for a priori power analysis. As an example, we analyzed the Rabusertib + Irinotecan combination experiment in the LS-1034 colon cancer model from Jaaks et al.¹⁴. In this experiment, the maximum follow-up time was 33 days (Fig. 6f), and a significant Bliss synergy was observed at this endpoint (Supplementary Fig. 9b). Using the estimated parameters of this model and a sample size of 10 animals per group (corresponding to the mean group size in the original experiment), we evaluated the a priori statistical power for Bliss synergy across different follow-up times. Our analysis suggested that 18 days of follow-up was already sufficient to achieve a statistical power of 0.8 (Fig. 6g). Furthermore, conducting seven or more evenly spaced measurements within 18 days of follow-up also yielded a statistical power exceeding 0.8, while four

measurements resulted in a statistical power close to 0.8 (Fig. 6h). The SynergyLMM R package allows the user to specify any of the model parameters for power analysis to detect synergy.

By offering flexible options for both retrospective post hoc and prospective a priori power analysis, SynergyLMM facilitates the optimization of experimental design. This includes precise estimation of the required sample size, follow-up duration, or measurement frequency needed to achieve sufficient statistical power for the *in vivo* synergy assessments.

Discussion

Combination therapies offer a promising solution to monotherapy resistance by targeting independent molecular pathways to address tumor complexity, overcome resistance mechanisms, and enhance treatment responses^{1,3}. However, despite the wide interest in drug combination discovery for cancer treatment, only a few methods are available for the assessment of drug combination effects in the pre-clinical setting, particularly for *in vivo* experiments, and there is still much room for improvement in combination analyses and development of community standards for drug combination synergy scoring, especially in the context of longitudinal and heterogeneous animal studies (Table 1). Here, we presented a comprehensive and robust method for the analysis of drug combination effects in animal studies. SynergyLMM is implemented as an R package, but is also available as an easy-to-use, interactive web-tool for those users without prior programming skills. We expect the SynergyLMM method and the guidelines from the case examples will improve the statistical rigor and experimental designs of preclinical combinatorial studies, and pave the way toward the development of community standards and best practices for drug combination synergy modeling and scoring for longitudinal and heterogeneous animal treatment studies. The improved design and analysis options are also closely related to the ongoing reproducibility crisis of preclinical cancer research^{30,31}, and complement the community guidelines for the reporting standards of the animal experiments (e.g., ARRIVE guidelines³²⁻³⁴).

Due to the lack of established and easily available statistical methods, most studies assessing drug combination effects *in vivo* use simple statistical tests, which do not properly address the particularities of animal experiments. For example, in the original analyses of Narayan et al.¹³, the authors used the Combination Index by considering only single doses of individual compounds, which can easily bias the evaluation of synergistic effects of drug combinations. Multi-dose experiments would ideally be needed, similar to the current practices in cell line experiments, but this would naturally require much more animals and resources, something that is not always

feasible, and may violate the 3R principles of animal studies (replacement, reduction, and refinement), if one is required re-testing of multiple concentration ranges.

Preclinical research carried out in experimental animal models is still an essential part of the drug development process³⁵; however, a number of reports have called into question the reproducibility, validity, and translatability of the preclinical findings due to limitations in their experimental design and statistical analysis³⁶. This does not only lead to loss of animal lives and research costs, but also delays the production of new and desperately-needed therapies for many human diseases³⁷. Preclinical synergy is cited frequently in phase I-II studies to justify the evaluation of a specific drug combination, but inappropriate methods for synergy evaluation have been used in most studies³⁸. Suggested improvements include better design of the preclinical experiments, in terms of power analyses and sufficient sample sizes, and more rigorous statistical analysis of the tumor growth data. SynergyLMM provides valuable tools for this purpose, not only by robust statistical analysis of combination effects, but also by offering post hoc power analysis to assess experiments retrospectively and a priori power analysis for designing optimal studies. The latter relies on parameter estimates from earlier similar studies or pilot experiments. Using insights from the post hoc evaluations, researchers can determine optimal sample sizes, follow-up durations, and measurement frequencies to ensure sufficient statistical power is reached in their experiments. This will lead to reduction in animal experiments by implementing improved experimental designs for in vivo experiments with minimal number of animals and measurements, yet still having enough statistical power to detect true synergistic effects.

Another common challenge in animal treatment experiments is biological variability among subjects, often leading to false-negative results when the sample sizes are too small. Ethical animal research necessitates minimizing such errors to obtain confident results and avoid unnecessary animal suffering. SynergyLMM supports these goals through time-dependent power analysis, outlier identification, and influential diagnostics, enabling researchers to optimize study designs and ensure robust and confident conclusions. For instance, SW837 experiment showed non-significant drug combination effect at day 25 (Fig. 5e). However, the post hoc power analysis at this time point indicated that the statistical power was only 0.46 (Fig. 6c, d). The a priori power results suggested that a sample size of 14 animals per group would be needed to reach 0.8 statistical power (Fig. 6e). On the other hand, the a priori power analysis for LS-1034 experiment showed that the time of follow-up could have been reduced to 18 days when performing 7 tumor measurements and still reach higher than 0.8 statistical power (Fig. 6g, h). These examples show the importance of study design in terms of optimal animal numbers and follow-up times. We note that the consistency of detecting treatment effects does not only depend on the sample size, but also on the time point for synergy evaluation, and robustness of the methods to tolerate outlier measurements, as was demonstrated in the simulation studies.

A unique feature offered by SynergyLMM is the statistical analysis of antagonism alongside synergy, while the previous methods for evaluating drug combinations in animal studies have primarily focused on synergy, often using one-tailed tests^{20,21}. However, drug combinations can result in additive, synergistic, or antagonistic effects, making two-sided statistical testing essential to avoid false-positive and false-negative conclusions. While much of the focus in combination therapy research has been on identifying synergistic interactions, assessing antagonism can be equally valuable for understanding tumor biology. Antagonistic effects may reveal previously unrecognized interactions between biological pathways in specific cancer types, potentially paving the way for therapeutic approaches. Indeed, some studies suggest that antagonistic combinations might paradoxically help prevent long-term drug resistance³⁹. Therefore, it is essential that statistical methods for evaluating drug combination effects enable the

exploration of all potential combination effects. The results from the experiments performed in MAS98.06 PDX model offer a clear example of this importance of antagonistic effects (Fig. 4), suggesting that drugs targeting HER3 might serve as additional strategies for the treatment of hormone treatment-resistant breast cancer²².

Another unique functionality offered by SynergyLMM is the possibility to analyze more than two drug combination experiments. The current version of SynergyLMM allows for the analysis of three-drug combinations only, as it is the next most frequent scenario in practice, after the two-drug combinations. However, the approach could be extended to any number of drug combinations (see a detailed description in Supplementary Methods).

