



Evaluating optimal therapy robustness by virtual expansion of a sample population, with a case study in cancer immunotherapy

Syndi Barish^{a,1}, Michael F. Ochs^a, Eduardo D. Sontag^{b,c}, and Jana L. Gevertz^{a,2}

^aDepartment of Mathematics & Statistics, The College of New Jersey, Ewing, NJ 08628; ^bDepartment of Mathematics, Rutgers University, Piscataway, NJ 08854; and ^cCenter for Quantitative Biology, Rutgers University, Piscataway, NJ 08854

Edited by Martin A. Nowak, Harvard University, Cambridge, MA, and accepted by Editorial Board Member Rakesh K. Jain June 15, 2017 (received for review February 27, 2017)

Cancer is a highly heterogeneous disease, exhibiting spatial and temporal variations that pose challenges for designing robust therapies. Here, we propose the VEPART (Virtual Expansion of Populations for Analyzing Robustness of Therapies) technique as a platform that integrates experimental data, mathematical modeling, and statistical analyses for identifying robust optimal treatment protocols. VEPART begins with time course experimental data for a sample population, and a mathematical model fit to aggregate data from that sample population. Using nonparametric statistics, the sample population is amplified and used to create a large number of virtual populations. At the final step of VEPART, robustness is assessed by identifying and analyzing the optimal therapy (perhaps restricted to a set of clinically realizable protocols) across each virtual population. As proof of concept, we have applied the VEPART method to study the robustness of treatment response in a mouse model of melanoma subject to treatment with immunostimulatory oncolytic viruses and dendritic cell vaccines. Our analysis (i) showed that every scheduling variant of the experimentally used treatment protocol is fragile (nonrobust) and (ii) discovered an alternative region of dosing space (lower oncolytic virus dose, higher dendritic cell dose) for which a robust optimal protocol exists.

robust therapies | cancer treatment | mathematical modeling | virotherapy | immunotherapy

Heterogeneity is a defining feature of cancer (1, 2). Interpatient heterogeneity manifests clinically in variable disease progression and treatment response between patients with the same diagnosis, whereas inpatient heterogeneity describes variations that exist between tumor cells in a single patient. Inpatient heterogeneity can be broken down further into intratumor heterogeneity, intrametastatic heterogeneity, intermetastatic heterogeneity, and temporal heterogeneity (2, 3). Intratumor heterogeneity is evident through the presence of multiple genetic subclones within a primary tumor (2), which have even been shown to exist in spatially distinct regions of the primary tumor (4). Intrametastatic heterogeneity is similar to intratumor heterogeneity, but describes heterogeneity within a single metastatic lesion instead of within the primary tumor (2). Intermetastatic heterogeneity, on the other hand, describes variations in subclones between different metastases in the same patient (2). Finally, temporal heterogeneity is defined as changes that take place in the tumor over time, whether they are a result of genomic instability, natural selection, non-Darwinian evolution, or selective pressures imposed by treatment (4–6). Note that, in each case, heterogeneity need not be genetic but may also be epigenetic, phenotypic, or microenvironmental (2, 7, 8).

For decades, cancer patients have been treated using standard of care, meaning they receive the best known treatment that has been deemed as efficacious and safe in epidemiological studies (9). However, in the face of such interpatient and inpatient heterogeneity, standard of care fails to induce a strong antitumor response in some patients, and often loses its efficacy with time (7, 9). The notion of personalized medicine has emerged as an

alternative to standard therapy, and holds the promise of improving clinical care by using patient-specific data to tailor treatment protocols (5, 7, 9–11). The potential benefits of this approach are multifaceted, and include the expectation of a strong patient response with a minimal toxicity profile (5, 9, 10).

As heterogeneity presents challenges for standard of care therapy, it also poses challenges for personalized therapy. If a therapeutic protocol is individually tailored based on measurements from one region of a patient's tumor at a particular pretreatment time point, it is nontrivial to determine how regions of the primary tumor distinct from the biopsy region, not to mention metastatic lesions, will respond to that protocol. Given that both standard of care and personalized cancer treatment protocols must confront the challenge of high levels of variability between and within patients, the design of robust therapeutic protocols is of the utmost importance. A robust treatment should result in the same qualitative response despite a reasonable level of uncertainty (12). In other words, it would be expected to exhibit an antitumor response across a large fraction of patients, and would be more likely to be effective across the range of spatial and temporal variations observed within an individual patient. In this work, we propose to study treatment robustness through a technique that we call Virtual Expansion of Populations for Analyzing Robustness of Therapies (VEPART, for short). VEPART provides a platform for assessing the robustness of treatment

Significance

A successful cancer therapy induces a strong antitumor response while causing minimal side effects. The heterogeneous nature of cancer observed across different regions of the primary tumor, across metastatic sites, across time, and across patients makes designing such a successful therapy challenging. Both standard of care and finely tailored treatment protocols run the risk of not exhibiting a robust antitumor response in the face of these uncertainties. Here we introduce a platform for exploring this robustness question using treatment response data from a sample population. Our method integrates these experimental data with statistical and mathematical techniques, allowing us to quantify therapeutic robustness. Using this approach, we identified a robust therapeutic protocol that combines oncolytic viruses with an immunotherapeutic vaccine.

Author contributions: E.D.S. and J.L.G. designed research; S.B. and J.L.G. performed research; S.B., M.F.O., E.D.S., and J.L.G. analyzed data; S.B. and J.L.G. wrote the paper; and M.F.O. oversaw statistical analyses.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. M.A.N. is a guest editor invited by the Editorial Board.

¹Present address: Graduate Program in Biological & Biomedical Sciences, Yale University, New Haven, CT 06520.

²To whom correspondence should be addressed. Email: gevertz@tcnj.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1703355114/-DCSupplemental.

protocols by coupling experimental data with statistical techniques and mathematical modeling.

As a methodology, VEPART is distinct from important efforts to analyze dynamical system robustness to perturbations, whether they be perturbations in parameter values, initial conditions, or the functional forms used in the mathematical model (12). Of these forms of robustness for dynamical systems, the parameter robustness problem is the most well studied; see, for instance, refs. 12–16. Parameter robustness in a mathematical model has important consequences, as it lends support that conclusions drawn from the model can be trusted in a real-world setting where noise is inevitable. Although not as well understood, others have considered the questions of robustness to change in initial conditions (17, 18), and to change in the dynamical functions (19–21). These considerations are also of paramount importance when assessing the reliability of model predictions.

Comparatively, VEPART is a data-driven approach for studying robustness to an external control, such as a treatment protocol. The starting ingredients for the VEPART method are time course data for multiple individuals in a sample population, together with a parametric mathematical model that has been validated by fitting to aggregate data from this population. VEPART proceeds by amplifying the sample population using nonparametric statistical techniques, eventually resulting in the generation of a large number of “virtual populations.” These virtual populations, each defined by a parameterization of the mathematical model, can be thought of as statistically plausible subsets of the full patient population.

Virtual populations have been used by others, particularly in the field of quantitative systems pharmacology (16, 22–30). Overwhelmingly (see, for instance, refs. 16, 23, 27, and 28), these virtual populations have been used to tackle the previously mentioned challenge of uncertainty in model parameters. From this perspective, patients in a virtual population can be used to understand how differences in disease manifestation occur, and to propose mechanisms for variable drug response. As an exam-

ple, virtual populations have been used to demonstrate the clinical importance of accounting for the correlated expression of different metabolic enzymes (26), and to provide mechanistic explanations for variations in response to a drug for rheumatoid arthritis (27).

In some cases, virtual populations have been exploited to make therapeutic design recommendations. For instance, Wang et al. (24) describe the use of virtual populations to select a dose for a phase 2b clinical trial, and Valitalo et al. (30) used virtual populations to propose a revised dosing guideline for the use of two antibiotics in neonates. Another particularly relevant example is the work of Kansal and Trimmer (22), which discusses how virtual populations (designed through weighting of virtual patients) can be used to optimize a target outcome across virtual populations. No details on how to practically achieve such a goal are given in ref. 22, although a case study is mentioned for using virtual populations to select clinical trial endpoints and to optimize clinical trial design.

These virtual population methods require dozens to thousands of samples to create virtual populations that are representative of the target population (23, 26, 28–30). On the other hand, the VEPART procedure proposed herein works with very small amounts of data, and “expands” the available data in an effort to model the heterogeneity expected in the full population. Not only can VEPART create virtual populations despite limited amounts of data, it can also handle the situation in which the data come from different patients and experimental trials. This ability is in contrast to many existing virtual population approaches in which all variables measured come from the same set of individuals.

In this paper, we describe how VEPART also goes beyond the standard use of virtual populations for studying parametric uncertainty, and instead utilizes these virtual populations to assist in designing therapeutic protocols. In particular, we will systematically detail how the VEPART procedure (summarized in Fig. 1) moves from a small amount of data to a virtual population pool to a therapeutic robustness analysis. This process

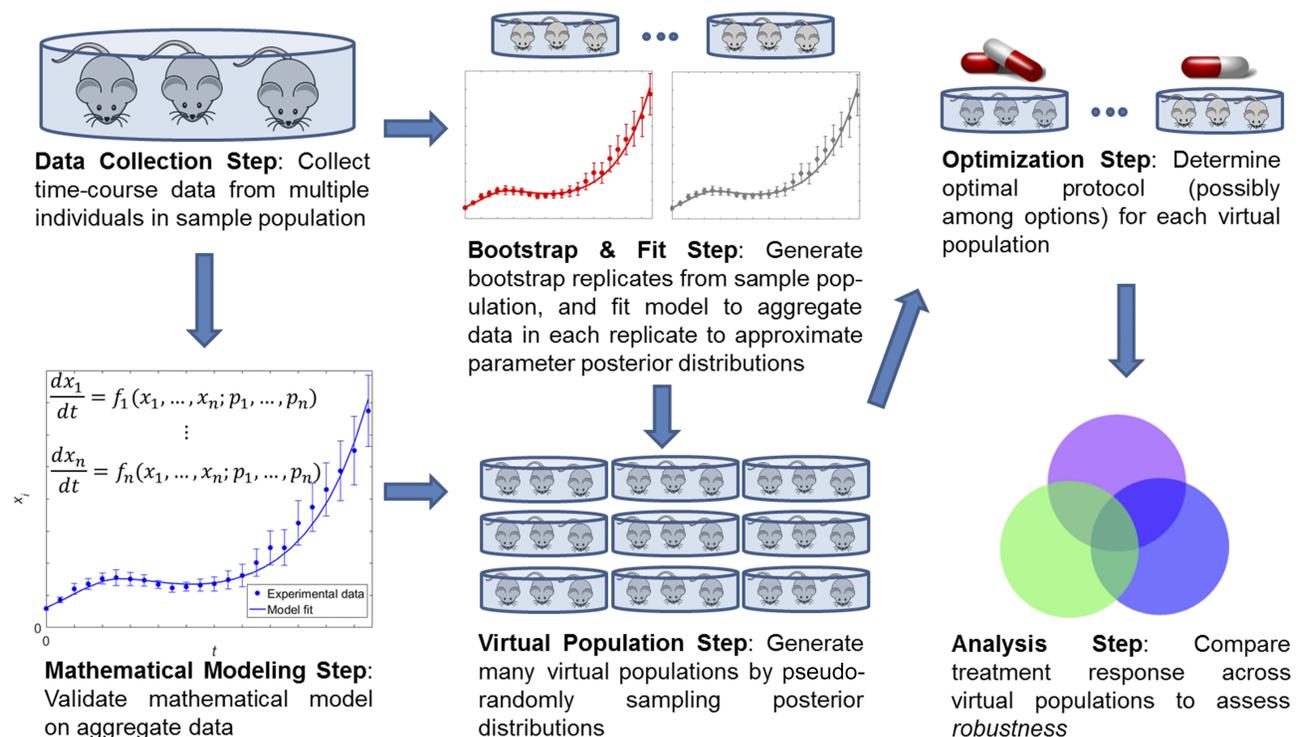


Fig. 1. Schematic of VEPART method for analyzing the robustness of optimal therapeutic protocols.

requires the identification of the optimal protocol (possibly among a list of clinically viable options) for each of the virtual populations, followed by an analysis of responses to the optimal protocol(s) across the virtual populations.

