



Published in final edited form as:

*Nat Biomed Eng.* 2021 April ; 5(4): 297–308. doi:10.1038/s41551-020-00662-0.

## A mathematical model for the quantification of a patient's sensitivity to checkpoint inhibitors and long-term tumour burden

Joseph D Butner<sup>1,†</sup>, Zhihui Wang<sup>1,2,†,\*</sup>, Dalia Elganainy<sup>3</sup>, Karine A Al Feghali<sup>3</sup>, Marija Plodinec<sup>4</sup>, George A Calin<sup>5</sup>, Prashant Dogra<sup>1</sup>, Sara Nizzero<sup>1</sup>, Javier Ruiz Ramírez<sup>1</sup>, Geoffrey V Martin<sup>3</sup>, Hussein A Tawbi<sup>6</sup>, Caroline Chung<sup>3</sup>, Eugene J Koay<sup>3</sup>, James W Welsh<sup>3</sup>, David S Hong<sup>7</sup>, Vittorio Cristini<sup>1,2,\*</sup>

<sup>1</sup>Mathematics in Medicine Program, Houston Methodist Research Institute, Houston, Texas 77030, USA

<sup>2</sup>Department of Imaging Physics, University of Texas MD Anderson Cancer Center, Houston, Texas 77230, USA

<sup>3</sup>Department of Radiation Oncology, University of Texas MD Anderson Cancer Center, Houston, Texas 77030

<sup>4</sup>Biozentrum and the Swiss Nanoscience Institute, University of Basel, Basel 4056, Switzerland

<sup>5</sup>Department of Experimental Therapeutics, University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

<sup>6</sup>Department of Melanoma Medical Oncology, University of Texas MD Anderson Cancer Center, Houston, Texas 77030

<sup>7</sup>Department of Investigational Cancer Therapeutics, University of Texas MD Anderson Cancer Center, Houston, Texas 77230, USA

### Abstract

A large proportion of patients with cancer are unresponsive to treatment with immune checkpoint blockade and other immunotherapies. Here, we report a mathematical model of the time-course of tumour responses to immune-checkpoint inhibitors. The model takes into account intrinsic tumour-growth rates, the rates of immune activation and of tumour-immune-cell interactions, and the efficacy of immune-mediated tumour killing. For 124 patients, four cancer types and two immunotherapy agents, the model reliably described the immune responses and final tumour

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:[http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

\*Corresponding authors, [zwang@houstonmethodist.org](mailto:zwang@houstonmethodist.org), [vcristini@houstonmethodist.org](mailto:vcristini@houstonmethodist.org).

†These authors contributed equally to this work.

Author contributions

J.D.B implemented the model and collected the data. Z.W., V.C. initiated the project, participated in data collection, and supervised model development and analysis. J.D.B., Z.W., M.P., P.D., V.C. performed model analysis. D.E., K.A.F., G.V.M., C.C., E.J.K., J.W.W., D.S.H. participated in data collection. J.D.B., Z.W., D.E., K.A.F., M.P., G.A.C., P.D., S.N., J.R.R., G.V.M., H.A.T., C.C., E.J.K., J.W.W., D.S.H., V.C. interpreted the data and model results. J.D.B., Z.W., M.P., E.J.K., V.C. wrote the manuscript.

Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41551-01X-XXXX-X>

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

burden across all different cancers and drug combinations examined. In validation cohorts from four clinical trials of checkpoint inhibitors (with a total of 177 patients), the model accurately stratified the patients according to reduced or increased long-term tumour burden. We also provide model-derived quantitative measures of treatment sensitivity for specific drug–cancer combinations. The model can be used to predict responses to therapy and to quantify specific drug–cancer sensitivities in individual patients.

---

Immunotherapy has shown great promise in the fight against cancer. Activation of the immune system against a patient’s tumours through various mechanisms, including the administration of cytokines or vaccination to increase desired T-cell counts, the administration of immune checkpoint inhibitor agents, of co-stimulator receptor agonists, of oncolytic viruses, or of adoptive-activated T-cell and natural-killer-cell therapy, has been shown to stimulate favourable cytotoxic effector-cell activation against tumour cells<sup>1,2</sup>. By activating the body’s own immune system, this next-generation set of cancer therapies offers the potential for highly targeted cancer treatment with significant clinical efficacy, reduced negative side effects (relative to traditional cancer therapies), and improved treatment of some types of cancer that are known to have poor responses to other treatment methods. For example, the checkpoint inhibitor ipilimumab (alone or in combination with nivolumab) has shown great success in treating late-stage melanoma, a disease that previously had high mortality rates and poor response to traditional treatment strategies<sup>2,3</sup>, and atezolizumab has shown strong potential for the treatment of urothelial cell carcinoma (UCC)<sup>4</sup>.

Despite significant advances, immunotherapy still presents notable challenges. Immunotherapy has been shown to be highly effective against certain types of cancer, while at the same time ineffective at treating others, with over 50% of all patients failing to respond to immunotherapy treatment<sup>5</sup>. In patients that do respond, it often occurs more slowly than when more traditional treatment regimens are used, resulting in new challenges towards identification of patients with favourable prognosis (or, perhaps of greater clinical urgency, identification of patients that are not responding and require a different clinical approach)<sup>6</sup>. To date, several promising biomarkers for immunotherapy response have been identified, including intratumoural programmed death ligand 1 (PD-L1) expression levels and immune presence (e.g., tumour infiltrating lymphocytes<sup>7,8</sup>), blood cell counts (e.g., neutrophils, lymphocytes, and CD8<sup>+</sup> T cells<sup>9–11</sup>), serum markers (e.g., LDH, VEGF, and IL-6<sup>12–14</sup>), and genetic mutation markers<sup>15,16</sup>. However, these biomarkers are often specific to a certain family of drug and disease combinations, presenting a significant clinical challenge due to the potential necessity for extensive and invasive diagnostic testing. For example, PD-L1 levels and genetic mutations may vary significantly within the same tumour and between primary and secondary tumours<sup>3,17,18</sup>, and blood cell and serum markers are often specific to certain drug or cancer types<sup>14,19</sup>. As such, identification of reliable prognostic biomarkers of immunotherapy response that may be broadly applied across many cancer and treatment types to provide timely identification of the most effective immunotherapy treatment protocol remains a significant clinical need.

Clinically, lesion response is often assessed using the RECIST v1.1 criteria<sup>20</sup> to evaluate the change in lesion volume after treatment initiation, as quantified by the sum of indexed

lesion diameters measured at a lesion's widest cross-section. Under these criteria, patient response is categorized as 1) progressive disease: > 20% summed diameter increase; 2) partial response: > 30% summed diameter decrease; 3) stable disease: sum of diameters remains stable, crossing neither threshold for progressive disease or partial response; and 4) complete response: lesion disappearance (longest diameter < 5mm)<sup>20</sup>. These definitions, while useful, are limited to only the *spatial* scale while neglecting the *time* scale, which may add valuable resolution in prognostic decisions and may be further indicative of the aggressiveness of the disease or the efficacy of the treatment and associated disease response (e.g., how fast the tumour grows or shrinks). These limitations, combined with the previously discussed limitations of current biomarkers for predicting immune response, further suggest the need for improved patient stratification methodologies.

Mathematical modelling has been extensively used in many areas of cancer research, including disease initiation, progression, and treatment (see<sup>21–28</sup> for further discussions). Likewise, mathematical modelling of immunotherapy has also demonstrated value through the prediction of treatment efficacy and validation with both experimental and clinical data, using many different modelling techniques, including time dependent continuum descriptions (ordinary differential equations)<sup>29</sup>, hybridized spatial and temporal continuum descriptions (partial differential equations)<sup>30</sup>, and agent-based modelling<sup>31–33</sup>. See Supplemental Materials for discussion of some interesting immunotherapy modelling work; for a more exhaustive review of these and other exciting advances, the interested reader is referred to<sup>34,35</sup>. Efforts are being undertaken with the goal of stratifying patient outcome from immunotherapy treatment based on important biological parameters such as neoantigen fitness<sup>36</sup> and autophagy compartments<sup>37</sup>. However, these modelling approaches rely on specific cell-level information that must be obtained invasively, which is not always easy or cost-effective to obtain for rapid transition to the clinic as a prognostic tool. Moreover, they are dependent on measured heterogeneities at the cell or subcellular scale within the tumour that may not be easily determined or accurately represented by minimally invasive clinical techniques (e.g., needle biopsy, which may only provide information at a specific region within a single tumour, and may not accurately represent tumour or system-wide heterogeneities)<sup>38</sup>. Unfortunately, mathematical modelling has yet to emerge as a powerful, predictive bedside tool to assist the physician, as these modelling efforts face significant challenges for transition to bedside use.

To address the pressing need for better assessment of patient outcome under immunotherapy, we have developed a mechanistic mathematical model for describing patient tumour burden over time under checkpoint-inhibitor treatment; details of the model can be found in Methods. We present here the parameterization of the model against data obtained from the literature using anti PD-1/PD-L1 checkpoint inhibitors ( $n = 124$ ), which is further validated against patient cohorts from four in-house clinical trials (NCT02239900,  $n = 93$ , using the anti-CTLA4 agent ipilimumab; a basket-study patient cohort ( $n = 28$ ); NCT02444741,  $n = 35$  patients with non-small cell lung carcinoma (NSCLC) treated with pembrolizumab; and lastly a cohort from the Institutional Review Board (IRB)-approved MD Anderson Melanoma Database, MelCore, of  $n = 21$  patients with melanoma metastatic to the brain treated with ipilimumab, nivolumab, or a combination of both; see Methods) in order to gain quantitative, mechanistic insights into effective immunotherapy. It is worth noting that

the specific approach our model has taken offers the valuable advantage of using only measured tumour dimensions from CT/MR imaging data (although we are also currently developing histological and blood marker surrogates for model calibration), which is already obtained non-invasively as part of the standard of care protocol, and that this model is not limited to the unique molecular mechanism of a specific checkpoint. This will help facilitate the process of translating the model into a clinically useful prospective tool for predicting patient outcome. Here, we briefly describe the model derivation and its key equations, and then focus on the quantification and validation of the model for predicting long-term patient responses to immunotherapy using only readily measurable diagnostic information. Our results provide evidence that two model parameters that quantify 1) a combination of the anti-tumour immune state and tumour cell kill rates and 2) tumour growth rate at first restaging (used individually or in combination) may function as universal biomarkers of immunotherapy response, independently of specific disease or drug.

## Results

Details of the literature-derived patient cohort for model calibration and our in-house clinical trial patient cohorts for model validation are provided in Methods. A flowchart of the model calibration (including model fitting and parameterization) and validation processes is provided in Fig. S1. An example subset of model fitting results (chosen from all four tumour types in the literature-derived calibration cohort to show a wide range of possible tumour burden changes under immunotherapy and the associated model behaviour) is presented in Fig. 1.

### Model parameters stratify patient response

Figure 2 shows the distributions of  $\lambda \times \mu$  and  $\alpha_{\text{imaging}}$  for all patients measured. We note that in this paper, we will refer to patients with last reported *measured* (via CT or MRI) and *normalized* (by tumour burden at start of treatment, e.g.,  $t = 0$ ) tumour burden (e.g.,  $\rho_{\text{measured}}^{\infty}$ ) larger than baseline ( $\rho_{\text{measured}}^{\infty} > 1.0$ ) as *non-responders*, and patients with reduced total tumour burden ( $\rho_{\text{measured}}^{\infty} < 1.0$ ) as *responders*; thus summing all RECIST v1.1<sup>20</sup> classifications into only two possible categories for the purposes of our analysis. Under these definitions, responder patients in the literature-derived calibration cohort (58 of 124 patients) were found to have significantly higher values of  $\lambda \times \mu$  but lower values of  $\alpha_{\text{imaging}}$  (Fig. 2A,B;  $P < 0.001$  by two-tail Wilcoxon rank sum) compared to non-responders (66 of 124 patients). This trend was confirmed in our IPI clinical trial cohort obtained from [NCT02239900](#) (Fig. 2E,F), the NSCLC clinical cohort from [NCT02444741](#) (Fig. 2G,H), and the brain mets clinical cohort (Fig. 2I,J); however, in the basket-study cohort, while  $\alpha_{\text{imaging}}$  showed significant separation between response types,  $\lambda \times \mu$  did not show significant separation (Fig. 2C,D; this is likely due to high patient heterogeneity and small sample size in this cohort). Consistent results were also observed with  $\rho_{\text{predicted}}^{\infty}$  ( $P < 0.01$  for all patient cohorts; this is shown in Fig. S6).