SynergyLMM implements three widely used reference models for defining drug interaction effects: Bliss independence, RA, and the HSA model. The choice of reference model remains a topic of ongoing debate, dating back to the 1992 Saariselkä Agreement. During this meeting, an expert committee convened to reconcile differing synergy principles, but ultimately failed to reach consensus. Instead, they recommended that studies clearly specify the reference model used to define synergy^{11,40}. Each of these models is based on distinct assumptions, which directly influence the interpretation of synergy or antagonism. In our framework, RA is the most stringent model for detecting synergy, whereas HSA is the most permissive. To illustrate these differences, consider the combination of Imatinib and Dasatinib in the BV-173 leukemia model. Both compounds are tyrosine kinase inhibitors targeting the BCR-ABL tyrosine kinase^{41,42}. The assumption of drug independence underlying the Bliss model may not be appropriate here, given that both drugs act on the same pathway and share molecular targets. Consistent with this, the Bliss model suggests that the observed combination effect is less than what would be expected if the drugs acted independently (Fig. 2c, d), indicating antagonism under this model. In contrast, the HSA model shows that the combination effect exceeds that of the most effective single agent—Dasatinib—thus indicating synergy under this less stringent criterion. These results are not contradictory; rather, they underscore that the interpretation of drug interactions depends on the chosen reference model. SynergyLMM users are encouraged to consider the underlying assumptions of each model when interpreting results. Additional guidance on reference model selection is provided in Supplementary Information S1.

It is also important to recognize that increased therapeutic activity does not necessarily imply synergy, at least not according to strict definitions based on models such as Bliss independence or RA. A drug combination can be clinically beneficial simply by achieving greater efficacy than either of the single drugs alone, regardless of whether it meets any formal synergy criteria. This has been highlighted in the work of the Palmer and Sorger laboratories, who introduced the concept of independent drug action (IDA)⁵⁻⁷. Their framework effectively aligns with the HSA model to evaluate the therapeutic benefit of drug combinations in cancer patient populations using clinical outcomes such as progression-free survival. The implementation of multiple synergy reference models in the SynergyLMM framework enables a systematic evaluation of drug combination effects, whether these arise from Bliss-type or other forms of synergy, or from IDA as quantified by the HSA model. This provides a more comprehensive view of the in vivo drug combination results, grounded in the assumptions specific to each reference model.

SynergyLMM also has some limitations that must be considered. Firstly, our method assumes that the tumor growth follows either an exponential or Gompertz growth kinetics. Previous evidence suggests that the exponential growth is an adequate assumption for in vivo tumor growth, particularly during early phases of treatment⁴³⁻⁴⁶, while the Gompertz tumor growth is one of the most commonly used models to describe tumor growth data⁴⁷. However, these assumptions may not hold for all preclinical cancer models and therapies, and

deviations could occur. To address this, SynergyLMM includes diagnostic tools, such as plots of observed versus predicted values, to help researchers verify whether the model assumptions are adequately satisfied. This is because reliable results from statistical tests require that model assumptions—such as normality and homoscedasticity of random effects and residuals—are met. LMMs have demonstrated remarkable robustness to a degree of violations of distributional assumptions, albeit this can come at the expense of less precise estimates⁴⁸. Our simulation studies using the exponential growth to analyze a simulated logistic growth data offer a clear example of this (Supplementary Figs. 13 and 14). Therefore, users are recommended to thoroughly evaluate the estimated model, as certain experimental datasets may result in highly imprecise estimates. To support this, SynergyLMM provides diagnostic plots and tests to assess potential violations, ensuring the validity of the model and its results.

Another limitation to note involves the assumptions made by SynergyLMM when follow-up periods differ across groups. The model accounts for uncertainty due to missing data through the standard errors of the estimated coefficients—that is, fewer data points in a group (e.g., due to early termination) lead to greater uncertainty, which is reflected in the statistical inference of synergy or antagonism. However, this approach implicitly assumes that the tumor would have continued to follow the same growth pattern over the later time points for which data are missing. SynergyLMM allows the user to freely select the endpoint of the time window in which to analyze the combination effects with sufficient measurements and certainty. SynergyLMM also assumes that both drugs are administered simultaneously when assessing synergy. This assumption stems from the fact that growth rates are calculated starting from the initial time point, i.e., the time at which treatment is initiated. However, our framework allows users to modify this starting time point, enabling the analysis to begin at a later time when any drug was first administered. While SynergyLMM can still be applied in contexts involving non-simultaneous drug administration, researchers should be aware of the limitations this introduces and how it may affect interpretation. Detailed guidelines and recommendations for such scenarios are provided in Supplementary Information S1.

The synergy hypothesis testing by SynergyLMM relies on asymptotic approximations that may fail when sample sizes are limited, which are common in animal studies. In a small-sample context, Bayesian approaches can potentially offer improved performance compared to frequentist methods⁴⁹, including hierarchical shrinkage methods that reduce the error and stabilize the estimates when sample sizes per group are small⁵⁰. However, Bayesian approaches also entail some practical considerations. The primary challenge with Bayesian approaches is the specification of prior distributions, which requires substantial domain expertise and careful consideration to avoid introducing unintended bias. The choice of priors can significantly influence results, particularly in small-sample settings where non-informative prior or default hyperparameters may lead to severely biased estimates^{49,51}. Computational considerations also play a significant role in the practical implementation of Bayesian methods. While modern Bayesian approaches have made these methods more accessible, Bayesian mixed-effects models can be computationally intensive and slow to converge, especially for complex models or large datasets, potentially limiting their practical application^{52,53}. Despite these limitations, Bayesian mixed-effect models represent a valuable alternative approach that merits consideration for future development.

Finally, for highly complex models—such as those incorporating within-group correlation structures or unequal variances, or when using the Gompertz model—SynergyLMM may encounter convergence issues. These challenges arise due to the large number of parameters to be estimated, which become particularly relevant in experiments with small sample sizes or when analyzing early time

points with limited tumor measurements. SynergyLMM flags for the convergence problems, which also acts as a safeguard, preventing unreliable results when insufficient data are available to support confident conclusions. SynergyLMM offers the possibility of modifying the parameters to overcome the convergence issues, such as increasing the number of algorithm iterations or selecting a different optimization algorithm. We have included a detailed description of the suggestions to deal with convergence issues in the Supplementary Information S1.

Despite these limitations, we have demonstrated the utility of SynergyLMM across a wide range of *in vivo* experiments. The case studies included various cancer models, tumor-related measurements (e.g., tumor volume and relative luminescence units), diverse mouse backgrounds (immunodeficient and immunocompetent), and different treatment classes (chemo-, targeted-, and immunotherapy) with heterogeneous combination effects. These case studies underscore the wide applicability and robustness of SynergyLMM as a versatile resource for analyzing drug combination effects *in vivo*. By improving the rigor and reliability of preclinical synergy analyses, SynergyLMM is expected to advance the field of preclinical cancer research and enable a faster and safer transition from preclinical studies to clinical combination trials.

Methods

Modeling of tumor growth dynamics

SynergyLMM implements two of the most common mathematical models of tumor growth^{54,55}, the exponential model⁵⁶ and the Gompertz model^{45,47}.

