Computational modeling of KRAS mutant resistance mechanisms to targeted therapy

Master thesis

by

Huub van der Ent

Prof. Lodewyk Wessels	Supervisor	Netherlands Cancer Institute TU Delft
Dr. Bram Thijssen	Daily supervisor	Netherlands Cancer Institute
Prof. Peter Bosman	Thesis committee	TU Delft CWI
Dr. Jana Weber	Thesis commitee	TU Delft

Submitted December 2023

Preface

This document signifies the final hurdle taken in completing the Computer Science master's program at the Delft University of Technology. It is a description of the most important research results from my time working at the Netherlands Cancer Institute. During this project I worked on elucidating the resistance mechanism of cancer cells with a mutation in the kirsten rat sarcoma virus gene to targeted therapy.

At the Netherlands Cancer Institute I always felt inspired and motivated by the work of the people around me. The computational biology group is filled with researchers who work on the cutting edge of science, and I felt grateful for being a part of it. I would especially like to thank Dr. Bram Thijssen for being my daily supervisor. He was always ready for all of my questions and showed unwavering support of my project. Without his guidance I would never have been able to do this work. I would also like to thank Prof. Lodewyk Wessels for providing me the opportunity to be a part of his group and for providing valuable feedback on my project. Lastly, I would like to thank my family, girlfriend and friends for supporting me throughout my time at university.

Huub van der Ent

Abstract

KRAS mutations are very common in several different types of cancer. A promising targeted combination therapy using a MEK and HER inhibitor was proposed based on in vitro finding. The clinical results of this combination were found to be lacking due to emergent treatement resistance. Here we investigate what mechanism is causing this emergent resistance in KRAS mutant cancers. We propose a novel ODE model of the MAPK pathway that can be used to infer kinetic parameter estimates from a population of KRAS mutant cells under drug perturbation. Parameter estimates inferred from FRET biosensor data correctly predict protein activity in an external CyTOF validation dataset. However, the parameter estimates did not recapitulate the known gain-of-function in RAS activity that we would expect. From this we conclude that more experimental observation are required to elucidate the inner working of the resistance mechanism of KRAS mutant cancer to the proposed combination therapy.

Contents

1

Introduction

The mitogen-activated protein kinase (MAPK) pathway is one of the major cellular signaling cascades. Cellular signaling cascades are central in the functioning of a cell as they transduce extracellular signals into the cell. The MAPK pathway (see Figure 1.1) is crucial in the regulation of several processes like differentiation, survival and proliferation [1, 2]. Given this central function, it is not surprising that mutations in this pathway are prone to cause cancerous growth of the affected cells. Two common examples are mutations in b rapidly accelerated fibrosarcoma (BRAF) and kirsten rat sarcoma gene (KRAS). KRAS gene mutations specifically are prevalent in pancreatic cancer (90%), non-small cell lung cancer (NSCLC) (35%), and colorectal cancer (CLC) (45%) [3–5]. Mutations in this gene alter the conformation of the KRAS protein in such a way that it will continuously promote downstream signaling activity (gain-of-function) [2]. Historically, a lot of effort has been exerted to find KRAS mutant-specific inhibitors. Currently, only one inhibitor has been approved for clinical use, namely Sotorasib which is specific for the G12C mutation in KRAS [6]. Recently, there were some promising results in this field of research [7], however these inhibitors are still in the preliminary phase of development.

The ineffectiveness of KRAS-specific inhibitors prompted the investigation of other methods to decrease MAPK pathway activity. One promising alternative was the inhibition of other crucial proteins in the cascade. Monotherapeutic approaches in this context commonly induced resistance due to alternative signaling regimes. For example, MAPKK (MEK) inhibitor treatment was proposed to induce human epidermal growth factor receptor 3 (HER3) upregulation via MYC. This finding prompted the addition of a general HER inhibitor to counter this effect. In vitro experiments for this combination showed synergistic results for both NSCLC and CLC [11]. However, the clinical benefit was deemed to be insufficient [3–5].

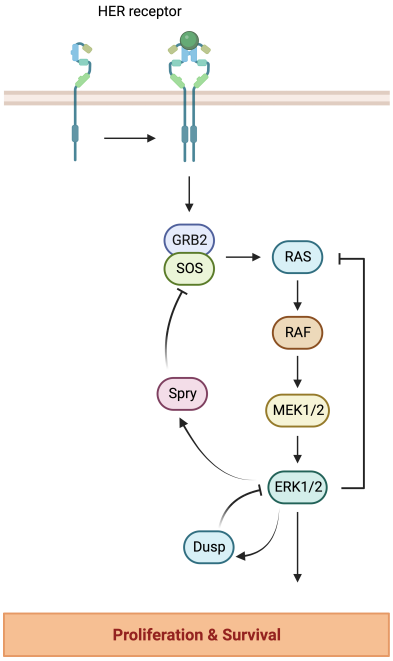

Figure 1.1: Schematic overview of central MAPK cascade and a subset of known feedback mechanisms reported in literature. [1, 8–10]. Created with BioRender.com.

1.1. Cellular heterogeneity and treatment resistance

The fact that most cancer treatment regimes can produce some form of resistance raises several questions. Centrally, we can wonder what molecular mechanisms cause some cells to survive treatment and others to adapt and proliferate. There could be some prior differences between these cells that control their probability of survival. In general, there are two possible types of heterogeneity in this context.

One possibility is that there might already be a genetically resistant subpopulation present in the tumor. There are several different types of mutations that could induce drug resistance. Some examples are that the drug target is altered, the drug is inactivated through interaction with altered effector proteins, or cell death is inhibited [12–14].

