How to use transcriptomic data for game-theoretic modeling of treatment-induced resistance in cancer cells? A case study in patient-derived glioblastoma organoids

Weronika Gąska*, Christer Lohk*, Maikel Verduin, Christopher Hubert, Marc Vooijs, Rachel Cavill**, Kateřina Staňková**

* joint first authors, ** joint last authors

Abstract

Game theory is a powerful theory to model strategic decision-making, but also complex biological systems, such as cancer. In the past decades, game-theoretical models have helped to understand cancer, its response to various treatments, and to design better therapies. However, to fully utilize the potential of game-theoretical modelling in designing better anti-cancer therapies, we need more information on cancer population (ecological) and strategy (evolutionary) dynamics in response to treatment. Here we explore how transcriptomics data can be utilized as inputs of game-theoretic models for predicting evolutionary response to irradiation using patient-derived glioblastoma organoids. For that purpose, we utilize both supervised and unsupervised methods to disentangle cell mixtures from publicly available datasets and identify proportions of relevant cancer cell types in patient-derived glioblastoma organoids treated with radiotherapy. We then fit these proportions to the replicator dynamics, the most classic evolutionary game dynamics, to predict both transient evolutionary dynamics and evolutionary stable cell proportions in these organoids. Hereby we demonstrate a methodology that can be utilized to design evolutionary therapies, i.e. therapies that anticipate and prevent evolution of therapy resistance in cancer cells. Our predictions in glioblastoma suggest that hypoxia is the most important factor in identifying short-term response to irradiation, while this seems much less relevant for the long-term response. Further, we conclude that supervised methods are the best way to estimate cancer evolutionary dynamics when therapy resistance is a qualitative evolutionary trait. We provide a pipeline for designing better therapies through testing evolutionary responses in patient-derived organoids, while in parallel the ecological response can be tracked through serum biomarkers and imaging in the corresponding patients.

Keywords: Qualitative resistance, evolutionary dynamics, game theory, transcriptomics, glioblastoma, patient-derived organoids, radioresistance

1 Introduction & Motivation

The biggest obstacle in cancer treatment is therapy resistance [37, 50, 41]. This is especially problematic in advanced and/or metastatic cancers where the cancer is more widespread and the most common cause of cancer death [5, 11, 21]. The discovery of strategies to combat this evolution of resistance in cancers, or to at least prolong life would be a major step forward in cancer treatment [30, 29].

Game theory can help us to design evolutionary (or adaptive) therapy, which is a therapy anticipating and preventing evolution of therapy resistance in cancer cells [12, 14, 51, 39]. Strategies based on game-theoretic foundations have already been shown to be effective in clinical trials for metastatic castrate-resistant prostate cancer [25, 55, 15, 6, 7, 54]. Based on the success of this trial, new clinical trials have been opened on this type of therapy, often led by game-theoretic modelling [51, 42, 54].

In Zhang's metastatic castrate-resistant prostate cancer trial, the prostate-specific antigen (PSA) was used as a proxy for the tumor size in the course of the treatment [15]. Because of the lack of information on the composition of the tumors, possible initial tumor compositions were estimated from in-vitro measurements and expected interactions among cancer cells of different types [25, 55]. This has led to 22 different possible parametrizations of the model, differing in their predictions regarding proportion of therapy resistant cells in the evolutionary stable equilibrium. These 22 cases were clustered into 3 different groups: best-responder, responder, and nonresponder [25, 55, 6, 7]. In simulations, Zhang's protocol was shown to be more effective than the maximum tolerable dose corresponding to the standard of care for all these parametrizations [55, 6, 7]. This was confirmed in the trial itself [15, 54]. In the absence of detailed information on tumor composition, the treatment protocol was designed such that it was effective for all possible initial tumor compositions. These different compositions had to be considered as tumors with the same size but different composition will likely respond to the same treatment differently [41]. More precise information on the tumor composition in the course of treatment is needed to design even more effective evolutionary therapies. A rough estimate of tumor composition may be made from cancer cells' response to initial treatment application, as proposed in [42]. Another way, explored in the current study, is to utilize transcriptomics data from patient-derived organoids to model and predict how the proportions of different cancer cell types change over time in response to the treatment.

Ideally, single cell technologies would give us this information, however these techniques are prohibitively expensive, both in monetary terms and in terms of the computational and expert bioinformaticians' time that are needed to deal with the vast amounts of data these techniques produce [20]. Despite this, upfront costs are dropping and some studies have been conducted using this technology to study particular cancers e.g. [44]. If costs either in processing time/effort or money were to remain high this would prevent their routine use in the clinic, especially in scenarios where patients need long term regular monitoring to track the status of their tumor.

Therefore, here we will focus on techniques to obtain this information from bulk sequencing data, where data is obtained from many cells simultaneously using standard omics techniques, such as transcriptomics [43]. These deconvolution methods come in two distinct classes, supervised - where the training set is furnished with the known proportions of the cell types, or the profile of each cell type is known in advance and unsupervised - where nothing is assumed about the proportions nor the profiles of the cell types in advance of the analysis [56]. In this paper we test a range of approaches from each class.

We compare supervised and unsupervised methods on the data from patient-derived glioblastoma organoids to obtain cell profiles and proportions for the cell types present in glioblastoma [18, 47]. The organoids provide a platform on which we can test potential treatments and observe the treatment effects *in vitro*. With organoids, where the total tumor cell population is kept constant, we can track how treatment impacts the cell type composition over time without needing invasive biopsies. The tumor burden change over time can be tracked in the patient blood biomarker and imaging data. This makes organoids an ideal testing ground for the game-theoretic models. After identifying different cell types in organoids and how their proportions change over time in response to radiotherapy, we fit the most classical game-theoretic model, replicator dynamics, to this time series data. We hereby provide a methodology to utilize bulk data for estimating cancer evolutionary dynamics, which will then lead to better evolutionary therapies.

2 Methods

2.1 Datasets

In this work we make use of several datasets, which we describe below.

2.1.1 TGCA Glioblastoma dataset

The Cancer Genome Atlas (TCGA) is a publicly available resource containing molecular data for over 10000 specimens, from at least 33 different types of cancer and metadata linking this information to clinical data for the patients [16] . We used the TCGA-GBM project, which focuses on glioblastoma samples. We used 168 samples with available RNA-SEQ data. For each sample we retrieved the FPKM-UQ normalised RNA-SEQ. These files provide measurements for over 60000 RNA sequences per sample.