Exponential model. According to the exponential model, the tumor growth follows an exponential kinetics, so the tumor volume at time t , $TV(t)$, is given by:

$$TV(t) = TV_0 \cdot e^{\beta t}, t \geq 0, \quad (1)$$

where TV_0 is the initial tumor volume, β is the tumor-specific growth rate, and t is time in days, weeks, etc. This exponential growth model is simple to interpret, and previous evidence has shown that it is adequate for the *in vivo* tumor growth, especially during the early phases of the treatment^{43–46}. Previous works using this model made the assumption that at the time of treatment start (baseline), the tumor volume is the same in all the treatment groups^{19,21}. However, this assumption is not usually met in the real settings, where the baseline tumor volume shows heterogeneity between animals. This problem is effectively solved if the tumor volume at time t is expressed as the relative tumor volume ($RTV(t) = \frac{TV(t)}{TV_0}$) with respect to the initial tumor volume. Then, Eq. (1) becomes:

$$RTV(t) = e^{\beta t}; t \geq 0. \quad (2)$$

After natural logarithm transformation, the tumor growth can be expressed linearly with time as:

$$\log RTV(t) = \beta \cdot t; t \geq 0. \quad (3)$$

One benefit of this transformation is that, by definition, at $t = 0$, the RTV for all subjects is $RTV(0) = 1$, and therefore:

$$\log RTV(0) = \log(1) = 0. \quad (4)$$

Linear mixed-effect model (LMM). In a typical animal study for assessing the synergy of a drug combination, the tumor volumes are measured over time in different treatment groups. A linear mixed-

effect model for the tumor volume can therefore be formulated as:

$$\log RTV_i(t) = \beta_{T_i} \cdot t + b_i \cdot t + \varepsilon_i(t), \quad (5)$$

where $\log RTV_i(t)$ denotes the value of the logarithm of the RTV measured for subject i at time t . The coefficient β_{T_i} represents the fixed effects for each treatment T_i , where $T_i \in \{\text{Control}, \text{DrugA}, \text{DrugB}, \dots, \text{Combination A+B+...}\}$, and b_i indicates the tumor-specific growth rate for each treatment group. Term $b_i \cdot t$ corresponds to the subject-specific random slope that takes into account the longitudinal nature of the data, where b_i is the random effect for subject i . $\varepsilon_i(t)$ is the residual random term for subject i at time t .

Gompertz model. According to the Gompertz model, the tumor volume at time t is given by:

$$TV(t) = TV_0 \cdot e^{\frac{r_0}{\rho}(1-e^{-\rho t})}; t \geq 0, \quad (6)$$

where TV_0 is the initial tumor volume, r_0 is the growth rate at $t=0$, and ρ is constant that accounts for the decrease in the growth rate with time. Equation (6) can also be expressed using $RTV(t)$ as:

$$RTV(t) = e^{\frac{r_0}{\rho}(1-e^{-\rho t})}; t \geq 0. \quad (7)$$

Similarly to the exponential model, for the Gompertz model at $t=0$:

$$RTV(0) = e^{\frac{r_0}{\rho}(1-e^{-\rho \cdot 0})} = e^0 = 1. \quad (8)$$

After natural logarithm transformation:

$$\log RTV_i(t) = \frac{r_0}{\rho}(1 - e^{-\rho \cdot t}); t \geq 0. \quad (9)$$

Non-linear mixed-effect model. A non-linear mixed-effect model for the Gompertz growth can be formulated as:

$$\log RTV_i(t) = \frac{r_{0,T_i} + b_{0,i}}{\rho_{T_i} + b_{1,i}}(1 - e^{-(\rho_{T_i} + b_{1,i}) \cdot t}) + \varepsilon_i(t), \quad (10)$$

where $\log RTV_i(t)$ denotes the value of the logarithm of the RTV measured for subject i at time t . $T_i \in \{\text{Control}, \text{DrugA}, \text{DrugB}, \dots, \text{Combination}\}$ is the treatment group of subject i . r_{0,T_i} and ρ_{T_i} are the fixed effects for treatment group T_i . $b_{0,i} \sim N(0, \sigma_{r_0}^2)$ is the random effect on r_0 for individual i . $b_{1,i} \sim N(0, \sigma_{\rho}^2)$ is the random effect on ρ for individual i . And $\varepsilon_i(t) \sim N(0, \sigma^2)$ the residual random term for subject i at time t .

Both in the exponential and Gompertz model, the parameter t typically corresponds to days but it can represent any time unit reflecting the timing of the measurements. Additionally, the time units can be transformed to improve the model fit and ensure that the model assumptions are satisfied, without altering the interpretation of the results. In our analysis of the MDA-MB-231 FM, BV-173-Gluc, CHL-1 FM, U87-MG FM, CRI197, and 4T1 cancer models, we applied a square root transformation to t .

The implementation of the (non-)linear mixed-effect model in SynergyLMM is done using R library nlme^{57,58}, which also allows for the specification of within-group correlations structures and/or unequal variances. These structures can capture, but are not limited to, heterogeneous variances per subject, time point, treatment, or combinations of these factors. More details of the modeling aspects are provided in the Supplementary Methods and the user guidelines in Supplementary Information S1.

Synergy calculation in SynergyLMM

For the calculation of combination synergy, we followed the statistical framework provided by Demidenko and Miller¹⁹. This framework assumes that tumor volume, or any other response variable, reflects the number of cancer cells. We can then define $S(t)$ as the proportion of surviving cells in a treatment group compared to the control group at time t . This proportion can be treated as the probability of survival of a single cell in the treatment group.

Exponential tumor growth kinetics synergy calculation

Bliss independence model. Following the Bliss independence model^{2,8} and the statistical determination proposed by Demidenko and Miller¹⁹, drugs A and B act independently if:

$$S_{\text{Combination}}(t) = S_{\text{DrugA}}(t) \cdot S_{\text{DrugB}}(t).$$

This can be extended to the case of 3-drug combination (or any higher order combinations):

$$S_{\text{Combination}}(t) = S_{\text{DrugA}}(t) \cdot S_{\text{DrugB}}(t) \cdot S_{\text{DrugC}}(t).$$

The additivity of a drug combination effect according to the Bliss independence model can be tested from the estimated growth rates (β) of the linear mixed-model described in Eq. (5) (see Supplementary Methods). In this case, the null hypothesis for Bliss independence can be expressed as:

2-drug combinations:

$$H_0 : \beta_{\text{Combination}} = \beta_{\text{DrugA}} + \beta_{\text{DrugB}} - \beta_{\text{Control}}. \quad (11)$$

3-drug combinations:

$$H_0 : \beta_{\text{Combination}} = \beta_{\text{DrugA}} + \beta_{\text{DrugB}} + \beta_{\text{DrugC}} - 2\beta_{\text{Control}}. \quad (12)$$

Highest single agent model. The HSA model assumes a positive combination effect when the drug combination exerts a higher response (i.e., fewer surviving cells) than the most effective single drug^{2,10}, which can be mathematically expressed as:

$$S_{\text{Combination}}(t) = \min(S_{\text{DrugA}}(t), S_{\text{DrugB}}(t)).$$

And for 3-drug combination:

$$S_{\text{Combination}}(t) = \min(S_{\text{DrugA}}(t), S_{\text{DrugB}}(t), S_{\text{DrugC}}(t)).$$

Again, this can be tested from the estimated growth rates of the linear mixed-model in Eq. (5) (see Supplementary Methods):

2-drug combinations:

$$H_0 : \beta_{\text{Combination}} = \min(\beta_{\text{DrugA}}, \beta_{\text{DrugB}}). \quad (13)$$