A more elusive type of heterogeneity can also be observed in genetically homogeneous tumor cell populations [15]. Pre-treatment protein expression levels are predictive of cell survival post-treatment. In human melanoma cells treated with a BRAF inhibitor, it was shown that a specific subset of genes are differentially expressed in the resistant melanoma cells. These genes induce epigenetic changes which were found to convert the cells to a resistant state by activating alternative signaling pathways [16]. In HeLa cells treated with TNF-

related apoptosis-inducing ligand (TRAIL), a similar effect was observed. Five differentially expressed genes in resistant cells were shown to correlate with the time to death after treatment. In contrast to the effect described in melanoma cells there was no epigenetic mode of action observed in these cells [17]. This indicates that there are several different ways in which pre-treatment protein levels or pathway activity can affect resistance.

The goal of work is to investigate the treatment resistance mechanism of KRAS mutant cells to combination therapy. We specifically want to infer from a population of cells what mechanism is activated during drug perturbation. This could indicate why some cells survive while others die.

1.2. Research aims

1. **What is the minimal mitogen-activated protein kinase pathway model that can describe KRAS mutant ERK protein activity in a population of cells?**
   The first research objective is to find a minimal model of the MAPK pathway that can describe the activity of ERK protein in single-cells. This model will indicate which parts of MAPK pathway are essential to describe its behavior. The usage of a minimal pathway model will also decrease unnecessary computational costs. In turn, this will provide the possibility to analyze cell population signaling kinetics.

2. **Can we infer the gain-of-function in a population KRAS mutant cells?**
   To confirm the biological relevance of this model the model will be assessed in its ability to detect the increased activity of RAS signaling due to the gain-of-function mutation in KRAS. This information will be inferred solely based on observed protein activity in single-cells without providing any information on the mutation a priori. This result should indicate how reliable the kinetic parameter estimates are of the used inference technique.

3. **Are any alternative pathways activated under perturbed conditions?**
   The main focus of this project is what the differences are in pathway activity between drug-perturbed cells and unperturbed cells. Specifically, the focus is the combination of a MEK + HER inhibitor in a population of KRAS mutant cells. For this purpose, the previously mentioned MAPK pathway model will be used to detect any kinetic parameter changes.

4. **Can the minimal MAPK pathway model predict protein activity in a population of perturbed KRAS mutant cells in an external validation dataset?**
   A secondary model relevance test will be performed by comparing the model simulations to an external validation dataset. Kinetic parameters will be inferred based on

protein activity measurements in single-cells. This estimate will then be used to simulate the treatment of KRAS mutant cells with a MEK inhibitor. We will compare the model predictions for both active ERK and MEK levels to cytometry by time of flight (CyTOF) observed levels. This result should give us more confidence on how accurate the protein activity inference is for unobserved.

1.3. Outline

Chapter 2 will introduce the necessary background information for the rest of this work. It gives background information on the two data modalities that are used in this research. Specifically, it will explain how a biosensor based on Förster resonance energy transfer (FRET) can be used to measure the activity of signaling proteins on a single-cell level. Additionally, it will give some background information on cytometry by time of flight (CyTOF). Secondly, the chapter will give some background on techniques that can be used to model cellular signaling cascades in silico. An overview will be given of some existing methods that use Bayesian inference for parameter estimation of ordinary differential equations.

Chapter 3 provides the methods section of this research. It explains the data preprocessing steps performed on the biosensor and CyTOF data. Subsequently, it will explain how we inferred information about pathway activity.

Chapter 4 is the results section of this research. It will present how a subset of possible pathway topologies was chosen for further analysis. Next, a comparison will be made between the most relevant models under perturbed and unperturbed conditions. This comparison should give insight into possible resistance mechanisms that are activated under perturbation. To confirm the biological relevance of this result two control experiments are presented.

Chapter 5 will provide the most important contributions of this research and refer back to the research aims. Subsequently, a discussion of the limitations of this research is presented and recommendations are made for future work.

2

Background

This chapter provides background information on two methods that can be used to measure protein activity on a single-cell level. Additionally, it also discusses methods to infer information about kinetic parameters from longitudinal single-cell protein activity data. Specifically, it gives an overview of existing methods to perform Bayesian inference for ordinary differential equation parameters.

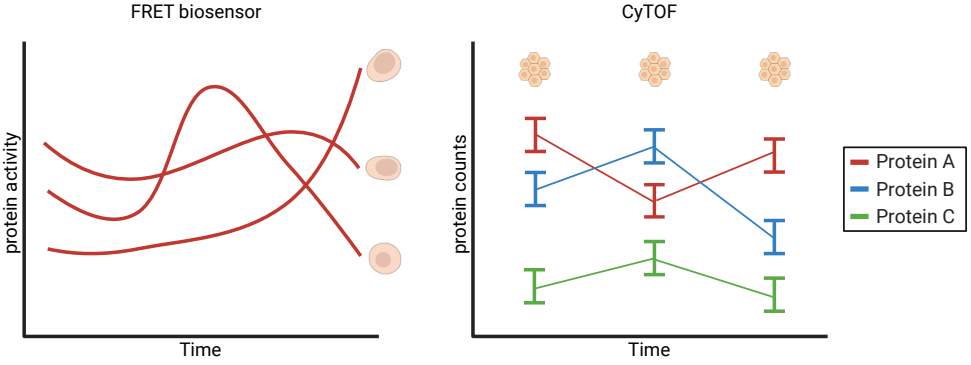

Figure 2.1: Overview of two single-cell protein activity measurement techniques. Gives an overview of the difference between Förster resonance energy transfer measurement and cytometry by time-of-flight measurements. Created with BioRender.com.