Case Study: Immunoenhanced Oncolytic Viruses with Dendritic Cell Vaccines

As a case study, we have applied this methodology to the experimental data of Huang et al. (31) on a mouse model of melanoma treated with oncolytic viruses (OVs) and dendritic cell (DC) vaccines. OVs are standard viruses genetically engineered to selectively replicate in cancer cells. Upon integration into a tumor cell, OVs replicate and, as part of their normal life cycle, lyse the cell (32). Lysis results in the release of more OVs that can infect additional tumor cells, spreading an infection throughout the tumor that, in theory, results in tumor regression while sparing normal cells. These OVs can be further genetically enhanced to act as a vector for delivering therapeutic genes to the tumor site (32). In ref. 31, an oncolytic virus, the adenovirus (Ad) in particular, is engineered to deliver genes that boost the immune system's ability to identify, target, and kill cancer cells. The transgenes of interest in this study are 4-1BB ligand (4-1BBL) and interleukin (IL)-12; 4-1BBL is a costimulatory molecule expressed on antigen presenting cells. Binding of 4-1BBL to its receptor, 4-1BB, promotes the development and expansion of type-1 T helper cells and cytolytic effector T cells (31). IL-12 is a cytokine that strongly promotes the differentiation of naïve CD4⁺ T cells to type-1 T helper cells (33).

Huang et al. (31) have demonstrated that oncolytic adenoviruses can successfully infect cancer cells, locally deliver therapeutic genes, and cause tumor regression in their mouse model (Fig. 2A). Further, the therapy has been shown to have minimal side effects, as lysis and high expression of therapeutic genes were restricted to cancer cells (31). The efficacy of this treatment protocol has been shown to be enhanced by the use of DC vaccines (Fig. 2B). DCs are antigen-presenting cells that stimulate naïve T cells and generate memory T cells (31). Huang et al. (31) harvested these DCs from the bone marrow of mice, pulsed them *ex vivo* with tumor-associated antigens, and delivered them intratumorally to melanoma-bearing mice. This treatment exhibited antitumor activity in isolation, and potent antitumor effects in combination with oncolytic adenoviruses carrying 4-1BBL and IL-12 (31), which we will refer to as Ad/4-1BBL/IL-12 for brevity (Fig. 2B).

We have previously developed a mathematical model that describes well the data from Huang et al. (31) in all experimentally explored situations (Fig. 2): treatment with only OVs, OVs enhanced with one or more immunostimulatory molecules (4-1BBL, IL-12, or both), DC vaccines, and DC vaccines coupled with Ad/4-1BBL/IL-12 (34, 35). In this prior work, we used the best-fit parameters to address the question of the optimal treatment ordering when administering three doses of Ad/4-1BBL/IL-12 with three doses of the DC vaccine. We predicted that exactly one strategy, front loading the cytokine-bearing OVs (OV-OV-OV-DC-DC-DC), resulted in tumor eradication (35). However, upon further exploration, we found that the model displayed some unintuitive responses to altering the dosing order. For instance, the ordering of the protocols from maximal to minimal tumor response displayed no discernible pattern (35). A striking example of this phenomenon was when the optimal protocol became the worst-case scenario by simply moving the DC dose on the last day of treatment to day one (DC-OV-OV-OV-DC-DC). This extreme sensitivity to dosing order could be explained by our finding that the doses of Ad/4-1BBL/IL-12 and DCs used in the experiments of Huang et al. (31) were near a bifurcation point in our mathematical model. As a result, slightly altering the dose or sequence drastically changed the efficacy of the protocol (35).

Herein, we aim to gain insight into the unintuitive behavior observed in our single-population optimization study by performing a robustness analysis using VEPART. Specifically, we generated a large number of bootstrap replicates by sampling with replacement the experimental data in Huang et al. (31). As detailed in *Computational Methods*, pseudorandom sampling of the posterior distributions for the fit parameters approximated from these bootstrap replicates was used to generate 1,000 virtual populations. The procedure for finding the optimal dosing sequence (among a subset of protocols) was repeated for each virtual population. This process allows us to determine the probability that different protocols will successfully result in tumor eradication, and further allows us to compute the likelihood that a particular sequencing of drugs will be optimal across the virtual populations. We undertook this analysis in three different regions of dosing space: (i) at the experimentally used dose (31), (ii) at a 50% higher dose of Ad/4-1BBL/IL-12 but a 50% lower dose of DC than used in ref. 31, and (iii) at a 50% lower dose of Ad/4-1BBL/IL-12 but a 50% higher dose of DC than used in ref. 31.

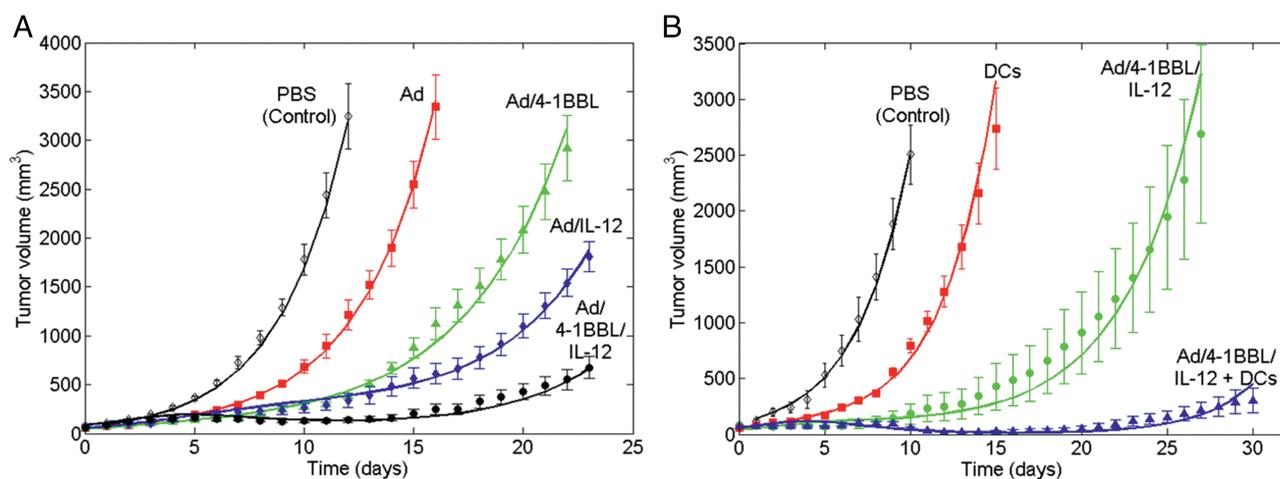


Fig. 2. Experimental data from Huang et al. (31) in which mice with B16-F10 subcutaneous tumors are intratumorally injected with different treatments. Data points represent mean tumor volume \pm SE in each group of six to nine mice. All injections occur on days 0, 2, and 4, unless otherwise specified. (A) Treatments include injection of PBS (control), injection of 10^{10} OVs, injection of 10^{10} OVs carrying the 4-1BBL transgene, injection of 5×10^9 OVs carrying the IL-12 transgene, and injection of 5×10^9 OVs carrying both the 4-1BBL and IL-12 transgene. (B) Treatments include injection of PBS, injection of 10^6 DCs on days 1, 3, and 5, injection of 2.5×10^9 Ad/4-1BBL/IL-12, and injection of 2.5×10^9 Ad/4-1BBL/IL-12 on days 0, 2, and 4 along with 10^6 DCs on days 1, 3, and 5. Also shown are the best-fit solution curves from the model presented in Eqs. 1–6 using the fitting procedure detailed in this work.

This robustness analysis uncovered something unexpected: Every scheduling variant of the experimentally used treatment protocol that involves administering three doses of Ad/4-1BBL/IL-12 and DCs is fragile (nonrobust). To detail, the protocol determined to be optimal at the experimentally used dose (OV–OV–OV–DC–DC) only results in tumor eradication in 30% of the virtual populations. All suboptimal protocols are even less effective. The regime of higher OV and lower DC dose also proved to be fragile, as the optimal protocol leads to tumor eradication in only 43% of the virtual populations. On the contrary, a robust optimal protocol exists when treating with higher doses of DCs and lower doses of OVs than those used by Huang et al. (31). In this dosing regime, the protocol of front loading the DCs (DC–DC–DC–OV–OV) not only proved to be optimal in all virtual populations but also resulted in tumor eradication in more than 84% of those virtual populations. This application serves as proof of concept that computational robustness studies, underpinned by experimental data, can aid in identifying treatment protocols predicted to have strong antitumor properties despite interpatient (and inpatient) heterogeneity.

Results

The full system of six differential equations described in Eqs. 1–6 includes 14 parameters. Previously, we found that the value of 7 of the 14 parameter values (viral production rate, infected cell lysis rate, viral decay rate, naive and cytotoxic T-cell decay rates, base T-cell killing rate, and T-cell differentiation rate) could be reasonably approximated from the literature (Table S1), whereas the other 7 could not. Therefore, it is these 7 parameters that were fit to give the best-fit solution curves to the experimental data, as shown in Fig. 2.

The best-fit parameters for the full model (Ad/4-1BBL/IL-12 + DCs), determined using the hierarchical scheme detailed in *Computational Methods*, are found in Table S2 in *SI Results*. For reference, Table S2 also compares the best-fit parameters found herein to those previously found using a different algorithm and a slightly different metric of fit (35). Finally, Table S2 includes the 95% credible interval for each fit parameter, as calculated from the approximated posterior distributions (Fig. S1). These 95% credible intervals are compared to the results of a local sensitivity analysis, as shown in Figs. S2 and S3 of *SI Results*.

Robustness at Intermediate OV and DC Dose. Here we focus on identifying and classifying the response and robustness to 20 treatment protocols (all possible combinations of Ad/4-1BBL/IL-12 with DCs given at three doses a piece, separated by 1 d) at the experimentally used dose of 2.5×10^9 OVs per dose and 10^6 DCs per dose (31); we will refer to this as the “intermediate OV/DC dose” going forward. Our previous analysis revealed extreme sensitivity to parameters at this dosage (35), so here we expand upon that observation and undertake an exploration of treatment robustness using the VEPART approach.

To this end, we generate 1,000 virtual populations by pseudo-randomly sampling the posterior distributions on the fit parameters (Fig. S1), as described in *Computational Methods*. These virtual populations are alternative parameterizations to the model in Eqs. 1–6 that, when pooled together, in theory capture the underlying heterogeneity in the full patient population (16). The 20 treatment orderings considered in ref. 35 were administered to each of these virtual populations, allowing us to determine a “population-level” response to drug ordering at the intermediate OV/DC dose (Fig. 3).

Fig. 3A demonstrates the extent to which the virtual populations respond differently to the same treatment protocol. For a fixed ordering of the drugs (specified on the x axis), we see the frequency with which each protocol ranks in positions 1 through 20, with 1 being the optimal treatment ordering and 20 being the worst. Little consistency is observed in the ranking of the different treatment protocols across virtual populations. For instance, 65% of the treatment protocols rank in both the top two and bottom two positions in some of the virtual populations with nonzero probability (Fig. 3A). That number grows to 90% if we consider those that can rank in both the top and bottom five positions (Fig. 3A).