3-drug combinations:

$$H_0 : \beta_{\text{Combination}} = \min(\beta_{\text{DrugA}}, \beta_{\text{DrugB}}, \beta_{\text{DrugC}}). \quad (14)$$

Response additivity model. The RA model considers a positive drug combination effect when the combination effect is greater than the expected additive effect given by the sum of the individual effects^{2,9}. The drug combination effect can be considered as the proportion of cells killed by the treatment, M , compared to the control group. Then, the RA model can be expressed as:

$$M_{\text{Combination}}(t) = M_{\text{DrugA}}(t) + M_{\text{DrugB}}(t).$$

And similarly, for the case of 3-drug combinations:

$$M_{Combination}(t) = M_{DrugA}(t) + M_{DrugB}(t) + M_{DrugC}(t).$$

The proportion of cells killed by the treatment can be expressed in terms of survival as $M = 1 - S$. In this way, the RA model can also be formulated in terms of the proportions of surviving cells:

2-drug combinations:

$$(1 - S_{Combination}(t)) = (1 - S_{DrugA}(t)) + (1 - S_{DrugB}(t)).$$

3-drug combinations:

$$(1 - S_{Combination}(t)) = (1 - S_{DrugA}(t)) + (1 - S_{DrugB}(t)) + (1 - S_{DrugC}(t)).$$

This can be tested with the estimated growth rates of the linear mixed-model in Eq. (5):

2-drug combinations:

$$H_0 : e^{\beta_{Combination}t} = e^{\beta_{DrugA}t} + e^{\beta_{DrugB}t} - e^{\beta_{Control}t}. \quad (15)$$

3-drug combinations:

$$H_0 : e^{\beta_{Combination}t} = e^{\beta_{DrugA}t} + e^{\beta_{DrugB}t} + e^{\beta_{DrugC}t} - e^{2\beta_{Control}t}. \quad (16)$$

A more detailed explanation can be found in Supplementary Methods.

Synergy hypothesis testing. To evaluate whether the effect of the drug combination is synergistic/antagonistic one can test a linear combination of the coefficients for Bliss and HSA models. If H_0 is rejected, synergism or antagonism are defined according to following criteria:

Bliss:

$$\hat{\beta}_{Combination} < \hat{\beta}_{DrugA} + \hat{\beta}_{DrugB} - \hat{\beta}_{Control} \rightarrow \text{Synergy}$$

$$\hat{\beta}_{Combination} > \hat{\beta}_{DrugA} + \hat{\beta}_{DrugB} - \hat{\beta}_{Control} \rightarrow \text{Antagonism}$$

HSA:

$$\hat{\beta}_{Combination} < \min(\hat{\beta}_{DrugA}, \hat{\beta}_{DrugB}) \rightarrow \text{Synergy}$$

$$\hat{\beta}_{Combination} > \min(\hat{\beta}_{DrugA}, \hat{\beta}_{DrugB}) \rightarrow \text{Antagonism}$$

Similarly, for the case of the 3-drug combinations as given in Eqs. (12) and (14). SynergyLMM uses marginal effects package⁵⁹ to conduct hypothesis tests on the estimated coefficients of the model, based on the delta method.

For the RA model, the time t cannot be canceled out from the null hypothesis expression (Eqs. 15 and 16, see also Supplementary Methods), and therefore the null hypothesis cannot be tested using linear hypothesis testing directly on the estimated coefficients. For this reference model, SynergyLMM tests the null hypothesis by comparing the area under the curve (i.e., cumulative effect from the beginning of a treatment to a time point of interest) obtained from each side of the equations. A complete description of the approach is provided in the Supplementary Methods.

To achieve temporal assessment of drug combination effects using the exponential model, SynergyLMM uses a separate model for each time point being analyzed, resulting in tumor growth rates that vary with time. The estimated coefficients from each model are then used for the synergy hypothesis testing, as described above, reporting the results for each time point. When using the exponential model, SynergyLMM assumes a constant tumor growth rate over the whole modeled time window. This guarantees that testing of drug synergy/

antagonism during the given time period can be performed directly from the estimated LMM coefficients.

Gompertz tumor growth kinetics synergy calculation. SynergyLMM uses the same principles for the synergy assessment using the Gompertz growth model. The null hypotheses for the different synergy reference models using the Gompertz model and 2-drug combinations can be expressed as:

Bliss independence model.

$$H_{0(\text{Bliss})} : \frac{r_{0, Combination}}{\rho_{Combination}} (1 - e^{-\rho_{Combination}t}) = \frac{r_{0, DrugA}}{\rho_{DrugA}} (1 - e^{-\rho_{DrugA}t}) + \frac{r_{0, DrugB}}{\rho_{DrugB}} (1 - e^{-\rho_{DrugB}t}) - \frac{r_{0, Control}}{\rho_{Control}} (1 - e^{-\rho_{Control}t}). \quad (17)$$

Highest single agent model.

$$H_{0(\text{HSA})} : \frac{r_{0, Combination}}{\rho_{Combination}} (1 - e^{-\rho_{Combination}t}) = \min \left[\frac{r_{0, DrugA}}{\rho_{DrugA}} (1 - e^{-\rho_{DrugA}t}), \frac{r_{0, DrugB}}{\rho_{DrugB}} (1 - e^{-\rho_{DrugB}t}) \right]. \quad (18)$$

Response additivity model.

$$H_{0(\text{RA})} : \frac{r_{0, Combination}}{\rho_{Combination}} (1 - e^{-\rho_{Combination}t}) = e^{\frac{r_{0, DrugA}}{\rho_{DrugA}} (1 - e^{-\rho_{DrugA}t})} + e^{\frac{r_{0, DrugB}}{\rho_{DrugB}} (1 - e^{-\rho_{DrugB}t})} - e^{\frac{r_{0, Control}}{\rho_{Control}} (1 - e^{-\rho_{Control}t})}. \quad (19)$$

The time t cannot be canceled out from the null hypothesis of any of these expressions, and therefore, the null hypothesis cannot be tested using linear hypothesis testing directly on the estimated coefficients. For the Gompertz model, SynergyLMM tests the null hypothesis using a similar approach that the one used for RA model and exponential growth: comparison of the area under the curve (i.e., cumulative effect from the beginning of a treatment to a time point of interest) obtained from each side of the equations. Further details are provided in the Supplementary Methods, including the extension to testing for 3-drug combination experiments.

Calculation of synergy score and combination index

The estimated value of the CI and SS can be directly obtained from the results of the combination synergy hypothesis testing. From the hypothesis testing, an estimated difference between the experimentally observed drug combination group and the theoretical drug combination additive effect is obtained, together with the standard error and confidence interval of the difference, and the corresponding p value. If there is a synergistic effect, the difference will be negative. On the other hand, if there is an antagonistic effect, the difference will be positive. From this, a definition of combination index and SS can be provided in line with previous works^{24,60,61}.

SS is defined as the excess combination response due to drug interaction compared to the reference model^{21,24,61}. Following this definition, $SS > 0$, $SS = 0$, and $SS < 0$ indicate synergistic, additive, and antagonistic effects, respectively. SynergyLMM uses an equivalent definition of SS calculated from the result of the synergy hypothesis testing. For the Bliss, HSA, and RA models, the SS is defined as:

$$SS = \frac{-\hat{\theta}}{SD(\hat{\theta})},$$

where θ represents the contrast defined by the null hypothesis and $SD(\hat{\theta})$ is the standard error of $\hat{\theta}$.