2.1.2 Neftel's single-cell data

We utilized Neftel et al.'s single-cell dataset acquired through full-length scRNA-SEQ (Smart-seq2) [36]. We downloaded this dataset from NIH Gene Expression Omnibus (GEO: GSE131928). It contains preprocessed single cell RNA-seq expression profiles of adult and paediatric IDH-wildtype Glioblastomas from 28 tumors with an average of 5730 genes detected per cell. From the initial 5730 genes, 289 were selected for cell type distribution analysis based on meta-module signatures. scRNA-SEQ data were complemented with metadata from Broad Institute Single Cell Portal (Study: Single cell RNA-seq of adult and pediatric glioblastoma) containing meta-module (NPC1, NPC2, MES1, MES2, OPC, AC) scores for 7930 cells. Cells were categorized according to their highest corresponding meta-module scores.

2.1.3 Organoid data

The patient-derived organoids were grown as per the protocol in [18] and then then treated with either 4 Gy or 10 Gy irradiation. Samples were collected pre-treatment, directly post-treatment, and at 24 and 72 hours post-treatment. RNA was isolated usin Nucleospin RNA (Macherey-Nagel) according to the manufacturers protocol. RNA concentration and integrity was confirmed using 2100 Bioanalyzer (Agilent Technologies) . RNA was hybridized to affymetrix human microarrays by Eurofins (Denmark). All conditions were tested in GBM organoids derived from 3 patients.

2.2 Unsupervised methods

For the unsupervised methods, we focus on non-negative matrix factorisation (NMF), a method which describes the data matrix as a product of two matrices, one describing the cell type proportions and the other describing the profiles of each cell type. This assumes that the data can be approximated simply by taking a weighted sum of the cell type profiles where the weights are the proportions of each cell type present [27].

The deconvolution of cancer cells can be seen as a specific case of the blind source separation (BSS) issue [57], i.e. separating a group of source signals from measurements of mixed signals. A cancer tumor, as a heterogeneous mixture of different cell types, can be analysed using a linear BSS model, as this model does not require knowledge of any predefined cell type signatures and can therefore be used to formulate the deconvolution problem. Whilst there are various methods which have been applied for unsupervised deconvolution of bulk cell profiles [3, 10, 24], in earlier unpublished work we showed that Non-Negative matrix factorisation (NMF) is by far the most promising approach that agrees with the literature findings [57, 24]. Therefore, here we focused on the family of non-negative matrix factorisation-based approaches. In this study we applied several NMF algorithms [4, 26, 38, 3, 23]. The number of cell types selected should correspond to the NMF rank for which the best results are obtained. To evaluate the obtained results, we used several classic indicators (Cophenetic coefficient [4], Dispersion coefficient [23], Residual Sum of Squares [19], Silhouette Coefficient [53]), whose values indicated the optimal factorisation rank.

2.3 Supervised methods

For the supervised approaches we use Scaden (Single-cell Assisted Deconvolutional Network), a reference-based deconvolution framework to determine the cellular composition of glioblastoma

patient-derived organoids from microarray profiles [31]. As a reference, we rely on single-cell profiles and a blueprint suggested by Neftel et al. that assigns cells into four meta-modules, each characterized by a unique gene expression signature and genetic alterations [36].

Scaden is a deep learning framework that utilizes scRNA-SEQ expression data to train a model for the approximation of cell-type proportions from an input bulk expression profile [31]. The training requires vast amounts of bulk RNA-SEQ expression profiles with corresponding cell type proportions as ground truth. This was achieved through random subsampling of 6750 single-cell profiles from Neftel's scRNA-SEQ dataset into 1000 pseudo-bulk RNA-SEQ gene expression profiles containing sum of expression values by gene. After the training set generation, genes with zero expression across all samples or with variance less than 0.1 were removed from bulk expression and training profiles. Furthermore, only genes that are present in both, training and inference datasets were used by the model. Next, the input data (both bulk expression profiles and training data) were transformed to the logarithmic space and scaled between 0 and 1. The model consists of three independent neural networks which are separately trained with preprocessed training data for 5000 training steps. The training was followed by inference to approximate cell type proportions from input bulk expression profiles (in addition to bulk RNA-SEQ, Menden et al. [31]) demonstrated reliable performance with a microarray dataset. The predictions for cell-type proportions were acquired from all three networks and averaged for the final output.

2.4 Game-theoretic model of glioblastoma

Changes in population characteristics such as size and density of cancer cells are associated with cancer ecological dynamics, while changes in heritable cell lineage characteristics (quantitative traits) or frequency of cancer cell types (qualitative traits) are considered evolutionary dynamics. This eco-evolution in the tumor can be captured with a use of various mathematical models. As the total cell population in organoids remains constant through the experiment, we analyze only changes in evolutionary dynamics. We utilize one of the most general approaches to integrate time scales of evolutionary dynamics; replicator dynamics [45, 17], which is the most classic game-theoretic dynamics used in modelling many biological processes as well as selection in population genetics. Given a finite set of cancer strategies $\mathcal{T}$, here understood as a set of cell types, their fitness matrix and their proportions at time t, the dynamics of cells of the same type can be described by deriving the following dynamics:

$$\dot{x}_i = x_i \left((Ax)_i - x^T Ax \right), \tag{1}$$

where x_i denotes the proportion of type $i \in \mathcal{T}$, and consequently the column vector of proportions of all types is denoted by $\mathbf{x}$. The $n \times n$ fitness matrix $A = (a_{ij})$ defines the fitness of the considered cell types. If the entries of this matrix are between 0 and 1, each value a_{ij} defines the probability that a focal cell of type $i \in \mathcal{T}$ will produce an offspring of its own type when interacting with a cell of type $j \in \mathcal{T}$. We can calculate the so-called evolutionary stable strategies (ESSs) of A, which maximize the fitness of the entire population and are uninvadable by a rare mutant with a different strategy. These ESSs may correspond to the attractors of the replicator dynamics (1).

