From Fitting the Average to Fitting the Individual: A Cautionary Tale for Mathematical Modelers

Michael C. Luo[1†], Elpiniki Nikolopoulou[2†] and Jana L. Gevertz[1*]

[1] Department of Mathematics and Statistics, The College of New Jersey, Ewing, NJ, United States, [2] School of Mathematical and Statistical Sciences, Arizona State University, Tempe, AZ, United States

An outstanding challenge in the clinical care of cancer is moving from a one-size-fits-all approach that relies on population-level statistics towards personalized therapeutic design. Mathematical modeling is a powerful tool in treatment personalization, as it allows for the incorporation of patient-specific data so that treatment can be tailor-designed to the individual. Herein, we work with a mathematical model of murine cancer immunotherapy that has been previously-validated against the average of an experimental dataset. We ask the question: what happens if we try to use this same model to perform personalized fits, and therefore make individualized treatment recommendations? Typically, this would be done by choosing a single fitting methodology, and a single cost function, identifying the individualized best-fit parameters, and extrapolating from there to make personalized treatment recommendations. Our analyses show the potentially problematic nature of this approach, as predicted personalized treatment response proved to be sensitive to the fitting methodology utilized. We also demonstrate how a small amount of the right additional experimental measurements could go a long way to improve consistency in personalized fits. Finally, we show how quantifying the robustness of the average response could also help improve confidence in personalized treatment recommendations.

Keywords: cancer, mathematical modeling, personalized therapy, immunotherapy, nonlinear mixed effects modeling

1 INTRODUCTION

The conventional approach for developing a cancer treatment protocol relies on measuring average efficacy and toxicity from population-level statistics in randomized clinical trials (1–3). However, it is increasingly recognized that heterogeneity, both between patients and within a patient, is a defining feature of cancer (4, 5). This inevitably results in a portion of cancer patients being over-treated and suffering toxicity consequences from the standard-of-care dose, and another portion being under-treated and not benefiting from the expected efficacy of the treatment (6).

For these reasons, in the last decade there has been much interest in moving away from this 'one-size-fits-all' approach to cancer treatment and towards personalized therapeutic design (also called predictive or precision medicine) (1, 2, 7). Collecting patient-specific data has the potential to improve treatment response to chemotherapy (6, 8–11), radiotherapy (12–14), and targeted

molecular therapy (11, 15–17). However, it has been proposed that personalization may hold the most promise when it comes to immunotherapy (18). Immunotherapy is an umbrella term for methods that increase the potency of the immune response against cancer. Unlike other treatment modalities that directly attack the tumor, immunotherapy depends on the interplay between two complex systems (the tumor and the immune system), and therefore may exhibit more variability across individuals (18).

Mathematical modeling has become a valuable tool for understanding tumor-drug interactions. However, just as clinical care is guided by standardized recommendations, most mathematical models are validated based on population-level statistics from preclinical or clinical studies (19). To truly realize the potential of mathematical models in the clinic, these models must be individually parameterized using measurable, patient-specific data. Only then can modeling be harnessed to answer some of the most pressing questions in precision medicine, including selecting the right drug for the right patient, identifying the optimal drug combination for a patient, and prescribing a treatment schedule that maximizes efficacy while minimizing toxicity.

Efforts to personalize mathematical models have been undertaken to understand glioblastoma treatment response (20, 21), to identify optimal chemotherapeutic and granulocyte colony-stimulating factor combined schedules in metastatic breast cancer (22), to identify optimal maintenance therapy chemotherapeutic dosing for childhood acute lymphoblastic leukemia (9), and to identify optimized doses and dosing schedules of the chemotherapeutic everolimus with the targeted agent sorafenib for solid tumors (23). Interesting work has also been done in the realm of radiotherapy, where individualized head and neck cancer evolution has been modeled through a dynamic carrying capacity informed by patient response to their last radiation dose (24).

Beyond these examples, most model personalization efforts have focused on prostate cancer, as prostate-specific antigen is a clinically measurable marker of prostate cancer burden (25) that can be used in the parameterization of personalized mathematical models. The work of Hirata and colleagues has focused on the personalization of intermittent androgen suppression therapy using retrospective clinical trial data (26, 27). Other interesting work using clinical trial data has been done by Agur and colleagues, focusing on individualizing a prostate cancer vaccine using retrospective phase 2 clinical trial data (25, 28), as well as androgen deprivation therapy using data from an advanced stage prostate cancer registry (29). Especially exciting work on personalizing prostate cancer has been undertaken by Gatenby and colleagues, who used a mathematical model to discover patient-specific adaptive protocols for the administration of the chemotherapeutic agent abiraterone acetate (30). Among the 11 patients in a pilot clinical trial treated with the personalized adaptive therapy, they observed the median time to progression increased to at least 27 months as compared to 16.5 months observed with standard dosing, while also using a cumulative drug amount that was 47% less than the standard dosing (17).

Despite these examples, classically mathematical models are not personalized, but are validated against the average of experimental data. In particular, modelers choose a single fitting methodology, a single cost function to minimize, and find the best-fit parameters to the average of the data. Using the best-fit parameters and the mathematical model, treatment optimization can be performed. Recognizing the limitations of this approach in describing variable treatment response across populations, modelers have begun employing virtual population cohorts (31–33). There is much value in this population-level approach to study variability, but it is not equivalent to looking at individualized treatment response.

In this work, we explore the consequences of performing individualized fits using a minimal mathematical model previously-validated against the average of an experimental dataset. In *Materials and Methods*, we describe the preclinical data collected by Huang et al. (34) on a mouse model of melanoma treated with two forms of immunotherapy, and our previously-developed mathematical model that has been validated against population-level data from this trial (35). Individual mouse volumetric time-course data is fit to our dynamical systems model using two different approaches detailed in *Materials and Methods*: the first fits each mouse independent of the other mice in the population, whereas the second constrains the fits to each mouse using population-level statistics. In *Results*, we demonstrate that the treatment response identified for an individual mouse is *sensitive to the fitting methodology utilized*. We explore the causes of these predictive discrepancies and how robustness of the optimal-for-the-average treatment protocol influences these discrepancies. We conclude with actionable suggestions for how to increase our confidence in mathematical predictions made from personalized fits.

2 MATERIALS AND METHODS

2.1 Data Set

The data in this study considers the impact of two immunotherapeutic protocols on a murine model of melanoma (34). The first protocol uses oncolytic viruses (OVs) that are genetically engineered to lyse and kill cancer cells. In (34) the OVs are immuno-enhanced by inserting transgenes that cause the virus to release 4-1BB ligand (4-1BBL) and interleukin (IL)-12, both of which result in the stimulation of the tumor-targeting T cell population (34). The preclinical work of Huang et al. has shown that oncolytic viruses carrying 4-1BBL and IL-12 (which we will call Ad/4-1BBL/IL-12) can cause tumor debulking *via* virus-induced tumor cell lysis, and immune system stimulation from the local release of the immunostimulants (34).

The second protocol utilized by Huang et al. are dendritic cell (DC) injections. DCs are antigen-presenting cells that, when exposed to tumor antigens ex vivo and intratumorally injected, can stimulate a strong adaptive immune response against cancer cells (34). Huang et al. showed that combination of Ad/4-1BBL/IL-12 with DC injections results in a stronger antitumor response than either treatment individually (34). Volumetric trajectories

of individual mice treated with three doses of Ad/4-1BBL/IL-12 on days 0, 2 and 4, and three doses of DCs on days 1, 3, 5, along with the average trajectory, are shown in **Figure 1**.

2.2 Mathematical Model

Our model contains the following five ordinary differential equations:

$$\frac{dU}{dt} = rU - \beta \frac{UV}{N} - (\kappa_0 + c_{kill}I)\frac{UT}{N}, U(0) = U_0, \quad (1)$$

$$\frac{dI}{dt} = \beta \frac{UV}{N} - \delta_I I - (\kappa_0 + c_{kill}I)\frac{IT}{N}, I(0) = 0, \quad (2)$$

$$\frac{dV}{dt} = u_V(t) + a\delta_I I - \delta_V V, V(0) = 0 \quad (3)$$

$$\frac{dT}{dt} = c_T I + \chi_D D - \delta_T T, T(0) = 0, \quad (4)$$

$$\frac{dD}{dt} = u_D(t) - \delta_D D, D(0) = 0 \quad (5)$$

where U is the volume of uninfected tumor cells, I is the volume of OV-infected tumor cells, V is the volume of free OVs, T is the volume of tumor-targeting T cells, D is the volume of injected dendritic cells, and N is the total volume of cells (tumor cells and T cells) at the tumor site. When all parameters and time-varying terms are positive, this models captures the effects of tumor growth and response to treatment with Ad/4-1BBL/IL-12 and DCs (35). By allowing various parameters and time-varying terms to be identically zero, other treatment protocols tested in Huang et al. (34) can also be described.

This model was built in a hierarchical fashion, details of which have been described extensively elsewhere (32, 35–37).

Here, we briefly summarize the full model. Uninfected tumor cells grow exponentially at a rate r, and upon being infected by an OV convert to infected cancer cells at a density-dependent rate $\beta UV/N$. These infected cells get lysed by the virus or other mechanisms at a rate of δ_I, thus acting as a source term for the virus by releasing an average of α free virions into the tissue space. Viruses decay at a rate of δ_v.