More details are provided in the Supplementary Methods. Note that the value of the SS reflects the excess (or lack) of response in the tumor growth due to the drug combination compared to the reference synergy model.

Combination index. According to the common definition of the CI, $CI < 1$, $CI = 1$, and $CI > 1$ indicate synergistic, additive, and antagonistic effects, respectively^{19,21,60}. The CI represents the proportion of tumor cell survival at time t in the drug combination group compared to the expected tumor cell survival according to the reference synergy model. Therefore, the CI provides information about the observed drug combination effect versus the expected additive effect provided by the reference synergy model.

Following the definition by Mao and Guo²¹, an equivalent definition of the CI is the ratio between the observed and the expected proportion of surviving cells given by the synergy model. Therefore, for Bliss, HSA, and RA models, the CI can be defined as:

$$CI_{Bliss} = \frac{S_{Combination}(t)}{S_{DrugA}(t) \cdot S_{DrugB}(t)}$$

$$CI_{HSA} = \frac{S_{Combination}(t)}{\min(S_{DrugA}(t), S_{DrugB}(t))}$$

$$CI_{RA} = \frac{(1 - S_{Combination}(t))}{(1 - S_{DrugA}(t)) + (1 - S_{DrugB}(t))}$$

Similarly, for the case of the 3-drug combinations. For the Bliss and HSA models using the exponential growth, the CI (and its confidence interval) can be obtained by substituting the values with the result of the hypothesis testing (see Supplementary Methods). For the RA model using the exponential growth, and for the Gompertz model, the CI is reported based on the distribution of the ratio of the area under the curve between the experimentally observed drug combination group (numerator) and theoretical drug combination additive effect given by the synergy reference model (denominator, see Supplementary Methods). The value of t used by SynergyLMM to calculate the CI corresponds to the experimental time point at which the synergy is reported.

In-house animal experiments

Wild-type BALB/c female mice, aged 8–16 weeks, were obtained from Japan SLC. All mice were used in accordance with protocols approved by the Institutional Animal Care and Use Committee at the Center for Cancer Immunotherapy and Immunobiology, Kyoto, Japan, and the study was approved by the Research Integrity Division, Kyoto University. Since 4T1 is a breast cancer cell model, only female mice were used. The mice were housed under specific pathogen-free (SPF) conditions, with ad-libitum water and food supply, with 12 h dark/light cycles, at room temperature (23–25 °C), and 45–65% humidity. First, mice were implanted subcutaneously with placebo or GABA pellets designed to release GABA over 21 days (31.5 mg per pellet; Innovative Research). Then, 4T1 cells (10^4) were injected intradermally (i.d.) into the mammary fat pad area. For combination treatment, mice were injected i.p. with anti-PD-L1 monoclonal antibody (clone 1-111 A, produced in-house, 50 µg/mouse) once on day 8 after tumor inoculation. This resulted in the four experimental groups: control ($n = 10$), GABA ($n = 10$), anti-PD1 ($n = 10$), and GABA + anti-PD1 ($n = 11$). Tumor volumes were measured every other day using electronic calipers, and the tumor volume was calculated using the formula: tumor volume = $\frac{\pi}{6}$ (length × width × height). The maximum allowed tumor size was 20 mm in any direction or 2000 mm³ volume. As the 4T1 tumor may also form spontaneous lung metastasis, mice were also

evaluated regularly on predefined humane endpoints (loss of 20% body weight, signs of distress, difficulty breathing or lack of activity). Upon reaching maximum allowed tumor size or humane endpoints, the mice were euthanized by carbon dioxide asphyxiation and cervical dislocation.

Statistics and reproducibility

For the in-house animal experiments, no sample size calculation was performed. The number of animals was determined based on the number of animals implemented in previously published papers. No data were excluded from the analysis, with each animal experiment performed with at least ten biological replicates. Age and sex matched mice were randomly allocated into the experimental groups, and blinding to the groups was performed in the measurements of the tumor size. Differences in initial tumor volumes between groups were assessed using Kruskal–Wallis test. Pairwise Wilcoxon rank sum test with Benjamini–Hochberg multiple testing correction was used to compare differences between groups when the Kruskal–Wallis test was significant.

Published drug combination experiments

The data from Narayan et al.¹³ in vivo experiments were obtained from the supplementary data available in Tellingen and Menezes⁶². We reanalyzed the following in vivo experiments performed in mice: MDA-MB-231 FM triple-negative breast cancer cell model treated with BRAF inhibitor AZD628 ($n = 7$), the nucleoside analog Gemcitabine ($n = 7$), vehicle ($n = 8$), or combination of AZD628 + Gemcitabine ($n = 8$); BV-173-Gluc chronic myeloid leukemia cell model treated with ABL inhibitor Imatinib ($n = 6$), ABL inhibitor Dasatinib ($n = 6$), vehicle ($n = 6$), or combination ($n = 6$); CHL1 FM orthotopic melanoma cell model treated with CDK4 inhibitor GCP-082996 ($n = 6$), nucleoside analog Gemcitabine ($n = 6$), vehicle ($n = 6$), or combination ($n = 6$); U87-MG FM orthotopic glioblastoma cell model treated with PI3K/mTOR inhibitor GNE-317 ($n = 4$), microtubule inhibitor Docetaxel ($n = 4$), vehicle ($n = 4$), or combination ($n = 4$); and U87-MG FM treated with EGFR inhibitor Osimertib ($n = 7$), mTORC1/2 inhibitor AZD2014 ($n = 7$), microtubule inhibitor Docetaxel ($n = 7$), vehicle ($n = 6$), or combination ($n = 7$).

We reanalyzed the following in vivo experiments with colon cancer cell lines engrafted in mice from the supplementary data available in Jaaks et al.¹⁴: LS-1034 treated with TOP1 inhibitor Irinotecan ($n = 11$), CHEK1 inhibitor Rabusertib ($n = 12$), vehicle ($n = 6$), or combination ($n = 12$); SW837 treated with Irinotecan ($n = 8$), Rabusertib ($n = 6$), vehicle ($n = 6$), or combination ($n = 4$); and SNU-81 treated with Irinotecan ($n = 10$), Rabusertib ($n = 6$), vehicle ($n = 5$), or combination ($n = 5$).

The data for the in vivo drug combination study using colon PDX model CR1197 were obtained from Mao and Guo²¹ paper GitHub (https://github.com/maobinchen/invivoSyn_manuscript). We reanalyzed the in vivo CR1197 experiment evaluating the effect of treatment with CDK4/6 inhibitor Palbociclib ($n = 9$), EGFR inhibitor Cetuximab ($n = 8$), vehicle ($n = 10$), or combination ($n = 10$).

The data for the in vivo experiments performed in MAS98.06 luminal-like breast cancer PDX model described in Gilfillan et al.²² were kindly provided by the authors. We reanalyzed the data from the MAS98.06 PDX model treated with HER3 agonist Heregulin ($n = 5$), EGFR agonist EGF ($n = 4$), HER2 inhibitor Trastuzumab ($n = 5$), ER antagonist Fulvestrant ($n = 5$), vehicle ($n = 5$), Heregulin + Fulvestrant combination ($n = 8$), EGF + Fulvestrant combination ($n = 5$), Fulvestrant + Trastuzumab combination ($n = 5$), Heregulin + Trastuzumab combination ($n = 6$), or Heregulin + Fulvestrant + Trastuzumab combination ($n = 5$).