2.1. Single-cell protein activity measurement

There are several experimental techniques to observe protein activity on a single-cell level. Protein activity is defined as the kinetic activity of a protein. It is closely linked to the concept of protein phosphorylation and complex formation as most proteins are activated by phosphorylation of specific sites or protein-protein interactions [1]. If possible we will refer to protein activity instead of phosphorylation state and complex formation. This work uses such measurements of protein activity to infer information about the dynamics of cellular signaling pathways under different conditions. Specifically, we used two different data modalities. Förster resonance energy transfer (FRET) can be used in the form of a biosensor to observe longitudinal changes in the activity of proteins on a single-cell level. Due to its reliance on

fluorescence, it is increasingly hard to measure the activity of multiple proteins at the same time [18]. Cytometry by time-of-flight (CyTOF) can be used to generate counts for multiple proteins at specific cross-sections on a single-cell level. Choosing a specific panel of protein phosphorylation or complex formation states allows for the inference of protein activity. As this technique is destructive, it cannot be used to observe the same cell multiple times (see Figure 2.1).

2.1.1. Förster resonance energy transfer

Fluorescent molecules can transfer energy via Förster resonance when they are in close vicinity to each other. This energy transfer occurs through the resonance of the outermost electron shells (see Figure 2.2). It can only occur at very small distances (1̃0 nm) and the amount of transfer is dependent on this distance:

$$E_{FRET} = \frac{R_0^6}{R_0^6 + r^6} \tag{2.1}$$

, where E_{FRET} represents the amount of energy transfer given by the ratio between the donor excitation energy and the acceptor emission energy. R_0 represents the distance at 50% energy transfer. r represents the distance between the two fluorescent molecules. Thus, we can use this technique to determine distances between molecules at a very small scale.

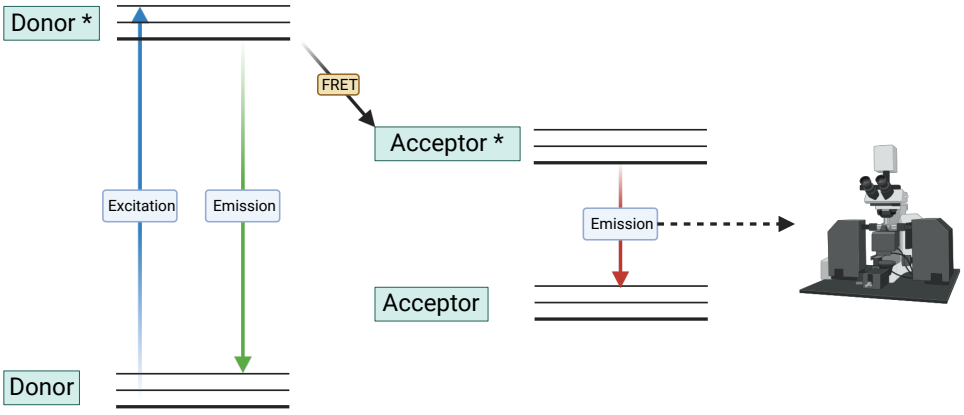

Figure 2.2: Jablonski plot of Förster resonance energy transfer. The donor is excited (donor*) with radiation. Subsequently, there is a simultaneous energy transfer through Förster resonance energy transfer (FRET) and photonic emission. The resonance energy transfer excites the acceptor, which emits a photon. This photon is detected to quantify the distance between donor and acceptor [19]. Created with BioRender.com.

2.1.2. Single-molecule FRET in proteins

When two fluorescent molecules are attached to a protein they can be used to observe conformational changes over time by measuring the resonance energy transfer. When this technique

is combined with single-molecule fluorescence measurement it gives sufficient accuracy to determine conformational states in single-cells. This concept can be utilized to measure activity by constructing a protein with a target domain for the protein of interest. If this domain is phosphorylated then the biosensor conformation changes, which changes the position of two integrated fluorophores. These changes in distance can then be quantified with FRET to determine protein activity [19, 20].

2.1.3. Cytometry by time-of-flight

Cytometry by time-of-flight (CyTOF) is a measurement technique that can quantify a large panel of proteins in single-cells. Cells are incubated with antibodies that are tagged with lanthanide metal isotopes. The cell suspension is then nebulized such that every droplet can only contain a single-cell. These cells are then ionized and all the naturally occurring ions are removed. This only leaves the artificially added lanthanide ions. These are then identified and quantified by time-of-flight mass spectrometry to recover the original protein counts in single-cells (see Figure 2.3). CyTOF can be used for protein panels of up to 60 in size [21].

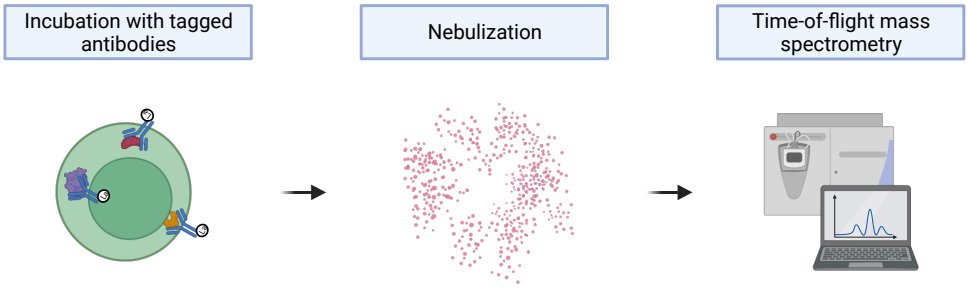

Figure 2.3: Overview of CyTOF workflow. Cells are incubated with lanthanide antibodies. These cells are nebulized and ionized. Subsequently, the ion content of every cell is measured with time-of-flight mass spectrometry [21]. Created with BioRender.com.

2.2. Inferring cellular signaling dynamics for heterogeneous cell populations

This section will give a short overview of some existing methods to infer cellular signaling behavior from protein activity observations. It will first describe how the kinetics of a single-cell can be inferred. Subsequently, a short overview will be given of some methods to infer signaling kinetics for heterogeneous cell populations.