The activation/recruitment of tumor-targeting T cells can happen in two ways: 1) stimulation of cytotoxic T cells due to 4-1BBL or IL-12 (modeled through I, at a rate of c_T, as infected cells are the ones to release 4-1BBL and IL-12), and 2) production/recruitment due to the externally-primed dendritic cells at a rate of χ_D. These tumor-targeting T cells indiscriminately kill uninfected and infected tumor cells, with the rate of killing that depends on IL-12 and 4-1BBL production (again, modeled through I in the term $(\kappa_0 + c_{\kappa ill}I)$), and they can also experience natural death at a rate of δ_T. The time-dependent terms, $u_v(t)$ and $u_D(t)$, represent the source of the drug and are determined by the delivery and dosing schedule of interest.

2.3 Fitting Methodologies

For both fitting methodologies, the full set of model parameters $\{r,\beta,\alpha,\delta_v,\kappa_0,\delta_T,\chi_D,\delta_I,c_{kill},c_T,\delta_D,U_0\}$, which includes the initial uninfected tumor volume, is fit to each individual mouse.

2.3.1 Independently Fitting Individuals

Our first attempt at individualized fitting is to find the parameter set that minimizes the square of the ℓ^2-norm between the model and the individual mouse data:

$$\zeta = \sum_{t=0}^{n}(V_{model}(t) - V_{data}(t))^2 \quad (6)$$

where $V_{model}(t) = U(t) + I(t)$ is the volumetric output predicted by our model in eqns. (1)-(5), $V_{data}(t)$ represents the volumetric
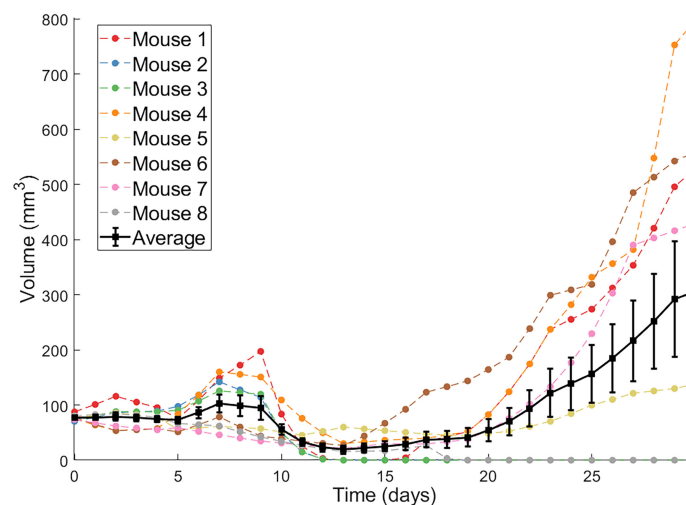

FIGURE 1 | Individual volumetric trajectories are shown for eight mice treated with Ad/4-1BBL/IL-12 (on days 0, 2, 4) + DCs (on days 1, 3, 5). The average, with standard error bars, is also shown in black (34).

data for an individual mouse, and n is the last time point at which the volume is measured in the experiments.

To independently fit an individual mouse, 12-dimensional space is first quasi-randomly sampled (with each point sampled in the range [0,1]) using high-dimensional Sobol' Low Discrepancy Sequences (LDS). LDS are designed to give rise to quasi-random numbers that sample points in space as uniformly as possible, while also (typically) having faster convergence rates than standard Monte Carlo sampling methods (38). Each randomly sampled point is then scaled to be in a biologically plausible range for the corresponding parameter value. For those parameters that were previously-fit to the average of the experimental data $(r,\beta,\chi_D,c_T,c_{kill})$, the range was set using the lower and upper-bound of the 95%-credible interval for the parameter, as determined in (35). For parameters not fit to the average in prior work, the minimum and maximum values were set to 50% and 200% of the value the parameter was fixed to for the average, respectively. See **Supplementary Table 1** for details.

After the best-fit parameter set has been selected among the 10^6 randomly sampled sets chosen by LDS, the optimal is refined using simulated annealing (39). Having observed that the landscape of the objective function near the optimal parameter set does not contain local minima, we randomly perturb the LDS-chosen parameter set, and accept any realistic parameter changes that decrease the value of the objective function - making the method equivalent to gradient descent. We consider a parameter set realistic at this stage if all parameter values are non-negative. This random perturbation process is repeated until no significant change in ζ can be achieved, which we defined as the relative change in ζ for the last five accepted parameter sets being less than 10^{-5}. We call this final parameter set the optimal parameter set. More details can be found in **Supplementary Algorithm 1**.

It is important to note that, by approaching fitting in this way, the parameters for Mouse i depend only the volumetric data for Mouse i; that is, the volumetric data for the other mice are not accounted for.

2.3.2 Fitting Individuals with Population-Level Constraints

Nonlinear mixed effects (NLME) models incorporate fixed and random effects to generate models to analyze data that are non-independent, multilevel/hierarchical, longitudinal, or correlated (40). Fixed effects refer to parameters that can generalize across an entire population. Random effects refer to parameters that differ between individuals that are randomly sampled from a population. To employ NLME for our mathematical model, for each mouse i we define the structural model $T(t_{ij},\psi_i)=U(t_j)+I(t_j)$. We assume that each parameter $\psi_{i,k}$ in the parameter set ψ_i is lognormally distributed with mean $\overline{\psi}_{i,k}$ and standard deviation $\omega_{i,k}$:

$$log(\psi_{i,k}) \sim \mathcal{N}(log(\overline{\psi}_{i,k}, \omega_{i,k}^2)) \quad (7)$$

We also assume that the error is a scalar value proportional to our structural model. Our resultant mixed effects model is:

$$y_{ij} = T(t_{i,j}, \psi_i) + bT(t_{i,j}, \psi_i)\varepsilon_{i,j}, i = 1, ..., M, j = 0, ..., n_i - 1, \quad (8)$$

where y_{ij} is the predicted tumor volume at each day j for each individual i (that is, at time t_{ij}), $M = 8$ is the number of mice, $n_i = 31$ is the number of observations per mouse, and ε_{ij} is the random noise term, which we assume to follow a standard normal distribution.

Typically, NLME models attempt to maximize the likelihood of the parameter set given the available data. There does not exist a general closed-form solution to this maximization problem (41), so numerical optimization is often needed to find a maximum likelihood estimate. In this work, we employ Monolix (42), which uses a Markov Chain Monte Carlo method to find values of the model parameters that optimize the likelihood function. To implement NLME in Monolix, we first processed and arranged our experimental data consisting of tumor volume and dosing schedule in a Monolix-specified spreadsheet. The data is then censored to avoid overfitting very small tumor volumes, as detailed in (43). To understand why this overfitting occurs in uncensored data, consider the scenario where the model predicts a volume 10^{-4} mm^3 at a time point whereas the experimental measurement is 0 mm^3. The parameter set corresponding to this prediction is assigned a lower likelihood, despite the fact that 10^{-4} is a perfectly reasonable model prediction of an experimental measurement of 0. To avoid penalizing insignificant prediction errors at very small tumor volumes, the data has been censored so that when the negative log likelihood is computed, instead of calculating the likelihood the model gives exactly the value of 0, it computes the likelihood the model predicts a value between 0 and 1. While this censoring was necessary to prevent NLME from over-fitting data points of volume zero at the expense of the fits to the nonzero data points, such censoring was not required for the independent fitting approach, as there we are just minimizing the sum of the square error. That is, in the independent fitting approach, when the model predicts a very small volume and the experimental measurement is 0, the contribution to the sum of the square error is negligible and thus censoring is not needed.

In order to solve this NLME model in Monolix, initial guesses are needed for the mean and standard deviation of our lognormally-distributed parameters. Based on previous fits to the average of the data in (37), we used the following set of initial guesses for the mean of each parameter:

$$[r, \beta, \alpha, \delta_v, \kappa_0, \delta_T, \chi_D, \delta_I, c_{kill}, c_T, \delta_D, U_0]$$
$$= [0.32, 1, 3, 2.3, 2, 0.35, 5.5, 1, 0.51, 1.2, 0.35, 55.6],$$

and after numerical exploration, we ended up choosing the initial standard deviations as:

$$[\omega_r, \omega_\beta, \omega_\alpha, \omega_{\delta_v}, \omega_{\kappa_0}, \omega_{\delta_T}, \omega_{\chi_D}, \omega_{\delta_I}, \omega_{C_{kill}}, \omega_{C_T}, \omega_{\delta_D}, \omega_{\kappa_0}]$$
$$= [0.25, 0.5, 1, 0.1, 1, 0.1, 0.25, 0.1, 0.5, 0.5, 0.1, 5]$$

2.4 Practical Identifiability *via* the Profile Likelihood Method

It is well-established that estimating a unique parameter set for a mathematical model can be challenging due to the limited

availability of often noisy experimental data (44). A non-identifiable model is one in which multiple parameter sets give "good" fits to the experimental data. Here, we will study the practical identifiability of our system in eqns. (1) - (5) using the profile likelihood approach (45, 46).

