Integration of transcriptomics data into agent-based models of solid tumor metastasis

Jimmy Retzlaff[1, 2, 3], Xin Lai[1, 2, 3], Carola Berking[1, 2, 3] and Julio Vera[1, 2, 3]

1. Department of Dermatology, Universitätsklinikum Erlangen and Friedrich-Alexander-Universität Erlangen-Nürnberg, Laboratory of Systems Tumor Immunology, Erlangen, DE

2. Deutsches Zentrum Immuntherapie, Erlangen, DE

3. Comprehensive Cancer Center Erlangen-EMN, Erlangen, Bayern, DE

Abstract

Most of the recent progress in our understanding of cancer relies in the systematic profiling of patient samples with high throughput techniques like transcriptomics. This approach has helped in finding gene signatures and networks underlying cancer aggressiveness and therapy resistance. However, -omics data alone is not sufficient to generate insights into the spatiotemporal aspects of tumor progression. Here, multi-level computational models are promising approaches, which would benefit from the possibility to integrate in their characterization the data and knowledge generated by the high throughput profiling of patient samples.

We present a computational workflow to integrate transcriptomics data from tumor patients into hybrid, multi-scale models of cancer. In the method, we employ transcriptomics analysis to select key differentially regulated pathways in therapy responders and non-responders and link them to agent-based model parameters. We next utilize global and local sensitivity together with systematic model simulations to assess the relevance of variations in the selected parameters in triggering cancer progression and therapy resistance. We illustrate the methodology with a *de novo* generated agent-based model accounting for the interplay between tumor and immune cells in melanoma micrometastasis. Application of the workflow identifies three different scenarios of therapy resistance.

1 Introduction

In the last decade we have progressed remarkably in our understanding of cancer pathogenesis and metastasis, and this has had positive consequences in our ability to diagnose, stratify and treat metastatic tumors. In line with this, immune evasion is a hallmark of metastatic cancer (Hanahan and Weinberg 2011) and our deep understanding on its mechanisms of action has been vital to the development of therapies such as immune checkpoint inhibitors (ICI) for aggressive tumors like advanced melanoma (Wei et al. 2018), which greatly increased patient survival in metastatic melanoma patients (Larkin et al. 2015). A significant fraction of the progress in cancer research is due to the characterization of tissue samples from large cohorts of patients through genomics, transcriptomics, proteomics and/or epigenomics analysis. These techniques give access to quantitative data describing the activation and expression of (all) genes in cancer, thereby providing the information necessary

to investigate the genetic landscape of cancer progression (Akbani et al. 2015) and to reconstruct and dissect the gene regulatory networks underlying cancer pathogenesis and therapy response (Dreyer et al. 2018). However, -omics data alone cannot account for some relevant levels of (de)regulation happening in cancer, which are linked to spatiotemporal variations in the tumor's molecular and cellular composition, as well as to the existence of nonlinear regulatory structures like feedback and feedforward loops (Lai et al. 2016).

In this context, mathematical modeling and in particular multi-level spatial computational models are a viable method as they allow to investigate the dynamic behavior of the tumor microenvironment (TME) from hypothesized cell behavior and to evaluate therapeutic strategies (Metzcar et al. 2019). These models can describe and simulate the dynamics of cancer-deregulated intracellular gene circuits (Kirouac et al. 2017), but they can also be used to integrate genes and gene circuits activity into tissue-scale models of cell-to-cell interactions (Vera et al. 2013).

In such agent-based models (ABM), cells act as discrete individuals according to their set of rules. ABMs of cancer immune environments have been recently reviewed by Norton et al. (2019). There are some challenges to devising these computational models in the context of cancer: First, there is a trade-off between detailed modeling of biological features with different scales occurring in cancer progression and keeping the model simple enough to allow interpretability and reasonable computing effort. Second, these model types include a large amount of model parameters that require diverse, quantitative data to be characterized. In the case of tissue modeling, many parameters can only be calibrated indirectly, as the experimental modalities to observe single cell behavior *in vivo* are missing. We also need domain knowledge to decide which parameters are relevant for model simulations and investigating hypotheses, allowing to propose or optimize therapies.

In this paper, we propose to integrate transcriptomics analysis from cancer patient cohorts in the design and characterization of agent-based model simulations. To this end, we describe a method for analyzing transcriptomics data, ranking and selecting key gene sets underlying a condition of interest and linking these to selected agent-based model parameters. We utilize this information to prioritize parameters for investigation of the model behavior, as an analysis of the whole parameter space is computationally infeasible.~~,~~ The influence of these priorized parameters is investigated via global sensitivity analysis and massive, systematic model simulations. We exemplify the use of the method for a case study on melanoma metastasis and immunotherapy resistance. To this end, we built an agent-based model accounting for the interplay between tumor and immune cells in a micrometastasis.

2 Materials and Methods

In this work we followed a workflow sketched in Fig. 1. It contains several steps including model construction, exploration and calibration, linking enriched gene sets to parameters, and analysis. The steps we take are discussed in detail below.

2.1 Multi-level melanoma immunology model

The general modeling concept we used is to create a spatial agent-based model of the TME that interacts with a systemic compartment. This is a modeling approach that has been suggested in recent reviews (Norton et al. 2019; Metzcar et al. 2019) and was followed in other multiscale models as well (Santos and Vera 2020; Gong et al. 2017). We built our model based on knowledge of tumor immunology and signaling in cancer and melanoma (Abbas et al. 2014; Marzagalli et al. 2019). The model contains immune and tumor cells in the melanoma TME and their interactions, which can be either based on cell-cell contact or on intercellular communication through cytokines.

The model accounts for parts of the innate and adaptive immunity to tumors including immunosurveillance without considering memory and long-term immunity. For the immunosurveillance we assumed that the tumor antigens are not yet detected by the adaptive immune system. A sketch of the model and in particular the considered cell interactions are shown in Fig. 2. Specifically, the immune cells accounted for are cytotoxic T lymphocytes (CTLs), T helper cells (Th), B cells, regulatory T cells (Tregs), dendritic cells (DCs), macrophages and myeloid-derived suppressor cells (MDSCs).

We assumed that the communication between immune cells is mediated by cytokines and chemokines. The cell behavior is modeled as logical rules. Further, our model includes helper cells and suppressor cells as abstract cell types that account for immune cells that primarily have a regulatory role, such as CD4+ T cells, B cells and MDSCs. These cells influence the immune response via secreting cytokines.

We labeled the involved cytokines based on whether they have a primarily immunoenhancing or immunosuppressive effect and modeled two abstract surrogate cytokines accordingly: immunoenhancing (ENH) and immunosuppressive cytokine (SUP). ENH accounts for cytokines that increase the effectiveness of cytotoxic mechanisms like IFN-γ, as well as for chemoattractants for cytotoxic cells such as CCL3, CCL4, CCL5, CXCL9 or CXCL10. Examples for molecular species that have immunosuppressive effects are IL-10, TGF-β and IDO. The surrogate cytokines keep the model simpler by assuming that the cytokines do not have pleiotropic effects, although it has been shown that some cytokines may trigger both immunosuppressive and immunoenhancing effects (Donia et al. 2016). For instance, IFN-γ, a key regulator of the adaptive immune response, can trigger both the expression of major histocompatibility complex I (MHC-I) and of PD-L1. The former increases the recognition of cancer cells by T cells, while the latter inhibits the effector mechanism of T cells.

2.1.1 Cytokine diffusion

Cytokines are modeled using a continuum model that tracks cytokine concentrations rather than discrete molecules. Their diffusion is described by Fick's second law

$$\frac{\partial \boldsymbol{c}}{\partial t} = \boldsymbol{D}\Delta \boldsymbol{c} + \boldsymbol{f}$$

with the compounds' concentrations $\boldsymbol{c}$, the diffusion constants $\boldsymbol{D}$, and a term $\boldsymbol{f}$ accounting for secretion and degradation. Cytokine diffusion is solved using a finite difference method with the Euler forward method.

2.1.2 *In situ* cell populations

Cancer cells. A cancer cell population of 125 cells is seeded at the lattice center at the beginning of a simulation. As the tumor cells migrate and divide, they will spread over the tissue in the course of a simulation.

Cell motility is implemented as a random walk, allowing a cell to move to neighboring lattice positions with a probability $p_{\text{migration}}$. If the chosen position is already occupied by another cell, the cells either swap places or stay at their positions with equal probability. In general, it is assumed that cancer cells are less motile than immune cells.