3 Results

3.1 Results of unsupervised methods

As explained in Section 2.2, when applying the BSS model to define a deconvolution problem there is no prerequisite on prior knowledge regarding cell type signatures. In the absence of this information, one has to start the deconvolution by determining the number of cell types in the cancer mixture and only then determine their profiles and proportions. Here we created several simulated data sets and tested them with different NMF algorithms [4, 26, 38, 3, 23]. We used measures detailed in Section 2.2 to evaluate performance of these algorithms. These indicators on the simulated dataset mainly pointed to the actual value of the factoring rank. To disentangle

Cell type	Pathway	Adjusted p-value
1	Cytokine-cytokine receptor interaction	$2.86 \cdot 10^{-9}$
	Viral protein interaction with cytokine and cytokine receptor	$1.40 \cdot 10^{-7}$
	Staphylococcus aureus infection	$1.37 \cdot 10^{-5}$
2	Salmonella infection	$2.10 \cdot 10^{-5}$
	Endocytosis	$5.53 \cdot 10^{-5}$
	MicroRNAs in cancer	$6.41 \cdot 10^{-5}$
	Cell cycle	$1.76 \cdot 10^{-4}$
	Human T-cell leukemia virus 1 infection	$2.02 \cdot 10^{-4}$
	Protein processing in endoplasmic reticulum	$2.37 \cdot 10^{-4}$
	Amyotrophic lateral sclerosis	$4.72 \cdot 10^{-4}$
3	Neuroactive ligand-receptor interaction	$1.26 \cdot 10^{-19}$
	Olfactory transduction	$5.99 \cdot 10^{-6}$

Table 1: KEGG pathways with an adjusted $p < 0.001$ after over-representation analysis using ConsensusPathDataBase [22] with the genes from the NMF derived cell-type profiles and a background list of all genes used to generate the proflies

the tumor mixture in glioblastoma, determine the number of component cancer cell types and determine their profiles, we compared results of the different NMF algorithms on the TCGA collection 2.1.1. We concluded that the measured gene expression values consist of three clusters, and therefore we used rank 3 of the factorisation. Subsequently, we approximated gene profiles of the clusters found in the tumor mixture. We then identified the proportion of each subtype (designated 1, 2 and 3) in the mixture before treatment was applied and tracked their changes in time after irradiation.

The changes in the proportions observed while the total population of cancer cells remains constant led us to a hypothesis that there are three types of cells in the mixture: The first type is sensitive to treatment (proportion in the mixture decreases), the second type is treatment-resistant (proportion in the mixture increases) and the for the last one the proportions remains unchanged with treatment.

To test this for each cell type obtained through NMF, we calculated Log2FoldChange (LFC) with respect to the other two cell types (one expression value subtracted from the other) for every gene. We then defined a threshold (LFC > 2). If the LFC was above 2 for both cases (LFCs calculated with respect to the other cell types), the gene was assigned to the signature of the corresponding cell type. We found that type 1, type 2 and type 3 signatures consist of 408, 15566, and 987 genes, respectively.

Taking these signature lists we adopted an over-representation approach [22], looking for which KEGG pathways are over-represented in the list of genes each signature. The most significant pathways (adjusted $p < 0.001$) from the KEGG database are shown in Table 1. Here we can see that the significant pathways in type 1 cells are related to infections, indicating genes involved in immune processes are important for those cells, thus type 1 may be immune cells. The corresponding processes could be then radiation-induced immunogenic cell death [46]. For type 2 cells we see some cancer pathways and cell cycle, indicating that the genes defining the type 2 cell profile are related to the tumor, typically proliferative cancer cells. Finally, for type 3 cells we see that the neuroactive ligand-receptor interaction pathway is the most significant hit, indicating that we are finding genes relating to more normal neuronal cells.

As a further analysis we compared these genes to the genes from the meta-modules suggested by Neftel et al. We found that these were mostly covered by the type 2 cell type signature from the unsupervised model, as shown in Table 2

	MES1	MES2	AC	OPC	NPC1	NPC2	G1/S	G2/M
V1	0/50	0/50	0/39	0/50	0/50	0/50	0/29	0/45
V2	43/50	46/50	36/39	42/50	45/50	41/50	22/29	41/45
V3	0/50	0/50	0/39	0/50	0/50	1/50	0/29	0/45

Table 2: Overlap of signature genes between Neftel's meta-modules and cell types obtained through NMF. (overlapping genes / total number of genes in the given Neftel's meta-module). Signature genes for meta-modules have been made available by Neftel et al [36].

3.2 Results of supervised methods

In the supervised methods cell type proportions were inferred from microarray data using Scaden (see Section 2.3) The corresponding cell type proportions are shown for patient-derived organoids from three patients under two different conditions (4 and 10 Gy irradiation) both untreated and at 3 treated time points (Figure 1). For each patient, the first pre-exposure timepoint is shared between two conditions and the next three timepoints diverge due to differences in radiation doses. DNN-based model predictions indicate that more than 96% of all organoids (at all timepoints) consist of mesenchymal-like cells while the remaining 4% (or less) belong to other cell types.

For patient 1 organoid, we observe a slight decline in the fraction of hypoxia-independent cells (MES-like 1) and an increase for hypoxia-dependent fraction post-radiation. Furthermore, a lower dose of radiation shows a greater effect on changes in cell-type proportions directly after the irradiation.

Patient 1's organoid has the least varying proportions of different cell types.

The most visible changes in cell populations were recorded for patient 3. There is a clear trend of hypoxia-dependent cells (MES-like 2) overtaking hypoxia-independent cells (MES-like 1) as the primary cell type 72 hours post irradiation.