A single parameter is profiled by fixing it across a range of values, and subsequently fitting all other model parameters to the data (44). To execute the profile likelihood method, let p be the vector that contains all parameters of the model, θ be one parameter of interest contained in the vector p. The profile likelihood PL for the parameter θ is defined in (47) as:

$$PL(\theta) = \min_{p \in \{p | p_k = \theta\}} \left(\sum_{t=0}^{n} \left(\frac{V_{data}(t) - V_{model}(t;p)}{\sigma(t)} \right)^2 \right) \quad (9)$$

where $V_{model}(t;p) = U(t) + I(t)$ is the volumetric output predicted by our model for parameter set p, and $\overline{V}_{data}$ represents the average volume across all mice at that time point with corresponding standard deviation $\sigma(t)$. For normally distributed observational noise this corresponds to the maximum likelihood estimate of θ:

$$PL(\theta) = \min_{p \in \{p | p_k = \theta\}} \left(-2LL(p; V_{data}(0), ..., V_{data}(n)) \right) \quad (10)$$

where $LL(p; V_{data}(0), ..., V_{data}(n))$ is the log of the likelihood function. The likelihood function represents the likeliness of the measured data V_{data} given a model with parameters p (48). This likelihood is higher for a parameter set that is more likely given the available data, and it is smaller for parameter sets that are less likely given the data. The profile likelihood curve for any parameter of interest θ is found using the following process:

1. Determine a range for the parameter values of θ.
2. Fix $\theta = \theta^*$ at a value in the range.
3. Find the value of the non-fixed parameters that minimize the objective function in eqn. (9). The quasi-random Monte Carlo method with gradient descent was used for the fitting, as detailed previously.
4. Evaluate the objective function at those optimum values for the fixed value of θ^*.
5. Repeat the process described in steps 2-4 for a discrete set of values in the range of the parameter θ. This yields the profile likelihood function for the parameter θ.

This process results in a population-level (not individual) profile likelihood curve for each parameter. Once $PL(\theta)$ is determined, the confidence interval for θ at a level of significance α can be computed using:

$$PL(\theta) - 2LL(p_k^*) \leq \Delta_\alpha \quad (11)$$

where Δ_α denotes the α quantile of the χ^2 distribution with df degrees of freedom (which represents the number of fit model parameters when calculating $PL(\theta)$) (44). We use $\alpha = 0.95$ for a 95% confidence interval. The intersection points between the threshold $2LL(p_k^*) + \Delta_\alpha$ and $PL(\theta)$ result in the bounds of the confidence interval. A parameter is said to be practically identifiable if the shape of the profile likelihood plot is close to

quadratic on a finite confidence interval (49). Otherwise, a parameter is said to be practically unidentifiable.

3 RESULTS

3.1 Personalized Fits

The individual mouse data in response to treatment with Ad/4-1BBL/IL-12 + DCs (34) is fit using the two methodologies discussed previously: 1) quasi-Monte-Carlo method with gradient descent in which each mouse is fit independently (which we will call the "QMC" method for short), and 2) nonlinear mixed effects modeling in which population-level statistics constrain individual fits. In **Figure 2**, we can see the best-fit for each mouse using the two fitting approaches.

We do observe some shortcomings in the fits, particularly at earlier time points. These are most-pronounced in Mouse 1 and 4, where the model cannot capture the early-time decrease in tumor volume. This highlights that a model validated against the average of a dataset may not be fully sufficient at describing individual trajectories. That said, we overall find that the model is able to capture the trends in the volumetric data despite the heterogeneity across individual mice. While a more detailed model could potentially pick up some early-time trends our model did not capture, this would come at the expense of introducing more (likely non-identifiable) parameters.

For each mouse, the QMC algorithm results in a fit that more accurately captures the dynamics in the experimental data. The differences between the two fitting methodologies explain why this is occurring. NLME assumes each parameter is sampled from a lognormal distribution whose mean and variance are determined by the full population of mice. The estimated lognormal distributions for each model parameter are shown in **Figure S1**. On the other hand, the QMC algorithm fits each mouse independently, and despite the initial bounds set on the parameters when sampling parameter space, gradient descent relaxes these constraints and the end result is that non-negativity is the only constraint imposed. This allows the QMC algorithm to explore a larger region of parameter space, resulting in better fits. The potential downside, as we will show, is that the QMC algorithm can select parameter values that deviate more significantly from the average value. This variability may represent the true variability across individual mice, or may be a consequence of doing independent fits.

In **Figure 3** and **Figure S2** we show the best-fit parameter value for each mouse and fitting methodology relative to the best-fit parameter value for the average mouse. For example, the best-fit value of the tumor growth rate r to the average of the control data has been shown to be $r = 0.3198$ (35). Since Mouse 1 has a relative value of 1.0916 when fitting is done using QMC, the value of r predicted for that Mouse is 9.16% larger than the value for the average mouse, meaning QMC predicts $r = 0.3491$ for Mouse 1. On the other hand, the relative value is 0.7512 when fitting is done using NLME, meaning the predicted value is $r = 0.2402$, which is 24.88% less than the value for the average mouse.
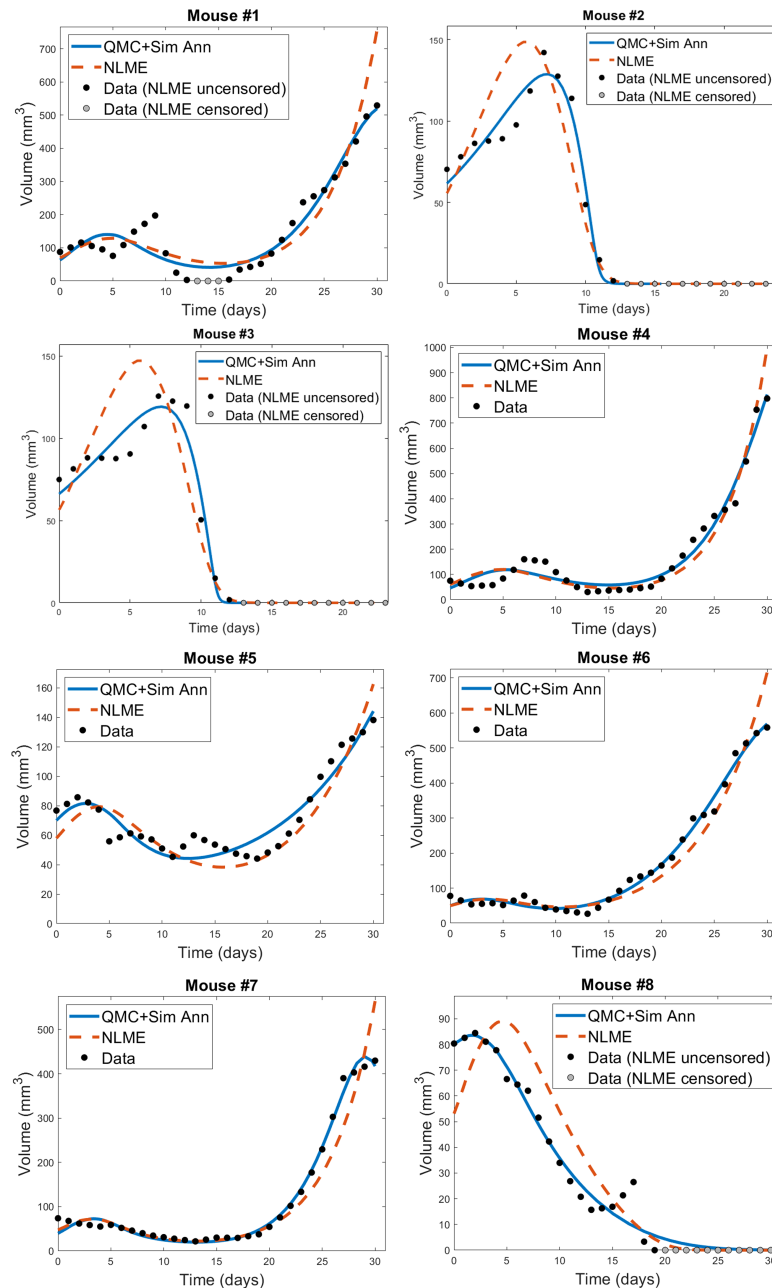
FIGURE 2 | Best-fit for each mouse treated with Ad/41BBL/IL-12 and DCs in the order VDVDVD at a dose of 2.5×10^9 OVs and 10^6 DCs (34). The QMC fits (in which each mouse is treated independently of the others) are shown in blue, and the NLME fits are shown in red. The experimental data (black if uncensored for NLME fitting, grey if censored) is also provided on each plot.