Cancer cells can die with a probability p_{death}, which leads to their removal from the lattice. Dead cancer cells will leave debris that can be collected by dendritic cells and facilitate an immune response.

Cancer cells are considered to have uncontrolled replication potential and will attempt to divide after a fixed length of time $t_{\text{proliferation}}$ has passed, which accounts for cell growth and cycle. Dividing cells are temporarily immobile for the time step where the cell division occurs. Cell division can only take place if there is a vacant neighboring position that a daughter cell can occupy. This rule implicitly models cell contact inhibition, a trait that cancerous cells usually lose (Hanahan and Weinberg 2011). However, this assumption is in line with previous studies that use lattice-based models (e.g. Wang et al. 2013). It simplifies the modeling of cell mechanics, and evades calculations such as equilibria of forces between cells and the consequential possibility of cells pushing each other away.

Cancer cells may present one or multiple tumor-specific antigens depending on their mutations. We modeled only passenger mutations, meaning if a cancer cell mutates, it may start to present another antigen. This can induce an adaptive immune response specific to that antigen. Cancer cells mutate at each time step with a probability p_{mutation}. The mutations that are modeled are non-driver mutations, as they only affect the antigen pool that a particular cancer cell presents, which influences its susceptibility towards clearance by CTLs: more mutations lead to a larger antigen pool and recognition by different CTL clones. The mutations are modeled as a finite allele model, with 32 possible mutations. This allows to track the individual mutations, but has the disadvantage that it is not realistic compared to the near-infinite mutations possible in a real human genome, potentially leading to artifacts, e.g. through the possibility of reverse mutations.

Cytotoxic T lymphocytes. Cytotoxic T lymphocytes patrol the tumor site and are able to induce apoptosis in cancer cells upon contact. A cell is considered to be in contact with another if it is present in its Moore neighborhood (i.e. adjacent cell including diagonal adjacency). A

CTL recognizes an antigen that is specific to its receptor and kills cancer cells presenting the antigen. It probes its neighborhood for a recognizable cancer cell in random order. The randomness is introduced to avoid a direction bias that might lead to simulation artifacts. If the CTL recognizes a cancer cell, it kills it with a probability modeled as

$$p_{\text{kill}} = \left(\left(p_{\text{base}} + (1 - p_{\text{base}}) \cdot \left(1 - e^{-\tau_{\text{ENH}}[\text{ENH}]}\right)\right) \cdot e^{-\tau_{\text{SUP}}[\text{SUP}]}\right)^{g}$$

that depends on a base killing probability p_{base}, the local concentrations of SUP [SUP] and ENH [ENH], respective rate constants τ_{SUP} and τ_{ENH}, and the influence of anti PD1 checkpoint inhibitor therapy g. The probability is modeled in such a way that killing of tumor cells by CTLs becomes more effective in the presence of ENH and less effective in the presence of SUP. Anti PD1 therapy is modeled as a power law influence (Vera et al. 2007) and is set to 1 (no therapy) and can be toggled to 0.1 during a simulation (application of therapy). At this abstraction level of the model it is indifferent whether the drug targets PD1 which may be expressed by CTLs or its ligand PDL1 which may be expressed by the cancer cells, as it only influences the killing mechanism on contact. The killing has a duration t_{kill}, in which the CTL becomes immobile and unable to kill other neighboring cancer cells.

CTLs undergo apoptosis after a fixed life span $t_{\text{life,CTL}}$ expires. Note that CTL expiration accounts for different cell fates including exhaustion, apoptosis or leaving the TME. We do not model CTL proliferation in the TME, although it has been reported in cases of combination therapy (Spranger et al. 2014). Unlike cancer cells, a CTL follows its migration rule at every time step unless it is in an immobile state. Therefore, its motility depends only on the cell density in its vicinity. It is capable of performing both random walk and chemotactic migration, following the ENH gradient. By default, CTLs perform random walk and they change to chemotactic migration if the concentration of surrounding ENH cytokines exceeds a threshold. We modeled this threshold to prevent the CTLs from being sensitive to very low ENH concentrations. The CTLs moving in the chemotactic mode are in an activated state. The activated CTLs secrete ENH cytokines with a fixed rate r_{ENH}. ENH increase their cytotoxic capabilities and attract other CTLs to their vicinity, allowing fast finding and clearance of cancer cell colonies.

Dendritic cells. DCs are antigen-presenting cells to immune effector cells such as T cells. In the model, DCs function as probes for tumor cells, and they move inside the lattice at each simulation step if a vacant position in the vicinity is available. We did not consider apoptosis of DCs in the TME, as we assume that DCs are either tissue-resident or filtrated through the TME during their lifetime. A DC will collect all cancer cell debris it encounters. It then starts to present the antigens it processed. Furthermore, it becomes activated and leaves the tumor site. Once a DC has left the tumor site, it increases a signal that leads to a delayed recruitment of CTLs that are specific to the antigens it now presents. This way we implicitly modeled the homing of DCs to the tumor-draining lymph node. We assume homeostasis of the DC

187 population, and for every DC that leaves the tumor microenvironment, a new DC will be
188 recruited.

189 **Helper and Suppressor cells.** Helper cells account for CD4+ helper T cells as well as tumor
190 infiltrating B cells. They constantly secrete ENH cytokines (r_{ENH}). Suppressor cells primarily
191 account for regulatory T cells (Treg) and myeloid derived suppressor cells (MDSCs). Analog to
192 helper cells, they constantly secrete SUP cytokines. Both helper and suppressor cells perform
193 a random walk during their lifetime, which is fixed to $t_{\mathrm{life,helper}}$ and $t_{\mathrm{life,suppressor}}$, respectively.

194 **Macrophages.** Macrophages have cytotoxic capabilities, and can secrete both ENH and SUP
195 cytokines. Similar to CTLs, their cytotoxicity is influenced by cytokines. Their cytokine
196 secretion rates depend on the ratio of the concentrations of local SUP and ENH cytokines:

197 $$r_{\mathrm{SUP}} = \frac{[\mathrm{SUP}]}{[\mathrm{SUP}] + [\mathrm{ENH}]} \cdot r_{\mathrm{cytokine}}$$

198 and

199 $$r_{\mathrm{ENH}} = \frac{[\mathrm{ENH}]}{[\mathrm{SUP}] + [\mathrm{ENH}]} \cdot r_{\mathrm{cytokine}}.$$

200 The equations result in positive feedback loops, making the secretion rates of SUP and ENH by
201 macrophages positively correlate with their own concentrations. The feedback loops imitate
202 an environment-dependent phenotype plasticity that resembles M1 and M2 phenotype
203 activation described in the literature (Kim and Bae 2016).

204 2.1.3 Cell recruitment

205 Newly recruited cells appear on a free position at the border of the TME lattice based on the
206 assumption that recruited cells enter from nearby blood vessels. The cell types recruited to
207 the TME are CTLs, DCs, macrophages, helper and suppressor cells.

208 The recruitment of CTLs is preceded by DC-induced clonal expansion and differentiation in
209 the lymphatic tissues, which introduces a delayed response. Therefore, CTLs are recruited to
210 the TME with an antigen specific rate r_{CTL} that depends on delayed tumor detection of DCs.

211 The delay is modeled as a queue with a fixed size t_{delay} for each antigen. At each simulation
212 step, the oldest value will be dequeued and leads to recruitment of CTLs, while a new value
213 will be enqueued and initialized to zero. Every DC presenting the respective antigen that leaves
214 the tumor site at the simulation step will add a number to the new value in the queue, leading
215 to immune cell recruitment in the future. Helper cells are recruited alongside CTLs with a fixed
216 ratio of 1:1, which approximates reported data (Hernberg 1996).

217 Recruitment of suppressor cells and macrophages depend on CTL recruitment with fixed
218 ratios $q_{<\text{cell type}>}$ e.g.