2.2.1. Ordinary differential equations model of cell signaling

Systems of ordinary differential equations (ODEs) are commonly applied to model physical systems. In this research, we will apply ODEs to simulate the chemical reactions that occur

during cellular signaling. In this context, ODEs are used to describe the speed of change in the concentration of proteins (species):

$$\frac{dP}{dt} = \alpha - \beta P \tag{2.2}$$

, where $\frac{dP}{dt}$ describes the speed of change in concentration of species P, kinetic parameter α is the rate of production, and kinetic parameter β is the degradation rate. A system of such functions can be solved over time to acquire the concentration of all species over time. This does require an initial concentration for every species and the kinetic parameter values. The full system of ODEs for a unique single-cell can be described as follows:

$$\frac{d\boldsymbol{x}_i(t)}{dt} = f(\boldsymbol{x}_i(t), \varphi_i) \tag{2.3}$$

, where $\boldsymbol{x}_i(t)$ is the vector of species concentrations at time t for cell i. φ_i represents the vector of model parameters (kinetic parameters & initital concentrations) for cell i [22].

2.2.2. Bayesian inference

Bayesian inference is a method to infer the distribution of model parameters, unlike maximum likelihood estimation, which aims to find a single optimal value [23]. This parameter distribution can be defined in the context of cellular signaling ODEs as follows:

$$P(\phi_i|Y_i, M) = \frac{P(Y_i|\phi_i, M)P(\phi_i)}{\int P(Y_i|\phi_i, M)P(\phi_i)d\phi_i} \tag{2.4}$$

, where $P(\phi_i|Y_i, M)$ (posterior) is the probability of the inferred parameter set ϕ_i given an ODE model M and the matrix of observed species over time Y_i for a cell i. If the vector φ_i is fully inferred then $\varphi_i \subset \phi_i$. $P(Y_i|\phi_i, M)$ (likelihood) is the probability of observing Y_i given a specific parameter set ϕ_i and ODE model M. $P(\phi_i)$ (prior) is the a priori probability of the parameter set ϕ_i. $\int P(Y_i|\phi_i, M)P(\phi_i)d\phi_i$ represents the marginal distribution, which can be interpreted as the likelihood of this model given the observations.

Prior

The prior $P(\phi_i)$ in the context of computational pathway analysis is a probability distribution that can usually be based on literature. Normal and uniform distributions are commonly used for this purpose.

Likelihood

Usually, there will just be a subset of species that are observed. $y_{i,j,t}$ is defined as the activity of a species j for cell i at timepoint t. $z_{i,j,t}$ is defined as the ODE activity of a species j for a cell i at timepoint t for a given ϕ_i. To get these matched values the system of ODEs is implicitly solved for a given ϕ_i over a specific timecourse to match our number of timepoints T. This provides $x_{i,j,t}$, which is the concentration of a species j for a cell i at a timepoint t. To obtain

activity values from these concentrations a function $g(.)$ is defined, where $z_{i,j,t} = g(x_{i,j,t})$. The full likelihood can now be defined as follows:

$$P(Y_i|\phi_i, M) = \prod_{t=1}^{T}\prod_{j=1}^{S} N(y_{i,j,t}|g(f(\varphi_i)), \sigma) \tag{2.5}$$

, where S is the number of observed species, T is the number of observed timepoints. σ represents the standard deviation of the error for this specific data, which is usually inferred $\sigma \in \phi_i$ [22].

Marginal

The marginal can be defined as an integration of the likelihood over all possible values of ϕ_i. This integration is not tractable, thus the posterior cannot be calculated analytically. However, it can be estimated by employing sampling algorithms.

Sampling methods

The simplest algorithm to sample the posterior is the Metropolis-Hastings algorithm:

$$\phi_{i+1} = N(\phi_i, \sigma) \tag{2.6}$$

$$r = \frac{P(\phi_{i+1}|\sigma)P(Y_i|\phi_{i+1}, M)}{P(\phi_i|\sigma)P(Y_i|\phi_i, M)} \tag{2.7}$$

, where r represents the acceptance probability for a new sample ϕ_{i+1}. $N(\phi_i, \sigma)$ represents the proposal distribution with σ as its standard deviation. This method provides the possibility to get samples from the posterior. Parallel tempering Markov chain Monte-Carlo is used in this work [22].

2.2.3. Cell population likelihood

Several methods have been proposed to extend this Bayesian inference process to also allow for simultaneous inference of model parameters for a cell population. Hasenauer et al. described the usage of a mixture of ansatz functions as a prior, to be able to infer multimodal posteriors from cross-sectional data. The different modes represent populations of cells with similar kinetics [24]. Dixit et al. proposed a non-parametric approach to infer multimodal posteriors from cross-sectional data [25]. Loos et al. describe a hierarchical approach that can infer multimodal posteriors without having to define ansatz functions from cross-sectional data. However, it does require an assumption on the number of cell clusters [26]. Dharmarjan et al. proposed a method to fit a linear mixed-effects model for longitudinal data [27]. Thijssen et al. a posterior distribution from longitudinal data [22], which will described in depth in the following section.

2.2.4. Thijssen et al.

To fit multiple cells Thijssen et al. assume that some of the kinetic parameters can vary between cells. The kinetic variables that vary between cells are calculated as follows:

$$\varphi_{i,v} = 2^{\alpha_{i,v}}\mu_v \tag{2.8}$$

$$\alpha_{i,v} \sim N(0, \sigma_v) \tag{2.9}$$

$$\mu_v \sim P(\mu_v) \tag{2.10}$$

,where $\varphi_{i,v}$ represents the kinetic parameter v for cell i. μ_v is the mean value of this parameter for all cells. This mean value is varied by transformation on a log scale with $\alpha_{i,k}$ to create a distinct single-cell. $\alpha_{i,k}$ is sampled from a normal distribution with a standard deviation of σ_k. Both σ_k and μ_k are usually defined to have a prior distribution. Simulating multiple distinct cells can be achieved by sampling various values for $\alpha_{i,k}$.