A study of the values a parameter can take on across methodologies reveals that while most values are of the same order of magnitude, differences can exist across methodologies. As expected due to the constraining lognormal distribution, NLME-associated parameters exhibit smaller variations from the best-fit parameter for the average mouse than QMC-associated parameters. Generally speaking, the variation seen

could be explained by a heterogeneous response to the treatment protocol across mice. For instance, a very small value of χ_D in Mouse 3 indicates the DCs are not successfully stimulating the production of tumor-targeting T cells. As another example, a very small value of κ_0 in Mouse 2 indicates that in the absence of immunostimulation, the T cells are unable to target and destroy cancer cells. There is one scenario that emerges in both Mouse 2
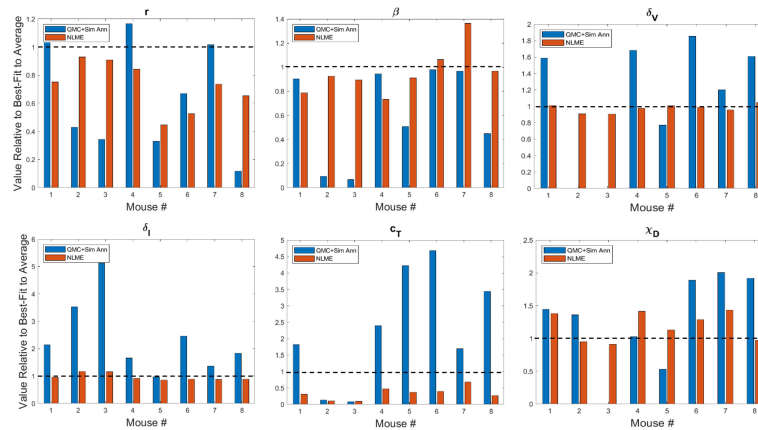
FIGURE 3 | Best-fit values of tumor growth rate parameter r, virus infectivity parameter β, viral decay rate δ_v, infected cell lysis rate δ_I, T cell stimulation term by immunostimulants c_T, and T cell stimulation term by DCs χ_D. The best-fit values are shown for each mouse and are presented relative to the best-fit value of the parameter in the average mouse (35). Therefore, a value of 1 (shown in the dashed black line) means the parameter value is equal to that in the average mouse, less than 1 is a smaller value, and greater than 1 is a larger value. Values for other model parameters are shown in **Figure S2**.

and 3, however, that cannot be explained by a heterogeneous treatment response. In particular, the QMC approach predicts that these mice have $\delta_v = 0$, indicative that the virus will not decay over the 30-day experimental time period. As this scenario is highly unlikely, we also refit these mice using the QMC approach, assuming a (somewhat arbitrary) lower bound on the viral decay rate of $\delta_v = 0.46$ day^{-1}, which assumes the decay rate can never be smaller than a quarter of the average value of δ_v = 2.3 day^{-1} (37). Refitting both mice with the QMC algorithm and this additional constraint resulted in the best-fit value of δ_v being this strict lower bound. All treatment predictions presented in this manuscript were identical whether Mouse 2 or 3 was analyzed using the parameter set with $\delta_v = 0$ or the parameter set where δ_v was a quarter the maximum value.

Looking across methodologies, parameter disparities are the most pronounced in c_T, the rate of cytotoxic T cell stimulation from 4-1BBL and IL-12. The QMC-predicted parameters cover a much larger range of values relative to the average mouse. According to the QMC fits, c_T can range anywhere from 92.15% below the value in the average mouse to 4.69 times higher than the value in the average mouse. Compare this to the NLME-predicted values of c_T, which can range from 90.29% below the value in the average mouse to 31.87% below the value for the average mouse. What is clear from looking at the best-fit parameter values across methodologies is that it is not differences in a single or small set of parameter values that explain the difference in fits. The nonlinearities in the model simply do not allow the effects of one parameter to be easily teased out from the effects of the other parameters.

3.2 Personalized Treatment Response at Experimental Dose

Here we seek to determine if the two sets of best-fit parameters for a single individual yield similar personalized predictions about tumor response to a range of treatment protocols. The

treatment protocols we consider are modeled after the experimental work in (34). Each day consists of only a single treatment, which can be either an injection of Ad/4-1BBL/IL-12 at 2.5×10^9 viruses per dose, or a dose of 10^6 DCs. Treatment will be given for six consecutive days, with three days of treatment being Ad/4-1BBL/IL-12, and three days being DCs. If only one dose can be given per day, there are exactly 20 treatment protocols to consider. The 20 protocols are shown on the vertical axis in **Figure 4**, where V represents a dose of Ad/4-1BBL/IL-12, and D represents a dose of dendritic cells.

To quantify predicted tumor response, we will simulate mouse dynamics using the determined best-fit parameters for each of the 20 6-day protocols. Unless otherwise stated, we will use the predicted tumor volume after 30 days, $V(30)$, to quantify treatment response. For each fitting methodology, mouse, and protocol we display the log ($V(30)$) in a heatmap (as in **Figure 4**). For all $V(30) \leq 1$ mm^3, we display the logarithm as 0, as showing negative values would hinder cross-methodology comparison and overemphasize insignificant differences in treatment response. We consider any tumor with $V(30) < 1$ mm^3 to be effectively treated by the associated protocol. Any nonzero values correspond to the value of log($V(30)$) when $V(30) > 1$ mm^3, and we assume these tumors have not been successfully treated. The resulting heatmap at the experimental dose of 2.5×10^9 viruses per dose, and 10^6 DCs per dose is shown in **Figure 4**.

Ideally, we would find that treatment response to a protocol for a given mouse is independent of the fitting methodology utilized, at least in the binary sense of treatment success or failure. However, that does not generally appear to be the case for our data, model and fitting methodologies, as we elaborate on here.

- **Cumulative statistics on consistencies across methodologies.** As shown in **Figure 4**, the two fitting methodologies give the
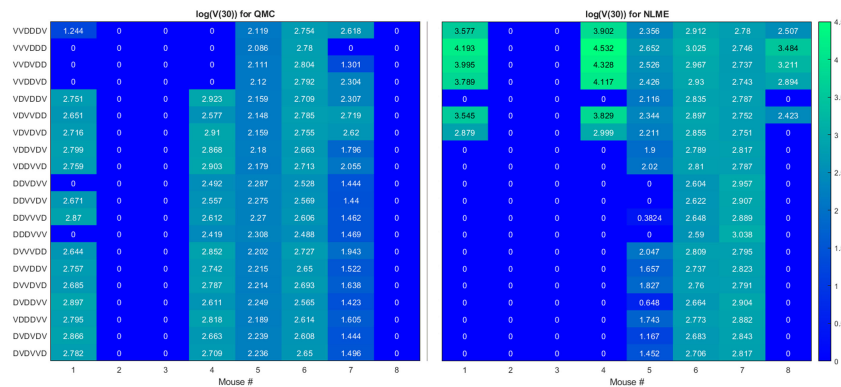
log(V(30)) for QMC

Protocol	1	2	3	4	5	6	7	8
VVDDDV	1.244	0	0	0	2.119	2.754	2.618	0
VVVDDD	0	0	0	0	2.086	2.78	0	0
VVDVDD	0	0	0	0	2.111	2.804	1.301	0
VVDVDV	0	0	0	0	2.12	2.792	2.304	0
VDVDDV	2.751	0	0	2.923	2.159	2.709	2.307	0
VDVVDD	2.651	0	0	2.577	2.148	2.785	2.719	0
VDVDVD	2.716	0	0	2.91	2.159	2.755	2.62	0
VDDVDV	2.799	0	0	2.868	2.18	2.663	1.796	0
VDDVVD	2.759	0	0	2.903	2.179	2.713	2.055	0
DDVDVV	0	0	0	2.492	2.287	2.528	1.444	0
DDVVDV	2.671	0	0	2.557	2.275	2.569	1.44	0
DDVVVD	2.87	0	0	2.612	2.27	2.606	1.462	0
DDDVVV	0	0	0	2.419	2.308	2.488	1.469	0
DVVVDD	2.644	0	0	2.852	2.202	2.727	1.943	0
DVVDDV	2.757	0	0	2.742	2.215	2.65	1.522	0
DVVDVD	2.685	0	0	2.787	2.214	2.693	1.638	0
DVDDVV	2.897	0	0	2.611	2.249	2.565	1.423	0
VDDDVV	2.795	0	0	2.818	2.189	2.614	1.605	0
DVDVDV	2.866	0	0	2.663	2.239	2.608	1.444	0
DVDVVD	2.782	0	0	2.709	2.236	2.65	1.496	0

Mouse #

log(V(30)) for NLME

Protocol	1	2	3	4	5	6	7	8
VVDDDV	3.577	0	0	3.902	2.356	2.912	2.78	2.507
VVVDDD	4.193	0	0	4.532	2.652	3.025	2.746	3.484
VVDVDD	3.995	0	0	4.328	2.526	2.967	2.737	3.211
VVDVDV	3.789	0	0	4.117	2.426	2.93	2.743	2.894
VDVDDV	0	0	0	0	2.116	2.835	2.787	0
VDVVDD	3.545	0	0	3.829	2.344	2.897	2.752	2.423
VDVDVD	2.879	0	0	2.999	2.211	2.855	2.751	0
VDDVDV	0	0	0	0	1.9	2.789	2.817	0
VDDVVD	0	0	0	0	2.02	2.81	2.787	0
DDVDVV	0	0	0	0	0	2.604	2.957	0
DDVVDV	0	0	0	0	0	2.622	2.907	0
DDVVVD	0	0	0	0	0.3824	2.648	2.889	0
DDDVVV	0	0	0	0	0	2.59	3.038	0
DVVVDD	0	0	0	0	2.047	2.809	2.795	0
DVVDDV	0	0	0	0	1.657	2.737	2.823	0
DVVDVD	0	0	0	0	1.827	2.76	2.791	0
DVDDVV	0	0	0	0	0.648	2.664	2.866	0
VDDDVV	0	0	0	0	1.743	2.773	2.882	0
DVDVDV	0	0	0	0	1.167	2.683	2.843	0
DVDVVD	0	0	0	0	1.452	2.706	2.817	0