Another indication of the nonrobust (fragile) response across virtual populations at this dose is that, on average, each of the 20 treatment protocols achieve over 17 different rankings. As an example, the previously found optimal protocol of OV–OV–OV–DC–DC achieves every rank between 1 and 20, with the exception of rank 13 (Fig. 3A). This protocol is most likely to rank in the top position, as it ranks in position 1 for 72.2% of the virtual populations (Fig. 3A). However, the next most likely ranking for this protocol is in the last position, as it ranks in position 20

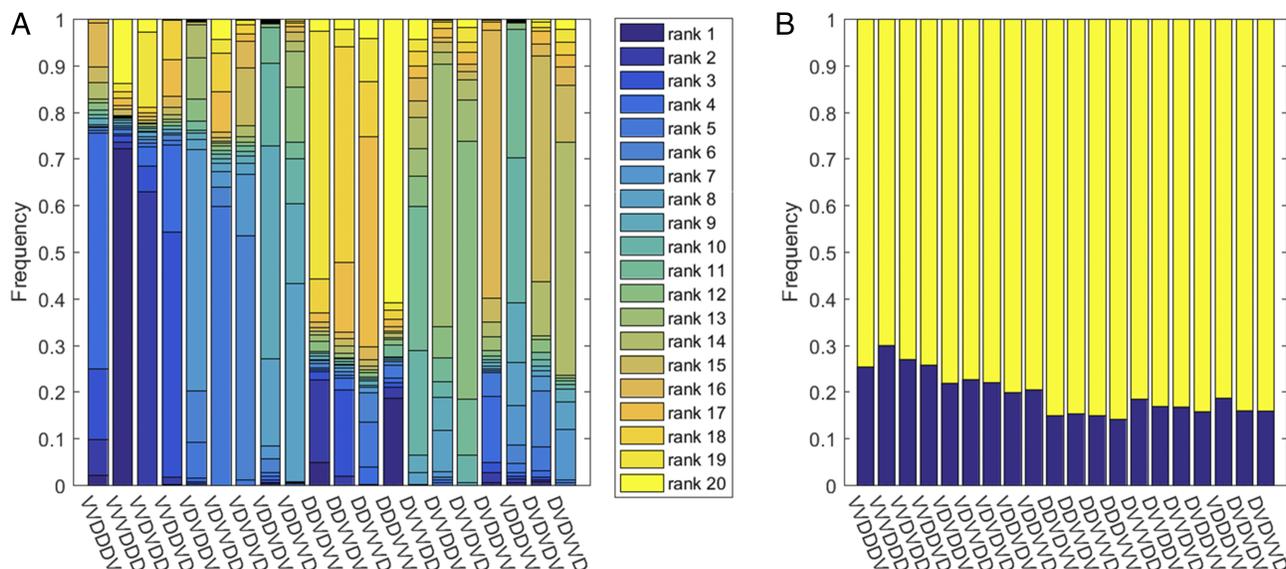


Fig. 3. VEPART output at intermediate OV and DC doses. The x axis indicates the treatment protocol, with “V” representing Ad/4-1BBL/IL-12 treatment and “D” representing DC treatment on a given day. (A) For each of the 20 treatment protocols, we see the frequency at which it ranks in positions 1 to 20. Observe that there is little consistency across the virtual populations: On average, each of the treatment protocols can achieve over 17 different ranks. (B) The frequency of virtual populations for which the specified treatment protocol leads to tumor eradication (blue) or tumor escape (yellow).

for 13.8% of the virtual populations (Fig. 3A). The fact that the same protocol is the best case for some virtual populations and yet is the worst case for others strongly suggests the originally identified optimal protocol is not robust.

Because there was minimal support for a given treatment protocol to achieve a consistent ranking across the virtual populations, we then sought to address a simpler question: Do certain protocols result in tumor eradication in a large fraction of the virtual populations, independent of where they appear on the rank-ordered list of schedules? Consistent with the lack of a robust response at this dose, we found no protocols that lead to eradication in a majority of the virtual populations. Instead, we observed that the treatment that ranked as the top protocol most frequently (OV–OV–OV–DC–DC–DC) only results in tumor eradication in 30% of the virtual populations (Fig. 3B). All other protocols have a lower likelihood of tumor eradication, lending further evidence to the claim that the tumor eradication response in this dosing regime is fragile.

Digging more deeply into the binary classification of treatment protocols, we found that treatment response is more closely tied to the virtual population under consideration than to the treatment protocol itself (order in which the doses are given). By sorting our data by virtual population (instead of treatment protocol, as done in Fig. 3), we find that just over 76% of the virtual populations have the same binary response (tumor eradication or tumor escape) to all 20 treatment protocols tested. In particular, our simulations reveal that tumor escape will occur independent of the treatment protocol used in just over 66% of the virtual populations. Further, tumor eradication will occur independent of the treatment protocol in 10% of the virtual populations. Because the majority of virtual populations have a binary response (eradication or escape) independent of the dosing order, we learn that the parameters in the virtual population, and not the treatment order itself, are really driving the predictions in the intermediate OV/DC region of dosing space.

To conclude, in this region of dosing space treatment response can vary significantly across virtual populations. None of the treatment protocols, including the optimal one of front loading the OVs, robustly result in tumor eradication across the virtual populations. The situation only gets messier when we consider

the frequency with which each protocol achieves different rankings (from top protocol to worst protocol). Finally, the observation that treatment response is driven by the parameters that describe a virtual population, rather than the dosing order itself, further suggests that optimality predictions at this dosage may be of limited value. Because the population-level eradication response is fragile, we conclude that the combination of giving three doses of 2.5×10^9 Ad/4-1BBL/IL-12 and three doses of 10^6 DCs is not the ideal drug dosage to administer in this experimental system.

Robustness at Higher OV and Lower DC Dose. Because the intermediate OV/DC dose was found to be fragile, we next sought to analyze the robustness of the 20 protocols at a 50% higher OV dose (3.75×10^9 versus 2.5×10^9 OVs per dose) and a 50% lower DC dose (0.5×10^6 versus 10^6 DCs per dose). Like before, robustness of a protocol is assessed by measuring treatment efficacy across all virtual populations generated by pseudorandomly sampling the approximated posterior distributions for the fit parameters, as detailed in *Computational Methods*.

Similar to what was observed in the intermediate OV/DC dose, no treatment protocols result in a robust eradication response across the virtual populations (Fig. 4A). Although the protocol of front loading OVs consistently ranks as the top treatment across all virtual populations (Fig. 4B), the likelihood of tumor eradication using this protocol is only 43%. This outperforms the same protocol (which was also optimal) in the intermediate OV/DC dosing regime by 13%, but it still leaves over half of the virtual populations unsuccessfully treated. Further, all other protocols have a lower likelihood of eradication (with OV–OV–DC–OV–DC–DC being the only other one to cross the 33% threshold), suggesting that the tumor eradication response in this dosing regime is fragile.

That said, some responses in this dosing regime are robust, just not when it comes to the desired outcome of tumor eradication. Sixty percent of the protocols result in tumor escape in more than 90% of the virtual populations, and front loading the DCs results in treatment failure in 100% of the virtual populations (Fig. 4A). Further, unlike in the intermediate OV/DC regime, there is little variation in the rank each treatment protocol can achieve

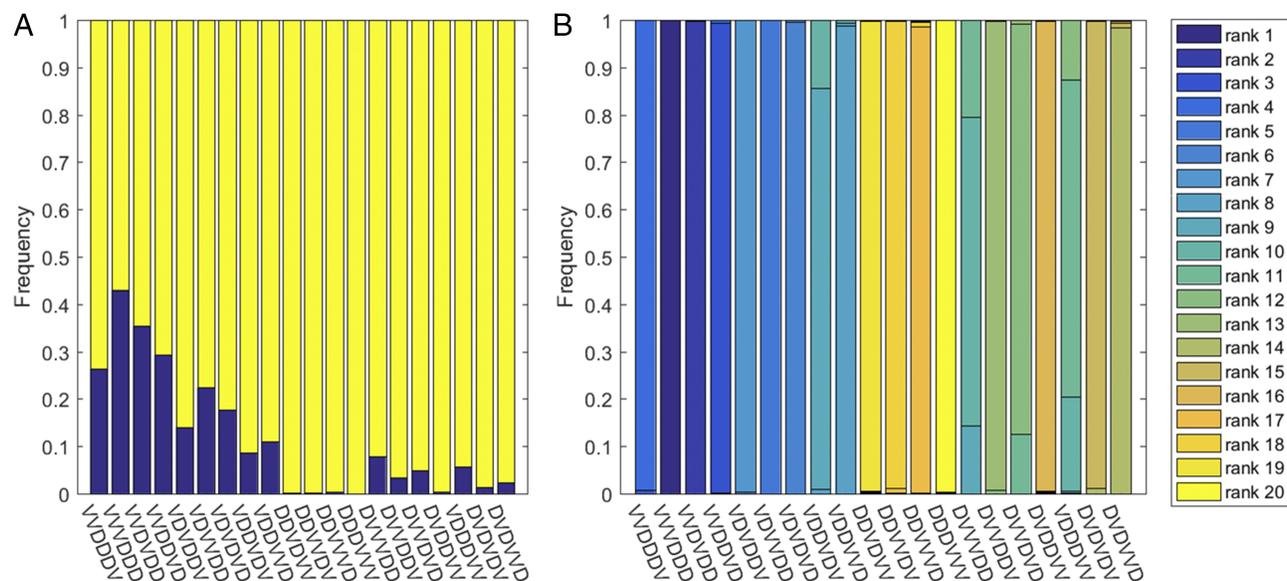


Fig. 4. VEPART output at high OV/low DC doses. The x axis indicates the treatment protocol. (A) The frequency of virtual populations for which the specified treatment protocol leads to tumor eradication (blue) or tumor escape (yellow). (B) For each of the 20 treatment protocols, we see the frequency at which it ranks in positions 1 to 20. Observe that there is strong consistency across the virtual populations, with each protocol achieving a dominant ranking and other rankings only near that dominant one.

across the virtual populations (Fig. 4B). This robust ranking can be explained by the fact that tumor response at this dosage is driven by the protocol itself, and not by the particular parameters in a virtual population. So, although several robust treatment protocols have been identified in this dosing regime, it is the undesirable tumor escape response that is robust. Therefore, in the high OV/low DC case, we predict that no treatment protocol will have robust antitumor activity in the face of interpatient and intrapatient heterogeneity.

Robustness at Lower OV and Higher DC Dose. We next sought to explore whether a robust tumor eradication response exists for any of the 20 protocols when treating at a 50% lower OV dose (1.25×10^9 versus 2.5×10^9 OVs per dose) and a 50% higher DC dose (1.5×10^6 versus 10^6 DCs per dose). As done in the two other regions of dosing space, robustness of a protocol is assessed by measuring treatment efficacy across all virtual populations.

We began by exploring whether any dosing protocols result in the eradication of a large fraction of the virtual populations. We find that 30% of the treatment protocols lead to tumor eradication in at least half of the virtual populations (Fig. 5A). This finding stands in stark contrast to both the intermediate OV/DC dose and the low OV/high DC dose. In those cases, not one protocol could result in tumor eradication in half of the virtual populations (Figs. 3B and 4A).

As our intention is to find protocols that result in a robust eradication response across virtual populations, we next narrowed our focus to protocols that cause tumor eradication in 70% or more of the virtual populations. Three protocols were found that satisfy this criterion, and they share one common feature: The first two doses are always DCs, meaning the protocol is of the form DC–DC–X–X–X–X. Further, the earlier the third DC dose was given in the treatment protocol, the better the protocol performed across the virtual populations. Front loading with DCs in the DC–DC–DC–OV–OV–OV protocol led to tumor eradication in 84.2% of the virtual populations (Fig. 5A). Moving the third DC dose to later in the schedule decreases the likelihood of tumor eradication by the protocol: A DC dose at day 4 results in 76.9% eradication, one at day 5

results in 70.2% eradication, and one at day 6 results in 61.9% eradication.

Not only do we see a robust eradication response for protocols of the form DC–DC–X–X–X–X in this dosing regime, we also find that the ranking of the protocols across the virtual populations is shockingly preserved (Fig. 5B). The most striking example is that the schedule of front loading with DCs ranks as the top protocol in all of the virtual populations. As can be calculated from the data in Fig. 5B, each treatment protocol, on average, will only rank in 4 of the 20 positions, and, overwhelmingly, the achievable ranks are consecutive (e.g., DC–DC–OV–OV–DC–OV most commonly ranks at position 3 but can also rank at positions 2 and 4). This finding is in stark contrast to the intermediate OV/DC dosing regime, in which each protocol ranks in 17 of 20 positions, on average. The lack of variation in the rank a treatment protocol can achieve across the virtual populations in the case of low OV/high DC further supports our claim that treatment protocols of the form DC–DC–X–X–X–X have a robust tumor eradication response in this dosing regime.

Although of less interest, a pattern also emerges among those protocols that administer Ad/4-1BBL/IL-12 on day 1. All of these protocols result in a robust escape response, with escape occurring in two thirds or more of the virtual populations (Fig. 5A). We observe that protocols that give OVs on the first 2 d perform very poorly, resulting in tumor eradication in less than 14% of the virtual populations (Fig. 5A). Further, with the exception of one protocol (OV–DC–DC–DC–OV–OV), all protocols that administer an OV on day 1 have a lower probability of tumor eradication than protocols that administer a DC on day 1. To understand why this response occurs, note that, in this regime, the dose of OV is 50% below the dose deemed to have a strong tumor-killing effect with a reasonable toxicity profile in the experimental system. Starting a treatment protocol with a relatively ineffective dose of OV increases the likelihood of tumor escape. On the other hand, the optimal protocol administers high doses of DCs early in treatment, and these high doses can compensate for the relatively low dose of OV administered. Although protocols that front-load OVs would be avoided due to limited antitumor efficacy, this finding further illustrates the robust behavior of several of the treatment protocols in this region of dosing space.