To calculate the full likelihood, each simulated cell is matched to an observed cell. To match simulated cells to observed cells all pairwise likelihoods are calculated between observed and simulated cells. This problem can then be rewritten as an assignment problem. The Hungarian algorithm is used to find the optimal pairing of simulated to observed cells. These pairs can then be used to calculate the full likelihood over all fitted cells as follows:

$$P(Y|\phi) = \prod_{i=1}^{N}\prod_{t=1}^{T}\prod_{j=1}^{S} N(y_{i,j,t}|g(f(\varphi_i)), \sigma) \tag{2.11}$$

To make the simulation deterministic a quasi-random normal is used to replace $N(0, \sigma_k)$. A Sobol sequence is used to generate this variable by converting it from a uniform range to a normal range by applying an inverse cumulative density function transform [22].

2.2.5. Marginal likelihood for model comparison

Models can be compared to each other by using the marginal likelihood $\int P(Y|\phi, M)P(\phi)d\phi$. However, as previously mentioned this marginal likelihood is not analytically solvable, thus it is estimated from the posterior samples [28]. Different models can be compared by calculating the Bayes factor (BF) with the marginal likelihood:

$$BF = \frac{\int P(Y|\phi, M_1)P(\phi|M_1)d\phi}{\int P(Y|\phi, M_2)P(\phi|M_2)d\phi} \tag{2.12}$$

Where M_1 and M_2 represent two different models. Table 2.1 shows how these Bayes factor values can be interpreted.

Table 2.1: Interpretation of the Bayes factor [29].

ln(BF)	Interpretation
0-1.1	hardly worth mentioning
1.1-3	positive support
3-5	strong support
>5	overwhelming support

3

Methods

This chapter gives an overview of the methods used to get to the results of this research. It starts by describing the FRET biosensor dataset and the preprocessing steps that were performed. Next, it describes the method that was used to infer the cell signaling kinetics from protein activity data in single-cells. Then the chapter will describe the CyTOF data set and the preprocessing steps that were performed. Lastly, the simulation study used to assess the biological relevance of the kinetic parameter estimates is discussed.

3.1. FRET biosensor dataset

To infer signaling dynamics we make use of the experimental observation reported by Ponsioen et al. The authors report the usage of a novel ERK biosensor based on FRET. This biosensor was engineered to reflect the activity of ERK in the distance between the donor and acceptor fluorophore. The sensor domain (cdc25) is a downstream target of ERK. When this domain is phosphorylated by ERK its affinity for the ligand domain increases. Subsequently, the distance between these two domains is decreased. This leads to an increased energy transfer between the two attached fluorophores, which can be quantified with fluorescence microscopy. This energy transfer can then be converted to the distance between the two fluorophores. In turn, this distance is indicative of the activity of the activity of ERK (Figure 3.1).

For this research, the data reported by Ponsioen et al in Figure 4a, 4c and extended Figure 7c was used [20]. All of these measurements were made in patient-derived organoids (PDOs) from colorectal tissue. Only the KRAS G12V mutation was analyzed as it is prevalent and was also analyzed in the external validation dataset. The biosensor data was first normalized to a [0-1] range using the superinhibition and superactivation regimes as min-max values. This was performed for every cell separately. Subsequently, only the time range of interest was extracted.

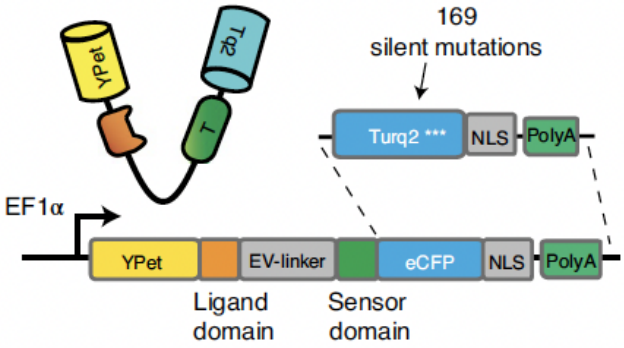

Figure 3.1: Förster resonance energy transfer biosensor EKAREN5. Contains a sensor domain which is a downstream target of activated ERK. The conformational changes due to activation are reflected in the distance between the YPet and Tq2 fluorophores. Figure from Ponsioen et al. [20].

3.2. Kinetic parameter estimation for cellular signaling

To infer probability distributions of the initial protein concentrations and the mean/variance of kinetic parameters the method proposed by Thijssen et al. was used [22]. Two novel features were incorporated to be able to answer the posed research questions. Both of these changes were applied in C++ in a forked version of the original method proposed by Thijssen et al. (see Github).

3.2.1. Inferring the ratio between active and inactive protein

The original method proposed by Thijssen et al. allows for the inference of initial species concentrations. Total concentration estimates are usually readily available for the most important signaling proteins. However, in this context, we are interested in inferring the amount of active and inactive protein over time. Thus, to reduce the number of inference variables a method was implemented to just infer the initial ratio between active and inactive species if the total concentration is provided.

3.2.2. Modeling drug perturbations

A new method was implemented to model the instantiation addition of drugs for perturbation. At the point of drug addiction, the ODE integration was halted. Subsequently, the ODE right-hand side and jacobian were recalculated. Next, the ODE integration was reinitiated. This process can take place multiple times. It was found that using such discontinuity integration did affect the sampling convergence. However, this method was found to be most accurate in terms of inferring parameter distributions that could recapitulate the observed signaling dynamics.