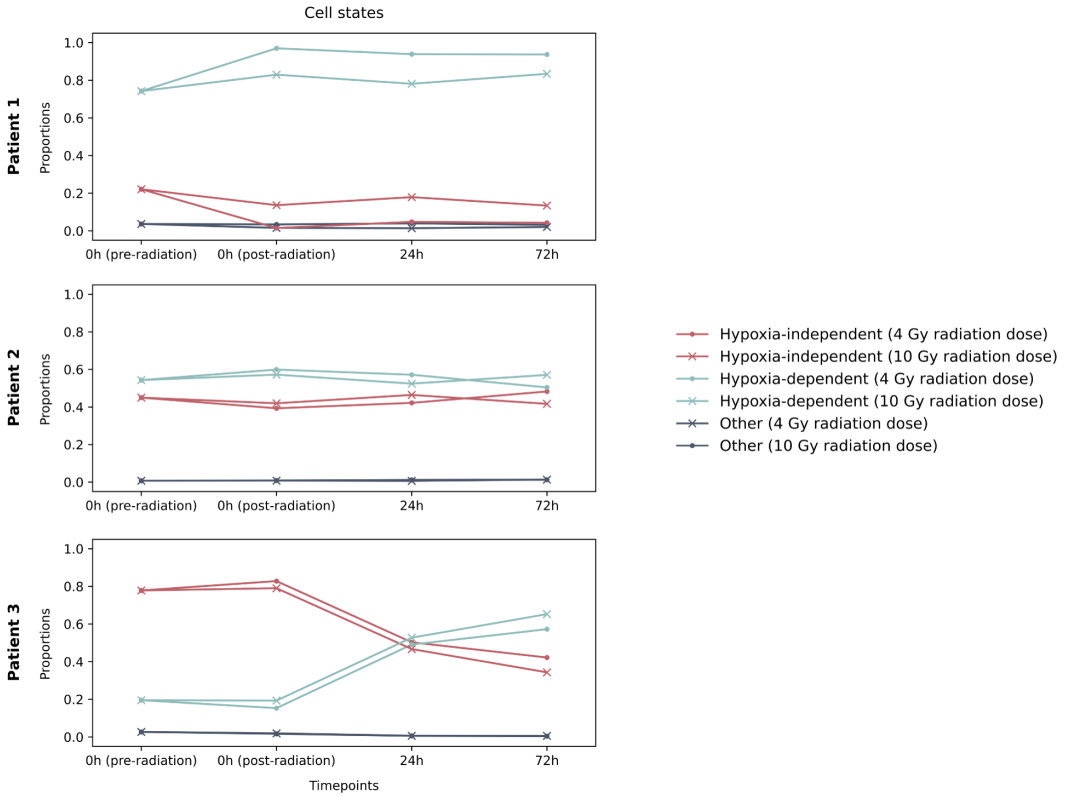

Figure 1: Population dynamics of mesenchymal-like and cycling cells in patient-derived glioma organoids under two divergent conditions. Hypoxia-independent cells are MES-like 1 and Hypoxia-depedent cells are MES-like 2 in Neftel's scheme.

The dynamics in MES-like cell type could be caused by the upregulation of hypoxia-response genes in the hypoxia-independent cell population. This would lead to a transition from one cell type to another without any changes in population size. Similar dynamics could also be explained through different resistance to irradiation. If hypoxia-dependent cell types showed higher probability of survival (fitness) upon irradiation, as suggested by Wang et al. [49], we would observe a similar effect with a decrease in a total number of tumor cells due to apoptosis or other forms of cell death [9].

3.3 Results of game-theoretic model of organoids

As discussed earlier, comparison of the type profiles found by unsupervised methods with the meta-models proposed by Neftel et al. showed that most genes are covered by cell signatures from one of the three types. Thus, the assumption that the unsupervised methods result in different cancer cell types appears to be incorrect, hence the game theoretic model was only fitted to the results obtained by the supervised methods. The results obtained with the DNN model showed that organoids are mainly composed of mesenchymal cells. In order to simplify the tumor dynamics model, a replicator population was considered, which consists of two types of mesanchymal cells: hypoxia-dependent (MES-like 2) and hypoxia-independent (MES-like 1). Given two cell types in a population, the payoff matrix A has dimension 2×2, and its elements a_{ij} represent the probability of reproduction of cell type i when interacting with type j. As adding any constant to any column of the fitness matrix A does not change the dynamics, a_{11} and a_{22} are set to zero. To determine the true frequency dynamics defined by the fitness matrix, the remaining matrix parameters are fitted using available measurements in both of the environments created by treatment initialisation. The

fitness matrix is assumed to have the following form:

	MES 1	MES 2
MES 1	0	a_{12}
MES 2	a_{21}	0

Table 3: Fitness matrix A for cancer populations with two types of cells.

For the found matrices, the residual sum of squares (RSS) between the available measurements and the fitted replicator dynamics is no more than 0.001 for each patient in both environments. The plots below 2 show the predicted dynamics of the two types (MES-like 1 and MES-like 2) and the relevant measurements. Horizontal lines in the figures represent evolutionary stable strategies (ESSs).

Patient	Radioexposure	(a_{12}, a_{21})	ESS
3565	Rx4	$(0.13, 0.18)$	$\left(\frac{13}{31}, \frac{18}{31}\right)$
3565	Rx10	$(0.1, 0.18)$	$\left(\frac{5}{14}, \frac{9}{14}\right)$
1919	Rx4	$(0.06, 1)$	$\left(\frac{3}{53}, \frac{50}{53}\right)$
1919	Rx10	$(0.19, 0.98)$	$\left(\frac{19}{117}, \frac{98}{117}\right)$
3128	Rx4	$(0.04, 0.04)$	$\left(\frac{1}{2}, \frac{1}{2}\right)$
3128	Rx10	$(0.74, 0.92)$	$\left(\frac{36}{83}, \frac{47}{83}\right)$

Table 4: Fitted parameters of matrix A and calculated ESSs for the cancer population with two types.

One of the characteristic features of advanced tumors is hypoxia, which is closely related to tumor radioresistance, as it can promote tumor progression and lead to poor clinical outcomes. In each patient, regardless of radiotherapy dose, the fitted model predicts in the long-term a majority of hypoxia-dependent cells, while both observed cancer cell types coexist at equilibrium.