Mouse #

FIGURE 4 | Heatmaps showing the log of the tumor volume measured at 30 days, at the OV and DC dose used in (34). If log(V (30))≤1, its value is shown as 0 on the heatmap. Left shows predictions when parameters are fit using QMC, and right shows NLME predictions.

same qualitative predictions for 73.75% (118/160) of the treatment protocols. Of the 118 agreements, 57 consistently predict treatment success whereas 61 consistently predict treatment failure. It is of note that these numbers only change slightly if we use *V(80)* as our measurement for determining treatment success or failure (81.875% agreement with 78/131 consistently predicting eradication and 53/131 consistently predicting failure - see **Figure S3**). Mouse 2, 3 and 6 have perfect agreement across fitting methodologies, and Mouse 7 has 95% agreement across methodologies. For these mice, treatment response is generally not dependent on dosing order. For instance, Mouse 2 and 3 are successfully treated by all twenty protocols considered, whereas Mouse 6 cannot be successfully treated by any protocol. In fact, *V(30)* for Mouse 6 is highly conserved across dosing order, suggesting that the ordering itself is having minimal impact on treatment response. While performing a bifurcation analysis in 11D parameter space is not feasible, what is clear is that for the mice with significant agreement across methodologies, the best-fit parameters must be sufficiently far from the bifurcation surface, as shown in the schematic diagram in **Figure 5**. As a result, predicted treatment response is not sensitive to changes in the parameter values that result from using a different fitting methodology. While not equivalent, they also do not appear to be sensitive to dosing order.

- **Cumulative statistics on inconsistencies across methodologies.** The two fitting methodologies give different qualitative predictions for 26.25% (42/160) of the treatment protocols (see **Figure 4**). Mouse 1 and 4 are largely responsible for these predictive discrepancies, with Mouse 1 having inconsistent predictions for 75% of protocols, and Mouse 4 having inconsistent predictions for 90% of protocols. Note that each methodology must agree for the protocol VDVDVD, as this was the experimental protocol that was used for parameter fitting. So, 95% is the maximum disagreement rate we can see across methodologies for a given mouse. We observe that the QMC-associated parameter set is much more likely to predict treatment failure for these mice, whereas the NLME parameter set is more
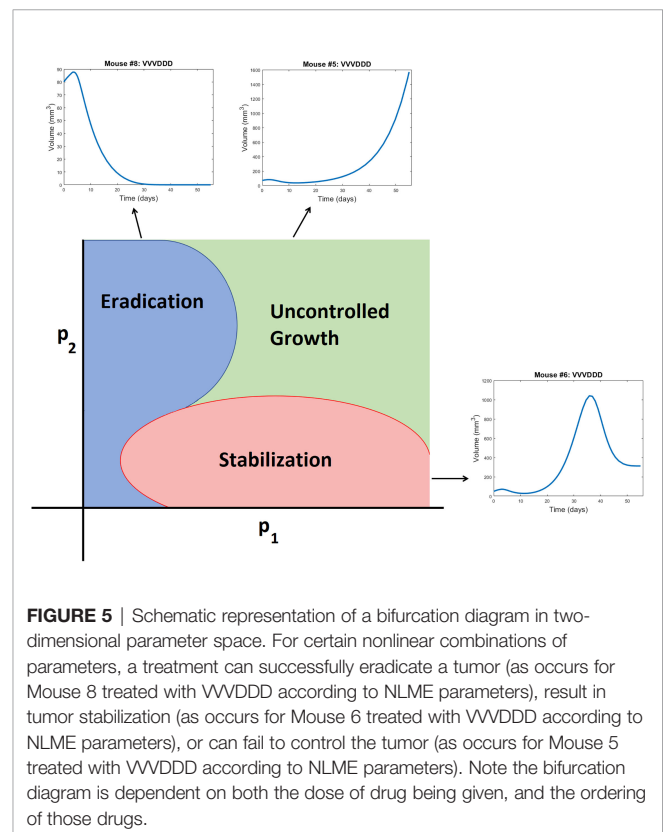

FIGURE 5 | Schematic representation of a bifurcation diagram in two-dimensional parameter space. For certain nonlinear combinations of parameters, a treatment can successfully eradicate a tumor (as occurs for Mouse 8 treated with VVVDDD according to NLME parameters), result in tumor stabilization (as occurs for Mouse 6 treated with VVVDDD according to NLME parameters), or can fail to control the tumor (as occurs for Mouse 5 treated with VVVDDD according to NLME parameters). Note the bifurcation diagram is dependent on both the dose of drug being given, and the ordering of those drugs.

likely to predict treatment success. Contrary to the mice for which there is significant cross-methodology agreement, we see a high dependency of treatment response to dosing order for Mouse 1 and 4. From the perspective of the high dimensional bifurcation diagram, these parameters must fall sufficiently close to the bifurcation surface so that parametric changes that result from using different fitting methodologies can lead to wildly different predictions about treatment response (see schematic in **Figure 5**).

In turn, this appears to make these mice significantly more sensitive to dosing order.

Though the results in this paper are presented for one best-fit parameter set per methodology, we have also explored how parametric uncertainty influences treatment predictions. In particular, for the QMC fitting method, for each mouse we identified suboptimal parameter sets by performing Sobol sampling in a 10% range about the optimal parameter set. Any sampled parameter set that gives a goodness-of-fit within 10% of the optimal is considered a suboptimal parameter set (see **Figure S4**). For all such suboptimal parameter sets, treatment response to the 20 protocols was determined. This allows us to study if binary treatment response is insensitive to the precise best-fit parameters used. In **Figure S5** we show the probability a treatment is effective for each mouse across all suboptimal parameter sets. Overwhelmingly, treatment response predicted for an individual mouse and protocol shows excellent agreement across suboptimal parameter sets. Besides treatment response to the protocols VDVVDD and VDVDVD for Mouse 7, predicted treatment response across suboptimal parameter sets agrees over a minimum of 95% of the suboptimal parameter sets. This is seen in **Figure S6** by the probabilities of an effective treatment being either >0.95 or <0.05. As small parametric perturbations that result in "good" fits to the data do not significantly influence predicted treatment response, we conclude it is reasonable to compare the prediction across methodologies using only the best-fit parameters.

3.3 Exploring Predictive Discrepancies Between Fitting Methodologies

The predictive discrepancies across fitting methodologies begs the question of whether the parameters we are fitting are actually practically identifiable given the available experimental data. To explore this question, we generated profile likelihood curves for fitting the *average* tumor growth data, following the

methodology detailed in Section 2. As a first step, we fixed the parameters whose values we could reasonably approximate from experimental data: $\delta_I = 1$, $\alpha = 3000$, $\delta_v = 2.3$, $\kappa_0 = 2$, $\delta_T = 0.35$, and $\delta_D = 0.35$ (37). This means we are using $df = 5$ in the calculation of the threshold, as the generation of each profile likelihood curve requires fitting four model parameters plus the initial condition $U(0)$.

The resulting profile likelihood curves in **Figure 6** show that, even under the assumption that six of the eleven non-initial condition parameters are known, several of the fit model parameters lack practical identifiability. The tumor growth rate r and the infectivity parameter β are both practically identifiable, ignoring slight numerical noise. The T cell activation parameters χ_D and c_T lack practical identifiability as they have profiles with a shallow and one-sided minimum (44). The profile for c_{kill} demonstrates that the model can equally well-describe the data over a large range of values for this enhanced cytotoxicity parameter. The flat likelihood profile is indicative of (local) structural unidentifiability, which also results in the parameter being practically unidentifiable (44). It is worth noting that the original work fitting to the average mouse was done in a *hierarchical* fashion (35, 37), and this circumvented the identifiability issues that emerge when doing simultaneous parameter fitting.

As we are unable to exploit the benefits of hierarchical fitting when performing personalized fits, this lack of practical identifiability poses significant issues for treatment personalization. We have already seen the consequences of this when we observed that despite both giving good fits to the data, QMC and NLME make consistent qualitative predictions in only 73.75% of the treatment protocols tested across all individuals. While the lack of practical identifiability helps explain why this can happen, it does not explain the mechanisms that drive predictive differences. To this end, we will now focus on the simulated dynamics of Mouse 4 in more detail, as this was the mouse with the most predictive discrepancies across methodologies.
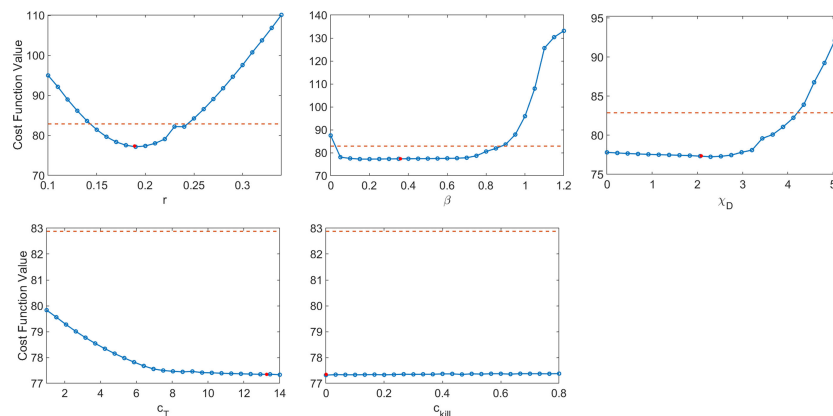

FIGURE 6 | Profile likelihood curves. Top row: tumor growth rate r, infectivity rate β, T cell activation rate by DCs χ_D. Bottom row: T cell stimulation rate by immunostimulants c_T, and rate at which immunostimulants enhance cytotoxicity of T cells c_{kill}. The threshold (red dashed line) is calculated using $df = 5$ and a 95% confidence interval.