219 $$r_{\mathrm{macrophage}} = q_{\mathrm{macrophage}} \cdot r_{\mathrm{CTL}}.$$

2.2 Model environment, simulation and parameterization

The lattice is modeled with cubic cells with a side length of $10\mu m$, with $100 \times 100 \times 100$ cells, representing a volume of $1 mm^3$. We set $\Delta t = 10\,\mathrm{min}$ for the duration of a simulation step, which is taken from a similar model by Gong et al. (2017). The fastest action a cell undertakes and that is affected by the step size is cell movement. The time step and cell length correspond with a maximum cell speed of $\sqrt{3}\,\mu m\,\mathrm{min}^{-1}$. The maximum cell speed is therefore about 10 times slower than the speed of neutrophils performing chemotaxis in a microfluidic device and about 5 times faster than H69 small cell lung cancer cells (Milo et al. 2010). We assume this depicts cell motion with an adequate speed assuming that leukocytes move more slowly in tissue than in a microfluidic device. Simulations start with a small homogeneous population of 125 cancer cells at the center of the lattice and random uniformly distributed populations of DCs and macrophages. We run simulations for a period of 100 days or 8401 steps, respectively. This period is chosen on the assumption that a successful immune response will clear the metastasis within 50 days, as is indicated for adaptive immune responses (Abbas et al. 2014). We doubled the simulation time to investigate the model progression of small residual cancer cell populations that many simulations showed at day 50. Further, a simulation will abort earlier if the cancer cell population grows larger than 800,000. This abortion condition is chosen to limit the computational effort of the simulations. We consider it justified as 80% of the lattice spaces will be occupied by cancer cells, effectively simulating a tumor expansion beyond the model space.

We used published experimental data to calibrate as many parameters as possible, whose annotation and nominal values are listed in supplementary Table S1. To determine the maximum recruitment rates, we use ratios of cell types that are described in literature, leaving but one recruitment rate uncharacterized. With this modeling choice we achieve that the simulated immune infiltrate resembles an infiltrate found in experiments over the course of a simulation.

Running a single simulation took about 1 hour and 20 minutes on our hardware (4 Intel Xeon E5-4660, 256GB RAM), requiring about 10 MB RAM. For the sensitivity analyses, we ran up to 64 simulations in parallel. We implemented our model in c++17, using HDF5 and json data formats for I/O. To account for the stochasticity in the model, we repeated each simulation multiple times with different seeds for the random number generator. To decide on a number of replicas to make for each simulation, we ran simulations with the nominal parameter configuration both with and without application of ICI therapy and compared the 95% confidence intervals of the expected cancer cell populations after 3 and 100 replications, and found that using a low number of replications is acceptable for our analysis (supplementary Fig. S1). For the local sensitivity we repeated each simulation 10 times, and for the global sensitivity analysis 3 times.

2.3 Linking differential regulation to model parameters

For computationally expensive model simulations, global sensitivity analysis is only feasible for a small subset of model parameters. Here we propose to base the selection on parameters linked to biologically relevant gene sets using a workflow as depicted in Fig 3. We performed a gene set enrichment analysis to identify and characterize a subset of them that is related to response to anti-PD1 treatment. First, we downloaded and processed the transcriptome data (GSE78220) of pre-treatment melanomas undergoing anti-PD1 checkpoint inhibition therapy. Second, we identified differentially expressed genes between responders and non-responders. Third, we performed a gene set enrichment analysis using the differentially expressed genes and identified gene sets in which the genes are involved. We assumed that the identified enriched gene sets are crucial for the pathogenesis and progression of melanoma, and therefore we manually annotated them with corresponding model parameters.

For the gene set enrichment analysis, we used the R package fgsea (Sergushichev 2016) that tested the enrichment of the identified differentially expressed genes using the MSigDB hallmark gene set collection (Liberzon et al. 2015) and cancer hallmark genes (CHG, Zhang et al. 2020). The fgsea algorithm searches for gene sets where highly ranked genes are enriched. It is given a ranked list of genes and a list of gene sets. We calculated gene ranks based on differential expression as

$$s_i = \mathrm{sign}(\mathrm{logFC}_i) \cdot (-\log_{10} p_i)$$

with the binary logarithmic fold change logFC_i and the p-Value p_i. To explore the enrichment, we examine the Benjamini-Hochberg adjusted p-value and list the significantly regulated gene set. These were manually annotated with model parameters they relate to, excluding those that could not be related (e.g. because of their generality or association to processes that are not modeled).

Next, we selected parameters for global sensitivity analysis based on the gene set enrichment and the local sensitivity analysis. In our case, we could afford to run a global sensitivity analysis for 5 parameters, while 6 parameters were identified with the gene set enrichment. To exclude one parameter from this selection, we performed a local sensitivity analysis (i.e. one at a time perturbation) and exluded the least influential parameter

2.4 Sensitivity analysis and decision tree based phenotype grouping

We selected five model parameters of interest for sensitivity analysis, which we set up in a quasi Monte Carlo fashion where we sample the selected parameter space using the Sobol' sampling sequence implementation of chaospy (Feinberg and Langtangen 2015). As boundaries for the parameter space we set (0, 2) times the nominal value. Assuming that about five simulations per parameter are needed to sufficiently cover the parameter space, we sample $5^5 = 3125$ parameter sets. As the model includes stochastic processes such as killing

or moving probabilities, we repeat each simulation 3 times, leading to a total number of 9,375 simulations for the sensitivity analysis.

To analyze the sensitivity of our parameter selection, we trained a decision tree as a meta-model, with the aim to find sensitive parameter subspaces and the parameter sensitivities on the simulation outcome (Hastie et al. 2009; Saltelli et al. 2008), a similar approach has been followed in an earlier work (Santos et al. 2018). To quantify the parameter sensitivities, we used partial rank correlation (Marino et al. 2008) and feature importances derived from the decision tree. As target variable we chose the cancer cell population at the end of the simulation, labeling the simulation results either as emerging metastasis (>700,000), complete remission (0) or else residual disease. This gives an indication of how good the immune response is in eliminating the emerging metastasis. There are some limitations though, as we can generally not assume that a simulation will reach a steady-state by its end.

Decision trees have the advantage that they can reproduce nonlinear and non-monotonous behavior, which is useful as we cannot assume a linear model behavior a priori. We use the scikit-learn implementation, which also calculates normalized parameter importances on the regression splits (Pedregosa et al. 2011). As an optimization criterion we chose the Gini impurity. The simulations were randomly split into a training and test data set (80:20) and 5-fold cross-validation was carried out on the training set yielding a mean accuracy of 0.9 and a standard deviation of 0.004. To avoid overfitting and to keep the decision tree easily human-interpretable we constrain it to a depth of 5 and a minimum split size of 3.3%.

3 Results

We developed a model of the immune reaction to melanoma that aims to account for the core cellular mechanisms influenced by the cytokine milieu. The TME is set up to simulate a newly seeded micrometastasis, where a cancer cell colony grows, is detected by immunosurveillance and challenged by both innate and adaptive immune responses, where CTLs are the cytotoxic actors.

3.1 Selection of the nominal model configuration

We calibrated most model parameters to data estimates from the literature (supplementary Table S1). To find values for the three parameters to which no data was found, we explored their parameter space to find a sensitive parameter set that we fixed as nominal values (supplementary Fig. S2).

In Fig. 4 we show simulations of the nominal parameter configuration and a configuration with reduced recruitment of immune cells with and without ICI treatment. This is motivated by findings that immune cell infiltration or CTL infiltration in particular is correlated with ICI treatment outcome (Li et al. 2021, Kümpers et al. 2019, Du et al. 2021, Nie et al. 2019), which we tested if our model would replicate. Model parameters have been randomly perturbed within the range of +/- 25% to account for patient diversity. It can be seen that for the nominal

parameter configuration, simulations without ICI lead to emerging metastases in almost any case. Simulations with low CTL infiltration and application of ICI lead to remission in 76/100 and to residual disease in 24/100 simulations. Simulations with ICI and the nominal (high) CTL infiltration lead to complete removal of cancer cells in 99/100 simulated cases and residual disease in one case. Taken together, the selection of parameter values for the nominal model configuration renders results that qualitatively match clinical evidence of patient response with high and low CD8 cell infiltration (Li et al. 2021, comp. Fig. 4C).

3.2 Transcriptomics data-driven selection of therapy-response related gene sets and their connection to model parameters

Gene set enrichment analysis using differential gene expression (anti-PD1 responders vs. non-responders) resulted in 24 significantly differentially regulated gene sets (Benjamini-Hochberg adjusted p-value <= 0.05). Fig. 5 demonstrates how we used this data to identify model parameters that are of particular interest, which we selected for the computationally expensive global analysis. For the parameter selection we considered the significantly differentially regulated gene sets. We annotated and mapped them to the corresponding model parameters. We excluded 14 gene sets that could not be linked to any model parameter either because they are generic, disease-specific, or not directly related to any modeled mechanism.