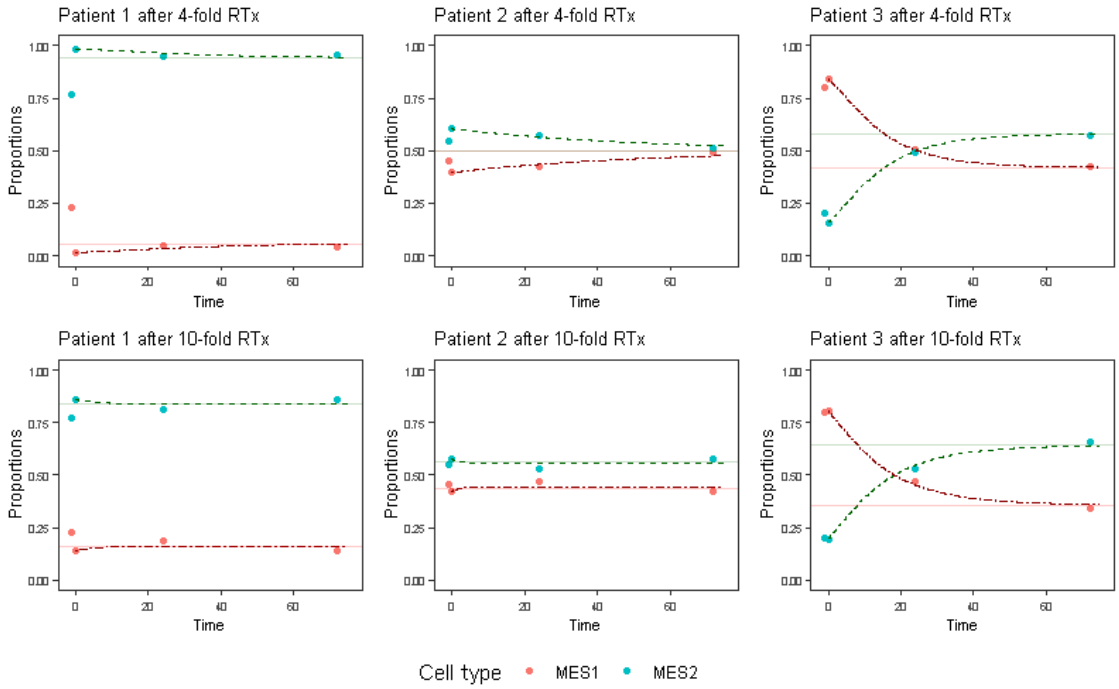

Figure 2: Fitted replicator dynamics equations vs. the measurements and ESS. First point (not fitted) indicates the measurement before the therapy is applied.

4 Discussion

Glioblastoma is a universally fatal disease for which there are no curative treatments and all patients eventually relapse after radiation and temozolomide chemotherapy. Treatment-induced resistance in cancer is thought to be caused by stochastic genetic or epigenetic driven phenotypes that preexist in small populations of tumour cells or evolve during the course of treatment [34, 8, 41, 39]. Tumor markers that predict the emergence of treatment refractory subpopulations in human tumors before and during the course of standard treatment may be exploited to adapt treatment and prevent relapse. Here we used machine-learning techniques to estimate how proportions of relevant glioblastoma cell types change over time under treatment from transcriptomics organoid data. We then fitted this data to the most common evolutionary game model, replicator dynamics, to predict long-term response to irradiation in these organoids.

To identify different cell subpopulations, we utilized both unsupervised and supervised machine learning methods. The unsupervised methods are attractive to researchers as they do not require any assumptions about the number of cell types present in the mixture, nor the profiles of said types. However, this makes them trickier to work with and less immediately interpretable. To overcome this shortcoming in interpretability, we selected genes with a high-fold change between their values in different profiles and compared them to both genes seen in signatures of published single cell analyses and genes from known pathways in standard pathway databases. Perhaps unsurprisingly, we found that the patient tumor samples were deconvoluted into immune cells, cycling cancer cells and healthy neuronal cells. This led to issues extrapolating from tumor profiles obtained *in vivo* to those *in vitro*, as not all cell types seen *in vivo* will be expected to be present in the *in vitro* situation. In the patient-derived organoids, for instance, immune cells, neurons, and blood vessels are missing.

Another issue with the unsupervised methods was the need to determine the number of cell types present *a priori*. If no assumptions about this are made, then extra steps are needed to estimate the optimal number of types empirically from the data. In our experience, both in this

work and earlier unpublished work, this step does not generally yield a clear result when working with real datasets.

We found the supervised methods more interpretable, as they immediately allowed us to extract and interpret key information about the expected cell types present in the mixture; a key advantage over the unsupervised methods.

However, there are issues extrapolating from single cell data to bulk data, as the difference in measurement platforms can cause systematic differences in measurement that would be seen in the cell profiles (similarly to batch effects between datasets measured on different platforms).

Further, the supervised analysis relies on there being appropriate single-cell datasets available for training, something which will not be the case for all scenarios where deconvolution results would be useful.

Another limitation of this study is that tumors have spatial aspects, for instance hypoxic cells are more common in the center/core of the tumor where the supply of oxygen is limited (diffusion-limited hypoxia) or in the neighborhood of defective vessels (acute hypoxia). The spatial aspects could be analysed through single-cell imaging data, for instance [2], and then using spatial game-theoretic models to predict further spatio-temporal response. Cells could then be assumed to only interact with the cells in their local neighbourhood, so local dynamics would become important, and tumors are not assumed to be well-mixed [25]. In case of using a replicator dynamics model, one could assume that different regions correspond to different fitness matrices, as done in [52], as cells' chances for proliferation and survival will vary with conditions in their neighborhood.

For prediction of the long-term response in organoids we fitted the identified proportions of different cell types to replicator dynamics [45, 17], which model classic and likely the most simple evolutionary dynamics. While the fit of data to replicator dynamics was very good, we could use different game-theoretic dynamics, such as Darwinian or adaptive dynamics [48, 33, 32]. As the total cell volume in organoids considered in the experiments carried out here was kept constant, to use these different dynamics we would need to assume that the population equilibrium was reached and fit only the evolutionary dynamics. Alternatively, we could combine volumetric data from the patients to fit the population dynamics, similarly as we did in [1, 13], while fitting the evolutionary dynamics that are very difficult to estimate *in vivo* from the patient-derived organoids. An advantage of this approach could be that we would not need to assume that there are fixed number of possible cell types with a fixed profile, but instead could trace radioresistance as an evolving trait, similarly to [39, 40].

While our game-theoretic models predicted coexistence of the two different tumor types in organoids, we did not collect additional measurements to validate whether these predictions are correct. This was also not the main goal of our study; rather we wanted to demonstrates all steps needed for utilization of transcriptomics data in predictive game-theoretic modeling. Validation of our predictions and comparing the fit and predictive capabilities of different game-theoretic dynamics will be our natural next step. More time series data will be collected to perform both data fitting and validation. Once we have a validated game-theoretic model of the radiotherapy response in organoids, we can predict the impact of alternative treatments (including their timing and dosing), in order to delay/predict therapy resistance in the corresponding patients. A huge advantage of organoids is that these predictions can be then validated through real experiments. This is because in many studies organoids have been shown to retain major patient characteristics that determine treatment resistance [28, 35, 47]. As such, we can utilize game-theoretic models validated through organoids to improve treatment in real patients.