As shown in **Figure 7**, when we simulate the model ten days beyond the data-collection window, we see that the QMC and NLME parameters fall on different sides of the bifurcation surface. In particular, in the QMC-associated simulation, at around 34 days the tumor exhibits a local maximum in volume and continues to shrink from there (**Figure 7**, left). This is in comparison to the NLME-associated simulation, where the tumor grows exponentially beyond the data-collection window. To uncover the biological mechanism driving these extreme differences, we look at the "hidden" variables in our model - that is, variables for which we have no experimental data. As shown in **Figure 7**, despite the similar fits to the volumetric data, the two parameters sets predict drastically different dynamics for the OVs and T cells. For the NLME-associated parameters, the virus and T cell population die out, eventually resulting in unbounded tumor growth. On the other hand, the virus and T cell population remain endemic throughout the simulation when using the QMC-associated parameters, driving the tumor population towards extinction.

It is common knowledge that more data improves parameter identifiability. Not all data is created equal, however. We could get a lot more time-course data on total tumor volume over the 30-day window, but that would not necessarily improve parameter identifiability. Instead, we have identified that the addition of a single data point, for the right variable, at the right time, could go a long way in reconciling predictive discrepancies across fitting methodologies. To make this concrete, suppose we had data that, for Mouse 4, no tumor-targeting T cells are detected at 30 days. If we introduced a modified cost function that weighed both the contribution of the tumor volume and this T cell measurement, the parameter set identified by QMC would no longer be optimal, as it predicts a T cell burden on the order of 10^8 (100×10^6). The optimal parameter set should be one for which $T\to 0$, and once this occurs, there is no mechanism to control the tumor in the long-term. As a result, the tumor must regrow, just as predicted for the NLME-associated parameters. While this thought experiment does not suggest all practical identifiability issues would be reconciled by having this one data point, it does indicate why the predictive discrepancies we see for Mouse 4 (and also Mouse 1) would be at least partially resolved by the addition

of a single data point on tumor-targeting T cell volume. This highlights that although one must be quite cautious in using mathematical models to make personalized predictions, models can help us determine precisely what additional data is needed so that we can have more trust in our mathematical predictions.

3.4 Personalized Treatment Response to the Optimal for Average Protocol

Ideally, when an optimal prediction is made for the average of a population, that optimal treatment protocol would also well-control the tumors of individual patients in the population. However, it is well known and supported by our earlier work with virtual populations that this is not necessarily the case. In (32) we showed that the experimental dose being considered herein is *fragile or non-robust*. We define a dosing regime as robust if virtual populations that deviate somewhat from the average population have the same qualitative response to the optimal-for-the-average protocol. Otherwise, a protocol is called fragile. The ability to classify fragility/robustness relies on the generation of a virtual population cohort that mimics a broad spectrum of individuals with different disease dynamics (31–33). By determining treatment response for each individual in the virtual cohort, we arrive at a statistic describing the likelihood the considered treatment is effective across heterogeneous individuals in the virtual population. We previously classified the optimal-for-the-average protocol of VVVDDD as fragile because this protocol eradicates the average tumor (37), yet only 30% of individuals in our virtual cohort were successfully eradicated by this treatment (32). Importantly, fragility is a probabilistic population-level descriptor, and not an individual descriptor. While it tells us that populations that deviate somewhat from the average are less likely to behave like the average, it tells us nothing about individuals, particularly if the individuals have behavior that deviates significantly from the average (which is often the case, as shown in **Figure 1**). Though, it seems natural to hypothesize that in such a fragile region we may have to be more careful about applying a prediction for the average of a population to any one individual in that population.

We will explore that hypothesis here by looking at statistics on how individual mice respond to VVVDDD, the predicted optimal
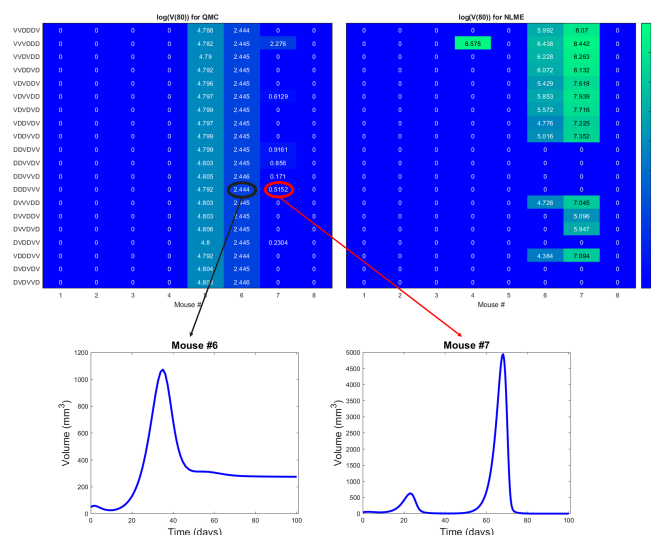

FIGURE 7 | Left: QMC and NLME-associated fits to Mouse 4 treated with VDVDVD, with model predictions extended 10 days beyond the data-collection window. Center and right: Predicted virus and T cell counts associated with each fitting methodology, respectively.
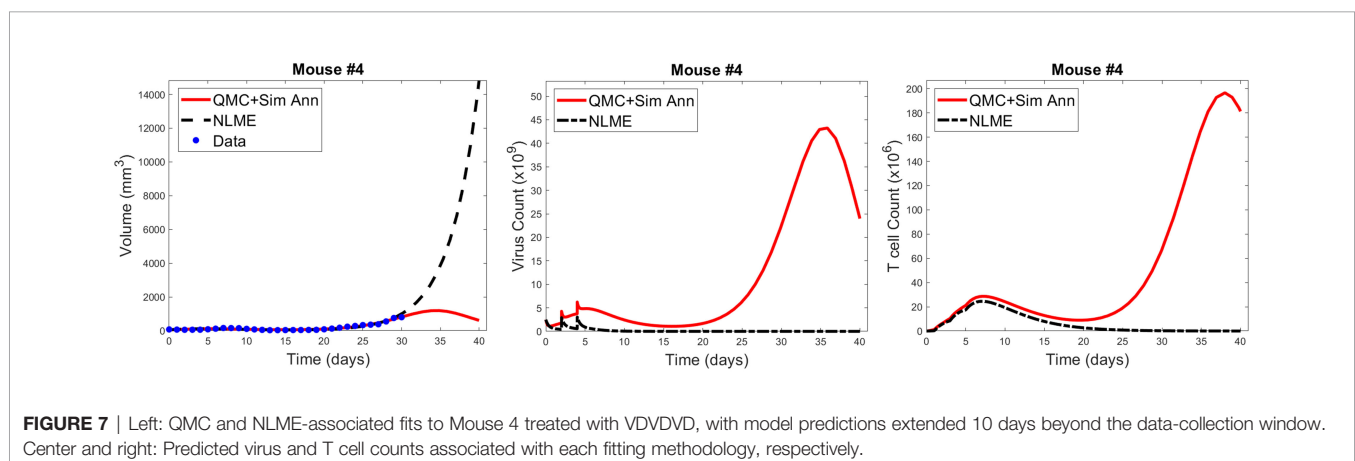
FIGURE 8 | Heatmaps showing the log of the tumor volume measured at 80 days, at the high DC (50% greater than experimental dose), low OV (50% lower than experimental dose) region of dosing space. Left shows predictions if parameters are fit using QMC, and right shows NLME predictions. Inserts show time course of predicted treatment response for Mouse 6 and 7 to the optimal-for-the-average protocol of DDDVVV.

treatment protocol for the average mouse. While this protocol was effectively able to eradicate the average tumor in the population, its success across individual mice varies significantly across fitting methodologies. For the QMC-associated predictions, this protocol eradicates tumors in 75% of the individual mice (second row of the heatmaps in **Figure 4**, left). Compare this to the NLME-associated predictions, in which this protocol eradicates tumors in only 25% of the individual mice (second row of the heatmaps in **Figure 4**, right). As shown in **Figure S3**, this prediction is unchanged if we determine treatment success or failure at day 80 instead of day 30.

We can also compare response to the optimal-for-the-average protocol across methodologies. We see a qualitative agreement across methodologies (eradication or treatment failure) in only 50% of the mice (Mouse 2, 3, 5, 6). Mouse 7 is particularly interesting, as there was 95% agreement across methodologies when using $V(30)$ to measure treatment success or failure, and the optimal for the average of VVVDDD is the only protocol for which treatment response differed (with QMC predicting tumor eradication, and NLME predicting treatment failure). As a further sign of caution, notice how for Mouse 1 and 4 (the cases with significant predictive discrepancies across methodologies), and Mouse 8 (intermediate case with 25% predictive discrepancies), VVVDDD eradicates the tumor with the QMC-associated parameters yet is the worst protocol that could be given (largest $\log(V(30))$) for the NLME-associated parameters. This is particularly unsettling as it means the population-level optimal treatment recommendation could be the worst protocol recommendation for some individuals. This confirms our hypothesis that a population-level prediction should be applied to individuals very cautiously when in a fragile region of dosing space.