We link one gene set, “genome instability and mutation”, to the mutation probability p_{mutation}. Another, “epithelial-mesenchymal transition”, can be linked to cancer cell motility $p_{\mathrm{migration}}$. We further link three gene sets to the cell cycle time $t_{\mathrm{proliferation}}$ and one to the influence of SUP τ_{SUP}. Another two we link to the CTL recruitment rate r_{CTL}. and yet another two to the influence of ENH τ_{ENH} .

3.3 Parameter sensitivity analysis indicates multiple mechanisms of therapy resistance

To consider conditions of limited computational power in our analysis that would be found in the analysis of any large-scale ABM, we constrained the global sensitivity analysis to five parameters, excluding the enriched parameter p_{mutation}, which is the least influential of the selected parameters in the local sensitivity analysis (cf. supplementary Fig. S3). Global sensitivity analysis showed that a large proportion of simulations ended either with an emerging metastasis or complete remission (Fig. 6A). We therefore categorized the labels in emerging metastasis, residual disease and complete remission. The influence of the parameters, either described as partial rank correlation coefficients (prccs) or parameter importance of the trained decision tree is shown in Fig. 6.B/C. Both metrics agree that cancer cell cycle time and motility are more influential than the immune response-related CTL recruitment rate and the cytokine influences. The signs of the prccs follow the intuitive interpretation: long cell cycle times, high CTL recruitment rates and higher influence of ENH tend to lead to better removal of cancer cells, while higher cancer cell motility and higher influence of SUP lead to worse removal.

The decision tree (Fig. 6.D, supplementary Fig. 4) indicates that, on the one hand the simulated ICI is effective in multiple conditions of the TME, while on the other hand there are different mechanisms of therapy resistance. For the ICI resistant regions, the mechanisms affected by their decision paths are shown colored in the model draft. The tumor tends to be ICI resistant in the following cases: aggressive tumors with short replication time and at least moderate motility, tumors with longer replication time but high motility with smaller CTL recruitment and a case with higher CTL recruitment and longer cell replication time but high influence of SUP and high cancer cell motility. These findings show why it is so difficult to predict ICI response (Morrison et al. 2018), as they indicate that there is a spectrum of different counter-balancing mechanisms that influence its effectiveness.

4 Discussion

The aim of this paper is to develop an approach to integrate transcriptomic data into computational models of cell-to-cell interactions in cancer. There is abundant published material about integrating these types of data into unsupervised and supervised machine learning models for the classification and prediction of cancer patient samples. However, to date little has been done regarding merging these data with tissue-level mechanistic computational models, allowing for computer model-supported interpretation of patient data. To this end, we implemented a hybrid, agent-based model describing the interplay between cancer and immune cells in melanoma micrometastasis. To build and characterize the model, we used knowledge of melanoma immunology and publicly available quantitative data describing the behavior of the melanoma cells and different immune cells infiltrating the TME.

There are similar cancer models proposed in the literature. Wang et al. (2013) developed an agent-based melanoma model accounting for cytokine mediated angiogenesis. Hatzikirou et al. (2012) modeled tumor invasion with a lattice gas cellular automaton. Gong et al. (2017) modeled the tumor immune response to PD-1/PD-L1 inhibition. They identified tumor mutational burden and antigen strength as key factors that influence the recruitment of immune cells. They simulate therapy with checkpoint inhibition by changing a model parameter (probability of T cell suppression) at a set time point during a simulation, the same approach we use to model therapy. Compared to the model proposed in this work we do not model CTL proliferation at the TME (cf 2.1.2). Instead, our model considered a greater extend of cell types, including DCs, helper and suppressor cells, and variability in tumor antigens. While this increases the complexity of the model, it hypothetically allows for a more detailed projection of the differential regulation data into the model. The model has some limitations though, which arise from abstractions and simplification, as well as from incomplete knowledge on the cellular mechanisms. For instance, the finite allele mutation model does not replicate the significance of mutational burden on the prognosis of ICI (Morrison et al. 2018).

Here we combined gene set enrichment analysis of cancer immunotherapy response data and global sensitivity analysis of systematic model simulations as a method to constrain the

406 analysis to select model parameters (and connected biological processes) in computational
407 models with large parameter spaces, which cannot be analyzed as a whole due to limitations
408 in available computational power. In our case, a systematic exploration of the entire parameter
409 space would require about 5^{28} simulations, while the application of the method allowed us to
410 reduce the effort to 9,375 simulations and about 175 hours of computation. In contrast to
411 previous approaches to analyze such models, where parameter selection is performed solely
412 hypothesis driven or by requirement as calibration data is missing, our approach offers to
413 perform this in a data driven fashion. Based on our analysis, we hypothesize a causal
414 relationship between given differentially regulated gene sets and cell functions and
415 phenotypes associated with selected cell types in the micrometastasis. In the case of our
416 model, investigation of the model behavior restricted to the selected parameters displayed 3
417 different mechanistic scenarios of ICI treatment resistance. Key players to these mechanisms
418 are cancer cell motility, which has been previously shown (Dreyer et al. 2018), CTL infiltration,
419 which corresponds to the "warm" vs "cold" tumor hypothesis (Maleki Vareki 2018), and
420 suppressive signaling (TGFβ: Zhao et al. 2018, IDO for non-small-lung-cancer: Botticelli et al.
421 2018). A limitation of the method is that the linkage between parameters and gene sets
422 remains a manual curation step and depends on the vast expert knowledge of the modelers.
423 Further, it is possible that the enrichment analysis renders gene sets as relevant that are not
424 linked to any parameter in the current instance of the model. In this regard, one can utilize
425 the approach as a method for data-driven, systematic model expansion. For instance, in our
426 analysis three metabolism related genes set are enriched between responders and non-
427 responders, which might encourage to model more details on the cells metabolism and
428 expanding the model with nutrients. This would give the possibility to capture the interplay
429 between immune and metabolic processes.

430 Furthermore, the linkage as described here is based solely on gene set enrichment analysis
431 and remains a qualitative step, yielding only categorical classification of parameters rather
432 than quantitative differentiation. The latter would enable to deduce parameter perturbations
433 from the data directly, while here we just narrow a selection of parameters, which is
434 subsequently investigated in further detail in a global sensitivity analysis. In this regard, we
435 think that the method can be expanded to generate a quantitative link of transcriptomic data
436 to parameters by calculating a magnitude of the differential regulation and mapping it as a
437 perturbation level to the respective parameters relative to their nominal calibration. This
438 however requires annotation of all parameters with associated gene sets, which poses
439 intensive manual work, but could be supported by automated methods such as text mining.

440 We think that the combination of enrichment analysis of transcriptomics data and global
441 sensitivity analysis can be applied generally to agent-based or ODE models reflecting cell-to-
442 cell and tissue interactions in cancer and other pathologies (cf. Fig. 3). To this end, it is
443 necessary to have a significant amount of annotated transcriptomics data reflecting the
444 investigated conditions or progression of the disease. When selecting the number of

parameters to be explored, one has to consider a trade-off between sufficient sampling of the chosen sub-parameter space and keeping the required computational load in control.

Acknowledgements

The authors thank Martin Eberhardt and Christopher Lischer for useful discussion.

Data Availability

The code of the model generated for this work is publicly accessible on https://zenodo.org/record/6393283

Ethics

This article does not contain any studies involving human or animal participants.

Funding

This work has been supported by the German Ministry of Education and Research (BMBF) through the initiatives e:Med-MelAutim on cancer and autoimmunity [01ZX1905A] and KI-VesD on computation modelling and artificial intelligence guided cancer diagnostics [031L0244A]. We also received funding for our computer model-based research in melanoma from the Manfred Roth-Stiftung, the Trunk-Stiftung and the Matthias-Lackas-Stiftung.

Conflicts of Interest

The authors declare no potential conflicts of interest.

References

Abbas, A. K., Lichtman, A. H., and Pillai, S. (2014). *Cellular and Molecular Immunology*. Elsevier Health Sciences, eighth edition.

Akbani, R., Akdemir, K. C., Aksoy, B. A., Albert, M., Ally, A., Amin, S. B., Arachchi, H., Arora, A., Auman, J. T., Ayala, B., et al. (2015). Genomic classification of cutaneous melanoma. *Cell*, 161(7), 1681–1696.