All the datasets with the tumor measurements can be found in Source Data S8.

SynergyLMM model definition, model selection, and synergy analysis

To identify the most appropriate model for analyzing each dataset with SynergyLMM, we fitted models based on both the exponential and Gompertz tumor growth functions. We then compared their diagnostic plots and model performance metrics (Supplementary Table 1). When both models converged successfully and exhibited satisfactory diagnostics, we selected the model with the lowest Bayesian Information Criterion value. Based on this analysis, we used the exponential growth model for all datasets, except for the CHL-1 FM and MDA-MB-231 FM cancer model, in which the Gompertz model was used.

To improve model linearity, stabilize variance and approximate normally distributed residuals, we applied a square root transformation to the time variable in the MDA-MB-231 FM, BV-173-Gluc, CHL-1 FM, U87-MG FM, CRI197, and 4T1 cancer model datasets.

To address heteroscedasticity, we chose different residual variance structures depending on the dataset. A treatment-specific variance structure was used in the BV-173-Gluc, MDA-MB-231 FM, and SW837 datasets. A sample-specific variance structure was specified for the CHL-1 FM, MAS98.06, and LS-1034 datasets, while a time point-specific variance structure was applied to the U87-MG and SNU-81 datasets.

For more complex variance patterns, we used combinations of two variance functions. In the U87-MG triple combination dataset, residual variance was modeled as varying independently across both individual samples and time points. In the CRI197 dataset, we combined subject-specific heteroscedasticity with a power-of-time variance function. Similarly, in the 4T1 dataset, we combined treatment-specific heteroscedasticity with a power-of-time variance function.

For those datasets in which the exponential growth model was used, the statistical assessment of SynergyLMM was performed using sandwich-based robust estimators with bias-reduced linearization small sample correction, except for U87-MG Docetaxel-GNE-317 experiment, in which the lack of enough data did not allow the use of robust estimators.

The synergy/antagonism p values for each day are provided without adjustments for multiple comparisons (see Supplementary Information S1). A full description of the models, including the variance structure arguments used in SynergyLMM, is provided in Supplementary Table 2.

Synergy analysis using invivoSyn

We installed the invivoSyn package²¹ from <https://github.com/maobinchen/invivoSyn>. For calculating the median AUC ratio, we set the parameters to 10000 bootstrap replicates and 95% bias-corrected and accelerated confidence interval. For calculating the AUC-based synergy, we set the parameters to 10000 bootstrap replicates, 95% bias-corrected and accelerated confidence interval, and we calculate the AUC-based synergy at each time point. By default, the synergy scores obtained with invivoSyn are provided as percentages. Therefore, we divided the results by 100 to make them comparable with the scores from the other methods. The synergy p values for each day are provided without adjustments for multiple comparisons.

Synergy analysis using CombPDX

The analysis of in vivo drug combination effects using CombPDX²⁰ was performed in the available web-tool (<https://licaih.shinyapps.io/CombPDX/>). CombPDX provides combination indexes on a continuous scale, where $CI < 0$, $CI = 0$, and $CI > 0$ indicate antagonistic, additive, and synergistic effects, respectively. Since this scale corresponds to the scale of synergy scores, to evaluate the different methods, we compared the CI results from CombPDX with the

synergy scores generated by invivoSyn and SynergyLMM. In addition to the local CI values, we also reported the global CI provided by CombPDX, which is derived by averaging the local CI values at individual time points within the study duration. The synergy p values for each day are provided without adjustments for multiple comparisons.

Simulation studies

Exponential growth simulations. We simulated tumor growth data for a two-drug combination experiment using functions implemented in SynergyLMM. The procedure involves a realistic simulation of tumor growth following exponential kinetics as specified in Eq. (1). Based on real animal experiments, we set an initial tumor volume (TV_0) of 200 mm³, the time of follow-up was 30 days, with measurements taken every 3 days, and a sample size of 5 subjects per group. The variability in tumor volume measurements was simulated by adding random noise from a normal distribution $N(1, \sigma^2)$, where σ^2 was set to 0.15. The growth rates (β coefficients) for the control, Drug A, and Drug B groups were set to 0.08, 0.07, and 0.065 ($\log \frac{RTV}{Day}$), respectively.

For the synergistic effect simulation, the growth rate for the combination group was set to 0.03 ($\log \frac{RTV}{Day}$). The outlier in this group was simulated adding a subject in which the tumor growth rate was 0.08 ($\log \frac{RTV}{Day}$).

For the additive effect simulation, the growth rate for the combination group was set to 0.055 ($\log \frac{RTV}{Day}$). Note that this is the value for the $\beta_{Combination}$ coefficient that coincides with the null hypothesis for the Bliss model defined in Eq. (6), i.e., representing additive effect, for the given β_{DrugA} , β_{DrugB} , and $\beta_{Control}$ values. The outlier in this group was simulated by adding a subject in which the tumor growth rate was -0.01 ($\log \frac{RTV}{Day}$).

Logistic growth simulations. We performed simulation experiments defining the tumor growth using a logistic growth model:

$$TV(t) = \frac{K}{1 + \frac{K-TV_0}{TV_0} \cdot e^{-rt}},$$

where $TV(t)$ is the tumor volume at time t , TV_0 is the initial tumor volume at $t=0$, r is the growth rate constant, and K is the carrying capacity (the maximum tumor volume that can be reached).

Similarly to the exponential model simulations, we set an initial tumor volume of 200 mm³, K was set to 1000 mm³, the time of follow-up was 30 days, with measurements taken every 3 days, and a sample size of 5 subjects per group. The variability in tumor volume measurements was simulated by adding random noise from a normal distribution $N(1, \sigma^2)$, where σ^2 was set to 0.1. The growth rates (r) for the control, Drug A, and Drug B groups were set to 0.25, 0.15, and 0.1, respectively. For the synergistic effect simulation, the growth rate for the combination group was set to 0.03. For the additive effect simulation, the growth rate for the combination group was increased to 0.06.

Model diagnostics

The two main distributional assumptions of linear mixed-effect models concern the normality of the random effects and the residual terms⁶³. SynergyLMM offers several diagnostic plots and tests for checking these assumptions (see Supplementary Methods). We used Q-Q plots and the Shapiro-Wilk normality test for checking the normality of the random effects and residuals. Plots of normalized residuals versus fitted values were examined to evaluate the heteroscedasticity of the residuals. Plots of observed versus fitted values were used to check the adequacy of the model for explaining the data. Cook's distances were used as metrics for the influential diagnostics, as described in Gałecki and Burzykowski⁶³. A detailed

description of their implementation in SynergyLMM is provided in the Supplementary Methods.

Power analysis

For a given tumor growth data, SynergyLMM allows for the calculation of the post hoc power of the synergy hypothesis testing for Bliss and HSA models. The power analysis is based on simulations of the dependent variable, following a similar approach as described in Gałecki and Burzykowski⁶³. A detailed description of the implementations is provided in the Supplementary Methods. For the power analyses presented in the manuscript, the post hoc power was evaluated for the Bliss model, using 1000 simulations, and a *p* value threshold of 0.05 for the synergy calculation to be considered statistically significant.