Despite its limitations, this study shows that it is possible to apply both supervised and unsupervised deconvolution techniques to bulk gene expression data from tumors and to use the information obtained on the changes in cell type proportions over time to fit game-theoretic models of the tumor eco-evolutionary response to treatment. This is an important step forward in the development and application of these models and their use with real data predicting both ecological and evolutionary cancer dynamics and designing evolutionary therapies, i.e. thereapies that anticipate and forestall treatment-induced resistance in cancer cells.

For our next steps we note that the TGCA dataset contains many different sets of measurements on the same patients. In this work we have only made use of the transcriptomics data, given all

these measurements are made on the same tumors, which can be assumed to be composed of the same types, one direction for further research is learning to leverage the additional information available in other measurement sets from these tumors to improve the deconvolution overall and to improve the fit and predictive capabilities of game-theoretic models.

We believe that using cancer data in a mathematical model and adapting treatment strategies to the resulting predictions has the potential to revolutionise personalised medicine. When facing such a heterogeneous formation as cancer, one must apply a therapy adapted to the targeted malignancy. Our ultimate goal is to provide a process which, using the available information about the tumour composition and its measurements combined with a game-theoretic model, will simplify the selection of a better cancer treatment strategy.

Acknowledgements

We would like to thank Dr. Jafar Rezaei for his valuable feedback that improved quality of this work. This research was supported by European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 955708, the Dutch National Foundation projects ENWPR.020.006, and OCENW.KLEIN.277 and Dutch Cancer Society (KWF 11698) and funding from StopHersenTumoren.

References

[1] V. Ardévol Martinez, N. G. Laleh, M. Salvioli, F. Thuijsman, J. S. Brown, R. Cavill, J. N. Kather, and K. Staňková. "Improving mathematical models of cancer by including resistance to therapy: a study in non-small cell lung cancer". In: *bioRxiv* (2021). DOI: 10.1101/2021.10.29.466444. eprint: https://www.biorxiv.org/content/early/2021/11/01/2021.10.29.466444.full.pdf. URL: https://www.biorxiv.org/content/early/2021/11/01/2021.10.29.466444.

[2] M. J. Baars, N. Sinha, M. Amini, A. Pieterman-Bos, S. van Dam, M. M. Ganpat, M. M. Laclé, B. Oldenburg, and Y. Vercoulen. "MATISSE: a method for improved single cell segmentation in imaging mass cytometry". In: *BMC Biology* 19.1 (2021), pp. 1–10. ISSN: 17417007. DOI: 10.1186/s12915-021-01043-y.

[3] L. Badea. "Extracting gene expression profiles common to colon and pancreatic adenocarcinoma using simultaneous nonnegative matrix factorization". In: *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing* (2008), pp. 267–278. ISSN: 2335-6928. URL: http://europepmc.org/abstract/MED/18229692.

[4] J.-P. Brunet, P. Tamayo, T. R. Golub, and J. P. Mesirov. "Metagenes and molecular pattern discovery using matrix factorization". In: *Proceedings of the National Academy of Sciences* 101.12 (2004), pp. 4164–4169. ISSN: 0027-8424. DOI: 10.1073/pnas.0308531101. eprint: https://www.pnas.org/content/101/12/4164.full.pdf. URL: https://www.pnas.org/content/101/12/4164.

[5] J. J. Cunningham. "A call for integrated metastatic management". In: *Nature ecology & evolution* 3.7 (2019), pp. 996–998.

[6] J. J. Cunningham, J. S. Brown, R. A. Gatenby, and K. Staňková. "Optimal control to develop therapeutic strategies for metastatic castrate resistant prostate cancer". In: *Journal of theoretical biology* 459 (2018), pp. 67–78.

[7] J. J. Cunningham, F. Thuijsman, R. Peeters, Y. Viossat, J. S. Brown, R. A. Gatenby, and K. Staňková. "Optimal Control to Reach Eco-Evolutionary Stability in Metastatic Castrate Resistant Prostate Cancer". In: *PLOS ONE* 15.12 (2020), pp. 1–24.

[8] L. A. Diaz Jr et al. "The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers". In: *Nature* 486.7404 (2012), pp. 537–540. DOI: 10.1038/nature11219.

[9] D. Eriksson and T. Stigbrand. "Radiation-induced cell death mechanisms". In: *Tumor Biology* 31.4 (May 2010), pp. 363–372. DOI: 10.1007/s13277-010-0042-8. URL: https://doi.org/10.1007/s13277-010-0042-8.

[10] A. Frigyesi and M. Höglund. "Non-Negative Matrix Factorization for the Analysis of Complex Gene Expression Data: Identification of Clinically Relevant Tumor Subtypes". In: *Cancer Informatics* 6 (2008). PMID: 19259414, CIN.S606. DOI: 10.4137/CIN.S606. eprint: https://doi.org/10.4137/CIN.S606. URL: https://doi.org/10.4137/CIN.S606.

[11] R. Gatenby. "A change of strategy in the war on cancer". In: *Nature* 459 (2009), 508?509.

[12] R. A. Gatenby, A. S. Silva, R. J. Gillies, and B. R. Frieden. "Adaptive therapy". In: *Cancer research* 69.11 (2009), pp. 4894–4903.

[13] N. Ghaffari Laleh et al. "Classical Mathematical Models for Prediction of Response to Chemotherapy and Immunotherapy". In: *Plos Computational Biology* (2022, in print). DOI: 10.1101/2021.10.23.465549.

[14] M. Gluzman, J. Scott, and A. Vladimirsky. "Optimizing adaptive cancer therapy: dynamic programming and evolutionary game theory". In: *bioRxiv (preprint)* (2018).

[15] *H. Lee Moffitt Cancer Center and Research Institute: Adaptive Abiraterone Therapy for Metastatic Castration Resistant Prostate Cancer (NCT02415621)*. https://www.clinicaltrials.gov. 2015–.

[16] K. A. Hoadley et al. "Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer." In: *Cell* 173.291-304 (2018), 291–304.e6. ISSN: 10974172. DOI: 10.1016/j.cell.2018.03.022.