3.3. CyTOF dataset

To check the biological accuracy of the inference process we make use of a second dataset. Brandt et al. report CyTOF measurement of a panel of proteins [30]. These measurements were also performed on PDOs from colorectal tissue with a KRAS G12V mutation. For this work, we are specifically interested in counts of p-ERK (phosphorylated ERK) and p-MEK (phosphorylated MEK) in Figure 7. These measurements were first arcsinh(count/5) normalized. This normalization is similar to a log transform, however it can also handle zero values which are common in count data [31]. They were then converted to activity values by calculating the min-max scaling over all measured values in all perturbed and unperturbed conditions.

3.4. CyTOF validation experiment

To check the biological relevance of the inferred signaling kinetics a simulation study was performed. In this simulation study, unperturbed ERK activity was used as an initial condition. For each of these values, a cell was simulated for 3h using the parameter set inferred from FRET biosensor data under perturbation. A cell was chosen with a uniform condition from all of the cells simulated for one parameter set with the method described by Thijssen et al. [22].

4

Results

This chapter gives an overview of the most important research results. Firstly, it describes the construction of a minimal model of the mitogen-activated protein kinase (MAPK) pathway that can be used for cell population analysis. Subsequently, it will compare several versions of this model that contain combinations of feedback mechanisms for unperturbed KRAS mutant cells. This comparison is made to find which model is most relevant under each condition. Next, the chapter will describe the comparison between the models fitted to KRAS mutant and wildtype cells, to determine if it is possible to detect the gain-of-function in KRAS activity. The chapter will then go on to compare the inferred kinetic parameters of KRAS mutant cells under perturbation with unperturbed cells. This should give some indication of what mechanism could be the cause of treatment resistance in KRAS mutant cells. Lastly, another validation experiment is described where the ability of the model to predict protein activity is evaluated by comparing it to a CyTOF dataset.

4.1. Novel minmal MAPK pathway model for cell population analysis

A novel MAPK pathway model was constructed by making use of ODEs. The model topology was based on several different reports on the most probable simplification of the signaling reality. The central signaling cascade is proposed to consist of RAS - RAF - MEK - ERK with a direct feedback inhibition by ERK on the phosphorylation of RAS [32–34]. To decrease model complexity RAF was omitted and summarized in the activation of MEK by RAS. Direct inhibition of RAS activation by ERK was introduced to replace the original feedback on RAF proposed by Kholodenko et al. [32]. Based on Lake et al. [8] two additional important feedback mechanisms were added to the model. DUSP is known to play an important role in the regulation of MAPK dynamics with its direct inhibition effect on the activation of ERK. Similarly, SPRY plays an important role by inhibiting the upstream activity of the MAPK pathway. SPRY was modeled to inhibit the activity of the GRB-SOS complex. Both of these proteins are transcriptionally induced, thus they were modeled to have a delayed inhibition effect as compared to the direct feedback mechanism on RAS activation [8, 33]. The full model is presented in Figure 4.1

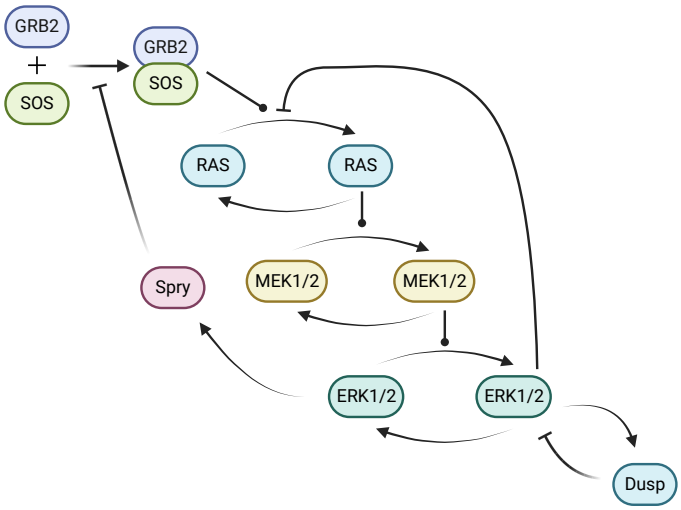

Figure 4.1: Full MAPK topology with ERK, SPRY and DUSP feedback mechanisms. Arrows represent species conversions. Lines with dots represent the activation of conversion. Lines with perpendicular lines represent inhibition of conversion. Created with BioRender.com.

4.1.1. Smallest model toplogy adequate for describing ERK protein activity in single cells

There is an array of other important proteins and mechanisms that contribute to MAPK dynamics. The proposed model is the result of an iterative process of elimination to obtain the minimal model to answer the research questions posed in this work. This final model is the 12th iteration, considering that an iteration means changing the graph topology. To illustrate this process one of these iterations is presented.

The central work on computational modeling of the MAPK pathway presented by Kholodenko et al. [32] includes three phosphorylation states of MEK and ERK. It was found that a model without these states like in Figure 4.2A is sufficient to describe oscillations that are generally observed in the activity of ERK in KRAS mutant cells. Figure 4.2B presents a single-cell posterior predictive distribution of the ERK activity in a KRAS mutant cell fitted with this simple model [20]. This model was fitted with the adapted version of the method presented by Thijssen et al. [22]. The initial conditions, kinetic parameters and data-specific variables are a subset of the ones used in the full model. A description of the system of ODEs, fixed parameters, and inferred parameters is given in the next section.

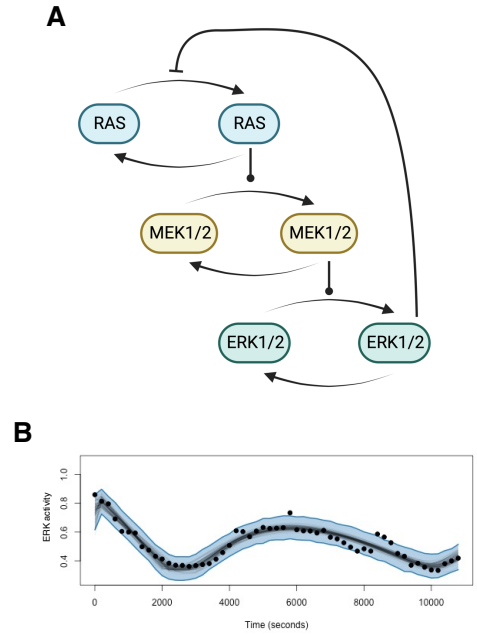

Figure 4.2: (A) Simple topology of MAPK pathway without double phosphorylation states. Created with BioRender.com (B) Single-cell ERK activity ppd with the simple MAPK pathway model. The model was fitted on ERK activity data presented by Ponsioen et al. [20] in Figure 4a.