The a priori power analysis implemented in SynergyLMM are based on *F*-tests of the fixed effects of the model as previously described^{63–65}. More details about the implementation are provided in the Supplementary Methods. Power analysis in SynergyLMM is only available for models built using the exponential growth model, and therefore, we only performed the power analysis for those (i.e., all datasets except for CHL-1 and MDA-MB-231 FM).

Ethical statement

The research described complies with all relevant ethical regulations. The animal experiments were performed in accordance with protocols approved by the Institutional Animal Care and Use Committee at the Center for Cancer Immunotherapy and Immunobiology, Kyoto, Japan, and the study was approved by the Research Integrity Division, Kyoto University.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All the datasets used for producing the results in the manuscript are available as Source Data. All the data and code needed to reproduce all the results presented in the manuscript are also deposited in the manuscript GitHub https://github.com/RafRomB/SynergyLMM_manuscript⁶⁶. Source data are provided with this paper.

Code availability

SynergyLMM is provided as an open-source R⁶⁷ package with user instructions available at <https://github.com/RafRomB/SynergyLMM> and The Comprehensive R Archive Network (CRAN) at <https://cran.r-project.org/package=SynergyLMM>. Codes to perform the analyses in the manuscript and reproduce all the figures are available at https://github.com/RafRomB/SynergyLMM_manuscript⁶⁶. The SynergyLMM web-tool is publicly available at <https://synergylmm.uiocloud.no/>.

References

1. Boshuizen, J. & Peepo, D. S. Rational cancer treatment combinations: an urgent clinical need. *Mol. Cell* **78**, 1002–1018 (2020).
2. Duarte, D. & Vale, N. Evaluation of synergism in drug combinations and reference models for future orientations in oncology. *Curr. Res. Pharmacol. Drug Discov.* **3**, 100110–100110 (2022).
3. Jin, H., Wang, L. & Bernards, R. Rational combinations of targeted cancer therapies: background, advances and challenges. *Nat. Rev. Drug Discov.* **22**, 213–234 (2023).
4. Mer, A. S. et al. Integrative pharmacogenomics analysis of patient-derived xenografts. *Cancer Res.* **79**, 4539–4550 (2019).
5. Palmer, A. C. & Sorger, P. K. Combination cancer therapy can confer benefit via patient-to-patient variability without drug additivity or synergy. *Cell* **171**, 1678–1691.e13 (2017).
6. Hwangbo, H., Patterson, S. C., Dai, A., Plana, D. & Palmer, A. C. Additivity predicts the efficacy of most approved combination therapies for advanced cancer. *Nat. Cancer* **4**, 1693–1704 (2023).
7. Plana, D., Palmer, A. C. & Sorger, P. K. Independent drug action in combination therapy: implications for precision oncology. *Cancer Discov.* **12**, 606–624 (2022).
8. Bliss, C. I. THE TOXICITY OF POISONS APPLIED JOINTLY. *Ann. Appl. Biol.* **26**, 585–615 (1939).
9. Slinker, B. K. The statistics of synergism. *J. Mol. Cell. Cardiol.* **30**, 723–731 (1998).
10. Berenbaum, M. C. What is synergy? *Pharmacol. Rev.* **41**, 93–141 (1989).
11. Tang, J., Wennerberg, K. & Aittokallio, T. What is synergy? The Saariselkä agreement revisited. *Front. Pharmacol.* **6**, 181 (2015).
12. Laajala, T. D. et al. Improved statistical modeling of tumor growth and treatment effect in preclinical animal studies with highly heterogeneous responses *in vivo*. *Clin. Cancer Res.* **18**, 4385–4396 (2012).
13. Narayan, R. S. et al. A cancer drug atlas enables synergistic targeting of independent drug vulnerabilities. *Nat. Commun.* **11**, 2935 (2020).
14. Jaaks, P. et al. Effective drug combinations in breast, colon and pancreatic cancer cells. *Nature* **603**, 166–173 (2022).
15. Chou, T. C. & Talalay, P. A simple generalized equation for the analysis of multiple inhibitions of Michaelis-Menten kinetic systems. *J. Biol. Chem.* **252**, 6438–6442 (1977).
16. Wu, J., Tracey, L. & Davidoff, A. M. Assessing interactions for fixed-dose drug combinations in tumor xenograft studies. *J. Biopharm. Stat.* **22**, 535–543 (2012).
17. Wu, J. Assessing interactions for fixed-dose drug combinations in subcutaneous tumor xenograft studies. *Pharm. Stat.* **12**, 115–119 (2013).
18. Zhao, W. et al. Detecting bliss synergy in *in vivo* combination studies with a tumor kinetic model. *Pharm. Stat.* **18**, 688–699 (2019).
19. Demidenko, E. & Miller, T. W. Statistical determination of synergy based on Bliss definition of drugs independence. *PLoS ONE* **14**, e0224137 (2019).
20. Huang, L. et al. CombPDX: a unified statistical framework for evaluating drug synergism in patient-derived xenografts. *Sci. Rep.* **12**, 12984 (2022).
21. Mao, B. & Guo, S. Statistical assessment of drug synergy from *in vivo* combination studies using mouse tumor models. *Cancer Res. Commun.* **3**, 2146–2157 (2023).
22. Gilfillan, S. et al. Breast tumors escape endocrine therapy by ER-independent mechanisms triggered by the coordinated activities of HER2/HER3 and deacetylated FOXA1. Preprint at <https://doi.org/10.1101/2020.09.22.308569> (2020).
23. Bergamaschi, A. et al. Molecular profiling and characterization of luminal-like and basal-like *in vivo* breast cancer xenograft models. *Mol. Oncol.* **3**, 469–482 (2009).
24. Ianevski, A., Giri, A. K. & Aittokallio, T. SynergyFinder 3.0: an interactive analysis and consensus interpretation of multi-drug synergies across multiple samples. *Nucleic Acids Res.* **50**, W739–W743 (2022).
25. O’Neil, J. et al. An unbiased oncology compound screen to identify novel combination strategies. *Mol. Cancer Ther.* **15**, 1155–1162 (2016).
26. Holbeck, S. L. et al. The National Cancer Institute ALMANAC: a comprehensive screening resource for the detection of anticancer drug pairs with enhanced therapeutic activity. *Cancer Res.* **77**, 3564–3576 (2017).
27. Turner, T. H., Alzubi, M. A. & Harrell, J. C. Identification of synergistic drug combinations using breast cancer patient-derived xenografts. *Sci. Rep.* **10**, 1493 (2020).

28. Zhang, B. et al. B cell-derived GABA elicits IL-10+ macrophages to limit anti-tumour immunity. *Nature* **599**, 471–476 (2021).

29. Huang, D. et al. Cancer-cell-derived GABA promotes β -catenin-mediated tumour growth and immunosuppression. *Nat. Cell Biol.* **24**, 230–241 (2022).

30. Mullard, A. Preclinical cancer research suffers another reproducibility blow. *Nat. Rev. Drug Discov.* **21**, 89 (2022).