Botticelli, A., Cerbelli, B., Lionetto, L. et al. Can IDO activity predict primary resistance to anti-PD-1 treatment in NSCLC? (2018). J Transl Med 16, 219. https://doi.org/10.1186/s12967-018-1595-3

Donia, M., Kjeldsen, J. W., and Svane, I. M. (2016). The controversial role of tnf in melanoma. *Oncoimmunology*, 5(4), e1107699.

Dreyer, F. S., Cantone, M., Eberhardt, M., Jaitly, T., Walter, L., Wittmann, J., Gupta, S. K., Khan, F. M., Wolkenhauer, O., Pützer, B. M., et al. (2018). A web platform for the network analysis of high-throughput data in melanoma and its use to investigate mechanisms of resistance to anti-pd1 immunotherapy. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1864(6), 2315–2328.

Feinberg, J. and Langtangen, H. P. (2015). Chaospy: An open source tool for designing methods of uncertainty quantification. *Journal of Computational Science*, 11, 46–57.

Gong, C., Milberg, O., Wang, B., Vicini, P., Narwal, R., Roskos, L., and Popel, A. S. (2017). A computational multiscale agent-based model for simulating spatiotemporal tumour immune response to pd1 and pdl1 inhibition. *Journal of The Royal Society Interface*, 14(134), 20170320.

Hanahan, D. and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *cell*, 144(5), 646–674.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer, second edition edition.

Hatzikirou, H., Basanta, D., Simon, M., Schaller, K., and Deutsch, A. (2012). ‚Go or grow': the key to the emergence of invasion in tumour progression?. *Mathematical medicine and biology: a journal of the IMA*, 29(1), 49-65.

Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook, J., Gillespie, M., Haw, R., et al. (2020). The reactome pathway knowledgebase. *Nucleic acids research*, 48(D1), D498–D503.

Kim, J. and Bae, J.-S. (2016). Tumor-associated macrophages and neutrophils in tumor microenvironment. *Mediators of inflammation*, 2016.

Kirouac, D. C., Schaefer, G., Chan, J., Merchant, M., Orr, C., Huang, S.-M. A., Moffat, J., Liu, L., Gadkar, K., and Ramanujan, S. (2017). Clinical responses to erk inhibition in braf v600e-mutant colorectal cancer predicted using a computational model. *NPJ systems biology and applications*, 3(1), 1–17.

Kümpers, C., Jokic, M., Haase, O., Offermann, A., Vogel, W., Grätz, V., Langan, E., Perner, S., Terheyden, P. (2019). Immune Cell Infiltration of the Primary Tumor, Not PD-L1 Status, Is Associated With Improved Response to Checkpoint Inhibition in Metastatic Melanoma. *Frontiers in Medicine,* 6, 2296-858X.

Lai, X., Wolkenhauer, O., & Vera, J. (2016). Understanding microRNA-mediated gene regulatory networks through mathematical modelling. *Nucleic acids research*, 44(13), 6019-6035.

Larkin, J., Chiarion-Sileni, V., Gonzalez, R., Grob, J. J., Cowey, C. L., Lao, C. D., Schadendorf, D., Dummer, R., Smylie, M., Rutkowski, P., et al. (2015). Combined nivolumab and ipilimumab or monotherapy in untreated melanoma. *New England journal of medicine*, 373(1), 23–34.

Li, F., Li, C., Cai, X., Xie, Z., Zhou, L., Cheng, B., ... & Liang, W. (2021). The association between CD8+ tumor-infiltrating lymphocytes and the clinical outcome of cancer immunotherapy: A systematic review and meta-analysis. *EClinicalMedicine*, 41, 101134.

513 Maleki Vareki, S (2018). High and low mutational burden tumors versus immunologically hot
514 and cold tumors and response to immune checkpoint inhibitors. *j. immunotherapy cancer* 6,
515 157. https://doi.org/10.1186/s40425-018-0479-7

516 Marino, S., Hogue, I. B., Ray, C. J., and Kirschner, D. E. (2008). A methodology for performing
517 global uncertainty and sensitivity analysis in systems biology. *Journal of theoretical biology*,
518 254(1), 178–196.

519 Marzagalli, M., Ebelt, N. D., and Manuel, E. R. (2019). Unraveling the crosstalk between
520 melanoma and immune cells in the tumor microenvironment. In *Seminars in cancer biology*,
521 volume 59, pages 236–250. Elsevier.

522 Masucci, M. T., Minopoli, M., and Carriero, M. V. (2019). Tumor associated neutrophils. their
523 role in tumorigenesis, metastasis, prognosis and therapy. *Frontiers in oncology*, 9.

524 Metzcar, J., Wang, Y., Heiland, R., and Macklin, P. (2019). A review of cell-based computational
525 modeling in cancer biology. *JCO clinical cancer informatics*, 2, 1–13.

526 Milo, R., Jorgensen, P., Moran, U.,Weber, G., and Springer, M. (2010). Bionumbers– the
527 database of key numbers in molecular and cell biology. *Nucleic Acids Res*, 38(Database issue),
528 750–753. BNID 106815, 112247.

529 Morrison, C., Pabla, S., Conroy, J.M., Nesline, M.K., Glenn, S. T., Dressmann, D., … & Ernstoff,
530 M.S. (2018). Predicting response to checkpoint inhibitors in melanoma beyond PD-L1 and
531 mutational burden. *Jounal of immunotherapy of cancer*, 6(1), 1-12.

532 Nie, R.-C., Yuan, S.-Q., Wang, Y., Chen, Y.-Y., Chen, S., Zhou, J., Chen, G.-M., Luo, T.-Q. Zhou,
533 Z.-W., and Li, Y.F. (2019). Robust immunoscore model to predict the response to anti-PD1
534 therapy in melanoma. *Aging (Albany NY)*, 11(23), 11576-11590.

535 Norton, K.-A., Gong, C., Jamalian, S., and Popel, A. S. (2019). Multiscale agentbased and hybrid
536 modeling of the tumor immune microenvironment. *Processes*, 7(1), 37.

537 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M.,
538 Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher,
539 M., Perrot, M., and Duchesnay, E. (2011). Scikitlearn: Machine learning in Python. *Journal of*
540 *Machine Learning Research*, 12, 2825–2830.

541 Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and
542 Tarantola, S. (2008). *Global sensitivity analysis: the primer*. John Wiley & Sons.

543 Santos, G. and Vera, J. (2020). The role of age, neutrophil infiltration and antibiotics timing in
544 the severity of streptococcus pneumoniae pneumonia. insights from a multi-level
545 mathematical model approach. *International Journal of Molecular Sciences*, 21(22), 8428.

Santos, G., Lai, X., Eberhardt, M., and Vera, J. (2018). Bacterial adherence and dwelling probability: two drivers of early alveolar infection by streptococcus pneumoniae identified in multi-level mathematical modeling. *Frontiers in cellular and infection microbiology*, 8, 159.

Sergushichev, A. A. (2016). An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *BioRxiv*, page 060012.

Smith, A. P., Hoek, K., and Becker, D. (2005). Whole-genome expression profiling of the melanoma progression pathway reveals marked molecular differences between nevi/melanoma in situ and advanced-stage melanomas. *Cancer biology & therapy*, 4(9), 1018–1029.

Spranger, S., Koblish, H. K., Horton, B., Scherle, P. A., Newton, R., and Gajewski, T. F. (2014). Mechanism of tumor rejection with doublets of ctla-4, pd-1/pd-l1, or ido blockade involves restored il-2 production and proliferation of cd8+ t cells directly within the tumor microenvironment. *Journal for immunotherapy of cancer*, 2(1), 3.

Vera, J., Balsa-Canto, E., Wellstead, P., Banga, J. R., and Wolkenhauer, O. (2007). Power-law models of signal transduction pathways. *Cellular signalling*, 19(7), 1531–1541.

Vera, J., Schmitz, U., Lai, X., Engelmann, D., Khan, F. M., Wolkenhauer, O., and Pützer, B. M. (2013). Kinetic modeling–based detection of genetic signatures that provide chemoresistance via the e2f1-p73/dnp73-mir-205 network. *Cancer research*, 73(12), 3511–3524.

Wang, J., Zhang, L., Jing, C., Ye, G.,Wu, H., Miao, H.,Wu, Y., and Zhou, X. (2013). Multi-scale agent-based modeling on melanoma and its related angiogenesis analysis. *Theoretical Biology and Medical Modelling*, 10(1), 41.

Wei, S. C., Duffy, C. R., and Allison, J. P. (2018). Fundamental mechanisms of immune checkpoint blockade therapy. *Cancer discovery*, 8(9), 1069–1086.