This raises the question: what if we were assessing individualized response to a protocol in a robust region of dosing space, wherein

treatment response across individuals in a virtual population is statistically similar to the treatment response in the population average? In (32), we previously classified the optimal-for-the-average protocol of DDDVVV as robust in the high DC (50% greater than experimental dose), low OV (50% lower than experimental dose) region of dosing space. It was classified as robust because this protocol eradicates the average tumor, and it also eradicates 84% of the individuals in our virtual cohort (32). This probabilistic population-level assessment of robustness naturally leads to the hypothesis that in a robust region of dosing space, we may have more success with the optimal-for-the-average treatment in individual mice. We will explore that hypothesis here.

The robust population-level optimal of DDDVVV yields a successful treatment response in all eight mice for the NLME-associated parameters. This holds whether we use $V(30)$, our original measure for establishing treatment success (as shown in **Figure S6**), or if we use $V(80)$ as shown in **Figure 8**. This is consistent with the robust nature of this region of dosing space, as the NLME-associated parameters are less likely to wildly deviate from the average mouse due to population-level distributions constraining the value of these parameters. In comparison, the QMC-associated predictions show that only 62.5% of the individual mice are successfully treated by the optimal for the average in an 80-day window (**Figure 8**, top left). That said, if we look at the data more closely, we can see that Mouse 7 has essentially been eradicated even though 80 days was not quite long enough to drive $V(80) < 1$ mm^3, our threshold for eradication. **Figure 8** also shows that the tumor volume for Mouse 6 has stabilized. Thus, we see that the QMC-associated predictions actually agree with the optimal-for-the-average response in 75% of cases (or, 87.5% if you consider the stabilization of Mouse 6 to be a "success" rather than a "failure").

In closing, we have confirmation of our hypothesis that there is a significant benefit to working with a robust optimal-for-the-average protocol, even in the absence of all model parameters being practically identifiable. In the presence of robustness, we predict that one could generally apply the optimal-for-the-average protocol and expect a qualitatively similar response in most individuals. While this does not mean each individual is treated with their personalized optimal protocol, this has important consequences for determining when a population-level prediction will be effective in an individual.

4 DISCUSSION

In this work, we demonstrated that computational challenges can arise when using individualized model fits to make treatment recommendations. In particular, we showed that treatment response can be sensitive to the fitting methodology utilized when lacking sufficient patient-specific data. We found that for our model and preclinical dataset, predictive discrepancies can be at least somewhat explained by the lack of practical identifiability of model parameters. This can result in the dangerous scenario where an effective treatment recommendation according to one fitting methodology is predicted to be the worst treatment option according to a different fitting methodology. This raises concerns regarding the utility of mathematical models in personalized oncology when individual data is limited.

While it is well-established that more data improves parameter identifiability, here we highlight how we can identify precisely what data would improve the reliability of model predictions. In particular, we see how having a single additional measurement on the viral load or T cell count at the end of the data collection window would go a long way to reducing the predictive discrepancies across fitting methodologies (**Figure 7**). While the full benefits of this observation are not realized in a retrospective study, they could be realized in a scenario where data collection and modeling are occurring simultaneously. In this scenario, an experimentalist could collect data on a small number of individuals (like the eight mice shown in **Figure 1**). A mathematical model validated against this data can be used to identify any predictive challenges that emerge within this dataset, and what data would be needed to overcome these predictive challenges. This would inform the experimentalist of what data to collect in the next cohort of individuals in order to have more confidence in personalized treatment predictions.

When additional data is not available, an alternative option to personalization is simply treating with the population-level optimal. Here we showed the dangers of applying the optimal-for-the-average for a fragile protocol, and we demonstrated that such a one-size-fits all approach is much safer to employ for a robust optimal protocol. Therefore, even when data is lacking to make personalized predictions, establishing the robustness of treatment response can be a powerful tool in predictive oncology.

It is of note that this study uses just one mathematical model, with one set of assumptions, to reach our cautionary conclusion regarding the fitting methodology utilized and the resulting biological predictions. And this model is quite a simple one,

ignoring many aspects of the immune system, and spatial aspects of immune infiltration (as done in (50), among many other references). The model used herein was chosen because it has been previously validated against the average of the available experimental data. A more complex model would be problematic here, as there is simply not the associated experimental data to validate such a model. While this study certainly does not guarantee that similar issues will arise when working with other models and datasets, it highlights the need for caution when using personalized fits to draw meaningful biological conclusions.

As we enter the era of healthcare where personalized medicine becomes a more common approach to treating cancer patients, harnessing the power of mathematical models will only become more essential. Understanding the identifiability of model parameters, what data is needed to achieve identifiability and/or predictive confidence, and whether treatment response is robust or fragile are all important considerations that can greatly improve the reliability of personalized predictions made from mathematical models.

DATA AVAILABILITY STATEMENT

Data and code availability requests can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

All authors substantially contributed to the analysis of the data in this work, and the drafting of the manuscript. Each author also provides approval for publication of the content.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.793908/full#supplementary-material

REFERENCES

1. Deisboeck TS. Personalizing Medicine: A Systems Biology Perspective. *Molec Sys Biol* (2009) 5:249. doi: 10.1038/msb.2009.8

2. Agur Z, Elishmereni M, Kheifetz Y. Personalizing Oncology Treatments by Predicting Drug Efficacy, Side-Effects, and Improved Therapy: Mathematics, Statistics, and Their Integration. *WIREs Syst Biol Med* (2014) 6:239–53. doi: 10.1002/wsbm.1263

3. Barbolosi D, Ciccolini J, Lacarelle B, Barlési F, André N. Computational Oncology - Mathematical Modelling of Drug Regimens for Precision Medicine. *Nat Rev Clin Oncol* (2016) 13:242–54. doi: 10.1038/nrclinonc.2015.204

4. Malaney P, Nicosia S, Davé V. One Mouse, One Patient Paradigm: New Avatars of Personalized Cancer Therapy. *Cancer Lett* (2014) 344:1–12. doi: 10.1016/j.canlet.2013.10.010

5. Bryne A, Alferez D. Amant Fea. Interrogating Open Issues in Cancer Precision Medicine With Patient-Derived Xenografts. *Nat Rev Cancer* (2017) 17:254–68. doi: 10.1038/nrc.2016.140

6. Engels F, Loos W, van der Bol J, de Bruijn P, Mathijssen R, Verweij J, et al. Therapeutic Drug Monitoring for the Individualization of Docetaxel Dosing: A Randomized Pharmacokinetic Study. *Clin Cancer Res* (2011) 17:353–62. doi: 10.1158/1078-0432.CCR-10-1636

7. Lorenzo G, Scott MA, Tew K, Hughes TJR, Zhange YJ, Liu L, et al. Tissue-Scale, Personalized Modeling and Simulation of Prostate Cancer Growth. *Proc Natl Acad Sci* (2016) 113:E7663–71. doi: 10.1073/pnas.1615791113

8. Walko C, McLeod H. Pharmacogenomic Progress in Individualized Dosing of Key Drugs for Cancer Patients. *Nat Clin Pract Oncol* (2009) 6:153–62. doi: 10.1038/ncponc1303

9. Noble S, Sherer E, Hammemann R, Ramkrishna D, Vik T, Rundell A. Using Adaptive Model Predictive Control to Customize Maintenance Therapy Chemotherapeutic Dosing for Childhood Acute Lymphoblastic Leukemia. *J Theor Biol* (2010) 264:990–1002. doi: 10.1016/j.jtbi.2010.01.031

10. Patel J. Personalizing Chemotherapy Dosing Using Pharmacological Methods. *Cancer Chemother Pharmacol* (2015) 76:879–96. doi: 10.1007/s00280-015-2849-x

11. Chantal P, Hopkins B, Prandi D, Shaw R, Fedrizzi T, Sboner A, et al. Personalized *In Vitro* and *In Vivo* Cancer Models to Guide Precision Medicine. *Cancer Discov* (2017) 7:462–77. doi: 10.1158/2159-8290.CD-16-1154

12. Ree A, Redalen K. Personalized Radiotherapy: Concepts, Biomarkers Andtrial Design. *Br J Radiol* (2015) 88:20150009. doi: 10.1259/bjr.20150009

13. Caudell J, Torres-Roca J, Gillies R, Enderling H, Kim S, Rishi A, et al. The Future of Personalised Radiotherapy for Head and Neck Cancer. *Lancet Oncol* (2017) 18:e266–73. doi: 10.1016/S1470-2045(17)30252-8

14. Sunassee E, Tan D, Ji N, Brady R, Moros E, Caudell J, et al. Proliferation Saturation Index in an Adaptive Bayesian Approach to Predict Patient-Specific Radiotherapy Responses. *Int J Radiat Biol* (2019) 95:1421–6. doi: 10.1080/09553002.2019.1589013

15. Kim E, Herbst R, Wistuba I, Lee J, Blumenschein G, Tsao A, et al. The BATTLE Trial: Personalizing Therapy for Lung Cancer. *Cancer Discov* (2011) 1:44–53. doi: 10.1158/2159-8274.CD-10-0010