ROC (i.e., receiver operator characteristic) analysis showed high accuracy of patient separation between response types in all patient cohorts examined (Fig. 3, Tables 1, S5) using both  $\lambda \times \mu$  and  $\alpha_{\text{imaging}}$  as binary classifiers (this was also found to be true for

$\rho_{\text{predicted}}^{\infty}$ , see Fig. S7). In the literature-derived calibration cohort, parameter  $\lambda \times \mu$  was found to have unique optimal cutoff values for each cancer-drug combination examined (determined by maximizing Youden's J statistic), while optimal cutoff for  $\alpha_{\text{imaging}}$  was consistently found to be 0.0 (having a biological meaning of tumour growing or shrinking at first restaging). Based on these findings, the cutoff value for  $\lambda \times \mu$  was determined individually for all validation cohorts as described above, while the cutoff value for  $\alpha_{\text{imaging}}$  was assigned to be 0.0. This result was supported in all validation sets, where ROC analysis revealed a cutoff value of 0.0 for  $\alpha_{\text{imaging}}$  significantly separated patient response ( $P = 0.0095$  for the basket-study cohort, with overall accuracy 0.821; and  $P < 0.001$  for the IPI cohort from [NCT02239900](#) with overall accuracy 0.849,  $P < 0.001$  with overall accuracy for the NSCLC clinical cohort [NCT02444741](#) of 0.800, and  $P = 0.0491$  with overall accuracy for the brain mets cohort of 0.714; Figs. 2, 3, Table 1, S5). Uniquely identified cutoff values for  $\lambda \times \mu$  also showed high accuracy of separating patient response in these validation cohorts, although with slightly reduced accuracy relative to  $\alpha_{\text{imaging}}$  (basket-study cohort: accuracy = 0.714; IPI clinical trial cohort: accuracy = 0.763; NSCLC clinical cohort: accuracy = 0.743; brain mets clinical cohort: accuracy = 0.810; Table S5).

### Model predictions of treatment outcome correlate with patient data

Model parameter  $\rho_{\text{predicted}}^{\infty}$  (projected long-term tumour burden) was found to significantly correlate with the final measured tumour burden ( $\rho_{\text{measured}}^{\infty}$ ) for all tumour and drug types together (Fig. 4A;  $P < 0.001$ ,  $Pearson = 0.723$ ), as well as individually for each cancer and drug type (Fig. 4B,C). This allows us to use  $\rho_{\text{predicted}}^{\infty}$  as a reliable surrogate for measured tumour at last restaging in our analysis. Correlation analysis between retrospectively determined tumour growth rate after immunotherapy treatment ( $\alpha_{\text{retro}}$ ) and the prospective measure of tumour growth rate calculated at first restaging ( $\alpha_{\text{imaging}}$ ) revealed similar results, with significant correlation between all values within the calibration cohort (Fig. 4D;  $P < 0.001$ ,  $Pearson = 0.762$ ), as well as when isolated on a per-tumour and per-drug basis (Fig. 4D–F). Accordingly, we may use our prospective measure  $\alpha_{\text{imaging}}$  as a reasonable surrogate for the long-term value  $\alpha_{\text{retro}}$ . The interested reader may also refer to Supplemental Materials for additional correlation analysis between  $\alpha_{\text{imaging}}$  and  $\rho_{\text{predicted}}^{\infty}$  (Fig. S8).

### Model parameters quantify treatment sensitivity for each drug-disease combination examined

In order to quantify the sensitivity of drug-induced immune response as a function of drug-disease combination, linear regression analysis of long-term tumour burden  $\rho_{\text{predicted}}^{\infty}$  against the inverse of our measure of strength of immune response  $1/(\lambda \times \mu)$  was performed for all drug and disease subsets within the literature-derived cohort, revealing unique sensitivity to drug treatment for each drug-disease combination examined. These are shown in Fig. 5A–C (note that all show the same data, but are colour coded by A) response, B) disease, and C) drug types). Larger slopes indicate greater sensitivity of that specific disease to the drug administered; i.e., with high sensitivity, smaller changes in  $1/(\lambda \times \mu)$  result in larger changes in tumour growth or reduction ( $\alpha_{\text{retro}}$ ) than in cases with lower sensitivity. In

all cases, to be clinically relevant, we set the  $y$ -intercept of the linear best fit to zero, as total long-term tumour burden ( $\rho_{\text{predicted}}^{\infty}$ ) goes to zero when tumour kill is complete (i.e.,  $\frac{a_{\text{retro}} - \mu}{\Lambda \times \mu} = -1$ , equation (3)); thus our measure of sensitivity represents the magnitude of the change in post-treatment tumour growth rate ( $a_{\text{retro}}$ ) as a function of change in strength of immunotherapy-mediated tumour response ( $\Lambda \times \mu$ ). We thus estimated the relationship between  $\rho_{\text{predicted}}^{\infty}$  and  $1/(\Lambda \times \mu)$  in the form of  $y = kx$ , where slope  $k$  represents how sensitive the treatment outcome (i.e.,  $\rho_{\text{predicted}}^{\infty}$ ) is to immunotherapy in each case; values for  $k$  and associated  $R^2$  for each fit are shown in Table 2. Finally, we repeated this analysis using the exact tumour burden measured at last restaging ( $\rho_{\text{measured}}^{\infty}$ ) on the  $y$ -axis; however the strength of correlation was further reduced in this case (Fig. 5D–F, Table 2).

We find that, with respect to cancer type in the literature-derived calibration cohort, sensitivity to immunotherapy treatment was found to be greatest in melanoma, followed by NSCLC, UCC and then RCC in decreasing order based on our analysis using  $\rho_{\text{predicted}}^{\infty}$ . With respect to drug type, treatment sensitivity was found to be higher with atezolizumab relative to nivolumab. Comparison of model parameters obtained from the clinical validation sets revealed that the outcome sensitivity to treatment estimated as  $y = kx$  was consistent with the values obtained for the specific cancer-drug combinations examined from the literature-derived cohort (Table 2). Within the validation cohorts, our results revealed that tumours treated with ipilimumab (IPI trial cohort;  $n = 93$ ) were most sensitive to therapy (Table 2;  $k = 0.007836$ ), followed by melanoma metastatic to the brain treated with ipilimumab and/or nivolumab ( $k = 0.007007$ ), then by our basket study cohort ( $k = 0.002563$ ), and lastly by the NSCLC clinical cohort ( $k = 0.002498$ ) when sensitivity was estimated using  $\rho_{\text{predicted}}^{\infty}$ . We note that, because of the low correlation strength observed between  $\rho_{\text{measured}}^{\infty}$  and  $1/(\Lambda \times \mu)$ , we have focused only on  $\rho_{\text{predicted}}^{\infty}$  here; this is due to the difference between measured (at last restaging) and predicted (to long-term) final tumour burden, where truncation in measured follow-up times may prevent the capture of behaviour that would be observed (and that the model predicts) if follow-up times were extended, and is further explored in Discussion.

Moreover, the drug-disease combination specific measure of sensitivity for immunotherapy response  $1/(\Lambda \times \mu)$  (Fig. 5) suggests a critical threshold for immunotherapy-mediated lesion volume reduction (i.e., the weakest immune response that may result in reduction of tumour burden) for each cancer and treatment type in the literature-derived calibration cohort; these are indicated by dashed lines in Fig. 5 (statistics are provided in Table 3; found to range from 136.73–802.52, depending on the cancer and treatment type). Taking the lower end of the identified cutoff range (e.g.,  $1/(\Lambda \times \mu) = 136.73$ , representing the strongest immune response cutoff we have identified for the calibration cohort) to be the critical threshold, we examined how well this single cutoff value would perform across all patients irrespective of drug or cancer type. In the calibration cohort, this threshold performed well to sort patients by response in all cancer and drug types examined (patient response was identified correctly with >80% success rate in all cases). When applied to the validation cohorts, the

patient response sensitivity threshold value identified in the calibration cohort ( $1/(\lambda \times \mu) = 136.73$ ) was found to sort patient response in all cohorts with satisfactory results, but was found to have higher success rates identifying responders than non-responders (correctly identifying 90% of responders vs. 69% of non-responders in the clinical trial cohort, 83% of responders vs. 64% of non-responders in the basket-study cohort, 87% of responders vs. 35% of non-responders in the NSCLC clinical cohort, and 75% of responders vs. 67% of non-responders in the brain mets clinical cohort; Table 3). More detailed analyses of how individual model parameters relate to tumour response are shown in Figs. S3–S5, where we observed that  $\rho_{\text{predicted}}^{\infty}$  decreases with increasing  $\mu$  and  $\lambda$  across all cohorts examined, but increases with  $A$ , as expected based on equation (2a).

## Discussion

We have presented the validation and analysis of a mathematical model of immunotherapy efficacy to mechanistically examine the time-course of patient responses to immunotherapy and to predict total long-term patient tumour burden. Model calibration was performed on a literature-derived cohort of 124 patients with 4 different cancer types treated with one of two different immunotherapy agents, and further validated against four additional in-house patient cohorts. The model parameter representing tumour growth rate at first restaging ( $\alpha_{\text{imaging}}$ ) was found to significantly sort patient response with a similar accuracy as the retrospectively determined strength of immune response ( $\lambda \times \mu$ ; see Figs. 2,3 and Table 1), and both are shown to be quantifiable with only CT/MR imaging data that are already obtained as standard-of-care. By including the time scale, we believe that our prognostic model can be more precise than a pure spatial-scale assessment (that is, tumour volume-only prognostics, e.g., RECIST), which we hope will serve to inspire the community at large to consider how inclusion of additional dimensions into response criteria can improve response assessment precision. Further, as  $\alpha_{\text{imaging}}$  may be measured no later than first restaging, it may serve as a *prognostic* clinical tool to identify patient response to treatment on a per-patient basis, which may be possible at earlier times than with RECIST, especially in patients that demonstrate delayed progression or response.

We have shown a model-derived method for quantifying the sensitivity of a cancer type to a specific drug, which may provide a quantitative tool to help select optimal drug choices based on the specific disease and immune health of a patient. The treatment sensitivities identified were further found to be in satisfactory agreement with literature-reported objective response for the drug-disease combinations examined (this is explored further in Supplemental Materials). Interestingly, we found that the strength of immune response ( $\lambda \times \mu$ ) necessary to achieve tumour reduction was unique for each drug-disease combination examined, while the sign of tumour growth rate at first restaging ( $\alpha_{\text{imaging}}$ ) performed well to identify patient response in all cases examined. As such, ( $\alpha_{\text{imaging}}$ ) may perform as a universal indicator of response across all patient cohorts examined herein, while  $\lambda \times \mu$  may be applied on a per drug/disease basis in cases where further resolution is needed. However, our examination of treatment sensitivity via  $1/(\lambda \times \mu)$  revealed that a single threshold value can perform satisfactorily across multiple drug and tumour types (Fig. 5), suggesting that a more universally-applicable guideline based on drug sensitivity

may be discovered. This method of predicting cancer-drug sensitivity was found to be more indicative of the model projected long-term tumour burden than the final measured tumour burden in many cases, indicating that our model may be capturing long-term outcome that may not always be observed clinically (due to patient dropout, changes in treatment interventions, or other events that reduce the duration of patient follow-up), potentially providing additional information that may not be directly observed in standard clinical practice.