[17] J. Hofbauer and K. Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge University Press, May 1998. ISBN: 052162570X. URL: http://www.amazon.ca/exec/obidos/redirect?tag=citeulike04-20%7B%5C&%7Dpath=ASIN/052162570X.

[18] C. G. Hubert, L. C. Spangler, Q. Wu, S. Mack, B. C. Prager, M. C. M, R. E. McLendon, A. E. Sloan, and J. N. Rich. "A Three-Dimensional Organoid Culture System Derived from Human Glioblastomas Recapitulates the Hypoxic Gradients and Cancer Stem Cell Heterogeneity of Tumors Found In Vivo". In: *Cancer Res.* 76.8 (2016), pp. 2465–77. DOI: 10.1158/0008-5472.CAN-15-2402.Epub2016Feb19.PMID:26896279;PMCID:PMC4873351.

[19] L. N. Hutchins, S. M. Murphy, P. Singh, and J. H. Graber. "Position-dependent motif characterization using non-negative matrix factorization". In: *Bioinformatics* 24.23 (Oct. 2008), pp. 2684–2690. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btn526. eprint: https://academic.oup.com/bioinformatics/article-pdf/24/23/2684/16885804/btn526.pdf. URL: https://doi.org/10.1093/bioinformatics/btn526.

[20] B. Hwang, J. Lee, and D. Bang. "Single-cell RNA sequencing technologies and bioinformatics pipelines". In: *Exp Mol Med* 50 (2018), pp. 1–14. DOI: 10.1038/s12276-018-0071-8.

[21] A. Jemal et al. "Annual Report to the Nation on the Status of Cancer, 1975?2014, Featuring Survival". In: *JNCI: Journal of the National Cancer Institute* 109.9 (Mar. 2017).

[22] A. Kamburov, C. Wierling, H. Lehrach, and R. Herwig. "ConsensusPathDB - A database for integrating human functional interaction networks". In: *Nucleic acids research* 37.Database issue (Jan. 2009), pp. D623–8. ISSN: 03051048. DOI: 10.1093/nar/gkn698. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2686562%7B%5C&%7Dtool=pmcentrez%7B%5C&%7Drendertype=abstract.

[23] H. Kim and H. Park. "Sparse Non-negative Matrix Factorizations via Alternating Non-negativity-constrained Least Squares for Microarray Data Analysis". In: *Bioinformatics (Oxford, England)* 23 (July 2007), pp. 1495–502. DOI: 10.1093/bioinformatics/btm134.

[24] M. Kim, H. J. Seo, J.-G. Joung, and J. H. Kim. "Comprehensive evaluation of matrix factorization methods for the analysis of DNA microarray gene expression data". In: *BMC bioinformatics* 12 Suppl 13 (Nov. 2011), S8. DOI: 10.1186/1471-2105-12-S13-S8.

[25] L. You, J. S. Brown, F. Thuijsman, J. J. Cunningham, R. A. Gatenby, J. Zhang, and K. Staňková. "Spatial vs. non-spatial eco-evolutionary dynamics in a tumor growth model". In: *Journal of Theoretical Biology* 435 (2017), pp. 78–97. DOI: 10.1016/j.jtbi.2017.08.022.

[26] D. Lee and H. S. Seung. "Algorithms for Non-negative Matrix Factorization". In: *Advances in Neural Information Processing Systems*. Ed. by T. Leen, T. Dietterich, and V. Tresp. Vol. 13. MIT Press, 2001, pp. 556–562. URL: https://proceedings.neurips.cc/paper/2000/file/f9d1152547c0bde01830b7e8bd60024c-Paper.pdf.

[27] J. Leuschner, M. Schmidt, P. Fernsel, D. Lachmund, T. Boskamp, and P. Maass. "Supervised non-negative matrix factorization methods for MALDI imaging applications". In: *Bioinformatics* 35.11 (Nov. 2018), pp. 1940–1947. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/bty909. eprint: https://academic.oup.com/bioinformatics/article-pdf/35/11/1940/28759300/bty909.pdf. URL: https://doi.org/10.1093/bioinformatics/bty909.

[28] M. L. Martin et al. "Organoids Reveal That Inherent Radiosensitivity of Small and Large Intestinal Stem Cells Determines Organ Sensitivity". In: *Cancer Research* 80.5 (2020), pp. 1219–1227. ISSN: 0008-5472. DOI: 10.1158/0008-5472.CAN-19-0312. eprint: https://cancerres.aacrjournals.org/content/80/5/1219.full.pdf. URL: https://cancerres.aacrjournals.org/content/80/5/1219.

[29] A. Marusyk, D. P. Tabassum, P. M. Altrock, V. Almendro, F. Michor, and K. Polyak. "Non-cell autonomous tumor-growth driving supports sub-clonal heterogeneity". In: *Nature* 514 (2014), pp. 54–58. DOI: 10.1038/nature13556.

[30] A. Marusyk, V. Almendro, and K. Polyak. "Intra-tumour heterogeneity: a looking glass for cancer?" In: *Nature Reviews Cancer* 12.5 (2012), pp. 323–334.

[31] K. Menden, M. Marouf, S. Oller, A. Dalmia, D. S. Magruder, K. Kloiber, P. Heutink, and S. Bonn. "Deep learning-based cell composition analysis from tissue expression profiles". In: *Science Advances* 6.30 (2020), eaba2619. DOI: 10.1126/sciadv.aba2619. eprint: https://www.science.org/doi/pdf/10.1126/sciadv.aba2619. URL: https://www.science.org/doi/abs/10.1126/sciadv.aba2619.

[32] J. Metz, K. Staňková, and J. Johansson. "The adaptive dynamics of life histories: From fitness-returns to selection gradients and Pontryagin's maximum principle". In: *Journal of Mathematical Biology* 72.4 (2016), pp. 1125–1152.

[33] J. A. Metz, S. A. Geritz, G. Meszéna, F. J. Jacobs, and J. S. Van Heerwaarden. "Adaptive dynamics: a geometrical study of the consequences of nearly faithful reproduction". In: *Stochastic and Spatial Structures of Dynamical Systems, Proceedings of the Royal Dutch Academy of Science (KNAW Verhandelingen), North Holland, Amsterdam* (1995), pp. 183–231.