4.1.2. Full model description

All reactions use Michealis-Menten kinetics for both catalyzed and uncatalyzed reactions. The effect of the MEK and HER inhibitors was modeled using IC50 values without a plateau value. The full system of ODEs is shown in Table 4.1.

Table 4.1: All species conversions included in the model. Both the forward and reverse differential of the system of ODEs are shown.

Reaction	Forward differential	Reverse differential
$ERK \leftrightarrows ERKpp$	$kcat_{mek-erk} * MEK * \frac{ERK}{ERK+km_{mek-erk}} * \frac{1}{1+\frac{DUSP}{kI_{dusp}}} * \frac{1}{1+\frac{MEKi}{IC50_{meki}}}$	$kcat_{erk} * \frac{ERKpp}{ERKpp+km_{erk}}$
$MEK \leftrightarrows MEKpp$	$kcat_{ras-mek} * RAS * \frac{MEK}{MEK+km_{ras-mek}}$	$kcat_{mek} * \frac{MEKpp}{MEKpp+km_{mek}}$
$RAS \leftrightarrows RASpp$	$kcat_{grb-ras} * GRBSOS * \frac{RAS}{RAS+km_{grb-ras}} * \frac{1}{1+\frac{ERK}{kI_{erk}}}$	$kcat_{ras} * \frac{RASpp}{RASpp+km_{ras}}$
$GRB + SOS \leftrightarrows GRBSOS$	$kcat_{grbsos_f} * \frac{GRBSOS}{GRBSOS+km_{grbsos_f}} * \frac{1}{1+\frac{HERi}{IC50_{heri}}} * \frac{1}{1+\frac{SPRY}{kI_{spry}}}$	$kcat_{grbsos_b} * \frac{GRBSOS}{GRBSOS+km_{grbsos_b}}$
$\emptyset \leftrightarrows DUSP$	$1 + dusp_{ind} * \frac{ERK^2}{ERK^2+k_{dusp}} * \frac{ln(2)}{t_{dusp}}$	$DUSP * \frac{ln(2)}{t_{dusp}}$
$\emptyset \leftrightarrows SPRY$	$1 + spry_{ind} * \frac{ERK^2}{ERK^2+k_{spry}} * \frac{ln(2)}{t_{spry}}$	$SPRY * \frac{ln(2)}{t_{spry}}$

Fixed model parameters

DUSP and SPRY were initialized at 0 nM. The total amount of ERK and MEK was set at 1000 nM [32]. The total amount of RAS was set at 100 nM [35]. The total amount of GRB-SOS was set at 34 nM [36]. The kinetic parameters for SPRY and DUSP conversion were based on ryu et al. [33]. The IC50 value of Cetuximab was fixed from literature [37]. The IC50 value of Trametinib was fixed from literature [38]. The IC50 value of Selumetinib was fixed from literature [39]. Table 4.2 gives an overview of all fixed parameters.

Table 4.2: Fixed parameters included in the full model.

Type	Name	Description	Value	Unit
Kinetic	k_m values	Michaelis-Menten constant	1	nM
Kinetic	$dusp_{\text{ind}}$	DUSP induction level	3	nM
Kinetic	t_{dusp}	DUSP activation time	5400	s
Kinetic	k_{dusp}	DUSP rate constant	0.1	nM
Kinetic	$spry_{\text{ind}}$	SPRY induction level	3	nM
Kinetic	t_{spry}	SPRY activation time	5400	s
Kinetic	k_{spry}	SPRY rate constant	0.1	nM
Kinetic	$kcat_{\text{\_grbsos\_f}}$	GRBSOS forward rate	0.001	1/s
Kinetic	$kcat_{\text{\_grbsos\_b}}$	GRBSOS backward rate	0.001	1/s
Kinetic	$IC50_{meki}$	IC50 value of MEKi (Trametinib)	0.8	nM
Kinetic	$IC50_{meki}$	IC50 value of MEKi (Selumetinib)	14.1	nM
Kinetic	$IC50_{heri}$	IC50 value of HERi (Cetuximab)	0.33	nM
Initial concentration	$dusp$	Initial concentration of DUSP	0	nM
Initial concentration	$spry$	Initial concentration of Spry	0	nM
Initial concentration	$Total\ grb\ \&\ sos$	Initial concentration of total GRB and SOS	34	nM
Initial concentration	ras	Initial concentration of Ras	1000	nM
Initial concentration	mek	Initial concentration of MEK	1000	nM
Initial concentration	erk	Initial concentration of ERK	1000	nM

Inferred model parameters

To obtain activity values from the ODE concentration values a linear model was used. This previously mentioned function $g(.)$ was defined as follows for this model:

$$g(x_{erk,i,t}) = x_{erk,i,t} * \beta + \alpha = z_{erk,i,t} \tag{4.1}$$

The offset α was defined on a positive range to ensure no negative values could arise. The scale β was chosen to center around 10^{-3} as ERK was assumed to be present at $\sim 10^3$ nM in the cell. The data-specific standard deviation of the error σ was assumed to have a gamma prior distribution with a mode around 0.1 as this is the expected standard deviation at this scale.