Our model analysis demonstrates that the change in tumour growth kinetics from before and after treatment (included in our model but not in RECIST) may be important to consider when evaluating the immunotherapy outcome for each individual patient. Even if a patient was identified to have stable disease on immunotherapy by RECIST criteria (an outcome that might be considered clinically favourable in current standard of care), it could be of clinical benefit for the patient to discontinue the current treatment protocol, or to take a revised protocol that includes additional treatment methods. Indeed, there persists an imperative need for improved clinical prognostic tools to assist physicians in the cases where RECIST fails to distinguish patient response. Our mathematical model will help make this valuable distinction, sort patients in this grey area at early times in the treatment protocol, and quantify strength of the immune response under immunotherapy, ultimately giving clinicians a valuable tool to predict and assess treatment efficacy, even in cases that are difficult to assess under current clinical practice.

We note that while we are able to fully quantify  $\alpha_{\text{imaging}}$  prospectively at time of first restaging using only CT/MR imaging, the analysis of  $\lambda \times \mu$  as presented here is dependent on retrospective quantification, using the full time-history of patient CT images. In our calibration cohort, estimation of tumour volumes from digitization of publication plots has likely resulted in differences in our lesion volume estimates vs. those measured directly from clinical CT/MRI scans. While potential errors in tumour burden data obtained in this way will result in slight variation of model parameter values, the spread of model parameters and the overall trends demonstrated herein remain consistent with data obtained directly from CT/MRI scans, as we have demonstrated with the four in-house validation cohorts. We note that variations in imaging protocols among different institutions and in tumour boundary identification among radiologists have not been investigated at this stage. Despite this challenge, we are encouraged that the model differentiated between patient response groups in all cohorts examined.

The threshold values for prediction of patient response via  $\lambda \times \mu$  and as related to treatment sensitivity via  $1/(\lambda \times \mu)$  we have shown are likely only specific to the drug and disease combinations examined here, and may change as we obtain larger patient data sets. This is further complicated because our calibration cohort was collected from a range of clinical trial phases (phase I<sup>3</sup>, phase I expansion<sup>39</sup>, phase II<sup>40</sup>, and phase III<sup>41</sup>), and the therapeutic protocols received by this cohort were not always reflective of current clinical standard-of-care. However, the model application to a wide variety of drug doses and dosing schedules presented here indicates that our model is robust to handle potential interpatient variations in drug delivery dosages and dosing schedules, as well as potential future changes to standard checkpoint inhibitor therapeutic protocols as the field advances.



The use of  $\alpha_{\text{imaging}}$  as a broadly-applicable indicator independent of tumour or drug type should also be further examined in additional patient cohorts. However, we present the results herein as significant evidence that these thresholds likely exist in other cancer and drug combinations, and, when identified, can be broadly applied to identify optimal drug selections for other cancer types. Moreover, as these are built on known biological quantities (see Methods, Supplemental Materials), we are currently working on methods of quantifying them prospectively in our laboratory. Because we measure response as a rate ( $\alpha_{\text{imaging}}$ ; that is, volume change is normalized by time between restagings) instead of a scalar quantity (e.g., RECIST), our early-time predictor of patient response was shown to perform reliably even in the presence of significant variation in time to first restaging (ranges 5–230 days in the data herein), which is often unavoidable in clinical practice due to a variety of factors (e.g., patient travel schedules, changes in patient health status, etc.).

The predictive power of any prognostic clinical measure is dependent on the quality and accuracy of the relevant measured biological parameters by which it is informed, and we expect this will also be true for model parameters  $\mu$  and  $\Lambda \times \mu$ . Because these parameters are built upon known biophysical factors, they should be calculable from measurements of this underlying biology, likely at early times in treatment. Towards this end, we are currently investigating additional measures which may help determine model parameters, including intratumoural receptor status (e.g., intratumoural PD-L1 staining) and drug dosing and dosing schedules for  $\mu$ , and blood markers (e.g., CTLA4+ lymphocyte counts, neutrophil to lymphocyte ratio), genetic markers (e.g., microsatellite instability), tumour infiltrating lymphocyte (TIL) counts within the tumour microenvironment, and nanomechanical measures of the tumour microenvironment (e.g., tumour phenotype counts, distribution and aggressiveness, immune cell counts before treatment and influx due to treatment, and tumour-immune cell interactions<sup>42</sup>) for  $\Lambda$  and  $\Lambda \times \mu$ . We are also developing new methodologies of more accurately estimating  $\mu$  and  $\Lambda \times \mu$  through inclusion of information about blood perfusion in the tumour, under the hypothesis that both drug and immune cells (and thus their therapeutic availability) are primarily delivered to the tumour through the blood supply. This assumption is supported by our previous modelling work on chemotherapy treatment<sup>43</sup>, which will also provide an additional validation mechanism for the new model presented here. In particular, by obtaining accurate measures of tumour blood supply perfusion via atomic force microscopy-obtained nanomechanical measurements of tumour biopsies (we have previously demonstrated that tissue nanostiffness is indicative of blood supply<sup>42</sup>), we will be able to further refine our quantification of  $\mu$  and  $\Lambda \times \mu$  for model-facilitated prediction of treatment outcome at or before the start of treatment. The results we have presented herein represent key model validation, testing, and calibration steps towards this ongoing project.

Clinically, many cancer therapies often cause significant side effects and patient discomfort, and often therapies are ramped up from least to most aggressive based on the strength of patient response, especially towards end of life in patients with poor prognosis<sup>44</sup>. Unfortunately, this approach may often select for more aggressive cancer phenotypes when only incomplete tumour cell kill is achieved<sup>45</sup>, as the weaker ones may be eliminated by the first round(s) of treatment. A better approach would be to identify the strength of treatment intervention necessary to eliminate the disease on a patient-specific basis (based

on mechanistic principles and the patient's individual, quantifiable growth kinetics) and adjust the treatment plan accordingly early in the course of clinical intervention. Here, we have shown a method to quantify this phenomenon using only readily-available standard of care clinical information, potentially providing a much-needed bedside tool for clinicians that can easily be implemented alongside other standard of care protocols, as we have come to realize our model will likely perform best when used alongside other clinical diagnostics. Many of the patients examined herein received numerous treatments prior to immunotherapy intervention, which is common in clinical practice, as immunotherapy is often used as a "last line of defence" after traditional therapy proves ineffective. Clinically, measures are taken to reduce potential detrimental effects of these prior therapies before immunotherapy treatment (e.g., ensuring patient immune system is rebounded); however, these confounding effects are often unavoidable in real world practice, and our model reliably provides prognostic insights on immunotherapy response despite this diversity.

Biologically, the term  $\Lambda \times \mu$  is proportional to the ratio of tumour volume to immunotherapy efficacy (multiplicative combination of immunotherapy dosing, immune cell fitness, and immunotherapy kill rate). Importantly, these parameters may all either be measured (tumour volume, cell kill rate) or affected clinically (dosing and immune cell fitness, which may be modified through processes like radiotherapy-induced increase in PD-1 expression<sup>46</sup>). Because this parameter is composed of only values which may be readily obtained or modified in the clinic *on a per-patient basis*, it can provide clinicians a powerful tool to guarantee effective immunotherapy results: *through maximization of  $\mu$  (and thus minimization of  $\alpha_{\text{imaging}}$ ) and  $\Lambda \times \mu$ , immunotherapy may be tuned in a patient-specific way to maximize clinical efficacy using only clinically accessible parameters*. Therefore, we believe that this tool will help physicians design individualized treatment plans, increasing the efficacy of immunotherapy and overall patient survival.

## Methods

### Mathematical model of immunotherapy response

Our model mechanistically describes the tumour-immune interaction process in order to better understand the effects of immunotherapy on the time-course of patient lesion burden and outcome. It was built using our significant experience gained from previous successes in modelling tumour kill by chemotherapy, which have been validated against patients with multiple varied cancer types, including glioblastoma, colorectal cancer metastatic to the liver, and hepatocellular carcinoma<sup>43,47–60</sup>. The final form of our model description of tumour burden over time during immunotherapy treatment is:

$$\frac{d\rho}{dt} = \rho(\alpha_{\text{baseline}} - \mu + \Lambda\mu) + \rho^2(-\Lambda\mu), \quad (1)$$

where  $\rho$  is the normalized tumour volume (normalized by the patient's tumour volume at  $t = 0$ ; thus  $\rho(0) = 1$  for all patients) as a function of time,  $\alpha_{\text{baseline}}$  is the intrinsic tumour growth kinetic (a measure of the tumour growth rate prior to and independent from any clinical intervention),  $\mu$  represents the kill rate of cancer cells by immune cells (e.g., the effect of immunotherapy activation of immune cells), and  $\Lambda$  is the intratumoural immune state under

immunotherapy intervention, representing the health of the immune system (mathematically, this is the ratio of tumour to immune cells within the tumour, scaled by a “fitness” factor representing the number of cancer cells each immune cell can kill). We note that  $\Lambda$  and  $\mu$  are both representative of a combination of several biological factors; these, as well as the full derivation of the model, including underlying assumptions and immune factors and mechanisms included, and the system of equations that leads to our master equation, are given in Supplemental Materials. In words, equation (1) states that the change in tumour volume over time is due to the intrinsic growth of the tumour, reduced by the kill rate of cancer cells due to immunotherapy treatment, as dependent on the ratio of initial tumour cells to immune cells, the number of immune cells required to kill one tumour cell, and the efficacy of immune cell activation.

Through making substitutions:

$$A = \alpha_{\text{baseline}} - \mu + \Lambda\mu, \quad (2a)$$

$$B = -\Lambda\mu, \quad (2b)$$

the long-term solution,  $\rho^\infty$ , and time-dependent solution,  $\rho(t)$ , are found to be

$$\rho^\infty = -\frac{A}{B} = \frac{\alpha_{\text{baseline}} - \mu}{\Lambda\mu} + 1, \quad (3)$$

and

$$\rho(t) = \frac{\rho^\infty}{1 - (1 - \rho^\infty)e^{-At}}, \quad (4)$$

respectively. We further denote  $\alpha_{\text{retro}}$  the long-term tumour growth rate after immunotherapy; this is calculated retrospectively by subtracting tumour kill rate ( $\mu$ ) from the intrinsic tumour growth rate ( $\alpha_{\text{baseline}}$ ):

$$\alpha_{\text{retro}} = \alpha_{\text{baseline}} - \mu. \quad (5)$$

We also designate the short-term solution (e.g., the tumour growth rate shortly after start of treatment) as

$$\rho(t) \approx e^{(\alpha_{\text{imaging}}) \cdot t}, \quad (6)$$

where  $\alpha_{\text{imaging}}$  is determined between time of treatment start (baseline) and first restaging ( $t^1$ ) from CT-derived *imaging* data according to

$$\alpha_{\text{imaging}} = \log(\rho(t^1))/t^1. \quad (7)$$

By substituting  $\alpha_{\text{retro}}$  into equation (3), which describes total long-term tumour burden as a function of post-treatment growth rate, immune-mediated tumour cell kill rates, and

intratumoural immune state, the overall patient response may be observed at early times during treatment from the sign of  $\alpha_{\text{retro}}$ :

$$\rho^{\infty} = \frac{\alpha_{\text{retro}}}{\Lambda\mu} + 1. \quad (8)$$

A complete synopsis of all model parameters and their biological definitions has been provided in Table S1, and examination of the model parameter space is provided in Figs. S9, S10. In the work presented here, equation (4) was fit numerically to time-course tumour burden data from one literature-derived cohort and four in-house clinical trial cohorts to quantify model parameters, which were then analyzed to determine their distinct signatures of disease response. In our analysis, tumour kill rate weighted by the patient immune state (i.e.,  $\Lambda\times\mu$ ) was found to significantly sort patients by response. Further, we demonstrate that  $\alpha_{\text{imaging}}$  (equation (7)) may serve an acceptable substitute for the full model fit parameter  $\alpha_{\text{retro}}$ , and thus function as a prospective measure of patient response with high accuracy by the time of first restaging.