[34] S. Misale et al. "Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer". In: *Nature* 486.7404 (2012), pp. 532–536. DOI: 10.1038/nature11156.

[35] P. Nagle. "The use of organoids in the study of radiation response and therapeutic window". English. PhD thesis. University of Groningen, 2018. ISBN: 978-94-034-0965-8.

[36] C. Neftel et al. "An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma". In: *Cell* 178.4 (Aug. 2019), 835–849.e21. DOI: 10.1016/j.cell.2019.06.024. URL: https://doi.org/10.1016/j.cell.2019.06.024.

[37] R. O'Connor, M. Clynes, P. Dowling, N. O'Donovan, and L. O'Driscoll. "Drug resistance in cancer–searching for mechanisms, markers and therapeutic agents". In: *Expert opinion on drug metabolism & toxicology* 3.6 (2007), pp. 805–817.

[38] A. Pascual-Montano, J. M. Carazo, K. Kochi, D. Lehmann, and R. D. Pascual-Marqui. "Nonsmooth nonnegative matrix factorization (nsNMF)". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28.3 (2006), pp. 403–415. DOI: 10.1109/TPAMI.2006.60.

[39] M. Pressley, M. Salvioli, D. B. Lewis, C. L. Richards, J. S. Brown, and K. Staňková. "Evolutionary dynamics of treatment-induced resistance in cancer informs understanding of rapid evolution in natural systems". In: *Frontiers in Ecology and Evolution* 9 (2021), p. 460.

[40] M. Salvioli, H. Garjani, J. S. Brown, J. L. A. Dubbeldam, and K. Staňková. "Stackelberg evolutionary games of cancer treatment: Tumor stabilization as an alternative to dynamic treatment protocols". In: *submitted* (2021).

[41] K. Staňková. "Resistance games". In: *Nature Ecology & Evolution* 3 (2019), pp. 336–337. DOI: 10.1038/s41559-018-0785-y.

[42] K. Staňková, J. S. Brown, W. S. Dalton, and R. A. Gatenby. "Optimizing cancer treatment using game theory: A review". In: *JAMA Oncology* 5.1 (2019), pp. 96–103.

[43] R. Star, M. Grzelak, and J. Hadfield. "RNA sequencing: the teenage years". In: *Nat Rev Genet 20* (2019), pp. 631–656. DOI: 10.1038/s41576-019-0150-2.

[44] X. Tang, Y. Huang, J. Lei, H. Luo, and X. Zhu. "The single-cell sequencing: new developments and medical applications". In: *Cell & bioscience* 9.1 (2019), pp. 1–9.

[45] P. Taylor and L. Jonker. "Evolutionary stable strategies and game dynamics". In: *Math Biosci* 40.1–2 (1978), pp. 145–156.

[46] R. D. W. Vaes, L. E. L. Hendriks, M. Vooijs, and D. D. Ruysscher. "Biomarkers of Radiotherapy-Induced Immunogenic Cell Death". In: *Cells* 10 (2021). DOI: 10.3390/cells10040930.

[47] M. Verduin, A. Hoeben, D. D. Ruysscher, and M. Vooijs. "Patient-Derived Cancer Organoids as Predictors of Treatment Response". In: *Front Oncol.* 11 (2021). p.641980. DOI: 10.3389/fonc.2021.641980.PMID:33816288;PMCID:PMC8012903.

[48] T. L. Vincent and J. S. Brown. *Evolutionary game theory, natural selection, and Darwinian dynamics*. Cambridge University Press, 2005.

[49] H. Wang, H. Jiang, M. Van De Gucht, and M. De Ridder. "Hypoxic Radioresistance: Can ROS Be the Key to Overcome It?" In: *Cancers* 11.1 (2019). ISSN: 2072-6694. DOI: 10.3390/cancers11010112. URL: https://www.mdpi.com/2072-6694/11/1/112.

[50] X. Wang, H. Zhang, and X. Chen. "Drug resistance and combating drug resistance in cancer". In: *Cancer Drug Resistance* 2.2 (2019), pp. 141–160.

[51] B. Wölfl, H. te Rietmole, M. Salvioli, A. Kaznatcheev, F. Thuijsman, J. S. Brown, B. Burgering, and K. Staňková. "The Contribution of Evolutionary Game Theory to Understanding and Treating Cancer". In: *Dynamic Games and Applications* (Aug. 2021). ISSN: 2153-0793. DOI: 10.1007/s13235-021-00397-w. URL: https://doi.org/10.1007/s13235-021-00397-w.

[52] L. You, M. von Knobloch, T. Lopez, V. Peschen, S. Radcliffe, P. Koshy Sam, F. Thuijsman, K. Staňková, and J. S. Brown. "Including Blood Vasculature into a Game-Theoretic Model of Cancer Dynamics". In: *Games* 10.1 (2019). DOI: 10.3390/g10010013.

[53] C. Yuan and H. Yang. "Research on K-Value Selection Method of K-Means Clustering Algorithm." In: *J* 2.2 (2019), pp. 226–235. DOI: https://doi.org/10.3390/j2020016.

[54] J. Zhang, J. J. Cunningham, J. S. Brown, and R. A. Gatenby. "Multidisciplinary analysis of evolution based Abiraterone treatment for metastatic castrate resistant prostate cancer". In: *biorXiv* (2021). DOI: 10.1101/2021.11.30.21267059.

[55] J. Zhang, J. J. Cunningham, J. S. Brown, and R. A. Gatenby. "Integrating evolutionary dynamics into treatment of metastatic castrate-resistant prostate cancer". In: *Nature communications* 8.1 (2017), p. 1816.

[56] W. Zhu, L. Xie, J. Han, and X. Guo. "The Application of Deep Learning in Cancer Prognosis Prediction". In: *Cancers* 12.3 (2020). ISSN: 2072-6694. DOI: 10.3390/cancers12030603. URL: https://www.mdpi.com/2072-6694/12/3/603.

[57] A. Zinovyev, U. Kairov, T. Karpenyuk, and E. Ramanculov. "Blind source separation methods for deconvolution of complex signals in cancer biology". In: *Biochemical and Biophysical Research Communications* 430.3 (2013), pp. 1182–1187. ISSN: 0006-291X. DOI: https://doi.org/10.1016/j.bbrc.2012.12.043. URL: http://www.sciencedirect.com/science/article/pii/S0006291X12023741.