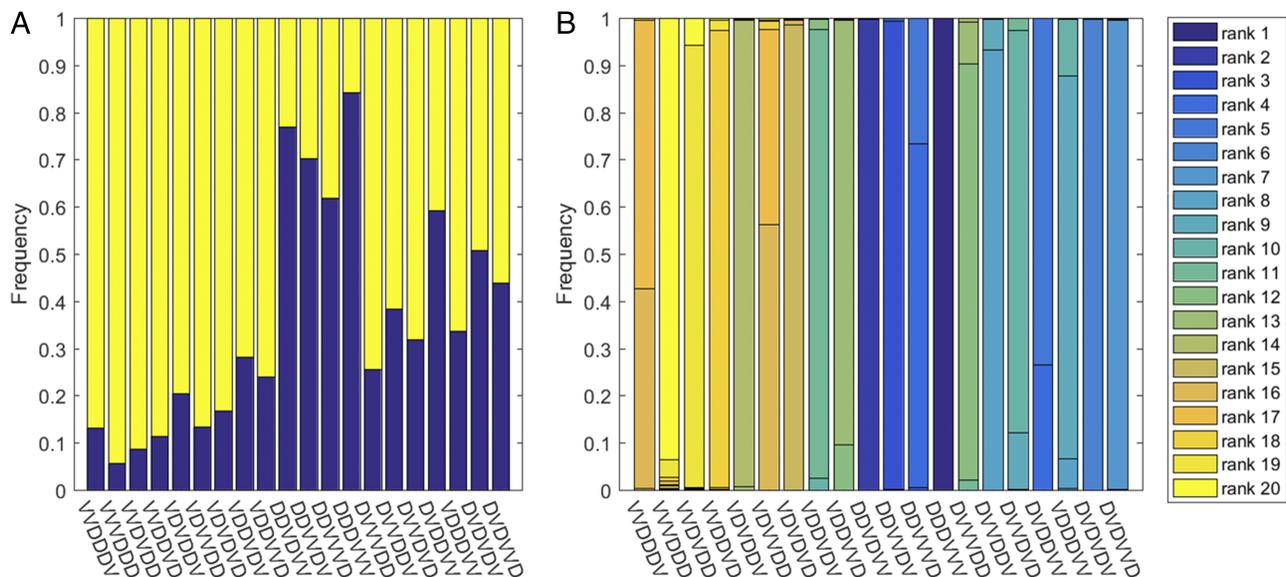


Fig. 5. VEPART output at low OV/high DC doses. The x axis indicates the treatment protocol. (A) The frequency of virtual populations for which the specified treatment protocol leads to tumor eradication (blue) or tumor escape (yellow). (B) For each of the 20 treatment protocols, we see the frequency at which it ranks in positions 1 to 20. Observe that there is strong consistency across the virtual populations, with each protocol achieving a dominant ranking and other rankings only near that dominant one.

Discussion

In this work, we introduced the VEPART platform for assessing the robustness of treatment protocols. VEPART implementation begins with an experimental dataset describing the temporal evolution of some variable (for instance, tumor volume) in response to treatment in a sample population. A mathematical model (for instance, a system of differential equations) is then developed and fit to the experimental data. Once a reasonable mathematical model is attained, nonparametric statistical techniques are used to amplify the sample population and generate a large number of virtual populations. One benefit of this approach is that a large sample size, which can be difficult to attain experimentally, is not required to create these virtual populations. At this point, treatment optimization can be performed for each virtual population.

Although optimization is restricted to a set of clinically realizable protocols in this work, once the virtual populations have been created, classical tools from optimal control theory could also be applied (36, 37). An excellent overview of applying optimal control theory to cancer chemotherapy planning can be found in ref. 38. The use of optimal control in the context of the VEPART method would allow a more general optimal protocol to be identified, where “optimal” could be defined in many ways; for instance, one could seek to minimize tumor volume at an endpoint, minimize tumor volume over a time horizon, minimize drug concentration, minimize toxicity, minimize some weighted average of the above, etc. (37–45). Further, this optimization can be performed subject to various types of constraints; for instance, toxicity could be introduced as a constraint on the amount of drug that can be administered (37, 46). These approaches can be used to find the optimal schedule for a single therapeutic agent, or for a mixture of drugs as done in refs. 43 and 47. Further, from any such optimal control problem, one could perform a sensitivity or elasticity analysis (similar to the one performed herein; see *SI Results*) to quantify how parametric changes influence the quality of an optimal treatment protocol, as in ref. 48.

A variety of solution methods exist for solving these optimal control problems (38). Analytical expressions for the optimal control can be obtained for simple models (see, e.g., ref. 49), and, when the model is too complex to obtain an analytic expression, approximation techniques can be used (see, for instance, ref. 43). When the optimal solution is computationally intractable, heuristic optimization algorithms, including genetic algorithm and simulated annealing can be used, as done in refs. 42, 47, and 50. Independent of how the optimization is carried out, an analysis of the optimal protocols across virtual populations using the VEPART method allows for the quantification of protocol robustness.

Related Work: Robust Cancer Therapies. In the cancer literature, examples of analogous efforts for designing robust treatment protocols can be found in radiotherapy (51–54). The challenges in designing a robust radiation treatment protocol can largely be attributed to geometric uncertainty and interpatient variability (54). Several studies have addressed the challenges posed by geometric uncertainty, including factors such as organ motion, variations in treatment setup, patient positioning errors, and fluctuations in machine output (54). For instance, Liu et al. (52) used “worst-case robust optimization” to identify intensity modulated proton therapy (IMPT) plans that are robust to uncertainties in the beamlet range and equipment setup. For the same beam arrangement, they computed a different dose distribution accounting for these uncertainties and compared this to the optimal dose distribution. They applied this approach to the lung, skull base and prostate and found that compared with IMPT plans optimized using conventional (nonrobust) methods, their method resulted in radiotherapy plans that are less sensitive to beamlet range and setup uncertainties (52). In a series of papers, robust optimization was used for planning intensity modulated radiation therapy for lung cancer in the face of uncertainty caused by breathing (51, 55, 56). This work introduces a data-driven model

of uncertainty to describe variations in breathing motion and allows the authors to determine the trade-off between ensuring the tumor receives sufficient radiation and minimizing the dose to normal tissue (51, 55, 56).

On the other hand, the work by Leder and colleagues (54) was the first to consider the challenge that interpatient variability poses to designing a robust radiotherapeutic protocol. They were particularly concerned about the potential risk that the optimal protocol (determined using the linear–quadratic model) may be weakly efficacious for some individuals, or may cause high toxicity to at-risk organs for others. To address this robustness concern, the authors used a stochastic optimization scheme in which the unknown parameters are assumed to be random variables with known distributions. The goal is to choose a dosing schedule (total dose, number of fractions, dose per fraction, and treatment duration) for which the objective function attains a high level with a given probability, and that the imposed constraints are also satisfied with a given probability. The authors found some important differences in the optimal schedule that accounts for parameter uncertainty compared with the optimal schedule found using the nominal parameter values in the absence of uncertainty (54). Our work is motivated by a similar robustness question to that considered by Leder and colleagues in ref. 54, although not restricted to radiotherapy.

Case Study: Immunoenhanced OV_s with DC Vaccines. As a proof of concept, we applied VEPART to explore the robustness of treatment protocols that combine three doses of immunostimulatory OV_s with three DC injections; to accomplish this, we worked with the experimental data from Huang et al. (31). The nature of this dataset previously led us to use a hierarchical method for fitting the parameter values (*Computational Methods*), which posed a challenge for directly fitting parameters for each of the virtual populations. To overcome this challenge, we began by bootstrapping the experimental data and using the best-fit parameters in each bootstrap replicate to approximate the posterior probability distribution for the parameters in our mathematical model (Eqs. 1–6). Pseudorandom sampling of these distributions, as detailed in *Computational Methods*, allowed for the creation of our 1,000 virtual populations.

In this case study, 20 treatment protocols (all possible combinations of Ad/4-1BBL/IL-12 with DCs given at three doses a piece, separated by 1 d) were ranked across the virtual populations in three regions of dosing space: (i) intermediate and experimentally used OV/DC dose, (ii) high OV/low DC, and (iii) high DC/low OV. VEPART revealed that the first two dosing regimes we considered are fragile. To detail, the protocol of administering Ad/4-1BBL/IL-12 on the first 3 d, and following up with a sequence of DCs, was optimal for both the intermediate dose of OV_s and DCs and the high OV/low DC dose. However, in neither case did this protocol cause a robust eradication response: The protocol only causes tumor eradication in 30% of the virtual populations at the intermediate dose, and in 43% of the virtual populations at the high OV/low DC dose. As a consequence of VEPART’s nonrobust prediction, it is not expected that any ordering of three OV_s and three DCs at these two doses will exhibit a strong antitumor response in the face of interpatient and inpatient heterogeneity.

On the other hand, in the high DC/low OV region of dosing space, a robust optimal protocol was identified. This protocol administers DCs on the first 3 d, and follows up with a sequence of Ad/4-1BBL/IL-12 on the last 3 d. Not only did this protocol rank as optimal in all virtual populations, it also resulted in tumor eradication in 84% of those virtual populations. Therefore, VEPART predicts that, in the high DC/low OV region of dosing space, the treatment that front-loads DCs will be effective across different individuals in a population, and will be more robust to spatial and temporal variations within an individual.

In future work, the use of approximate optimal control techniques or heuristic optimization algorithms (38, 42) can extend the current implementation of VEPART. In particular, we can optimize over a wider range of protocols (for instance, considering

a variable number of OV and DC doses, and/or allowing variable spacing between doses), and over a more complete range of dosing space. This extended analysis of our system could result in finding robust protocols beyond treating with DC–DC–DC–OV–OV–OV in the high DC/low OV region of dosing space.

This case study is intended to illustrate how the VEPART method can be used to assess therapeutic robustness. By combining experimental data with mathematical modeling and statistical analyses, the VEPART approach can contribute to tackling a significant problem clinicians face when treating cancer using standard of care or an individualized treatment protocol: high levels of interpatient and inpatient variability.

Computational Methods

In this section, we introduce our mathematical model that includes immunostimulated OVs and DC vaccines. Although this model has been successfully fit to the experimental data of Yun and colleagues (31, 35), we introduce an alternative approach to fitting the data that results in small but significant improvements in the model fits. To further study the relationship between the model and its parameters, in this section, we also describe two key steps of our proposed VEPART approach: (i) We detail the bootstrapping procedure that produces posterior distributions and credible intervals for each of the fit parameters, and (ii) we explain how these posterior distributions are used to generate virtual populations that provide the foundation of our robustness analysis.

Mathematical Model. In this study, the data-validated model from Wares et al. (35) was used. The model is represented by the following initial value problem:

$$\frac{dU}{dt} = rU - \beta \frac{UV}{N} - (k_0 + c_{kill}I) \frac{UT}{N} \quad U(0) = U_0, \quad [1]$$

$$\frac{dI}{dt} = \beta \frac{UV}{N} - \delta_I I - (k_0 + c_{kill}I) \frac{IT}{N} \quad I(0) = 0, \quad [2]$$

$$\frac{dV}{dt} = u_v(t) + \alpha \delta_I I - \delta_V V \quad V(0) = 0, \quad [3]$$

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

$$\frac{dA}{dt} = c_A I - \delta_A A \quad A(0) = 0, \quad [5]$$

$$\frac{dD}{dt} = u_D(t) - \delta_D D \quad D(0) = 0, \quad [6]$$

where U is the number of uninfected tumor cells (the only variable with a nonzero initial condition that gets fit at each step of the hierarchical fitting process), I is the number of tumor cells infected by the OV, V is the number of free virions (oncolytic adenoviruses), T is the number of tumor-targeted T cells, A is the number of naïve T cells, D is the number of injected DCs (note that endogenous DCs are not directly modeled), and N is the total number of cells (tumor cells and T cells) at the tumor site. The individual terms that contribute to the rate of change of each population are explained in more detail in *SI Computational Methods*.