### Data acquisition

The model was calibrated against time-course tumour burden datasets obtained from literature-reported measurements of tumour burden volume over time for 124 patients with either metastatic renal cell carcinoma (RCC) <sup>40</sup>, advanced nonsquamous non-small-cell lung cancers (NSCLC) <sup>41</sup>, or melanoma <sup>3</sup>, both during and after treatment with nivolumab, as well as urothelial cell carcinoma (UCC) treated with atezolizumab <sup>39</sup> (details shown in Table S3), together representing more than 6.5 years of clinical trials and forming our literature-derived calibration cohort. All patients had advanced disease (stage IIIB-IV) and were often significantly pretreated with a variety of traditional and targeted therapies (many having received more than 3 pretreatments; see discussion), thus composing a representative sample of highly aggressive and treatment-resistant cases. Combined, the dataset demonstrates the capability of immunotherapy in treating several clinically challenging cases, spans a range of potential patient tumour burden time-course profiles, and serves as a diverse and robust foundation on which to quantify our model across a wide variety of disease and immunotherapy treatment conditions. Patient tumour burden over time were extracted from published plots using WebPlotDigitizer <sup>61</sup>. In all literature-derived cohorts, these measurements were converted from measured lesion diameters (i.e., RECIST v1.1) to estimations of normalized lesion volume (estimated as 3D spheres, see Supplemental Materials).

Four in-house validation cohorts were obtained from patients treated at the University of Texas MD Anderson Cancer Center. The first constituted 93 patients with metastatic lesions to lung or liver treated with ipilimumab (IPI) at doses of 3 mg/kg every 21 days for a total of 4 doses (NCT02239900; 96 were obtained, but 3 were removed due to unavailability of pre-treatment lesion measurements). The second included 28 patients from various clinical trials treated with various checkpoint inhibitors (referred to as the “basket-study cohort” herein; see Table S4) (data for 58 patients were obtained, but 17 were removed due to having received non-immune checkpoint inhibitor immunotherapy, 11 were removed

due to receiving complementary non-immunotherapy or non-immune checkpoint inhibitor immunotherapy treatments concurrently with immune checkpoint inhibitor immunotherapy, and two were removed due to lack of pre-treatment measurements needed to quantify  $\alpha_{\text{baseline}}$  in a basket-study style cohort representing 18 cancer types and 7 different checkpoint inhibitor drugs (details in Table S4). A third in-house validation cohort of 35 patients with NSCLC treated with pembrolizumab (MK-3475) was also obtained (NCT02444741). All tumour measurements in the first 3 validation cohorts (baseline and restaging; restagings ranged 1–12, median = 2) were taken on post-contrast CT scans, with 2.5 mm thickness slices obtained for post-contrast phases. Selection and follow-up of indexed lesions was conducted according to RECIST 1.1 guidelines. At all restagings, long and short lesion diameter measurements were obtained, which were averaged and used to estimate tumour volume as 3D spheres (see Supplemental Materials for details). We note that our basket-study cohort was treated as a single group in this analysis, and is expected to display a median behaviour representing an average of the individual disease and drug combinations it contains.

Finally, model testing was performed on a fourth validation cohort of 21 patients with melanoma metastatic to the brain (referred to as “brain mets”, i.e., brain metastases, herein). Patients were included if they were diagnosed with melanoma brain metastasis between 2016 and 2018, and had a subsequent initial dose of either ipilimumab, nivolumab, or a combination of both agents. Of the 21 patients included in this cohort, 13 were treated with a combination of nivolumab (1 mg per kilogram of body weight) plus ipilimumab (3 mg per kilogram) every 3 weeks for two to four doses (according to tolerance), followed by single-agent nivolumab (3 mg per kilogram) every 2 weeks until progression or unacceptable toxic effects. The other eight patients were treated with either nivolumab or ipilimumab monotherapy. In this cohort post-contrast T1-weighted magnetic resonance (MR) sequences acquired longitudinally with at least 1-mm slice thickness were imported into RayStation Treatment Planning System. An experienced radiation oncologist volumetrically segmented all the brain metastases at each time point (baseline and follow-up restagings ranged 1–9, median = 3) for each patient to obtain the total volume of the intracranial metastatic disease.

### Model fitting

The time-dependent form of the model (equation (4)) was then fit to the normalized time-course data for each patient using Mathematica function NonLinearModelFit<sup>62</sup> to obtain the best-fit parameters ( $\rho_{\text{predicted}}^{\infty}$  and  $A$ ) specific to each patient. Additionally, short-term tumour growth rates after treatment ( $\alpha_{\text{imaging}}$ ) were determined exactly between baseline and first restaging (in the literature-derived cohort, this was represented by first available restaging data) via equation (7). First restaging times (median [range]) for each cohort were RCC: 41 days [9–89 days], literature-derived NSLC: 73 days [5–225 days], Melanoma: 55 days [44–230 days], UCC: 35 days [17–85 days], IPI clinical trial cohort: 53 days [17–91 days], basket-study cohort: 55 days [29–122 days], NSCLC clinical trial cohort: 55 days [25–82 days], and brain mets clinical cohort: 42 days [15–98 days].

Due to the unavailability of information about tumour growth before treatment initiation in the literature cohort, we estimated the intrinsic growth rate  $\alpha_{\text{baseline}}$  as being represented

by the fastest growing tumours in the dataset. Even though these are also receiving immunotherapy treatment, they showed the least (if any) reduction in tumour growth rate; thus it appears reasonable to take them to be the closest available representative of tumour growth without treatment effects. Along these lines, we estimated the pre-treatment growth rate  $\alpha_{\text{baseline}}$  as the average growth rate of the fastest 10% of progressing lesions in each patient set (estimated analogous to equation (7), but for restaging after treatment start; unique  $\alpha_{\text{baseline}}$  values were obtained for each of the 4 cancer types examined), measured at times no greater than 3 weeks after treatment initiation. Although our selection of the fastest 10% of progressing lesions is somewhat arbitrary, patients within this region were observed to have consistent tumour growth rates over time, indicating a lack of response to treatment; thus, we took the average of this group as a reasonable estimate of the intrinsic tumour growth rate. In the IPI clinical trial cohort, basket-study cohort, and, NSCLC clinical trial cohort,  $\alpha_{\text{baseline}}$  was calculated on a per-patient basis between the immediate restaging before start of treatment and restaging at treatment start ( $t = 0$ ), e.g., equation (7), however pre-treatment lesion measurements were unavailable in the brain mets cohort, so  $\alpha_{\text{baseline}}$  was estimated as in the literature cohort. When combined with  $\rho_{\text{predicted}}^{\infty}$  and  $A$  obtained as described earlier, we were able to obtain unique  $\mu$  and  $\lambda$  values (equations (2a, 2b); Supplemental Materials equations (S7, S10)) for each patient, and thus their combined effect  $\lambda \times \mu$  as well (representing tumour kill rate weighted by the patient intratumoural immune state). These are shown in Fig. S2 for the same patient subset shown in Fig. 1, as indicated by matching colours and symbols.

### Model validation

We performed a two-step process (Fig. S1) in order to verify the ability of model parameters ( $\lambda \times \mu$  and  $\alpha_{\text{imaging}}$ ) to uniquely identify patient response. First, model parameters for each patient in each cohort were uniquely determined as described above (*Model fitting* section), and then the values of model parameters that optimally sorted patient response in the calibration cohort were identified by optimizing Youden's J statistic via ROC analysis. Next, these threshold values were used to sort patients in each validation cohort based on the response predicted by each model parameter value examined. Finally, predicted responses were compared to measured responses (by CT at last follow-up) and assessed via calculation of sensitivity and specificity. We found that our model parameter  $\lambda \times \mu$  provides unique prediction threshold values for each cancer-tumour combination, due to variations in the mechanistic parameters unique to each combination; in this case, we repeated the process of optimizing Youden's J statistic via ROC analysis to determine the optimal threshold for response prediction for each unique cancer-drug combination.

### Reporting summary.

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

### Data availability.

The authors declare that all data supporting the results in this study are available within the paper and its Supplementary Information. The raw, de-identified patient data are available

from the corresponding author on reasonable request and subject to Institutional Review Board approval.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

The research reported in this manuscript was supported by the National Science Foundation Grant DMS-1930583 (Z.W., V.C.), the National Institutes of Health (NIH) Grants 1U01CA196403 (Z.W., E.J.K., V.C.), 1U01CA213759 (Z.W., V.C.), 1R01CA226537 (Z.W., V.C.), 1R01CA222007 (Z.W., G.A.C., V.C.), U54CA210181 (Z.W., E.J.K., V.C.), and the University of Texas System STARS Award (V.C.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Competing interests

J.W.W. reports research support from GlaxoSmithKline, Bristol Meyers Squibb, Merck, Nanobiotix, Mavu Pharma and Checkmate Pharmaceuticals. J.W.W. serves on the scientific advisory board for RefleXion Medical, MolecularMatch, OncoResponse, CheckMate, Mavu Pharmaceuticals, Alpine Immune Sciences. J.W.W. is co-founder of Healios Oncology, MolecularMatch, and OncoResponse and serves as an advisor to Astra Zeneca, Merck, MolecularMatch, Incyte, Aileron and Nanobiotix. J.W.W. has the following patents; MP470 (amuvatinib), MRX34 regulation of PDL1, XRT technique to overcome immune resistance. MD Anderson Cancer Center has a trademark for RadScopalTM.

## References

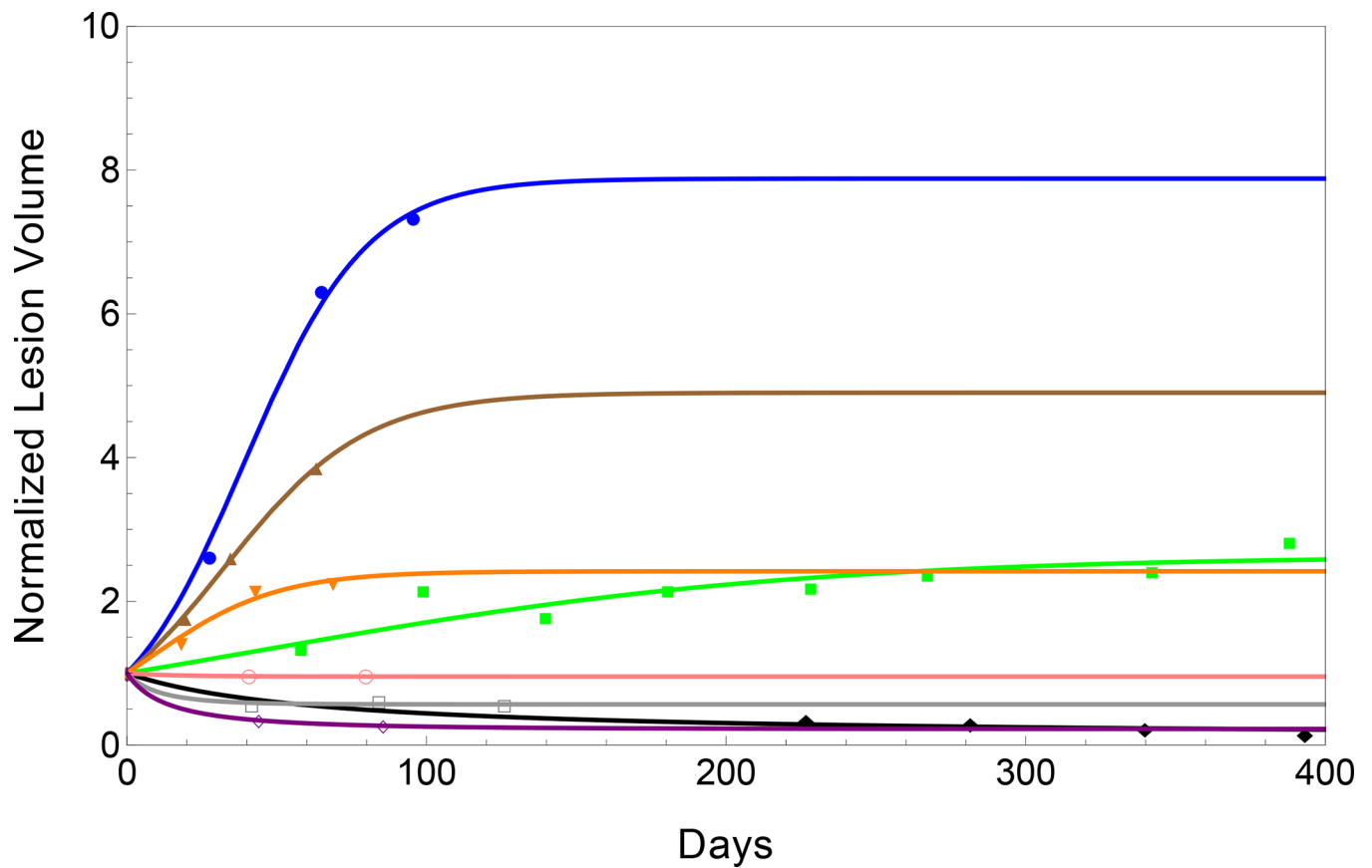
1. Disis ML Mechanism of action of immunotherapy. *Semin Oncol* 41 Suppl 5, S3–13 (2014).
2. Farkona S, Diamandis EP & Blasutig IM Cancer immunotherapy: the beginning of the end of cancer? *BMC Med* 14, 73 (2016). [PubMed: 27151159]
3. Topalian SL et al. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N Engl J Med* 366, 2443–54 (2012). [PubMed: 22658127]
4. Rosenberg JE et al. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. *The Lancet* 387, 1909–1920 (2016).
5. Schumacher TN, Kesmir C & van Buuren MM Biomarkers in cancer immunotherapy. *Cancer Cell* 27, 12–4 (2015). [PubMed: 25584891]
6. Chiou VL & Burotto M Pseudoprogression and Immune-Related Response in Solid Tumors. *J Clin Oncol* 33, 3541–3 (2015). [PubMed: 26261262]
7. Carthon BC et al. Preoperative CTLA-4 blockade: tolerability and immune monitoring in the setting of a presurgical clinical trial. *Clin Cancer Res* 16, 2861–71 (2010). [PubMed: 20460488]
8. Ahmadzadeh M et al. Tumor antigen-specific CD8 T cells infiltrating the tumor express high levels of PD-1 and are functionally impaired. *Blood* 114, 1537–44 (2009). [PubMed: 19423728]
9. Schmidt H et al. Pretreatment levels of peripheral neutrophils and leukocytes as independent predictors of overall survival in patients with American Joint Committee on Cancer Stage IV Melanoma: results of the EORTC 18951 Biochemotherapy Trial. *J Clin Oncol* 25, 1562–9 (2007). [PubMed: 17443000]
10. Phan GQ, Attia P, Steinberg SM, White DE & Rosenberg SA Factors associated with response to high-dose interleukin-2 in patients with metastatic melanoma. *J Clin Oncol* 19, 3477–82 (2001). [PubMed: 11481353]
11. Yuan J et al. Integrated NY-ESO-1 antibody and CD8+ T-cell responses correlate with clinical benefit in advanced melanoma patients treated with ipilimumab. *Proc Natl Acad Sci U S A* 108, 16723–8 (2011). [PubMed: 21933959]