31. Begley, C. G. & Ellis, L. M. Drug development: raise standards for preclinical cancer research. *Nature* **483**, 531–533 (2012).

32. Percie, D. et al. The ARRIVE guidelines 2.0: updated guidelines for reporting animal research. *PLoS Biol.* **18**, e3000410 (2020).

33. Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M. & Altman, D. G. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol.* **8**, e1000412 (2010).

34. Percie Du Sert, N. et al. Reporting animal research: explanation and elaboration for the ARRIVE guidelines 2.0. *PLoS Biol.* **18**, e3000411 (2020).

35. Honkala, A., Malhotra, S. V., Kummar, S. & Junnila, M. R. Harnessing the predictive power of preclinical models for oncology drug development. *Nat. Rev. Drug Discov.* **21**, 99–114 (2022).

36. Couzin-Frankel, J. When mice mislead. *Science* **342**, 922–925 (2013).

37. Freedman, L. P., Cockburn, I. M. & Simcoe, T. S. The economics of reproducibility in preclinical research. *PLoS Biol.* **13**, e1002165 (2015).

38. Ocana, A., Amir, E., Yeung, C., Seruga, B. & Tannock, I. F. How valid are claims for synergy in published clinical studies? *Ann. Oncol.* **23**, 2161–2166 (2012).

39. Saputra, E. C., Huang, L., Chen, Y. & Tucker-Kellogg, L. Combination therapy and the evolution of resistance: the theoretical merits of synergism and antagonism in cancer. *Cancer Res.* **78**, 2419–2431 (2018).

40. Greco, W. et al. Consensus on concepts and terminology for combined-action assessment: the Saariselkä agreement. *Arch. Complex Env. Stud.* **4**, 65–69 (1992).

41. Iqbal, N. & Iqbal, N. Imatinib: a breakthrough of targeted therapy in cancer. *Chemother. Res. Pract.* **2014**, 1–9 (2014).

42. Tsimberidou, A.-M. Dasatinib in chronic myeloid leukemia: a review. *Ther. Clin. Risk Manag.* 281–289, <https://doi.org/10.2147/tcrm.s3425> (2009).

43. Guo, S., Jiang, X., Mao, B. & Li, Q. X. The design, analysis and application of mouse clinical trials in oncology drug development. *BMC Cancer* **19**, 718 (2019).

44. Steel, G. G. & Lamerton, L. F. The growth rate of human tumours. *Br. J. Cancer* **20**, 74–86 (1966).

45. Laird, A. K. Dynamics of tumor growth. *Br. J. Cancer* **18**, 490–502 (1964).

46. Tubiana, M. Tumor cell proliferation kinetics and tumor growth rate. *Acta Oncol.* **28**, 113–121 (1989).

47. Tjørve, K. M. C. & Tjørve, E. The use of Gompertz models in growth analyses, and new Gompertz-model approach: an addition to the Unified-Richards family. *PLoS ONE* **12**, e0178691 (2017).

48. Schielzeth, H. et al. Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods Ecol. Evol.* **11**, 1141–1152 (2020).

49. Smid, S. C., McNeish, D., Miočević, M. & Van De Schoot, R. Bayesian versus Frequentist estimation for structural equation models in small sample contexts: a systematic review. *Struct. Equ. Model. Multidiscip. J.* **27**, 131–161 (2020).

50. Cao, Z. et al. Recalibrating single-study effect sizes using hierarchical Bayesian models. *Front. Neuroimaging* **2**, 1138193 (2023).

51. Goligher, E. C. & Harhay, M. O. What is the point of Bayesian analysis? *Am. J. Respir. Crit. Care Med.* **209**, 485–487 (2024).

52. Loh, P.-R. et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).

53. Bon, J. J. et al. Being Bayesian in the 2020s: opportunities and challenges in the practice of modern applied Bayesian statistics. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **381**, 20220156 (2023).

54. Gerlee, P. The model muddle: in search of tumor growth laws. *Cancer Res.* **73**, 2407–2411 (2013).

55. Heesterman, B. et al. Mathematical models for tumor growth and the reduction of overtreatment. *J. Neurol. Surg. B Skull Base* **80**, 072–078 (2019).

56. Collins, V. P., Loeffler, R. K. & Tivey, H. Observations on growth rates of human tumors. *Am. J. Roentgenol. Radium Ther. Nucl. Med.* **76**, 988–1000 (1956).

57. Pinheiro, J., Bates, D., & R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-168. Available at <https://CRAN.R-project.org/package=nlme> (2025).

58. Pinheiro, J. C. & Bates, D. M. *Mixed-Effects Models in S and S-PLUS* (Springer, 2000).

59. Arel-Bundock, V., Greifer, N. & Heiss, A. How to interpret statistical models using marginal effects for R and Python. *J. Stat. Softw.* **111**, 1–32 (2024).

60. Chou, T.-C. The combination index (CI<1) as the definition of synergism and of synergy claims. *Synergy* **7**, 49–50 (2018).

61. Ianevski, A., He, L., Aittokallio, T. & Tang, J. SynergyFinder: a web application for analyzing drug combination dose-response matrix data. *Bioinformatics* **33**, 2413–2415 (2017).

62. Van Tellingen, O. & De Menezes, R. X. Reanalysis of in vivo drug synergy validation study rules out synergy in most cases. *Nat. Commun.* **16**, 8534 (2025).

63. Gâtecki, A. & Burzykowski, T. *Linear Mixed-Effects Models Using R* (Springer, 2013).

64. Helms, R. W. Intentionally incomplete longitudinal designs: I. Methodology and comparison of some full span designs. *Stat. Med.* **11**, 1889–1913 (1992).

65. Verbeke, G. & Molenberghs, G. *Linear Mixed Models for Longitudinal Data*. 2nd edn, Springer Series in Statistics. (Springer, 2009).

66. Romero-Becerra, R. RafRomB/SynergyLMM_manuscript: Zenodo DOI (v1.1.0). Zenodo. <https://doi.org/10.5281/zenodo.17251200> (2025).

67. R Core Team. *R: a Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2024).

Acknowledgements

The authors thank Therese Sørli (Oslo University Hospital, Norway) and Antoni Hurtado (University of Barcelona, Spain) for the access to MAS98.06 PDX treatment data. This work was supported by the Norwegian Cancer Society (grants 216104 and 273810), South-Eastern Norway Regional Health Authority (grants 2020026 and 2023105), Radium Hospital Foundation, the Finnish Cancer Foundation, and the Sigrid Jusélius Foundation to T.A.

Author contributions

R.R.-B. and T.A. conceived the study. R.R.-B. implemented the methods, R package code and web-tool, performed the data analysis, and prepared the figures for the manuscript. R.R.-B. and T.A. wrote the manuscript, with input from all authors. Z.Z. supervised and reviewed the statistical analysis and R package code. D.N. performed the web-tool cloud deployment. E.M. performed the in-house animal experiments and acquired the data. E.M., H.B., and J.H.N. supervised data analysis and reviewed the manuscript and figures. T.A. supervised the study. All the authors reviewed the manuscript and approved its final version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-65218-9>.

Correspondence and requests for materials should be addressed to Rafael Romero-Becerra or Tero Aittokallio.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025