This model was hierarchically developed to fit the increasingly complex data in Fig. 2. Although more details are provided in *SI Computational Methods*, here we briefly note the different hierarchical stages of the model:

- i) Model 0 is a model of untreated tumor growth in the absence of any virions, T cells, or DCs.
- ii) Model 1 is a model of tumor growth and treatment with oncolytic virotherapy that does not carry any immunostimulatory transgenes. T cells and DCs are still not considered.
- iii) Model 2a is a model of tumor growth and treatment with the immunostimulatory oncolytic virotherapeutic Ad/4-1BBL. The immunostimulation results in the inclusion of tumor-

targeted T cells, but naïve T cells and DCs are still not considered.

- iv) Model 2b is a model of tumor growth and treatment with the immunostimulatory oncolytic virotherapeutic Ad/IL-12. The immunostimulation results in the inclusion of tumor-targeted T cells, including the naïve T cells, but DCs are still not considered.
- v) Model 3 is a model of tumor growth and treatment with the immunostimulatory oncolytic virotherapeutic Ad/4-1BBL/IL-12. The immunostimulation results in the inclusion of tumor-targeted T cells, including the naïve T cells, but DCs are still not considered.
- vi) Model 4 is a model of tumor growth and treatment with Ad/4-1BBL/IL-12 and DC vaccines. At this point, the model includes all equations and all terms within each equation, including the tumor-targeted T cells, the naïve T cells, and the DCs. A subcase of model 4 in which all virus-related parameters are set to zero gives us the case of treating with only the DC vaccine.

Fitting Model to Experimental Data. Just as the structure of the model was hierarchically developed, the parameter fitting scheme was also hierarchical in nature (34, 35). As illustrated in Fig. S4, parameters that could not be well-estimated from the literature were fit at each hierarchical model stage, and some subset of the fit parameters are inherited to the subsequent models in the hierarchical scheme. Details of this hierarchical fitting scheme can be found in *SI Computational Methods*. The initial condition for the number of uninfected tumor cells was always refit at each model stage, as each model was fit to a different data set. All other initial conditions are identically zero, as the administration of treatment is captured through the time-varying terms [$u_v(t)$ and $u_D(t)$] in Eqs. 1–6.

The parameters in models 1 through 4 were previously fit using the Levenberg–Marquadt algorithm as implemented by MATLAB's `lsqnonlin` command (35). Like most parameter fitting schemes, the Levenberg–Marquadt algorithm can get stuck at local minimum, and the algorithm's performance deteriorates as a function of the dimension of parameter space. To have more control over parameter fitting, and to attempt to avoid getting trapped at local minimum, here we instead use simulated annealing for model fitting.

Goodness of model fit to the experimental data is sought by minimizing the objective function S , as defined by

$$S = \sum_t v_t \frac{(v_t^{model} - v_t)^2}{\sigma_t^2}, \quad [7]$$

where v_t is the average experimental tumor volume at day t , v_t^{model} is the tumor volume at day t predicted by Eqs. 1–6, and σ_t^2 is the variance in the experimental tumor volume at day t . Within Eq. 7, the fractional term is a dimensionless measure of the error: The numerator gives the sum of the square errors between the model predictions and the (average) experimental data, and the denominator is the variance in the experimental data. By dividing the sum of the square errors by the variance, we impose the condition that, at time points where the variance is small, we want a better fit to the average volume than is required at time points where the variance is large, in accordance with the principle of maximum likelihood estimation (57).

Because volume measurements made using calipers are imprecise for smaller tumor sizes [instrumentation error in measuring length and width is independent of tumor volume (58, 59)], the dimensionless term in Eq. 7 is weighted by the average tumor volume at day t . Our weighting has the additional advantage of leading to a better-posed numerical optimization problem, as this objective function does not artificially bias the algorithm to fit well at small tumor sizes (when the variance is very small) at the expense of fitting well over a majority of the data points. Details on the simulated annealing algorithm being used to minimize Eq. 7 can be found in *SI Computational Methods* and Table S3 therein.

Analyzing Best-Fit Parameters. For each dataset (control, Ad only, etc.) consisting of N mice (N was between six and nine, depending on the dataset), 1,000 bootstrap replicates have been created. Each bootstrap replicate was created by sampling N mice from the original dataset with replacement (36) using the programming language R. The simulated annealing protocol was run on each bootstrap replicate, and the best-fit parameter values for each bootstrap replicate were binned in a histogram. This process provides a visualization of the estimated posterior marginal distribution on each fit parameter. Given the distributions were empirical and asymmetric, we defined the 95% credible interval for that parameter (the interval for which we can be 95% confident that the true value of the parameter occurs in) as the range that excluded 2.5% of the values on each end of the distribution. Bootstrapping has been used by others to explore parameter sensitivity in mathematical biology models with implications for cancer; see, for instance, refs. 60–62.

A local sensitivity analysis was also performed to give an alternative mechanism for analyzing the fit parameters. Details on this are found in *SI Computational Methods*.

Ranking and Robustness of Treatment Protocols. Previous work determined the optimal protocol among the 20 possible orderings of three OV injections and three DC injections at the experimentally used dose of 2.5×10^9 OVs and 10^6 DCs (35). To rank these 6-d-long protocols, tumor growth was tracked over 1 mo, and the volume predicted by the model after this period was recorded. That analysis revealed that the only ordering of the drugs that led to tumor eradication (defined as tumor volume dropping below the assumed volume of one cell, 10^{-6} mm³) was the one that front-loads the OV (that is, OV–OV–OV–DC–DC–DC) (35). However, further analysis revealed that the dose used in the experimental work was near a bifurcation point in the system, and small changes in the drug dose could result in a different optimal solution (35).

This finding motivated us to extend our analysis to investigate treatment robustness, as such an analysis gives more information about therapeutic efficacy than a single-population optimization study. To explore robustness, the data obtained from the bootstrap replicates were used to construct virtual populations, similar to the nonparametric approach described in ref. 29. In particular, herein, 1,000 virtual populations were created by pseudorandomly picking the six fit parameters (r , β , c_A , c_T , c_{kill} , and χ_D) from their posterior distributions approximated using the bootstrap replicates and the hierarchical fitting process. Two constraints were imposed when creating these virtual populations. First, to maintain the covariance structure between the parameters, those that were fit together were selected together to create a new virtual population. This criterion impacted the selection of parameters from model 3, and we implemented this parameter selection by randomly choosing a value of c_A from its approximated posterior distribution, and subsequently setting the values for c_T and c_{kill} to the best-fit value in the same bootstrap replicate that gave that value of c_A .

Note that an implicit assumption in creating virtual populations in this way is that parameters fit at different stages of the hierarchical fitting process are minimally correlated. This assumption is grounded in our fitting methodology, in which parameters fit at each step of the hierarchy are treated as independent biological phenomenon. For instance, the fitting of the tumor growth rate r at the first step of the hierarchy isolates the behavior of the tumor

without treatment, and the fitting of the viral infectivity parameter β at the next step of the hierarchy is meant to isolate the behavior of the OV. In effect, this setup establishes a regime of tumor growth and viral activity that the system responds to. Therefore, by independently sampling from the distribution for r and β to create a virtual population, we are trying to select a reasonable tumor growth rate and a reasonable viral infectivity rate, as seen in the experimental data. Others have given consideration to further preserve the covariance structure among all variables; see, for instance, refs. 16, 23, 25, 26, 29, and 30. The second constraint imposed is that only virtual populations whose parameter values are all within their respective 95% credible intervals were considered; this can be thought of as an “inclusion–exclusion” criterion that refines the virtual population pool to be statistically similar to the experimental population (29), including mirroring its heterogeneity. That said, this approach does constrain the virtual populations to statistically resemble the experimental data, which, because of the small sample size in our experiments, runs the risk of not resembling the population data (29). However, the experimental data considered herein inject genetically identical mice with the same number of cells from the same cancer cell line. Therefore, at least for these data, it is reasonable that the small sample can be expanded to result in virtual populations that accurately reflect the true heterogeneity in the population.

Therapeutic robustness was assessed by ranking the 20 protocols of interest for each of the virtual populations that were generated. Within a virtual population, a treatment was ranked using a two-component criterion. First, for tumors that were eradicated within 30 d (volume drops below 10^{-6} mm³), the time until eradication is used to measure the effectiveness of the protocol, with faster eradication times considered superior to slower ones. Second, for tumors that were not eradicated within 30 d, the volume after 30 d was used to measure the effectiveness of the treatment, with smaller volumes considered superior to larger ones. By analyzing the response to the 20 protocols across the 1,000 virtual populations, the robustness of each protocol can be studied.

Given that our previous work suggested that the experimentally used dose does not have a robust optimal solution, we performed this robustness study in three different regions of dosing space: (i) at the experimentally used dose (OV dose of 2.5×10^9 , DC dose of 10^6), (ii) at a 50% higher dose of OV but a 50% lower dose of DC (OV dose of 3.75×10^9 , DC dose of 0.5×10^6), and (iii) at a 50% lower dose of OV but a 50% higher dose of DC (OV dose of 1.25×10^9 , DC dose of 1.5×10^6). These two other regions of dosing space were chosen because they are sufficiently distinct from the experimental dose such that we could expect different optimal protocols and a different robustness profile. Further, both of these regions of dosing space involve increasing the dose of one drug and decreasing the dose of the other by the same relative amount; this was done in an attempt to preserve the toxicity profile across the three dosing regimes.

ACKNOWLEDGMENTS. J.L.G. thanks Chae-Ok Yun, Peter Kim, and Joanna Wares for their collaboration, and particularly for the exposure and access they provided to the rich dataset used in this work. M.F.O. was supported by National Institutes of Health (NIH) Grant R01LM011000. The work of E.D.S. was partially supported by NIH Grant 1R01GM100473, Air Force Office of Scientific Research Grant FA9550-14-1-0060, and Office of Naval Research Grant N00014-13-1-0074. The work of J.L.G. was partially done while visiting the Mathematics Department and Center for Quantitative Biology at Rutgers University, with partial support from NIH Grant 1R01GM100473.

- Menzies A, et al. (2014) Inter- and intra-patient heterogeneity of response and progression to targeted therapy in metastatic melanoma. *PLoS One* 9: e85004.
- Jamal-Hanjani M, Quezada S, Larkin J, Swanton C (2015) Translational implications of tumor heterogeneity. *Clin Cancer Res* 21:1258–1266.
- Schwarz R, et al. (2015) Spatial and temporal heterogeneity in high-grade serous ovarian cancer: A phylogenetic analysis. *PLoS Med* 12:e1001789.
- Ling S, et al. (2015) Extremely high genetic diversity in a single tumor points to prevalence of non-Darwinian cell evolution. *Proc Natl Acad Sci USA* 112:e6496–e6505.
- André F, et al. (2013) Personalized medicine in oncology: Where have we come from and where are we going? *Pharmacogenomics* 14:931–939.
- Pisco A, et al. (2013) Non-Darwinian dynamics in therapy-induced drug resistance. *Nat Commun* 4:2467.
- Cirkel G, Gadellaa-van Hooijdonk C, Koudijs M, Willems S, Voest E (2014) Tumor heterogeneity and personalized cancer medicine: Are we being outnumbered? *Future Oncol* 10:417–428.
- Heindl A, Nawax S, Yuan Y (2015) Mapping spatial heterogeneity in the tumor microenvironment: A new era for digital pathology. *Lab Invest* 95:377–384.
- Verma M (2012) Personalized medicine and cancer. *J Pers Med* 2:1–14.