12. Yuan J et al. Pretreatment serum VEGF is associated with clinical response and overall survival in advanced melanoma patients treated with ipilimumab. *Cancer Immunol Res* 2, 127–32 (2014). [PubMed: 24778276]
13. Blay JY et al. Serum level of interleukin 6 as a prognosis factor in metastatic renal cell carcinoma. *Cancer Res* 52, 3317–22 (1992). [PubMed: 1596890]
14. Kelderman S et al. Lactate dehydrogenase as a selection criterion for ipilimumab treatment in metastatic melanoma. *Cancer Immunol Immunother* 63, 449–58 (2014). [PubMed: 24609989]
15. Rizvi NA et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* 348, 124–8 (2015). [PubMed: 25765070]
16. Van Allen EM et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* 350, 207–211 (2015). [PubMed: 26359337]
17. Meng X, Huang Z, Teng F, Xing L & Yu J Predictive biomarkers in PD-1/PD-L1 checkpoint blockade immunotherapy. *Cancer Treat Rev* 41, 868–76 (2015). [PubMed: 26589760]
18. Madore J et al. PD-L1 expression in melanoma shows marked heterogeneity within and between patients: implications for anti-PD-1/PD-L1 clinical trials. *Pigment Cell Melanoma Res* 28, 245–53 (2015). [PubMed: 25477049]
19. Sabatino M et al. Serum vascular endothelial growth factor and fibronectin predict clinical response to high-dose interleukin-2 therapy. *J Clin Oncol* 27, 2645–52 (2009). [PubMed: 19364969]
20. Eisenhauer EA et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 45, 228–47 (2009). [PubMed: 19097774]
21. Deisboeck TS, Wang Z, Macklin P & Cristini V Multiscale cancer modeling. *Annu Rev Biomed Eng* 13, 127–55 (2011). [PubMed: 21529163]
22. Wang Z, Butner JD, Kerketta R, Cristini V & Deisboeck TS Simulating cancer growth with multiscale agent-based modeling. *Semin Cancer Biol* 30, 70–8 (2015). [PubMed: 24793698]
23. Cristini V, Koay E & Wang Z An Introduction to Physical Oncology: How Mechanistic Mathematical Modeling Can Improve Cancer Therapy Outcomes, 164 (CRC Press, 2017).
24. Wang Z, Butner JD, Cristini V & Deisboeck TS Integrated PK-PD and agent-based modeling in oncology. *J Pharmacokinet Pharmacodyn* 42, 179–89 (2015). [PubMed: 25588379]
25. Gillies RJ, Brown JS, Anderson ARA & Gatenby RA Eco-evolutionary causes and consequences of temporal changes in intratumoural blood flow. *Nat Rev Cancer* 18, 576–585 (2018). [PubMed: 29891961]
26. Wang Z & Deisboeck TS Dynamic Targeting in Cancer Treatment. *Front Physiol* 10, 96 (2019). [PubMed: 30890944]
27. Dogra P et al. Mathematical modeling in cancer nanomedicine: a review. *Biomed Microdevices* 21, 40 (2019). [PubMed: 30949850]
28. Dogra P et al. Image-guided mathematical modeling for pharmacological evaluation of nanomaterials and monoclonal antibodies. *Wiley Interdiscip Rev Nanomed Nanobiotechnol*, e1628 (2020).
29. Kronik N, Kogan Y, Vainstein V & Agur Z Improving alloreactive CTL immunotherapy for malignant gliomas using a simulation model of their interactive dynamics. *Cancer Immunol Immunother* 57, 425–39 (2008). [PubMed: 17823798]
30. Eikenberry S, Thalhauser C & Kuang Y Tumor-immune interaction, surgical treatment, and cancer recurrence in a mathematical model of melanoma. *PLoS Comput Biol* 5, e1000362 (2009). [PubMed: 19390606]
31. Wells DK et al. Spatial and functional heterogeneities shape collective behavior of tumor-immune networks. *PLoS Comput Biol* 11, e1004181 (2015). [PubMed: 25905470]
32. Palladini A et al. In silico modeling and in vivo efficacy of cancer-preventive vaccinations. *Cancer Res* 70, 7755–63 (2010). [PubMed: 20924100]
33. Pappalardo F, Lollini PL, Castiglione F & Motta S Modeling and simulation of cancer immunoprevention vaccine. *Bioinformatics* 21, 2891–7 (2005). [PubMed: 15817697]
34. Walker R & Enderling H From concept to clinic: Mathematically informed immunotherapy. *Curr Probl Cancer* 40, 68–83 (2016). [PubMed: 26645497]



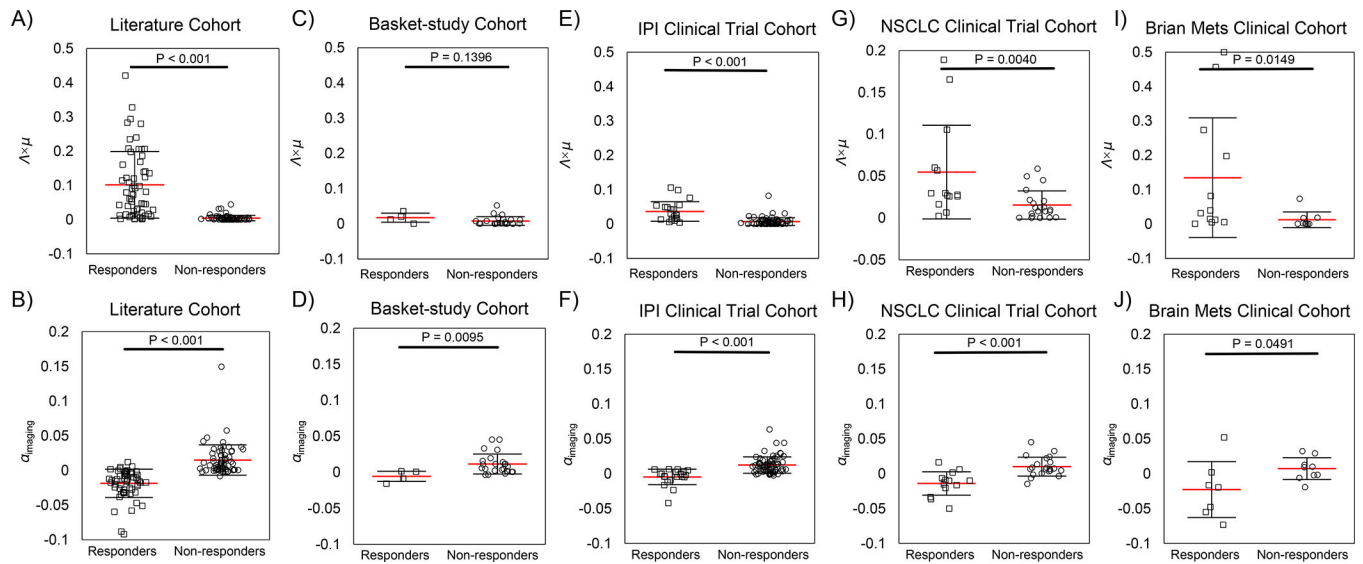
35. Konstorum A, Vella AT, Adler AJ & Laubenbacher RC Addressing current challenges in cancer immunotherapy with mathematical and computational modelling. *J R Soc Interface* 14, 20170150 (2017). [PubMed: 28659410]
36. Luksza M et al. A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy. *Nature* 551, 517–520 (2017). [PubMed: 29132144]
37. Kim E, Rebecca VW, Smalley KS & Anderson AR Phase I trials in melanoma: A framework to translate preclinical findings to the clinic. *Eur J Cancer* 67, 213–222 (2016). [PubMed: 27689717]
38. Cyll K et al. Tumour heterogeneity poses a significant challenge to cancer biomarker research. *Br J Cancer* 117, 367–375 (2017). [PubMed: 28618431]
39. Powles T et al. MPDL3280A (anti-PD-L1) treatment leads to clinical activity in metastatic bladder cancer. *Nature* 515, 558–62 (2014). [PubMed: 25428503]
40. Motzer RJ et al. Nivolumab for Metastatic Renal Cell Carcinoma: Results of a Randomized Phase II Trial. *J Clin Oncol* 33, 1430–7 (2015). [PubMed: 25452452]
41. Borghaei H et al. Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer. *N Engl J Med* 373, 1627–39 (2015). [PubMed: 26412456]
42. Plodinec M et al. The nanomechanical signature of breast cancer. *Nat Nanotechnol* 7, 757–65 (2012). [PubMed: 23085644]
43. Pascal J et al. Mechanistic patient-specific predictive correlation of tumor drug response with microenvironment and perfusion measurements. *Proc Natl Acad Sci U S A* 110, 14266–71 (2013). [PubMed: 23940372]
44. Earle CC et al. Aggressiveness of cancer care near the end of life: is it a quality-of-care issue? *J Clin Oncol* 26, 3860–6 (2008). [PubMed: 18688053]
45. De Larco JE, Wuertz BR, Manivel JC & Furcht LT Progression and enhancement of metastatic potential after exposure of tumor cells to chemotherapeutic agents. *Cancer Res* 61, 2857–61 (2001). [PubMed: 11306458]
46. van Meir H et al. Impact of (chemo)radiotherapy on immune cell composition and function in cervical cancer patients. *Oncoimmunology* 6, e1267095 (2017). [PubMed: 28344877]
47. Pascal J et al. Mechanistic modeling identifies drug-uptake history as predictor of tumor drug resistance and nano-carrier-mediated response. *ACS Nano* 7, 11174–11182 (2013). [PubMed: 24187963]
48. Das H et al. Impact of diffusion barriers to small cytotoxic molecules on the efficacy of immunotherapy in breast cancer. *PLoS ONE* 8, e61398 (2013). [PubMed: 23620747]
49. Koay EJ et al. Transport properties of pancreatic cancer describe gemcitabine delivery and response. *J Clin Invest* 124, 1525–36 (2014). [PubMed: 24614108]
50. Frieboes HB, Wu M, Lowengrub J, Decuzzi P & Cristini V A computational model for predicting nanoparticle accumulation in tumor vasculature. *PLoS ONE* 8, e56876 (2013). [PubMed: 23468887]
51. Wang Z et al. Theory and Experimental Validation of a Spatio-temporal Model of Chemotherapy Transport to Enhance Tumor Cell Kill. *PLoS Comput Biol* 12, e1004969 (2016). [PubMed: 27286441]
52. Brocato TA et al. Understanding the Connection between Nanoparticle Uptake and Cancer Treatment Efficacy using Mathematical Modeling. *Sci Rep* 8, 7538 (2018). [PubMed: 29795392]
53. Koay EJ et al. A Visually Apparent and Quantifiable CT Imaging Feature Identifies Biophysical Subtypes of Pancreatic Ductal Adenocarcinoma. *Clin Cancer Res* 24, 5883–5894 (2018). [PubMed: 30082477]
54. Dogra P et al. Establishing the effects of mesoporous silica nanoparticle properties on in vivo disposition using imaging-based pharmacokinetics. *Nat Commun* 9, 4551 (2018). [PubMed: 30382084]
55. Brocato TA et al. Predicting breast cancer response to neoadjuvant chemotherapy based on tumor vascular features in needle biopsies. *JCI Insight* 5(2019).
56. Goel S et al. Size-Optimized Ultrasmall Porous Silica Nanoparticles Depict Vasculature-Based Differential Targeting in Triple Negative Breast Cancer. *Small* 15, e1903747 (2019). [PubMed: 31565854]