Fei Zhao, Kathy Evans, Christine Xiao, Nicholas DeVito, Balamayooran Theivanthiran, Alisha Holtzhausen, Peter J. Siska, Gerard C. Blobe, Brent A. Hanks (2018); Stromal Fibroblasts Mediate Anti–PD-1 Resistance via MMP-9 and Dictate TGF β Inhibitor Sequencing in Melanoma. *Cancer Immunol Res*; 6 (12): 1459–1471. https://doi.org/10.1158/2326-6066.CIR-18-0086

Figures

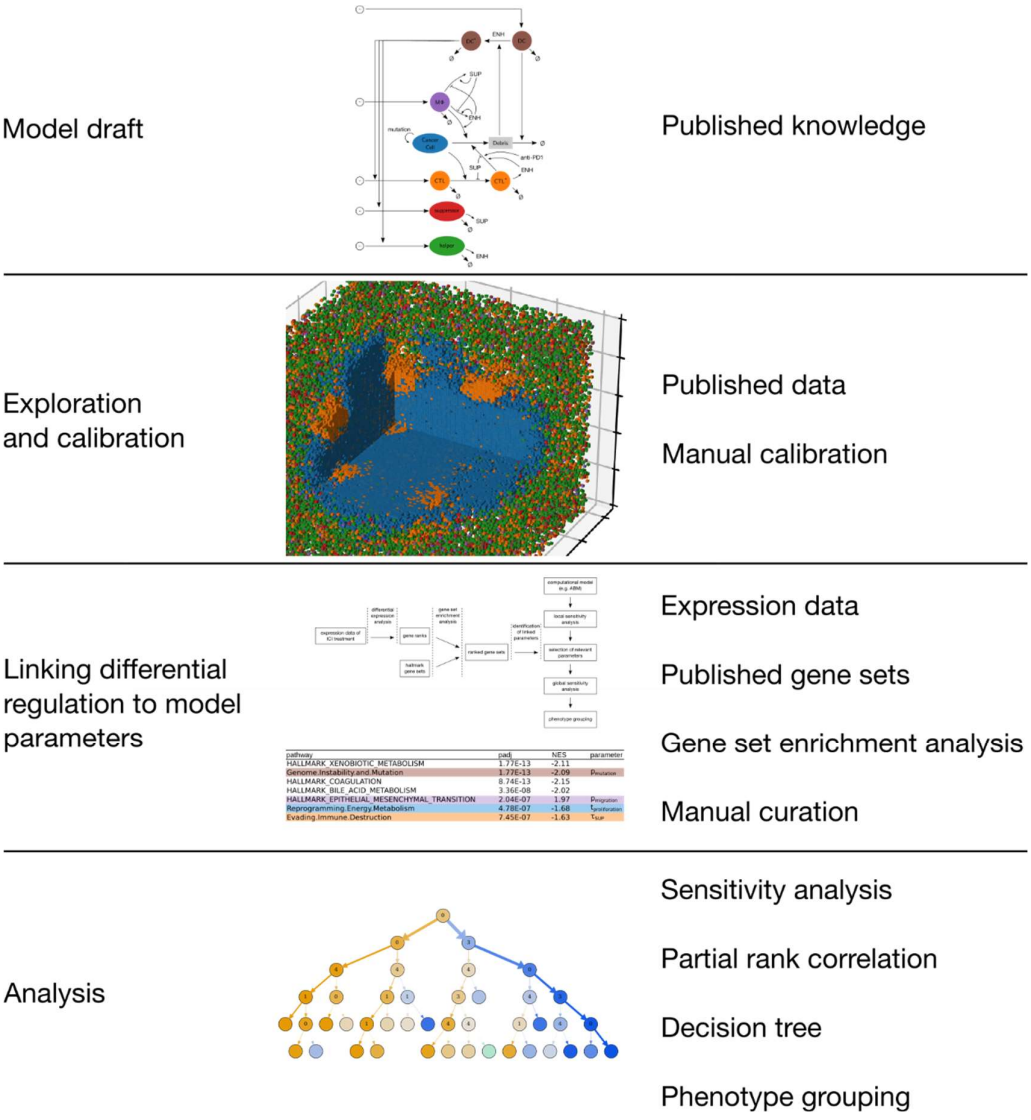
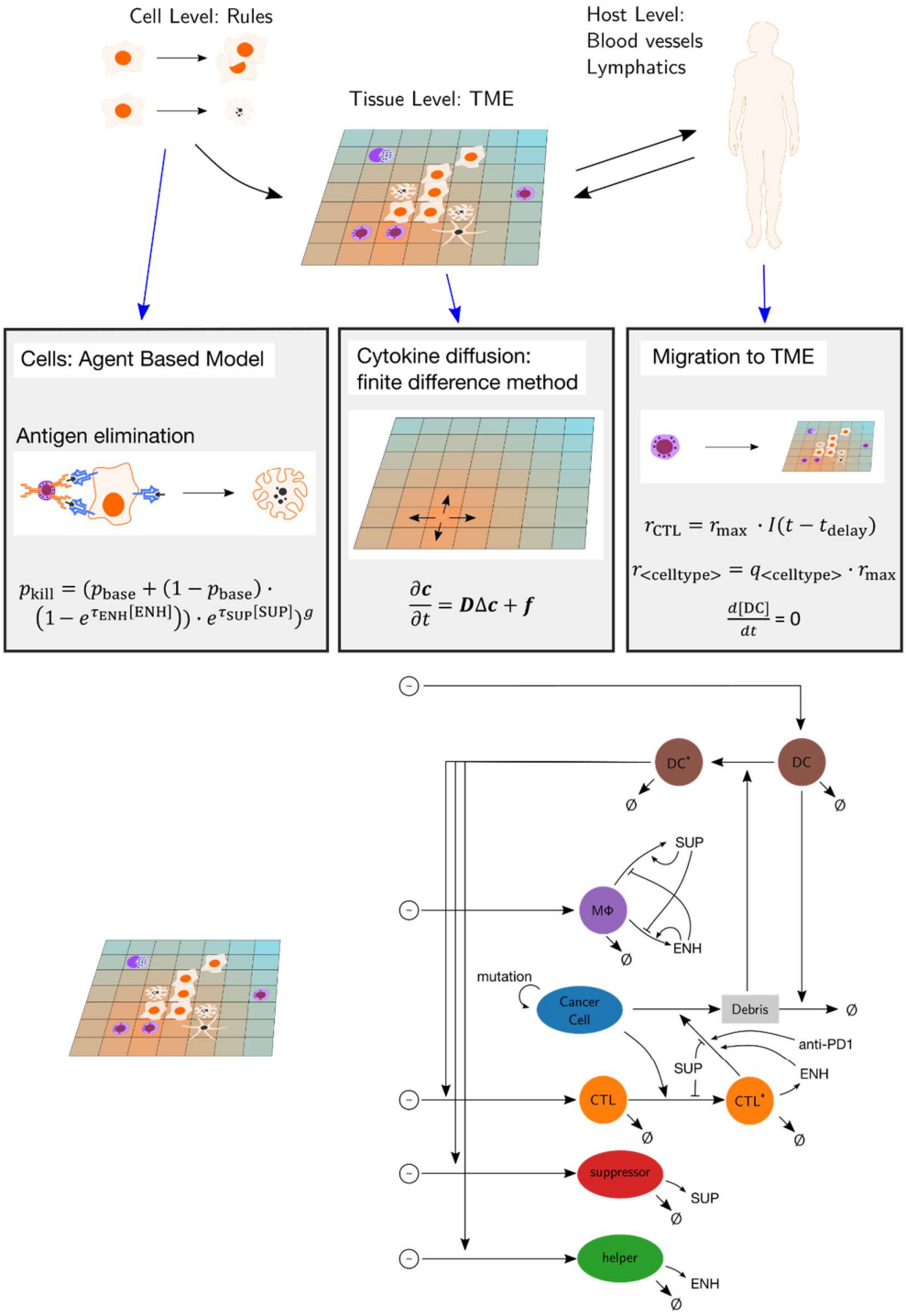


Fig. 1. Workflow of the study. On the left are labels of the four overarching steps, on the right is a brief list of the central materials and methods used in each step. We first created the model and set a nominal parameter configuration. Then we linked expression data to model parameters to narrow a selection of parameters whose influence we analyzed in more depth.