16. Tsimberidou AM, Iskander N, Hong D, Wheler J, Falchook G, Fu S, et al. Personalized Medicine in a Phase I Clinical Trials Program: The MD Anderson Cancer Center Initiative. *Clin Cancer Res* (2012) 18:6373–83. doi: 10.1158/1078-0432.CCR-12-1627

17. Zhang J, Cunningham J, Brown J, Gatenby R. Integrating Evolutionary Dynamics Into Treatment of Metastatic Castrate-Resistant Prostate Cancer. *Nat Commun* (2017) 8:1816. doi: 10.1038/s41467-017-01968-5

18. Agur Z, Halevi-Tobias K, Kogan Y, Shlagman O. Employing Dynamical Computational Models for Personalizing Cancer Immunotherapy. *Expert Opin Biol Ther* (2016) 16:1373–85. doi: 10.1080/14712598.2016.1223622

19. Agur Z, Vuk-Pavlović S. Mathematical Modeling in Immunotherapy of Cancer: Personalizing Clinical Trials. *Molec Ther* (2012) 20:1–2. doi: 10.1038/mt.2011.272

20. Kogan Y, Foryś U, Shukron O, Kronik N, Agur Z. Cellular Immunotherapy for High Grade Gliomas: Mathematical Analysis Deriving Efficacious Infusion Rates Based on Patient Requirements. *SIAM J Appl Math* (2010) 70:1953–76. doi: 10.1137/08073740X

21. Hawkins-Daruud A, Johnston S, Swanson K. Quantifying Uncertainty and Robustness in a Biomathematical Model–Based Patient-Specific Response Metric for Glioblastoma. *JCO Clin Cancer Inform* (2019) 3:1–8. doi: 10.1200/CCI.18.00066

22. Vainas O, Ariad S, Amir O, Mermershtain W, Vainstein V, Kleiman M, et al. Personalising Docetaxel and G-Csf Schedules in Cancer Patients by a Clinically Validated Computational Model. *Br J Cancer* (2012) 107:814–22. doi: 10.1038/bjc.2012.316

23. El-Madani M, Hénin E, Lefort T, Tod M, Freyer G, Cassier P, et al. Multiparameter Phase I Trials: A Tool for Model-Based Development of Targeted Agent Combinations–Example of Evesor Trial. *Future Oncol* (2015) 11:1511–8. doi: 10.2217/fon.15.49

24. Zahid MU, Mohsin N, Mohamed AS, Caudell JJ, Harrison LB, Fuller CD, et al. Forecasting Individual Patient Response to Radiotherapy in Head and Neck Cancer With a Dynamic Carrying Capacity Model. *Int J Radiat OncologyBiologyPhysics* (2021) 111(3):693–704 doi: 10.1016/j.ijrobp.2021.05.132

25. Kronik N, Kogan Y, Elishmereni M, Halevi-Tobias K, Vuk-Pavlović S, Agur Z. Predicting Outcomes of Prostate Cancer Immunotherapy by Personalized Mathematical Models. *PloS One* (2010) 5:e15482. doi: 10.1371/journal.pone.0015482

26. Hirata Y, Morino K, Akakura K, Higano CS, Bruchovsky N, Gambol T, et al. Intermittent Androgen Suppression: Estimating Parameters for Individual Patients Based on Initial Psa Data in Response to Androgen Deprivation Therapy. *PloS One* (2015) 10:e0130372. doi: 10.1371/journal.pone.0130372

27. Hirata Y, Morino K, Akakura K, Higano CS, Aihara K. Personalizing Androgen Suppression for Prostate Cancer Using Mathematical Modeling. *Sci Rep* (2018) 8:2563. doi: 10.1038/s41598-018-20788-1

28. Kogan Y, Halevi–Tobias K, Elishmereni M, Vuk-Pavlović S, Agur Z. Reconsidering the Paradigm of Cancer Immunotherapy by Computationally Aided Real-Time Personalization. *Cancer Res* (2012) 72:2218–27. doi: 10.1158/0008-5472.CAN-11-4166

29. Elishmereni M, Kheifetz Y, Shukrun I, Bevan GH, Nandy D, McKenzie KM, et al. Predicting Time to Castration Resistance in Hormone Sensitive Prostate Cancer by a Personalization Algorithm Based on a Mechanistic Model Integrating Patient Data. *Prostate* (2016) 76:48–57. doi: 10.1002/pros.23099

30. Gatenby R, Silva A, Gillies R, Frieden B. Adaptive Therapy. *Cancer Res* (2009) 69:4894–903. doi: 10.1158/0008-5472.CAN-08-3658

31. Allen R, Rieger T, Musante C. Efficient Generation and Selection of Virtual Populations in Quantitative Systems Pharmacology Models. *CPT Pharmacometrics Syst Pharmacol* (2016) 5:140–6. doi: 10.1002/psp4.12063

32. Barish S, Ochs M, Sontag E, Gevertz J. Evaluating Optimal Therapy Robustness by Virtual Expansion of a Sample Population, With a Case Study in Cancer Immunotherapy. *Proc Natl Acad Sci* (2017) 114:E6277–86. doi: 10.1073/pnas.1703355114

33. Cassidy T, Craig M. Determinants of Combination Gm-Csf Immunotherapy and Oncolytic Virotherapy Success Identified Through *In Silico* Treatment Personalization. *PloS Comput Biol* (2019) 15:1–16. doi: 10.1371/journal.pcbi.1007495

34. Huang JH, Zhang SN, Choi KJ, Choi IK, Kim JH, Lee M, et al. Therapeutic and Tumor-Specific Immunity Induced by Combination of Dendritic Cells and Oncolytic Adenovirus Expressing IL-12 and 4-1BBL. *Mol Ther* (2010) 18:264–274. doi: 10.1038/mt.2009.205

35. Gevertz J, Wares J. Developing a Minimally Structured Model of Cancer Treatment With Oncolytic Viruses and Dendritic Cell Injections. *Comp Math Meth Med* (2018) 2018:8760371. doi: 10.1155/2018/8760371

36. Kim P, Crivelli J, Choi IK, Yun CO, Wares J. Quantitative Impact of Immunomodulation Versus Oncolysis With Cytokine-Expressing Virus Therapeutics. *Math Biosci Eng* (2015) 12:841–58. doi: 10.3934/mbe.2015.12.841

37. Wares J, Crivelli J, Yun CO, Choi IK, Gevertz J, Kim P. Treatment Strategies for Combining Immunostimulatory Oncolytic Virus Therapeutics With Dendritic Cell Injections. *Math Biosci Eng* (2015) 12:1237–56. doi: 10.3934/mbe.2015.12.1237

38. Kucherenko S, Albrecht D, Saltelli A. Exploring Multi-Dimensional Spaces: A Comparison of Latin Hypercube and Quasi Monte Carlo Sampling Techniques. *arXiv* (2015) 1505:02350. doi: 10.48550/arXiv.1505.02350

39. Torquato S. *Random Heterogeneous Materials: Microstructure and Macroscopic Properties*. New York: Springer-Verlag (2002).

40. Olofsen E, Dinges D, Van Dongen H. Nonlinear Mixed-Effects Modeling: Individualization and Prediction. *Aviat Space Environ Med* (2004) 75:A134–140.

41. Myung JI. Tutorial on Maximum Likelihood Estimation. *J Math Psychol* (2003) 47:90–100. doi: 10.1016/S0022-2496(02)00028-7

42. Lixoft. *Monolix* (2021). Available at: https://lixoft.com/products/monolix/.

43. Lixoft. *Monolix* (2021). Available at: https://monolix.lixoft.com/data-and-models/censoreddata/.

44. Eisenberg M, Harsh E. A Confidence Building Exercise in Data and Identifiability: Modeling Cancer Chemotherapy as a Case Study. *J Theor Biol* (2017) 431:63–78. doi: 10.1016/j.jtbi.2017.07.018

45. Venzon D, Moolgavkar S. A Method for Computing Profile-Likelihood Based Confidence Intervals. *Appl Stat* (1988) 37:87–94. doi: 10.2307/2347496

46. Murphy S, van der Vaart A. On Profile Likelihood. *J Am Stat Assoc* (2000) 95:449–85. doi: 10.1080/01621459.2000.10474219

47. Raue A, Maiwald K, Bachmann J, Schilling M, Klingmuller U, Timmer J. Structural and Practical Identifiability Analysis of Partially Observed Dynamical Models by Exploiting the Profile Likelihood. *Bioinformatics* (2009) 25:1923–9. doi: 10.1093/bioinformatics/btp358

48. Sivia D, Skilling J. *Data Analysis: A Bayesian Tutorial*. New York: Oxford University Press (2006).

49. Maiwald T, Hass H, Steiert B, Vanlier J, Engesser R, Raue A, et al. Driving the Model to its Limit: Profile Likelihood Based Model Reduction. *PloS One* (2016) 11:1–18. doi: 10.1371/journal.pone.0162366

50. Zhang S, Gong C, Ruiz-Martinez A, Wang H, Davis-Marcisak E, Deshpande A, et al. Integrating Single Cell Sequencing With a Spatial Quantitative Systems Pharmacology Model Spqsp for Personalized Prediction of Triple-Negative Breast Cancer Immunotherapy Response. *ImmunoInformatics* (2021) 1–2: 100002. doi: 10.1016/j.immuno.2021.100002