57. Pichler M et al. Therapeutic potential of FLANC, a novel primate-specific long non-coding RNA in colorectal cancer. *Gut* (2020).
58. Butner JD et al. A Multiscale Agent-Based Model of Ductal Carcinoma In Situ. *IEEE Trans Biomed Eng* 67, 1450–1461 (2020). [PubMed: 31603768]
59. Butner JD et al. Mathematical prediction of clinical outcomes in advanced cancer patients treated with checkpoint inhibitor immunotherapy. *Sci Adv* 6, eaay6298 (2020). [PubMed: 32426472]
60. Dogra P et al. A mathematical model to predict nanomedicine pharmacokinetics and tumor delivery. *Comput Struct Biotechnol J* 18, 518–531 (2020). [PubMed: 32206211]
61. Rohatgi A *WebPlotDigitizer*. 4.0 edn (Austin, Texas, USA, 2017).
62. Wolfram Research. *Mathematica*. 11.2 edn (Champaign, Illinois, 2017).



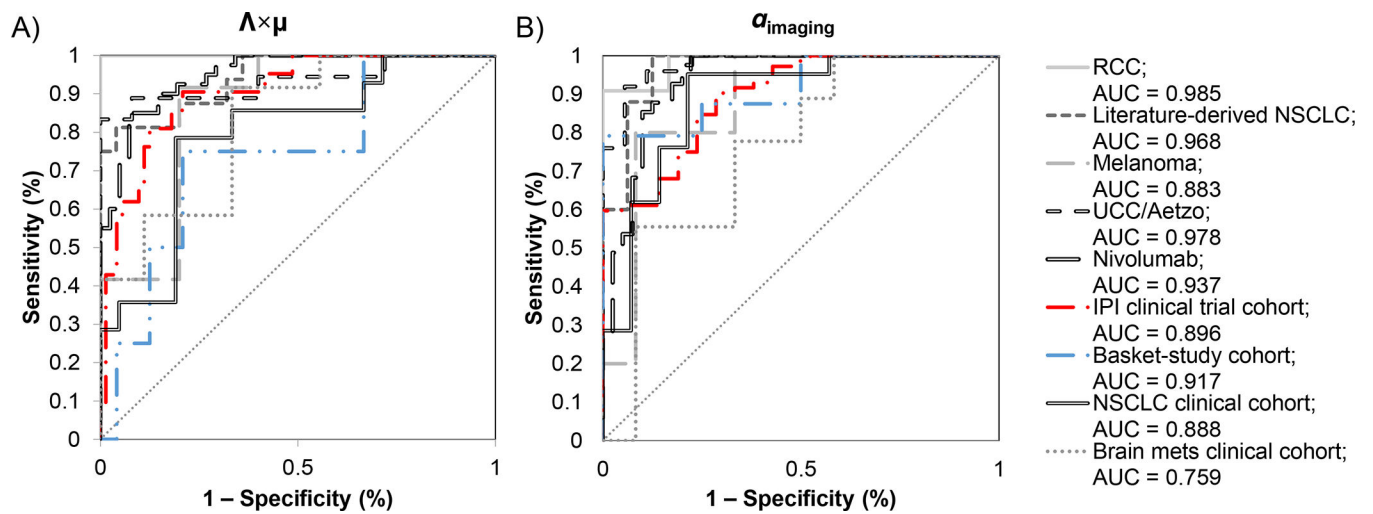
**Fig. 1 |**

Example model fits to clinical immunotherapy response. Curves show the time-dependent model solution (equation (4)) fit to normalized measured patient tumour burden under immunotherapy treatment, obtained from <sup>3,39–41</sup>; also see Table S3. Patients that experienced high therapy efficacy are seen to have reduced total tumour burden over time (and thus  $\rho_{\text{predicted}}^{\infty} < 1.0$ ) relative to tumour burden at time of patient on study ( $t = 0$ ), while patients with progressive disease are seen to have increased total tumour volume over time. The model shows two distinct long-term solution profiles, where tumour burden asymptotes either to a stabilized larger or smaller final tumour burden at long-term. Nonlinear fit statistics are provided in Table S2.



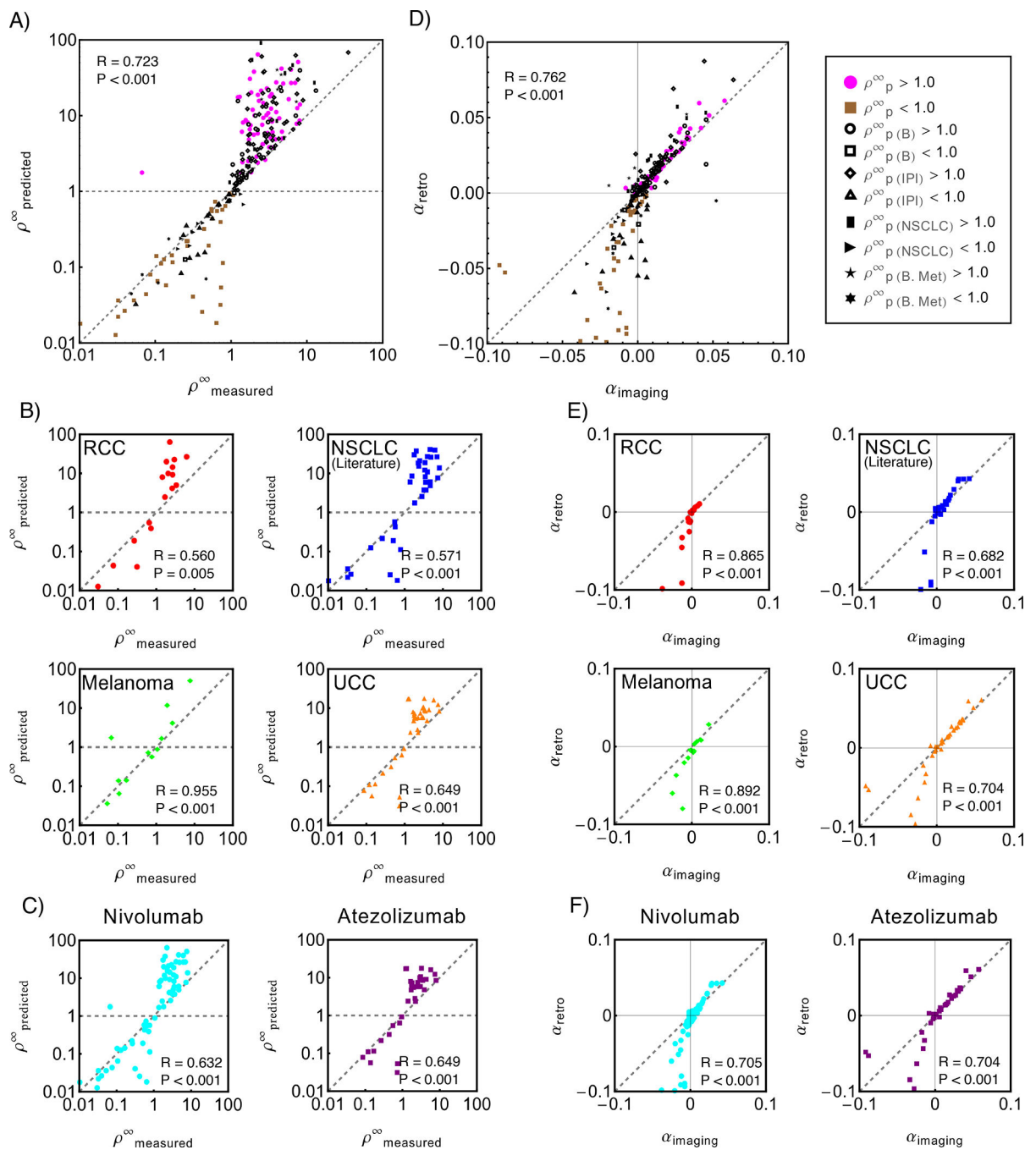
**Fig. 2 |**

Distribution of model parameters against immune response (responders: tumour burden reduced, or non-responders: tumour burden increased). A–B) Literature cohort ( $n = 124$ ) shows significant separation between both the retrospective measure of immunotherapy efficacy  $\Lambda \times \mu$  and our prospective estimate of tumour growth rate at first restaging  $\alpha_{\text{imaging}}$ . C–D) Analysis of our institutional basket study cohort ( $n = 28$ , Table S4) showed significant separation of  $\alpha_{\text{imaging}}$  values between responder and non-responder patients; however, separation of  $\Lambda \times \mu$  was found to be insignificant in this set. E–F) Further analysis of our larger and more cohesive IPI clinical trial cohort ( $n = 93$ ) revealed significant separation between both  $\Lambda \times \mu$  and our measure of growth at first restaging  $\alpha_{\text{imaging}}$ , suggesting that lack of significant separation in panel C is likely due to the small patient sample size in this cohort. This was further confirmed in G–H) the NSCLC clinical trial cohort ( $n = 35$ ) and I–J) the brain mets clinical cohort ( $n = 21$ ). In all panels, per-patient values are shown with mean and standard deviations (red and black bars, respectively). Means  $\pm$  standard deviation (responders, non-responders): A)  $0.1015 \pm 0.0973$ ,  $0.0044 \pm 0.0079$ ; B)  $-0.0186 \pm 0.0205$ ,  $0.0150 \pm 0.0219$ ; C)  $0.0176 \pm 0.128$ ,  $0.0081 \pm 0.124$ ; D)  $-0.0056 \pm 0.0071$ ,  $0.0114 \pm 0.0138$ ; E)  $0.0366 \pm 0.0284$ ,  $0.0071 \pm 0.0113$ ; F)  $-0.0047 \pm 0.0111$ ,  $0.0124 \pm 0.0017$ ; G)  $0.0547 \pm 0.0560$ ;  $0.0155 \pm 0.0169$ ; H)  $-0.0140 \pm 0.0167$ ,  $0.0101 \pm 0.0135$ ; I)  $0.1348 \pm 0.1741$ ,  $0.0126 \pm 0.0226$ ; J)  $-0.0228 \pm 0.0400$ ,  $0.0072 \pm 0.0156$ . Significance of separation was determined by Wilcoxon rank sum (two tail). Plots showing distribution of  $\rho_{\text{predicted}}^{\infty}$  by response category are provided in Fig. S6.