10. Ginsburg G, McCarthy J (2001) Personalized medicine: Revolutionizing drug discovery and patient care. *Trends Biotech* 19:491–496.
11. Emmert-Streib F (2012) Personalized medicine: Has it started yet? A reconstruction of the early history. *Front Genet* 3:313.
12. Gunawardena J (2010) *Models in Systems Biology: The Parameter Problem and the Meanings of Robustness*, eds Lodhi H, Muggleton S (Wiley, New York), Vol 1, pp 19–47.
13. Walter E, Pronzato L (1997) *Identification of Parametric Models from Experimental Data* (Springer, New York).
14. Varma A, Morbidelli M, Wu H (2005) *Parameter Sensitivity in Chemical Systems* (Cambridge Univ Press, New York).
15. Benzekry S, et al. (2014) Classical mathematical models for description and prediction of experimental tumor growth. *PLoS Comput Biol* 10:e1003800.
16. Allen R, Rieger T, Musante C (2016) Efficient generation and selection of virtual populations in quantitative systems pharmacology models. *CPT Pharmacometrics Syst Pharmacol* 5:140–146.
17. Barkai N, Leibler S (1997) Robustness in simple biochemical networks. *Nature* 387: 913–917.
18. Batchelor E, Goulian M (2003) Robustness and the cycle of phosphorylation and dephosphorylation in a two-component regulatory system. *Proc Natl Acad Sci USA* 100:691–696.
19. Thom R (1975) *Structural Stability and Morphogenesis* (W. A. Benjamin, Reading, MA).
20. Mallavarapu A, Thomson M, Ullian B, Gunawardena J (2009) Programming with models: Modularity and abstraction provide powerful capabilities for systems biology. *J R Soc Interf* 6:257–270.
21. Komarova N, Wodarz D (2010) Ode models for oncolytic virus dynamics. *J Theor Biol* 263:530–543.
22. Kansal A, Trimmer J (2006) Application of predictive biosimulation within pharmaceutical clinical development: Examples of significance for translational medicine and clinical trial design. *Syst Biol* 152:214–220.
23. Klinke D, II (2008) Integrating epidemiological data into a mechanistic model of type 2 diabetes: Validating the prevalence of virtual patients. *Ann Biomed Eng* 36:321–334.
24. Wang Y, et al. (2008) Leveraging prior quantitative knowledge to guide drug development decisions and regularly science recommendations: Impact of FDA pharmacometrics during 2004–2006. *J Clin Pharmacol* 48:146–156.
25. Jamei M, Dickinson G, Rostami-Hodjegan A (2009) A framework for assessing inter-individual variability in pharmacokinetics using virtual human populations and integrating general knowledge of physical chemistry, biology, anatomy, physiology and genetics: A tale of ‘bottom-up’ vs ‘top-down’ recognition of covariates. *Drug Metab Pharmacokin* 24:53–75.
26. Barter Z, et al. (2010) Determination of a quantitative relationship between hepatic CYP3A5*1/*3 and CYP3A4 expression for use in the prediction of metabolic clearance in virtual populations. *Biopharm Drug Dispos* 31:516–563.
27. Schmidt B, Casey F, Paterson T, Chan J (2013) Alternate virtual populations elucidate the type I interferon signature predictive of the response to rituximab in rheumatoid arthritis. *BMC Bioinf* 14:221.
28. Brown D, et al. (2015) Trauma in silico: Individual-specific mathematical models and virtual clinical populations. *Sci Trans Med* 7:285Ra61.
29. Teutonico D, et al. (2015) Generating virtual patients by multivariate and discrete re-sampling techniques. *Pharm Res* 32:3228–3237.
30. Valitalo P, et al. (2015) Novel model-based dosing guidelines for gentamicin and tobramycin in preterm and term neonates. *J Antimicrob Chemother* 70:2074–2077.
31. Huang JH, et al. (2010) Therapeutic and tumor-specific immunity induced by combination of dendritic cells and oncolytic adenovirus expressing IL-12 and 4-1BBL. *Mol Ther* 18:264–274.
32. Wong HH, Lemoine NR, Wang W (2010) Oncolytic viruses for cancer therapy: Overcoming the obstacles. *Viruses* 2:78–106.
33. Colombo M, Trinchieri G (2002) Interleukin-12 in anti-tumor immunity and immunotherapy. *Cytokine Growth Factor Rev* 13:155–168.
34. Kim P, Crivelli J, Choi IK, Yun CO, Wares J (2015) Quantitative impact of immunomodulation versus oncolysis with cytokine-expressing virus therapeutics. *Math Biosci Eng* 12:841–858.
35. Wares J, et al. (2015) Treatment strategies for combining immunostimulatory oncolytic virus therapeutics with dendritic cell injections. *Math Biosci Eng* 12:1237–1256.
36. Sontag E (1998) *Mathematical Control Theory: Deterministic Finite Dimensional Systems* (Springer, New York).
37. Schättler H, Ledzewicz U (2010) *Optimal Control for Mathematical Models of Cancer: An Application of Geometric Methods* (Springer, New York).
38. Shi J, Alagoz O, Erenay F, Su Q (2014) A survey of optimization models on cancer chemotherapy treatment planning. *Ann Oper Res* 221:331–356.
39. De Pillis L, Radunskaya A (2001) A mathematical tumor model with immune resistance and drug therapy: An optimal control approach. *J Theor Med* 2:79–100.
40. Ledzewicz U, Schättler H (2002) Optimal bang-bang controls for a two-compartment model in cancer chemotherapy. *J Optim Theory Appl* 114:609–637.
41. Świński A, Ledzewicz U, Schättler H (2003) Optimal control for a class of compartmental models in cancer chemotherapy. *Int J Appl Math Comput Sci* 13:357–368.
42. Agur Z, Hassin R, Levy S (2006) Optimizing chemotherapy scheduling using local search heuristics. *Oper Res* 54:829–846.
43. Nanda S, Moore H, Lenhart S (2007) Optimal control of treatment in a mathematical model of chronic myelogenous leukemia. *Math Biosci* 210:143–156.
44. Aïnseba B, Benosman C (2010) Optimal control for resistance and suboptimal response in CML. *Math Biosci* 229:81–93.
45. Engelhart M, Lebiez D, Sager S (2011) Optimal control for selected cancer chemotherapy ODE models: A view on the potential of optimal schedules and choice of objective function. *Math Biosci* 229:123–134.
46. Coldman A, Murray J (2000) Optimal control for a stochastic model of cancer chemotherapy. *Math Biosci* 168:187–200.
47. Vainas O, et al. (2012) Personalizing docetaxel and G-CSF schedules in cancer patients by a clinically validated computational model. *Br J Canc* 107:814–822.
48. Hofman A, Scherrer A, Küfer KH (2015) Analyzing the quality robustness of chemotherapy plans with respect to model uncertainties. *Math Biosci* 215:55–61.
49. Murray J (1997) The optimal scheduling of two drugs with simple resistance for a problem in cancer chemotherapy. *Math Med Biol* 14:283–303.
50. Gevertz J (2012) Optimization of vascular-targeting drugs in a computational model of tumor growth. *Phys Rev E* 85:041914.
51. Chang T, Mišić V (2013) Adaptive and robust radiation therapy optimization for lung cancer. *Eur J Oper Res* 231:745–756.
52. Liu W, Zhang X, Li Y, Mohan R (2012) Robust optimization of intensity modulated proton therapy. *Med Phys* 39:1079–1091.
53. Betts J, et al. (2015) Optimised robust treatment plans for prostate cancer focal brachytherapy. *Procedia Comput Sci* 51:914–923.
54. Badri H, Watanabe Y, Leder K (2016) Optimal radiotherapy dose schedules under parametric uncertainty. *Phys Med Biol* 61:338–364.
55. Chan T, Bortfeld T, Tsitsiklis J (2006) A robust approach to IMRT optimization. *Phys Med Biol* 51:2567–2583.
56. Bortfeld T, Chan T, Trofimov A, Tsitsiklis J (2008) Robust management of motion uncertainty in intensity-modulated radiation therapy. *Oper Res* 56:1461–1473.
57. Pan J, Fang K (2002) *Maximum Likelihood Estimation* (Springer, New York), pp 77–158.
58. Feldman J, Goldwasser R, Mark S, Schwartz J, Orion I (2009) A mathematical model for tumor volume evaluation using two-dimensions. *J Appl Quant Methods* 4: 455–462.
59. Ayers G, et al. (2010) Volume of preclinical xenograft tumors is more accurately assessed by ultrasound imaging than manual caliper measurements. *J Ultrasound Med* 29:891–901.
60. Kirk P, Stumpf P (2009) Gaussian process regression bootstrapping: Exploring the effects of uncertainty in time course data. *Bioinformatics* 25:1300–1306.
61. Lodhi H, Gilbert D (2011) *Bootstrapping Parameter Estimation in Dynamic Systems*, eds Elomaa T, Hollmén J, Mannila H (Springer, Berlin), pp 194–208.
62. Wang Z, Bordas V, Deisboeck T (2011) Identification of critical molecular components in a multiscale cancer model based on the integration of Monte Carlo, resampling, and ANOVA. *Front Physiol* 2:35.
63. Efron B (1979) Bootstrap methods: Another look at the jackknife. *Ann Stat* 7:1–26.
64. Arsenio J, Metz P, Chang T (2015) Asymmetric cell division in T lymphocyte fate diversification. *Trends Immunol* 36:670–683.
65. Gerritsen B, Pandit A (2016) The memory of killer T cell: Models of CD8+ T cell differentiation. *Immunol Cell Biol* 94:236–241.
66. Chang J, et al. (2007) Asymmetric T lymphocyte division in the initiation of adaptive immune responses. *Science* 315:1687–1691.
67. Metz P, et al. (2015) Regulation of asymmetric division and CD8+ T lymphocyte fate specification by protein kinase Czeta and protein kinase C-lambda/iota. *J Immunol* 194:2249–2259.
68. Reynolds J, Coles M, Lythe G, Molina-Paris C (2013) Mathematical model of naive T cell division and survival IL-7 thresholds. *Front Immunol* 4:434.
69. Chen Y, et al. (2001) CV706, a prostate cancer-specific adenovirus variant, in combination with radiotherapy produces synergistic antitumor efficacy without increasing toxicity. *Cancer Res* 61:5453–5460.
70. Jogler C, et al. (2006) Replication properties of human adenovirus in vivo and cultures of primary cells from different animal species. *J Virol* 80:3549–3558.
71. Li HL, et al. (2008) Pharmacokinetic and pharmacodynamic study of intratumoral injection of an adenovirus encoding endostatin in patients with advanced tumors. *Gene Ther* 15:247–256.
72. De Boer RJ, et al. (2001) Recruitment times, proliferation, and apoptosis rates during the CD8(+) T-cell response to lymphocytic choriomeningitis virus. *J Virol* 75:10663–10669.
73. de Pillis LG, Radunskaya AE, Wiseman CL (2005) A validated mathematical model of cell-mediated immune response to tumor growth. *Cancer Res* 65:7950–7958.

Supporting Information

Barish et al. 10.1073/pnas.1703355114

SI Results

Analysis of Best-Fit Parameters: 95% Credible Intervals. To better assess parameter variability across the full population, in the first step of VEPART, we use bootstrapping (63) to form 1,000 new samples (“bootstrap replicates”) that are of the same size as the original sample population. The aggregate data in each of these bootstrap replicates is fit to the mathematical model in Eqs. 1–6, giving a parametric description of each simulated cohort. Note that this process was repeated for each data set in Fig. 2, allowing us to approximate the posterior distributions on the parameters (Fig. S1) while fitting to a minimal subset of the parameters. From these posterior distributions (Fig. S1), we then calculate the 95% credible interval for each of the fit parameters (Table S2). For the parameter values that were fit in multiple models (c_A , c_{kill} , and c_T), the credible intervals presented were calculated from model 3, because model 3 is the one that gets inherited into the full model involving treatment with both OV_s and DC_s (model 4).

We find that the fit parameters can be divided into two classes. In the first class, we have the parameter values with relatively tight 95% credible intervals, meaning points in that interval do not vary by even an order of magnitude from the best-fit parameter value determined by fitting to the original dataset. The second case includes those parameters for which the 95% credible interval contains points that vary by orders of magnitude in comparison with the best-fit parameter value.

The fit parameters with tight credible interval are the net tumor growth rate (r), the virus infectivity rate (β), and the rate of T-cell stimulation by DC_s (χ_D). We find that values of r within the 95% credible interval can vary by, at most, 14% from the best-fit value. For β , variations can be no greater than 6.1%, and, for χ_D , they can be no greater than 81% (Table S2).

The other class of parameters can be classified as not having a tight credible interval. These parameters include the 4-1BBL-induced T-cell recruitment rate c_T , IL-12-induced naive T-cell recruitment rate c_A , and the enhanced T-cell cytotoxicity term c_{kill} . We find that values of c_T and c_A within the 95% credible interval can vary by three orders of magnitude from the best-fit value, whereas c_{kill} values can vary by four orders of magnitude (Table S2).