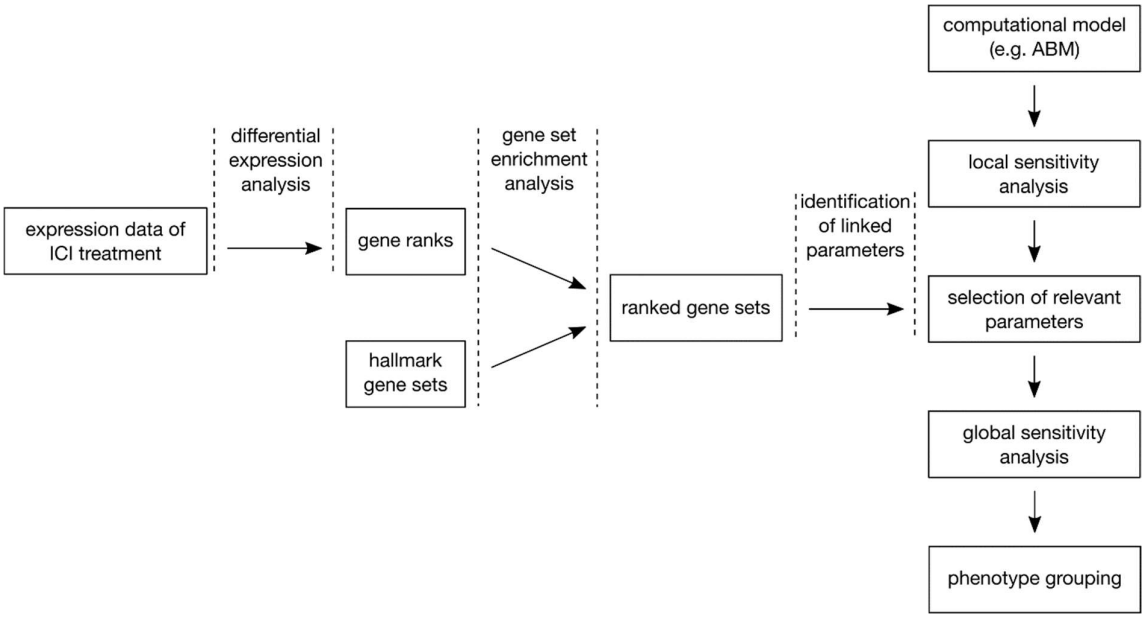


581

582 Fig. 2: Concept of the model. On the cellular level, an agent-based model is used, where cell
583 physiology is described as cell type specific rules. This is coupled on the level of the tumor-
584 microenvironment with the cytokine diffusion solver and the recruitment model of immune

585 cells. Cancer cell debris is detected by DCs which will attract CTLs, suppressors and helpers
586 after a delay. CTLs that contact cancer cells switch to an activated state and can kill the cancer
587 cell. The killing probability is influenced by cytokines and applied anti-PD1 therapy.
588 Abbreviations: ENH: immunoenhancing cytokine, SUP: immunosuppressive cytokine, DC:
589 dendritic cell, MΦ: macrophage, CTL: cytotoxic T lymphocyte.

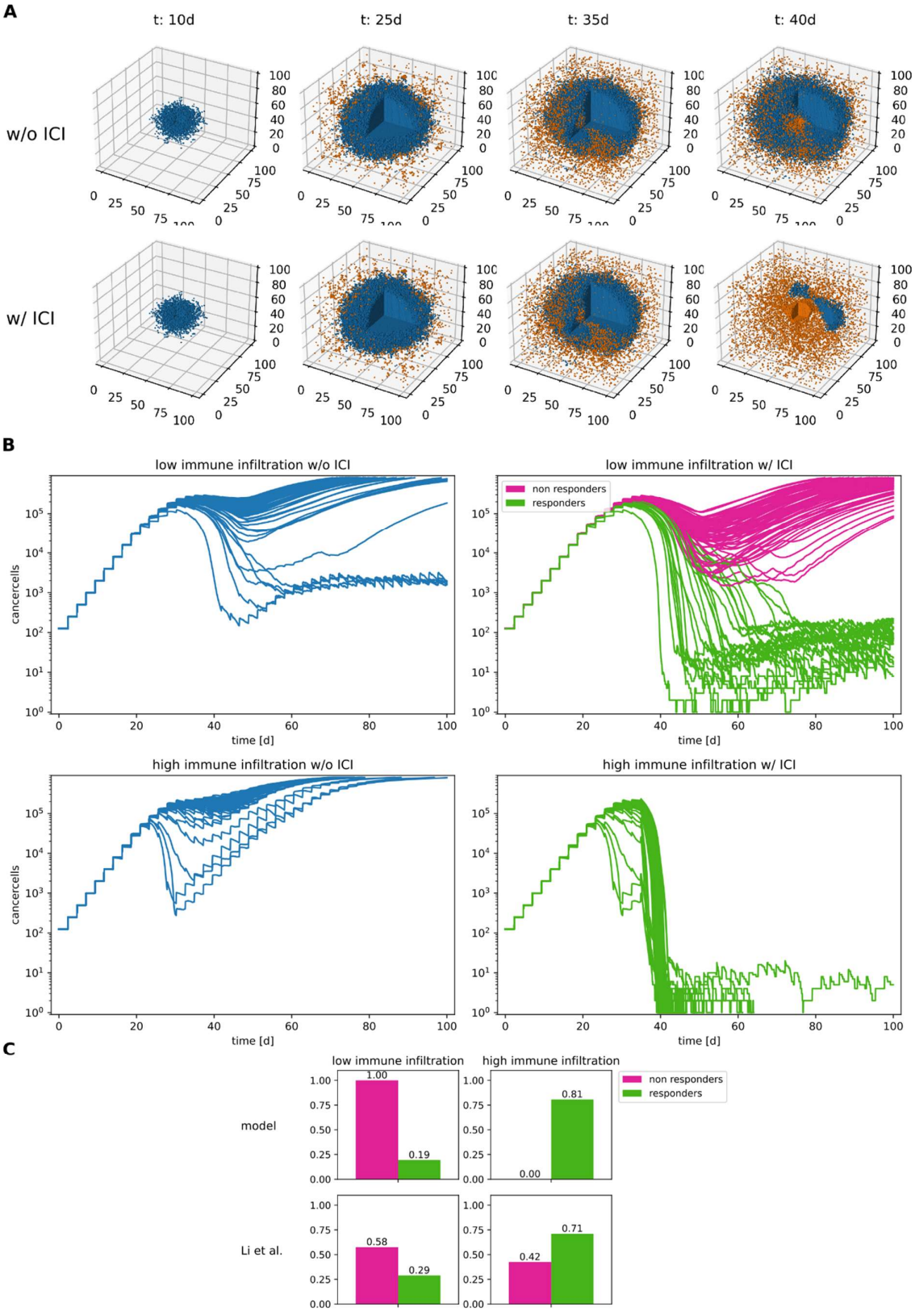
590



591

592 Fig. 3. Method of linking expression data to model parameters. Besides performing a local
593 sensitivity analysis to preselect a set of parameters for global analysis, we propose to enrich
594 expression data of different conditions to link them to parameters of potential biological
595 relevance.

596



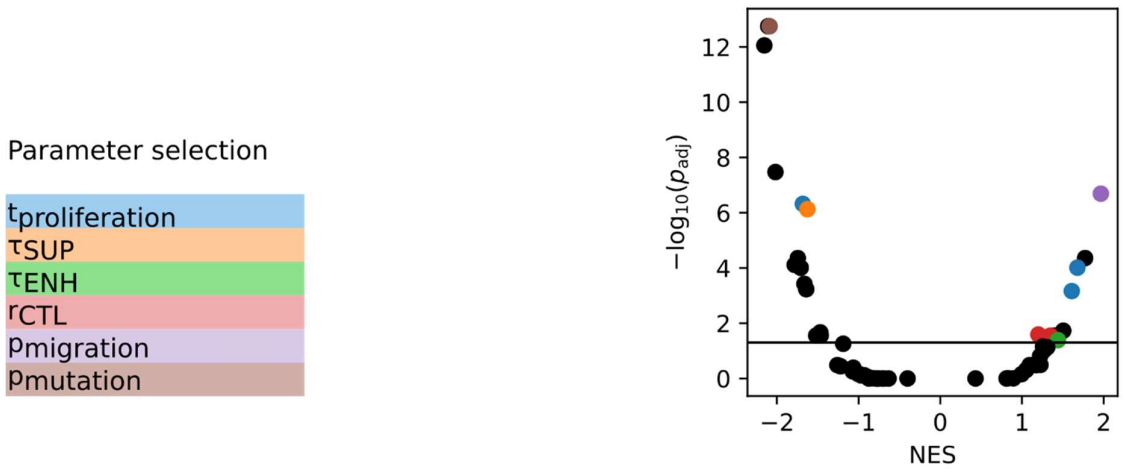
597

598 Fig. 4. Simulations of the nominal parameter configuration with and without ICI therapy. A:
599 comparison of the cell lattices over the course of a simulation. Only cancer cells (blue) and
600 CTLs (orange) are shown for clarity B. Cancer cell populations of 100 simulations per condition

601 with randomly perturbed model parameters (+/- 25%). High immune infiltration marks the
602 nominal configuration, low immune infiltration simulations have a 3-fold reduced immune cell
603 recruitment rate. The simulations with ICI are designated “responders” or “non-responders”
604 depending on their final cancer cell population. C. Qualitative comparison of the conditions
605 with fractions of anti-PD1 responders and non-responders with high and low CD8 infiltrates
606 respectively.