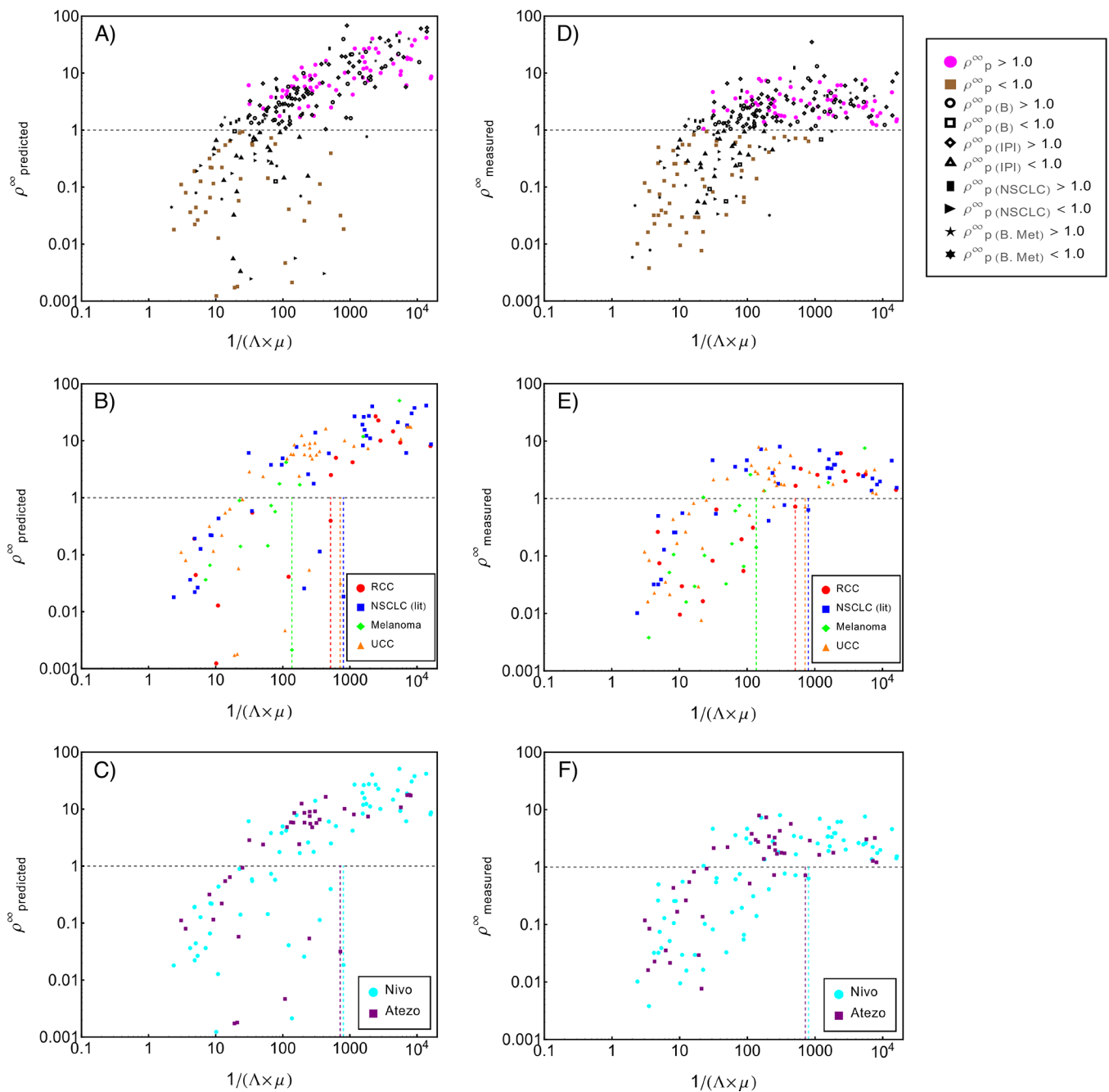
**Fig. 3 |**

ROC analysis showed high accuracy of patient response classification when sorted by A)  $\Lambda \times \mu$  (*retrospective* quantification of strength of immune response) and B)  $\alpha_{\text{imaging}}$  (*prospective* estimate of tumour growth rate after treatment) across all patient cohorts examined. Our prospective estimate of tumour response ( $\alpha_{\text{imaging}}$ ) shows consistent performance in response assessment with the retrospectively determined value,  $\Lambda \times \mu$ . Note both UCC and Atezolizumab (Atezo) represent the same cohort, and are shown as the same curve for clarity. Detailed statistics are provided in Tables 1 and S5, and ROC by  $\rho_{\text{predicted}}^{\infty}$  is shown in Fig. S7. Independent patient counts: RCC,  $n = 23$ ; Lit-derived NSCLC,  $n = 41$ ; Melanoma,  $n = 17$ ; UCC/Atezo,  $n = 43$ ; Nivolumab,  $n = 81$ ; IPI clinical trial cohort,  $n = 93$ ; Basket-study cohort,  $n = 28$ ; NSCLC clinical cohort,  $n = 35$ ; Brian mets clinical cohort,  $n = 21$ .

**Fig. 4 |**

Correlation analysis of literature-derived calibration cohort reveals significant correlation between measured values and model-predicted variables. A–C) Model-predicted long-term tumour burden  $\rho_{\text{predicted}}^{\infty}$  shows significant correlation with measured tumour burden at last restaging ( $\rho_{\text{measured}}^{\infty}$ ), sorted by A) response type, B) cancer type, and C) drug type. Data for  $n = 124$  patients (calibration cohort) with 4 different cancer types under one of two types of immunotherapy treatment are shown (coloured points; see Table S3), and compared to

all in-house validation cohorts (black points): basket-cohort ( $\rho_{p(B)}^{\infty}$ ;  $n = 28$ ), IPI clinical trial cohort ( $\rho_{p(IPI)}^{\infty}$ ;  $n = 93$ ) cohort, NSCLC clinical trial cohort ( $\rho_{p(NSCLC)}^{\infty}$ ;  $n = 35$ ), and brain mets cohort ( $\rho_{p(B. Mets)}^{\infty}$ ;  $n = 21$ ) in panel A. D–F) Correlation analysis between retrospective tumour growth rate after immunotherapy ( $\alpha_{retro}$ ) and prospective tumour growth rate at first restaging ( $\alpha_{imaging}$ ), again sorted by D) response, E) cancer, and F) drug type, with values from basket-study and clinical trial validation cohorts overlaid for comparison in D). In all cases, Person correlation coefficient (R) and associated P-values are shown, inset. Figure legends: P = predicted, L = literature cohort, B = basket-study cohort, IPI = in-house ipilimumab clinical trial cohort, NSCLC = NSCLC in-house clinical cohort, B. Mets = brain metastasis in-house clinical cohort. Independent patient counts in B, E): RCC,  $n = 23$ ; NSCLC (Literature),  $n = 41$ ; Melanoma,  $n = 17$ ; UCC,  $n = 43$ ; and in C, F): Nivolumab,  $n = 81$ ; Atezolizumab,  $n = 43$ . Additional analysis of tumour growth rate at first restaging ( $\alpha_{imaging}$ ) vs.  $\rho_{predicted}^{\infty}$  is provided in Fig. S8.



**Fig. 5 |**

Long-term tumour burden predicted by the model ( $\rho^\infty_{\text{predicted}}$ ; A–C) and measured at last restaging ( $\rho^\infty_{\text{measured}}$ ; D–F) is plotted against inverse immune response strength  $1/(\lambda \times \mu)$  for  $n = 124$  patients (calibration cohort) with 4 different cancer types under one of two types of immunotherapy treatment (coloured points; see Table S3), sorted by A,D) response type, B,E) cancer type (independent patient counts: RCC,  $n = 23$ ; NSCLC (lit),  $n = 41$ ; Melanoma,  $n = 17$ ; UCC = 43), and C,F) immunotherapy drug (independent patient counts: Nivo,  $n = 81$ ; Atezo,  $n = 43$ ). Maximum measured values for  $1/(\lambda \times \mu)$



with  $\rho^\infty < 1.0$  are indicated by dashed lines (colors correspond to figure legend); 515.80 for RCC, 802.52 for literature(lit)-derived NSCLC, 136.73 for melanoma, and 719.21 for UCC; and 802.52 for nivolumab, 719.21 for atezolizumab; more details and statistics are provided in Table 3. For all cancer and immunotherapy agent types examined,  $\rho_{\text{predicted}}^\infty$  shows a linear relationship to strength of immune response ( $1/(A \times \mu)$ ) by linear regression analysis, however this correlation is significantly diminished for and  $\rho_{\text{measured}}^\infty$  (see Table 2). Data from the validation sets (black points) are consistent with the literature-derived cohort dataset (Basket-study ( $\rho_{(B)}^\infty$ ,  $n = 28$ ; see Table S4), IPI clinical trial ( $\rho_{(IPI)}^\infty$ ,  $n = 93$ ), NSCLC in-house clinical trial ( $\rho_{(NSCLC)}^\infty$ ,  $n = 35$ ), and brain metastasis ( $\rho_{(B. Met)}^\infty$ ,  $n = 21$ ) validation cohorts are shown in A,D). Further breakdown of the relationship between  $\rho_{\text{predicted}}^\infty$  and other model parameters  $\mu$ ,  $A$ , and  $A$  are shown in Figs. S3–S5. Figure legends: P = predicted, m=measured, NSCLC (lit) = literature-derived NSCLC cohort, B = basket-study cohort, IPI = ipilimumab clinical trial cohort, NSCLC = NSCLC clinical trial cohort, B. Met = melanoma brain metastasis clinical trial cohort.

**Table 1:**

Statistics and (95% confidence intervals) for all patient cohorts. Cutoff values for  $\Lambda \times \mu$  were identified via ROC analysis by maximizing Youden's J statistic, while cutoffs for  $\alpha_{\text{imaging}}$  were taken to be 0.0 for all cases due to the physical meaning of  $\alpha_{\text{imaging}} = 0.0$  separating tumours that are growing or shrinking at first restaging. (lit) denotes the literature-derived NSCLC cohort, and NSCLC trial denotes the in-house cohort.