A possible interpretation of these results is that, for parameters with a tight credible interval (those that are tightly regulated in the population), the model is sensitive to the value of these parameters. Similarly, we could expect that the model is less sensitive to parameters with nontight credible intervals. These hypotheses have been confirmed through a local sensitivity analysis, as detailed below.

Analyzing Best-Fit Parameters: Local Sensitivity Analysis. To explore the hypothesis that model fit to the data is sensitive to the parameters with tight credible intervals, we performed a local sensitivity analysis on the corresponding model to identify all points within parameter space that give a fit within 10% of the optimal. The results for the parameters with tight credible intervals are shown in Fig. S2. This analysis confirms our hypotheses: Model fit is quite sensitive to parameters with tight credible intervals. β can vary by, at most, 0.7% and still give a fit within 10% of optimal (Fig. S2B), r can vary by, at most, 3% (Fig. S2A), and χ_D can only vary by 7% (maximum variation observed in Fig. S2C and D).

Taken together, the 95% credible intervals and the sensitivity analysis imply that we can have high confidence in our opti-

mal values of r , β , and χ_D . The posterior distributions determined via bootstrapping reveal that minimal variation from the predicted optimal would be found in the population (Table S2). The sensitivity analysis shows that, if these parameters are varied slightly from their optimal values, the model is no longer a good fit to the data.

As we did with the parameters with a tight credible interval, we also performed a local sensitivity analysis on the parameters without a tight credible interval. These three parameters happen to be simultaneously fit in the Ad/4-1BBL/IL-12 model (model 3). Therefore, this analysis will allow us to determine not only sensitivity to each of these parameters but also correlations between the parameters (Fig. S3). Correlations between these parameters would not be surprising, as each of them impacts the killing of cancer cells by cytotoxic T cells. Further, correlations between the parameters could partially explain why the 95% credible interval is so large for these parameters, as one parameter could compensate for another and still give a good fit to the data.

We find that the model is least sensitive to the c_A parameter, as it can vary by over 17,300% and still attain a goodness of fit value (S) within 10% of that attained by the optimal parameters (Fig. S3A). Conversely, c_T can only vary by 47.2%, and c_{kill} can only vary by 46.3% (Fig. S3), although these variations are much larger than those observed for the parameters with tight credible intervals. The 3D sensitivity plot shown in Fig. S3, along with the three 2D cross-sections, also visualizes the correlations between the parameters. Pairwise correlation analysis reveals a high degree of dependency between these parameters, and further emphasizes why the parameters fit to model 3 must be selected together when designing the virtual populations. The c_T and c_{kill} are the most highly correlated parameters, having a correlation coefficient of $\rho_{c_T, c_{kill}} = -0.9876$. The other parameters are also highly correlated, with c_A and c_{kill} having a correlation coefficient of $\rho_{c_A, c_{kill}} = -0.8376$, and c_T and c_A having a correlation coefficient of $\rho_{c_T, c_A} = -0.7869$. These data can be exploited in the future to ensure we have developed a minimally parameterized model with the minimal number of variables required to describe the data.

SI Computational Methods

Hierarchical Model Development. The hierarchical nature of the experimental data (increasing from simple to more complex) allowed the mathematical model presented in Eqs. 1–6 to be developed hierarchically (34, 35). Therefore, it is ideal to explain the model terms, and the parameter fitting, from the perspective of how each piece of the model builds off the simpler pieces.

Model 0: Untreated tumor growth. In the absence of any treatment, the model reduces to a single population of uninfected tumor cells, U . As shown in Eq. 1, it is assumed that these cells grow exponentially at rate r , an assumption supported by the control data set on the time scales of these experiments (Fig. 2A and B, black and white diamond curves). All other parameters in Eq. 1 are set to zero in the absence of treatment, making model 0 a simple exponential growth model for uninfected tumor cells.

Model 1: Ad only. In the absence of immunostimulation, the oncolytic adenovirus is capable of infecting uninfected cells, eventually resulting in infected cell lysis and the release of free virions into the tissue space. To expand the untreated tumor growth model to include the action of the OV_s, we include the free virion population (V) and the infected cell population (I). Free virions are injected into the system, and this is modeled

through the $u_v(t)$ term in Eq. 3. These virions infect uninfected cells at a rate of β , and infection is modeled using a frequency-dependent term. By dividing the infectivity rate β by the total population size (N), the model indirectly accounts for spatial effects. In particular, the rate of movement from the uninfected to infected pool is larger when U/N is large (as it is easier for the free virions to find the uninfected cells), and the rate of movement is smaller when U/N is small (as it is harder for the free virions to encounter uninfected cells). As shown in Eq. 2, it is assumed that the infected cells lyse at rate δ_I , and, when a cell lyses, α virions are released into the tumor environment (Eq. 3). It is assumed that virions decay exponentially at rate δ_V (Eq. 3). Therefore, model 1 includes only Eqs. 1–3, without the terms that consider the T-cell population; this is because we assume there are no T cells in the tumor tissue in the absence of immunostimulation, an observation supported by the experimental data in Huang et al. (31). Model 1 has been developed to describe the red square data shown in Fig. 2A.

Model 2a: Ad/4-1BBL. In model 2a, we expand upon model 1 to include the impact that the 4-1BBL transgene has on tumor killing. Experiments demonstrate that 4-1BBL increases T-cell infiltration at the tumor site, and increases the cytotoxicity of the T cells (31), and therefore we aim to include these phenomena in model 2a. However, instead of directly modeling the immunostimulatory effects of 4-1BBL, we use the fact that this gene is transcribed and translated inside of infected cells, and therefore the infected cell population can be used as a proxy for modeling the actions of 4-1BBL (34, 35). We thus model 4-1BBL-induced T-cell production as proportional to the infected cell population with production/recruitment rate c_T , as shown in the one new equation that gets introduced as we move from model 1 to model 2a: Eq. 4. T cells are assumed to kill tumor cells independent of their infected status (that is, the T cells recognize tumor antigens and not viral antigens) using a frequency-dependent term with rate of killing given by $k_0 + c_{kill}I$. This rate of T-cell-induced killing accounts for the fact that T cells have some baseline activity against the tumor cells (k_0), and that the cytotoxicity of the T cells increases in the presence of 4-1BBL (modeled indirectly through the $c_{kill}I$ term), an assumption supported by experimental data (31). The only other dynamics the T cells have in model 2a is their natural decay. Therefore, model 2a includes all of Eqs. 1–3, along with the T-cell source term and decay term in Eq. 4. Model 2a has been developed to describe the green triangle data shown in Fig. 2A.

Model 2b: Ad/IL-12. Model 2b also builds upon model 1 (note that it does not build upon model 2a), to include the impact that the IL-12 transgene has on tumor killing. Experiments demonstrate that, like 4-1BBL, IL-12 increases T-cell infiltration at the tumor site, while also increasing the cytotoxicity of those T cells (31). However, from what is known about the mechanism of IL-12 action, this increased infiltration of the cytotoxic T cells is indirectly driven by IL-12 through the naïve T-cell population. As with 4-1BBL, IL-12 will not be modeled directly, but will be assumed to be proportional to the infected cell population.

To model the effects of IL-12 through the naïve T-cell population, model 1 is expanded upon to include two new equations. First, the naïve T-cell population is introduced in Eq. 5, and it is assumed that IL-12 (through the infected cell population) recruits naïve T cells at rate c_A . These naïve T cells are assumed to divide asymmetrically (64), meaning that at the same time a naïve T cell is lost, a new one is gained; the result is that there is no net change in the naïve T-cell population due to asymmetric division. However, at the same time, a cytotoxic T cell is also produced at rate χ_A , and this is represented in the second population that gets introduced as we go from model 1 to model 2b: the cytotoxic T-cell population. Note that the timing and mechanisms underlying the division and differentiation of naïve T cells is still being actively studied (65). As our experimental data only

describe temporal changes in tumor volume, and give us no information about the differentiation pattern of T cells, we assumed asymmetric division, as it is consistent with some experimental observations (64, 66, 67). As an added benefit, this assumption minimizes the number of parameters we need to model naïve T-cell division. Others have considered modeling the division of naïve T cells in more detail; see, for instance, ref. 68.

Just as in model 2a, the cytotoxic T cells can indiscriminately kill tumor cells, and a frequency-dependent kill term with rate $k_0 + c_{kill}I$ is used. As before, k_0 represents the baseline kill rate of the T cells. However, in this case, c_{kill} represents the enhanced cytotoxicity of the T cells due to the presence of IL-12, not 4-1BBL. The only other dynamics the naïve and cytotoxic T cells have in model 2b is their natural decay. Therefore, model 2b includes all of Eqs. 1–5, with the exception that Eq. 4 does not include the first term (no direct recruitment of cytotoxic T cells by infected cells) or the last term (no DCs D are injected). Model 2b has been developed to describe the blue diamond data shown in Fig. 2A.

Model 3: Ad/4-1BBL/IL-12. No new terms were introduced to model treatment with Ads expressing 4-1BBL and IL-12. Instead, model 3 is simply the union of models 2a and 2b, and therefore consists of all of Eqs. 1–5, with the exception that Eq. 4 does not include the second-to-last term, as there are no DCs used during this treatment protocol. It is of note that, in model 3, the c_{kill} parameter now represents the enhanced cytotoxicity of the T cells due to the effects of both 4-1BBL and IL-12. Model 3 has been developed to describe the black circle data shown in Fig. 2A, and the red square data shown in Fig. 2B.

Model 4: Ad/4-1BBL/IL-12 and DCs. Model 4 introduces the injected DC population. These DC vaccines are developed by harvesting DCs from the bone marrow of tumor-bearing mice and allowing them to be pulsed ex vivo with tumor associated antigens for 8 d until maturation (31). These tumor-primed DCs are then injected directly into the tumor site. It is these DCs that we seek to model. Although we recognize that there may be some endogenous DCs at the tumor site, we work under the assumption that the effect of such DCs is wrapped up in the tumor growth rate term r in the absence of any treatment. We are particularly interested in modeling the impact of the treatment protocol, and hence only focus on the injected DCs. Again aiming to keep the model simple, only one new equation is introduced, and it is the equation for the injected DC population. As shown in Eq. 6, there is a source term for injected DCs, and the DCs natural decay rate is modeled. The DCs stimulate/recruit tumor-targeted cytotoxic T cells at a rate of χ_D , which is the only other new term as we expand from model 3 to model 4 (it accounts for the $\chi_D D$ term in Eq. 4). Model 4 has been developed to describe the blue triangle data shown in Fig. 2B. Note that a subcase of model 4 in which all virus-related parameters are set to zero gives us the case of treating with only the DC vaccine; this describes the green circle data shown in Fig. 2B.

Hierarchical Model Fitting to Experimental Data. Fig. S4 summarizes the hierarchical nature of the parameter fitting process. Fitting begins with model 0 (no treatment), in which the intrinsic tumor growth rate r was fit to the aggregate experimental data. Unless otherwise stated, for the purposes of this study, r was kept at the value determined previously in ref. 35. The hierarchical nature of parameter fitting implies that this value of r is not only used in model 0 but is used in all of the models that build off of model 0 (models 1 through 4).

Model 1 (Ad only) introduced four new parameter values, three of which (the infected cell lysis rate, number of virions released by a lysed cell, and virion decay rate) can be readily estimated from experiments (35). As the r parameter value is inherited from model 0, this leaves only two parameters to fit in model 1: the viral infectivity term β and the initial tumor size U_0 . The

initial conditions for all other variables are set to zero, meaning we assume there is no naïve or cytotoxic T-cell infiltration at the tumor site before treatment, a claim supported by the data (31). As U_0 is a function of the dataset and not the model itself, naturally, this parameter must be refit to each dataset, and therefore never gets inherited from one model to the next.

Focusing on β , the other undetermined parameter value in model 1, the best-fit value is attained using simulated annealing, and this value gets inherited into all subsequent models. The parameter inheritance scheme falls apart between models 2 and 3. In model 2a (Ad/4-1BBL), the noninitial condition parameters that are fit to the data are the enhanced cytotoxicity of T cells due to immunostimulation by 4-1BBL, c_{kill} , and the activation rate of T cells by 4-1BBL, c_T . The c_{kill} parameter cannot be inherited into model 2b, as, in model 2b, it represents the enhanced cytotoxicity due to IL-12 (not 4-1BBL), and it also cannot be inherited into model 3, as, in model 3, we need to consider how 4-1BBL and IL-12 work together to enhance the T-cell cytotoxicity. The c_T parameter also cannot be inherited into model 3, as different viral doses were used in the experiments giving Ad/4-1BBL compared with experiments giving Ad/4-1BBL/IL-12 (31).