607

pathway	padj	NES	parameter
HALLMARK_XENOBIOTIC_METABOLISM	1.77E-13	-2.11	
Genome.Instability.and.Mutation	1.77E-13	-2.09	$p_{mutation}$
HALLMARK_COAGULATION	8.74E-13	-2.15	
HALLMARK_BILE_ACID_METABOLISM	3.36E-08	-2.02	
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	2.04E-07	1.97	$p_{migration}$
Reprogramming.Energy.Metabolism	4.78E-07	-1.68	$t_{proliferation}$
Evading.Immune.Destruction	7.45E-07	-1.63	τ_{SUP}
HALLMARK_P53_PATHWAY	4.37E-05	-1.74	
HALLMARK_E2F_TARGETS	4.37E-05	1.77	
HALLMARK_FATTY_ACID_METABOLISM	7.72E-05	-1.78	
HALLMARK_COMPLEMENT	9.74E-05	-1.71	
HALLMARK_KRAS_SIGNALING_DN	9.74E-05	-1.71	
HALLMARK_G2M_CHECKPOINT	9.74E-05	1.68	$t_{proliferation}$
HALLMARK_MYOGENESIS	3.77E-04	-1.66	
HALLMARK_ESTROGEN_RESPONSE_LATE	5.89E-04	-1.64	
HALLMARK_MITOTIC_SPINDLE	6.85E-04	1.61	$t_{proliferation}$
HALLMARK_UV_RESPONSE_DN	1.87E-02	1.51	
HALLMARK_ESTROGEN_RESPONSE_EARLY	2.14E-02	-1.47	
Tumor.Promoting.Inflammation	2.57E-02	1.20	r_{CTL}
HALLMARK_APICAL_JUNCTION	2.78E-02	-1.46	
HALLMARK_TNFA_SIGNALING_VIA_NFKB	2.79E-02	1.40	τ_{ENH}
HALLMARK_PEROXISOME	2.80E-02	-1.52	
HALLMARK_INFLAMMATORY_RESPONSE	2.82E-02	1.35	r_{CTL}
HALLMARK_INTERFERON_ALPHA_RESPONSE	4.10E-02	1.44	τ_{ENH}



608

609 Fig. 5. Central step of the proposed method to select model parameters that are related to
610 differential regulation in benign vs. malignant tissue. Top: Significantly differentially regulated
611 gene sets in melanoma sample of different ICI treatment response that could be linked to
612 model parameters. See full list in supplementary Table 2. Bottom left: List of the identified
613 parameters connected to the selected gene sets. Bottom right: volcano plot of the adjusted p
614 value against the normalized enrichment score. Gene sets accounting for parameters are
615 marked with corresponding colors.

616

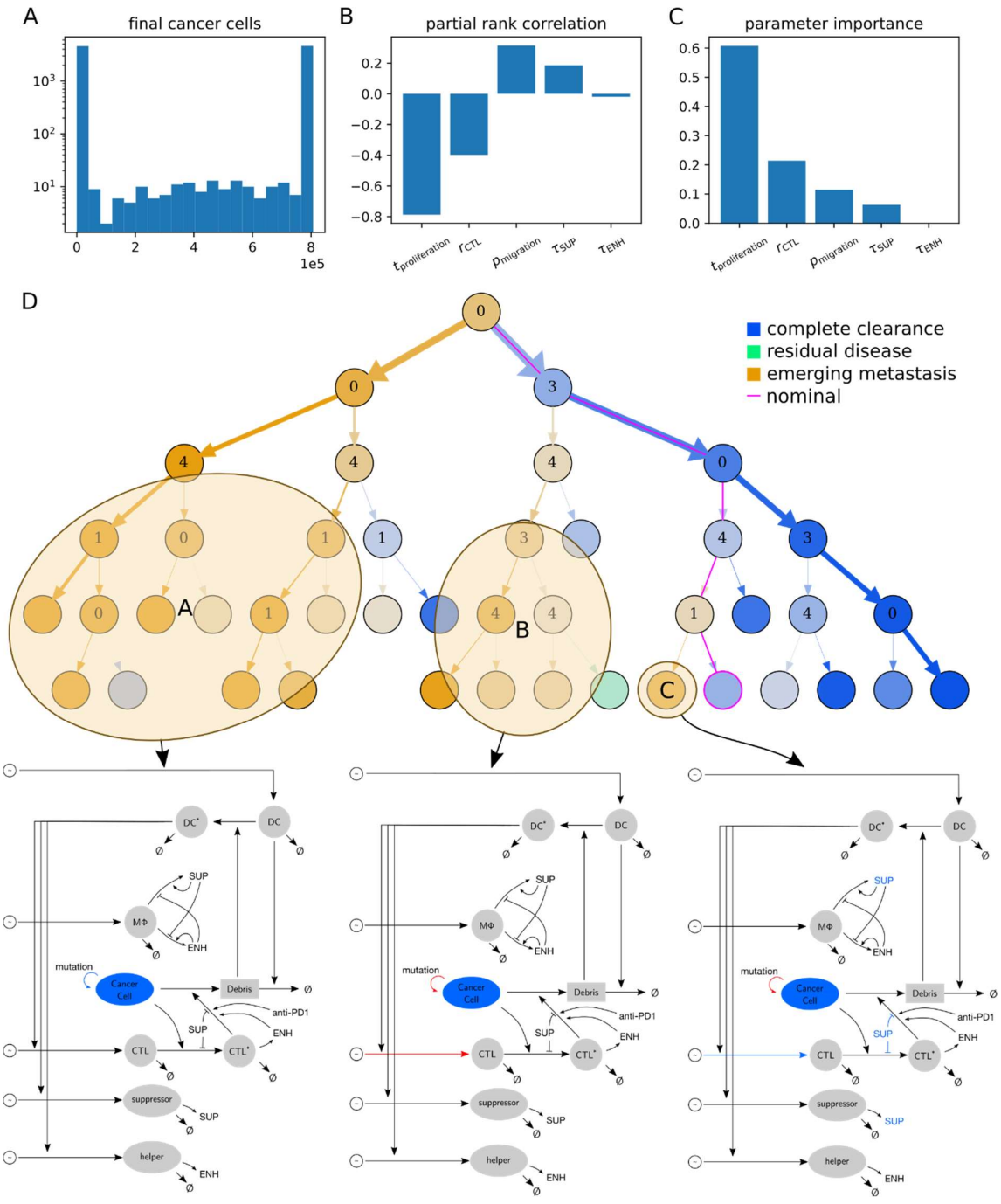


Fig. 6. A: Distribution of final cancer cell populations. B: Partial rank correlation coefficients. C: Parameter importances of the decision tree. D: Decision tree for the final cancer cell population. Each node predicts a simulation outcome for a region of the sampled parameter space. At a branch parent node, the parameter space is split along a threshold for a split parameter. The numbers in the parents refer to split parameters: 0: cell cycle time of cancer cells $t_{\text{proliferation}}$, 1: influence of immunosuppressive cytokines τ_{SUP}, 3: CTL recruitment rate r_{CTL}, 4: migration probability of cancer cells $p_{\text{migration}}$. The color saturation indicates node impurity. Model drafts are shown for highlighted ICI resistant parameter subspaces with color

626 indicating model components influenced by parameter deviations from the nominal
627 configuration (blue: lower/ red: higher values).

628

629