Cancer Type (patient count)	Classifier	Cutoff	AUC	Sensitivity *	Specificity **
<b>Literature cohorts</b>					
RCC ( $n = 23$ )	$\Lambda \times \mu$	0.0019	1.0 (1.0 – 1.0)	1.0 (1.0 – 1.0)	1.0 (1.0 – 1.0)
RCC ( $n = 23$ )	$\alpha_{\text{imaging}}$	0.0	0.984 (0.933 – 1.036)	0.909 (0.739 – 1.079)	1.0 (1.0 – 1.0)
NSCLC (lit) ( $n = 41$ )	$\Lambda \times \mu$	0.0042	0.943 (0.872 – 1.013)	0.875 (0.713 – 1.037)	0.800 (0.604 – 0.995)
NSCLC (lit) ( $n = 41$ )	$\alpha_{\text{imaging}}$	0.0	0.968 (0.904 – 1.031)	0.880 (0.753 – 1.007)	0.875 (0.745 – 1.0043)
Melanoma ( $n = 17$ )	$\Lambda \times \mu$	0.0073	0.867 (0.646 – 1.087)	1.0 (1.0 – 1.0)	0.600 (0.323 – 0.877)
Melanoma ( $n = 17$ )	$\alpha_{\text{imaging}}$	0.0	0.883 (0.720 – 1.046)	0.800 (0.449 – 1.151)	0.750 (0.370 – 1.129)
UCC ( $n = 43$ )	$\Lambda \times \mu$	0.0088	0.933 (0.858 – 1.009)	0.889 (0.744 – 1.034)	0.920 (0.795 – 1.045)
UCC ( $n = 43$ )	$\alpha_{\text{imaging}}$	0.0	0.978 (0.928 – 1.027)	0.880 (0.753 – 1.007)	0.944 (0.855 – 1.034)
<b>Validation cohorts</b>					
IPI trial * ( $n = 93$ )	$\Lambda \times \mu$	0.0080	0.904 (0.843 – 0.965)	0.905 (0.779 – 1.031)	0.722 (0.531 – 0.914)
IPI trial * ( $n = 93$ )	$\alpha_{\text{imaging}}$	0.0	0.896 (0.802 – 0.989)	0.903 (0.834 – 0.971)	0.667 (0.558 – 0.776)
Basket-study ** ( $n = 28$ )	$\Lambda \times \mu$	0.0086	0.740 (0.507 – 0.972)	0.750 (0.326 – 1.174)	0.708 (0.263 – 1.154)
Basket-study ** ( $n = 28$ )	$\alpha_{\text{imaging}}$	0.0	0.917 (0.723 – 1.110)	0.875 (0.743 – 1.007)	0.500 (0.300 – 0.700)
NSCLC trial † ( $n = 35$ )	$\Lambda \times \mu$	0.0145	0.793 (0.645 – 0.940)	0.857 (0.674 – 1.040)	0.667 (0.420 – 0.914)
NSCLC trial † ( $n = 35$ )	$\alpha_{\text{imaging}}$	0.0	0.888 (0.764 – 1.011)	0.810 (0.642 – 0.977)	0.786 (0.610 – 0.961)
Brain Mets †† ( $n = 21$ )	$\Lambda \times \mu$	0.0033	0.824 (0.631 – 1.016)	0.916 (0.760 – 1.073)	0.667 (0.400 – 0.933)
Brain Mets †† ( $n = 21$ )	$\alpha_{\text{imaging}}$	0.0	0.759 (0.552 – 0.966)	0.556 (0.231 – 0.880)	0.833 (0.590 – 1.077)
<b>Drug type (patient count)</b>					
<b>Literature cohorts</b>					
Nivolumab ( $n = 81$ )	$\Lambda \times \mu$	0.0063	0.949 (0.898 – 0.999)	0.900 (0.807 – 0.992)	0.854 (0.744 – 0.963)
Nivolumab ( $n = 81$ )	$\alpha_{\text{imaging}}$	0.0	0.937 (0.881 – 0.993)	0.878 (0.778 – 0.978)	0.875 (0.773 – 0.976)
Atezolizumab ( $n = 43$ )	$\Lambda \times \mu$	0.0088	0.933 (0.858 – 1.009)	0.889 (0.744 – 1.034)	0.920 (0.795 – 1.045)
Atezolizumab ( $n = 43$ )	$\alpha_{\text{imaging}}$	0.0	0.978 (0.928 – 1.027)	0.880 (0.753 – 1.007)	0.944 (0.855 – 1.034)
<b>Validation cohorts</b>					
IPI trial † ( $n = 93$ )	$\Lambda \times \mu$	0.0080	0.904 (0.843 – 0.965)	0.905 (0.779 – 1.031)	0.722 (0.531 – 0.914)
IPI trial † ( $n = 93$ )	$\alpha_{\text{imaging}}$	0.0	0.896 (0.802 – 0.989)	0.903 (0.834 – 0.971)	0.667 (0.558 – 0.776)
Basket-study †† ( $n = 28$ )	$\Lambda \times \mu$	0.0086	0.740 (0.507 – 0.972)	0.750 (0.326 – 1.174)	0.708 (0.263 – 1.154)
Basket-study †† ( $n = 28$ )	$\alpha_{\text{imaging}}$	0.0	0.917 (0.723 – 1.110)	0.875 (0.743 – 1.007)	0.500 (0.300 – 0.700)

NSCLC trial <sup>§</sup> ( <i>n</i> = 35)	$A \times \mu$	0.0145	0.793 (0.645 – 0.940)	0.857 (0.674 – 1.040)	0.667 (0.420 – 0.914)
NSCLC trial <sup>§</sup> ( <i>n</i> = 35)	$\alpha_{\text{imaging}}$	0.0	0.888 (0.764 – 1.011)	0.810 (0.642 – 0.977)	0.786 (0.610 – 0.961)
Brain Mets <sup>#</sup> ( <i>n</i> = 21)	$A \times \mu$	0.0033	0.824 (0.631 – 1.016)	0.916 (0.760 – 1.073)	0.667 (0.400 – 0.933)
Brain Mets <sup>#</sup> ( <i>n</i> = 21)	$\alpha_{\text{imaging}}$	0.0	0.759 (0.552 – 0.966)	0.556 (0.231 – 0.880)	0.833 (0.590 – 1.077)

\* Sensitivity:  $A \times \mu$  = number of responders identified correctly;  $\alpha_{\text{imaging}}$  = number of non-responders identified correctly.

\*\* Specificity:  $A \times \mu$  = number of non-responders identified correctly;  $\alpha_{\text{imaging}}$  = number of responders identified correctly.

<sup>†</sup> IPI trial: drug type = ipilimumab, cancer type = metastatic to lung or liver, see Methods

<sup>††</sup> Basket-study cohort represents 18 cancer types and 7 drug types (see Table S4)

<sup>§</sup> NSCLC trial: drug type = pembrolizumab; cancer type = NSCLC

<sup>#</sup> Brain Mets: drug type = ipilimumab, nivolumab, or ipilimumab plus nivolumab (see Methods), cancer type = melanoma metastatic to the brain.

**Table 2:**

Best fit parameters for least squares linear regression estimation (in the form of  $y = kx$ ) of how sensitive a certain cancer type is to treatment by the specified drug type; larger  $k$  values indicate larger sensitivity. Higher coefficients of determination were found when using predicted long-term tumour burden ( $\rho_{\text{predicted}}^{\infty}$ ), however these were diminished when measured tumour burden at last restaging ( $\rho_{\text{measured}}^{\infty}$ ) was used. P-values for the regression models were determined by F-test (right tail). (lit) denotes the literature-derived NSCLC cohort, and NSCLC trial denotes the in-house cohort.

Cancer type (patient count)	$\rho_{\text{predicted}}^{\infty}$			$\rho_{\text{measured}}^{\infty}$		
	$k$	$R^2$	$p$	$k$	$R^2$	$p$
<b>Literature cohorts</b>						
RCC ( $n = 23$ )	0.00062	0.561326	< 0.001	0.00004	0.137486	0.074
NSCLC (lit) ( $n = 41$ )	0.00259	0.486478	< 0.001	0.00031	0.165020	0.008
Melanoma ( $n = 17$ )	0.00927	0.992614	< 0.001	0.00140	0.873101	< 0.001
UCC ( $n = 43$ )	0.00249	0.552252	< 0.001	0.00039	0.106262	0.031
<b>Validation cohorts</b>						
IPI trial <sup>†</sup> ( $n = 93$ )	0.007836	0.284626	< 0.001	0.00048	0.177689	< 0.001
Basket-study <sup>††</sup> ( $n = 28$ )	0.002563	0.226202	< 0.001	0.00054	0.126117	0.059
NSCLC trial <sup>§</sup> ( $n = 35$ )	0.002498	0.666749	< 0.001	0.00011	0.064815	0.134
Brain Mets <sup>#</sup> ( $n = 21$ )	0.007007	0.771479	< 0.001	0.00088	0.434810	< 0.001
Drug type	$k$	$R^2$	$p$	$k$	$R^2$	$p$
<b>Literature cohorts</b>						
Nivolumab ( $n = 81$ )	0.000786	0.324897	< 0.001	0.00006	0.064571	0.021
Atezolizumab ( $n = 43$ )	0.002489	0.552252	< 0.001	0.00039	0.106262	0.031
<b>Validation cohorts</b>						
IPI trial <sup>†</sup> ( $n = 93$ )	0.007836	0.284626	< 0.001	0.00048	0.177689	< 0.001
Basket-study <sup>††</sup> ( $n = 28$ )	0.002563	0.226202	< 0.001	0.00054	0.126117	0.059
NSCLC trial <sup>§</sup> ( $n = 35$ )	0.002498	0.666749	< 0.001	0.00011	0.064815	0.134
Brain Mets <sup>#</sup> ( $n = 21$ )	0.007007	0.771479	< 0.001	0.00088	0.434810	< 0.001

<sup>†</sup>IPI trial: drug type = ipilimumab, cancer type = metastatic to lung or liver, see Methods

<sup>††</sup>Basket-study cohort represents 18 cancer types and 7 drug types (see Table S4)

<sup>§</sup>NSCLC trial: drug type = pembrolizumab; cancer type = NSCLC

<sup>#</sup>Brain Mets: drug type = ipilimumab, nivolumab, or ipilimumab plus nivolumab (see Methods), cancer type = melanoma metastatic to the brain.

**Table 3:**

Highest observed values for  $1/(A \times \mu)$  with  $\rho_{\text{predicted}}^{\infty} < 1.0$  (responders) by cancer and treatment type. In each case, values above those reported were always found to have unfavorable response to immunotherapy with increasing tumour burden over time (non-responders). Threshold values identified in the calibration cohort are indicated by dashed lines in Fig. 5B,C, with colours corresponding to cancer and treatment types. Percent and total counts of patients sorted correctly by the threshold  $1/(A \times \mu) = 136.73$  are listed. (lit) denotes the literature-derived NSCLC cohort, and NSCLC trial denotes the in-house cohort.

Cancer Type	Highest $1/(A \times \mu)$ with $\rho_{\text{predicted}}^{\infty} < 1.0$	$1/(A \times \mu)$ threshold	$\rho_{\text{predicted}}^{\infty} < 1.0$ and below threshold	$\rho_{\text{predicted}}^{\infty} > 1.0$ and above threshold
<b>Literature cohorts</b>				
RCC ( $n = 23$ )	515.80	136.73	92% (11/12)	100% (11/11)
NSCLC (lit) ( $n = 41$ )	802.52	136.73	81% (13/16)	84% (21/25)
Melanoma ( $n = 17$ )	136.73	136.73	100% (12/12)	100% (5/5)
UCC ( $n = 43$ )	719.21	136.73	89% (16/18)	84% (21/25)
<b>Validation cohorts</b>				
IPI trial <sup>†</sup> ( $n = 93$ )	235.44	136.73	90% (19/21)	69% (50/72)
Basket-study <sup>††</sup> ( $n = 28$ )	77.81	136.73	83% (5/6)	64% (14/22)
NSCLC trial <sup>§</sup> ( $n = 35$ )	417.17	136.73	87% (13/15)	35% (7/20)
Brain Mets <sup>#</sup> ( $n = 21$ )	1769.94	136.73	75% (9/12)	67% (6/9)
Drug type	Highest $1/(A \times \mu)$ with $\rho_{\text{predicted}}^{\infty} < 1.0$	$1/(A \times \mu)$ threshold	$\rho_{\text{predicted}}^{\infty} < 1.0$ and below threshold	$\rho_{\text{predicted}}^{\infty} > 1.0$ and above threshold
<b>Literature cohorts</b>				
Nivolumab ( $n = 81$ )	802.52	136.73	90% (36/40)	90% (37/41)
Atezolizumab ( $n = 43$ )	719.21	136.73	89% (16/18)	84% (21/25)
<b>Validation cohorts</b>				
IPI trial <sup>†</sup> ( $n = 93$ )	235.44	136.73	90% (19/21)	69% (50/72)
Basket-study <sup>††</sup> ( $n = 28$ )	77.81	136.73	83% (5/6)	64% (14/22)
NSCLC trial <sup>§</sup> ( $n = 35$ )	417.17	136.73	87% (13/15)	35% (7/20)
Brain Mets <sup>#</sup> ( $n = 21$ )	1769.94	136.73	75% (9/12)	67% (6/9)

<sup>†</sup>IPI trial: drug type = ipilimumab, cancer type = metastatic to lung or liver, see Methods

<sup>††</sup>Basket-study cohort represents 18 cancer types and 7 drug types (see Table S4)

<sup>§</sup>NSCLC trial: drug type = pembrolizumab; cancer type = NSCLC

<sup>#</sup>Brain Mets: drug type = ipilimumab, nivolumab, or ipilimumab plus nivolumab (see Methods), cancer type = melanoma metastatic to the brain.