In model 2b (Ad/IL-12), the noninitial condition parameters that are fit to the data are c_{kill} and the naïve T-cell recruitment rate, c_A . Similar to what occurs with model 2a, the best-fit parameters are not inherited from model 2b to model 3. The reason c_{kill} cannot be inherited is the same as above. The c_A parameter cannot be inherited into model 3 because IL-12 expression was found to differ from the Ad/IL-12 case to the Ad/4-1BBL/IL-12 case (31).

In model 3 (Ad/4-1BBL/IL-12), two datasets (for two viral doses) are simultaneously fit to determine the following noninitial condition parameters: c_{kill} , c_T , and c_A . Because these best-fit parameters values are properties of the virus with both 4-1BBL and IL-12, these parameters are also subsequently inherited into model 4 (Ad/4-1BBL/IL-12 + DCs). The aggregate data with DCs only and the data set for Ad/4-1BBL/IL-12 with DCs are finally simultaneously fit to the one noninitial condition parameter: the rate at which DCs stimulate the production of T cells, χ_D .

Details on Fitting Scheme: Simulated Annealing. Simulated annealing is a stochastic optimization scheme where parameters are changed through random perturbations until, ideally, the globally optimal parameters are found. For this application, the algorithm seeks to minimize the objective function S , as defined in Eq. 7.

The simulated annealing algorithm starts with an initial set of parameters, determines the value of S at those parameters, and then randomly perturbs those parameters (see Δ_{max} in Table S3) to generate a temporary set of new parameters. Looking at an arbitrary iteration of simulated annealing, the value of the

objective function is compared across two parameter sets (the last accepted set of parameters, and the new “temporary” set of parameters that is a random perturbation of the current set) to compute the change in the objective function, ΔS . Whether the new parameter set gets accepted or rejected is probabilistically determined using the Metropolis acceptance rule (50),

$$p = \begin{cases} 1, & \Delta S \leq 0 \\ \exp\left(-\frac{\Delta S}{T}\right), & \Delta S > 0, \end{cases} \quad [\text{S1}]$$

where T is a fictitious “temperature.”

Eq. S1 says that, if the new parameter set fits better than the prior one ($\Delta S < 0$), the new parameter set is accepted (as being closer to the best-fit parameters). If the new parameter set does not fit as well as the prior one ($\Delta S > 0$), there is a nonzero probability of accepting this new parameter set (en route to finding the best fit parameters). Accepting “uphill” parameter changes in this way is essential for simulated annealing to avoid getting trapped at local minimum. T is chosen to be a monotonically decreasing function that approaches zero as the number of annealing steps k increases (Table S3). In this way, the likelihood of accepting parameters that give a worse fit to the data decreases as simulated annealing progresses.

For each dataset in Fig. 2 (control, Ad only, etc.), the appropriate model was fit to the data using simulated annealing. Through experimentation with the data, we arrived at the following sufficient stopping criterion for the algorithm: The best-fit parameters have been found when K_{stop} new parameter sets, generated from random perturbations of a previously accepted parameter set, in a row have been rejected ($K_{stop} = 8,000$, as detailed in Table S3). However, using this stopping criterion, the optimal solution arrived at through simulated annealing could potentially depend on the chosen initial conditions. To avoid this issue, simulated annealing was run five times per data set ($N_{sim} = 5$ in Table S3), with each run of simulated annealing starting with a different set of initial conditions. In particular, the initial conditions were selected by randomly perturbing the initial parameter values shown in Table S3 by, at most, $\Delta IC = 50\%$. The best-fit parameter set was selected as the one among the five runs that gave the absolute minimum value of the objective function.

Analyzing Best-Fit Parameters: Local Sensitivity Analysis. For each of the models, we also performed a local sensitivity analysis to complement the results of the credibility interval analysis. This analysis allows us to study correlations between parameters, and how much parameter values can vary from the best-fit values and still give a reasonable fit to the average data. In particular, we performed an exhaustive search in parameter space about the best-fit values to the aggregate data for each model; this allows us to identify all parameter sets that give a fit within 10% of optimal to the original data (as measured by the value of the objective function, S).

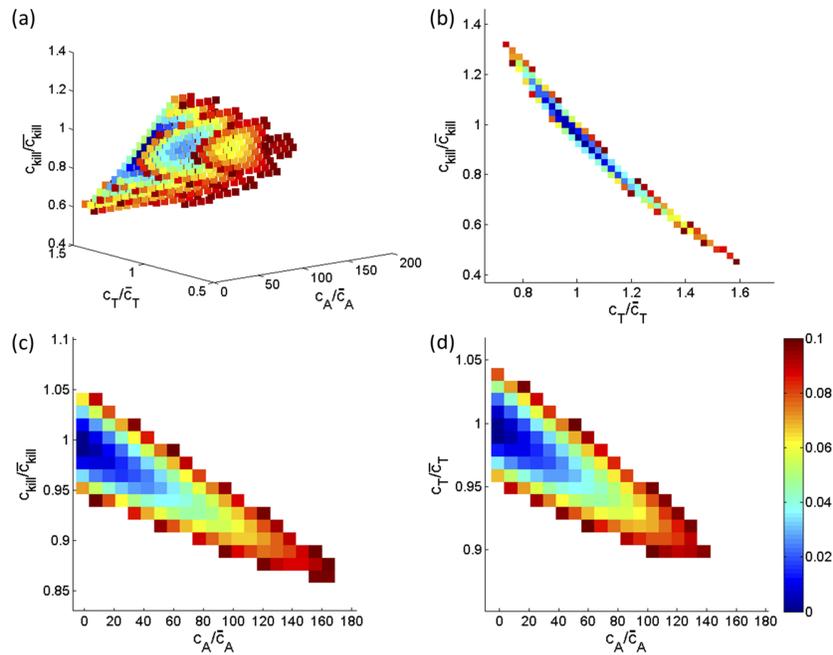


Fig. S3. Local sensitivity analysis about optimal parameters in model 3, fixing U_0 at its optimal value. Note that the values of the parameters shown on each axis are scaled by the optimal value (indicated with a bar). (A) Sensitivity in c_T - c_A - c_{kill} space. (B) Sensitivity in c_T - c_{kill} space, with c_A fixed at its optimal value. (C) Sensitivity in c_A - c_{kill} space, with c_T fixed at its optimal value. (D) Sensitivity in c_A - c_T space, with c_{kill} fixed at its optimal value.

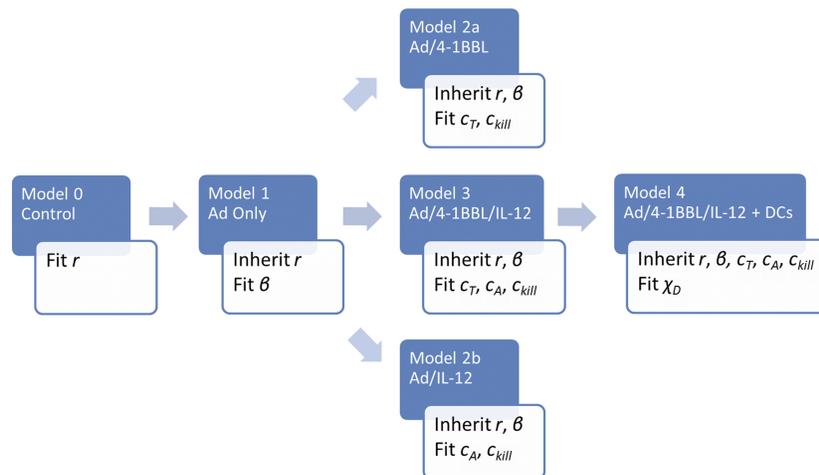


Fig. S4. Summary of hierarchical noninitial condition parameter fitting.

Table S1. Parameter values set based on literature estimates before the fitting process

Parameter	Description	Value	Ref.
α	Viral production	3,000 virions	69
δ_I	Infected lysis rate	1 d ⁻¹	70
δ_V	Viral decay rate	2.3 d ⁻¹	71
δ_T (δ_A)	T-cell (naïve) decay rate	0.35 d ⁻¹	72
k_0	Base T-cell killing rate	2 d ⁻¹	35, 73
χ_A	T-cell differentiation rate	1 d ⁻¹	35

Table S2. Best-fit parameters using the original fitting algorithm (Orig) and the fitting algorithm described herein (Curr)

Parameter	Description	Best fit (Curr)	Best fit (Orig)	95% credible interval	Model
r	Net tumor growth rate, d^{-1}	N/A	$\bar{r} = 0.3198$	$[0.928\bar{r}, 1.135\bar{r}]$	0
β	Infection rate d^{-1}	$\bar{\beta} = 0.00100854$	0.00100843	$[0.939\bar{\beta}, 1.050\bar{\beta}]$	1
c_A	Naïve T-cell recruitment rate, d^{-1}	$\bar{c}_A = 0.000517$	0*	$[0.542\bar{c}_A, 3321\bar{c}_A]$	3
c_T	T-cell recruitment rate, d^{-1}	$\bar{c}_T = 1.6984$	1.1707 [†]	$[0.005\bar{c}_T, 2.445\bar{c}_T]$	3
c_{kill}	Enhanced T-cell cytotoxicity, per cell per day	$\bar{c}_{kill} = 5.954 \times 10^{-7}$	5.1×10^{-7} *	$[1.7 \times 10^{-4}\bar{c}_{kill}, 1.895\bar{c}_{kill}]$	3
χ_D	Rate of T-cell simulation by DCs, d^{-1}	$\bar{\chi}_D = 4.6754$	5.5388	$[0.856\bar{\chi}_D, 1.803\bar{\chi}_D]$	4

Also shown is the 95% credible interval for each parameter, as determined by the bootstrapping analysis. The model in which the parameters were fit and bootstrapping was performed is shown in the last column.

*These parameters were not fit with high resolution: To handle the numerical limitations of fitting too many parameters using the original fitting routine, these parameters were optimized by confining the parameter to be within a fixed interval, and looking at discrete values (over a fixed step size) within that interval.

[†]This parameter was simultaneously fit under the constraints imposed on the * parameters. On the other hand, no constraints were imposed when fitting using the routine described in this work.

Table S3. Simulated annealing parameters

Parameter	Definition	Value
Δ_{max}	Maximum random perturbation that can be made to a parameter	$\Delta_{max} = 0.1$
T	“Cooling” schedule/temperature	$T = 43(0.96)^k$, k is annealing step
K_{stop}	Number of rejected random perturbation (annealing steps) until convergence is declared	$K_{stop} = 8,000$
N_{sim}	Number of times simulated annealing procedure is repeated per dataset	$N_{sim} = 5$
$\beta^{(0)}$	Baseline initial parameter guess for β for fitting model 1	$\beta^{(0)} = 0.1$
$c_T^{(0)}$	Baseline initial parameter guess for c_T	Model 2a: $c_T^{(0)} = 0.01$ Model 3: $c_T^{(0)} = 1$
$c_{kill}^{(0)}$	Baseline initial parameter guess for c_{kill}	Model 2a: $c_{kill}^{(0)} = 0.03$ Model 2b: $c_{kill}^{(0)} = 0.1$ Model 3: $c_{kill}^{(0)} = 0.45$
$c_A^{(0)}$	Baseline initial parameter guess for c_A	Model 2b: $c_A^{(0)} = 1$ Model 3: $c_A^{(0)} = 2$
$\chi_D^{(0)}$	Baseline initial parameter guess for χ_D in model 4	$\chi_D^{(0)} = 5$
$U_0^{(0)}$	Baseline initial parameter guess for U_0	Model 1: $U_0^{(0)} = 60$ Model 2a: $U_0^{(0)} = 62.4$ Model 2b: $U_0^{(0)} = 55$ Model 3 (Fig. 1A): $U_0^{(0)} = 59.3$ Model 3 (Fig. 1B): $U_0^{(0)} = 30$ Model 4 (for both protocols): $U_0^{(0)} = 60$
Δ_{IC}	Maximum percent variation of initial simulated annealing parameters from baseline	$\Delta_{IC} = 50\%$