Finally, we also pick a set of parameters that vary between cells (see Background). The deactivation of ERK was chosen to vary between cells. This choice was based on single-cell fits which indicated that this kinetic parameter varies significantly between cells.

Secondly, the entry time was chosen to vary between cells. The entry time indicates the time point where the likelihood evaluation starts. This variance is important as not all cells are

synchronized in their ERK activity dynamics. For example, one cell might be observed starting at the peak of oscillation while a second cell starts at the trough. An accurate model might be able to explain the ERK oscillations but it could be constrained to this oscillation starting at the midline. Thus, this variance term can correct this problem. All of these inferred variables are shown in table 4.3.

Table 4.3: Inference variables of full MAPK model.

Type	Name	Description	Distribution	Lower	Upper	Unit
Data-specific	α	Intercept of g(.)	Uniform	0.0	0.4	no unit
Data-specific	β	Slope of g(.)	Uniform	$10^{-4.0}$	$10^{-2.0}$	no unit
Variability parameter	$\sigma_{\text{entry time}}$	Entry Time Variability	Uniform	10^{0}	$10^{3.69897}$	no unit
Kinetic parameter	$kcat_{\text{grb-ras}}$	Activation of Ras by GRBSOS	Uniform	10^{-4}	10^{2}	1/s
Kinetic parameter	$kcat_{\text{ras}}$	Deactivation of Ras	Uniform	10^{-4}	10^{2}	1/s
Kinetic parameter	$kcat_{\text{ras-mek}}$	Activation of MEK by RAS	Uniform	10^{-4}	10^{2}	1/s
Kinetic parameter	$kcat_{\text{mek}}$	Deactivation of MEK	Uniform	10^{-4}	10^{2}	1/s
Kinetic parameter	$kcat_{\text{mek-erk}}$	Activation of ERK by MEK	Uniform	10^{-4}	10^{2}	1/s[Your Unit]
Kinetic parameter	$kcat_{\text{erk}}$	Deactivation of ERK	Uniform	10^{-4}	10^{2}	1/s
Kinetic parameter	kI_{erk}	Inhibition Constant for ERK	Uniform	10^{-2}	10^{10}	nM
Kinetic parameter	kI_{dusp}	Inhibition Constant for DUSP	Uniform	10^{-2}	10^{10}	nM
Kinetic parameter	kI_{spry}	Inhibition Constant for Spry	Uniform	10^{-2}	10^{10}	nM
Kinetic parameter	$ratio_{erk}$	Ratio between active and inactive ERK at baseline	Uniform	0	1	no unit
Type	Name	Description	Distribution	κ	θ	Unit
Data-specific	σ	Standard deviation of the error	Gamma	10	0.01	no unit
Type	Name	Description	Distribution	μ	σ	Unit
Variability parameter	$\sigma_{\text{ERK deactivation}}$	ERK Deactivation Variability	Normal	0	1	no unit

Model versions

Several different versions of this full model were defined. Each of these versions has a different combination of feedback mechanisms. These different versions were used to find the minimally sufficient model to describe a population of KRAS mutant cells with or without perturbations. They are defined as follows:

1. No feedback inhibition
2. ERK feedback inhibition
3. DUSP feedback inhibition
4. SPRY feedback inhibition
5. ERK + DUSP feedback inhibition
6. ERK + SPRY feedback inhibition
7. DUSP + SPRY feedback inhibition
8. ERK + DUPS + SPRY feedback inhibition

4.1.3. Conclusion

Many MAPK pathway models have been proposed in literature [32–34]. The novel model proposed in this section takes inspiration from these models but was optimized to be as minimalistic as possible. This was to enable cell population analysis by decreasing the model complexity. It was shown that only two phosphorylation states of ERK and MEK are required to describe the activity of ERK protein in a single-cell.

4.2. ERK feedback is sufficient to fit unperturbed KRAS signaling dynamics

The first step in the analysis was to find which model version is sufficient to describe the dynamics of the MAPK pathway. To make this comparison the different versions of the model were fitted to ERK protein activity time series of nine KRAS mutant cells presented by Ponsioen et al. [20]. 18 cells were simulated in total.

Figure 4.3A shows the ln(BF) between all model permutations and the worst-performing model. The worst-performing model was the model with ERK + SPRY feedback inhibition. The model with just ERK, ERK + DUSP, and ERK + DUSP + SPRY were found to perform best.

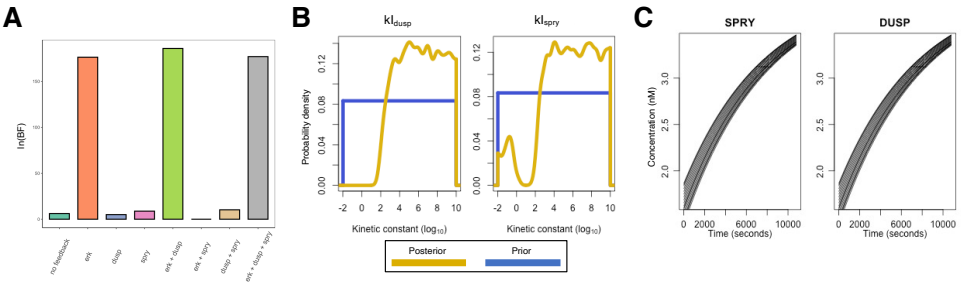

Figure 4.3: Comparison of all eight different models. Calculated from model fits to 9 KRAS mutant cells reported by Ponsioen et al. [20]. (A) ln(BF) between all models and the worst performing model. (B) Marginal posterior distribution of DUSP and SPRY inhibition for the full model version. (C) Simulated DUSP and SPRY concentrations for the full model version.

Figure 4.4A shows that the model without a feedback mechanism can describe the trend in ERK protein activity. However, Figure 4.4B shows that the model with ERK inhibition is also able to describe the oscillatory behavior in the ERK activity.

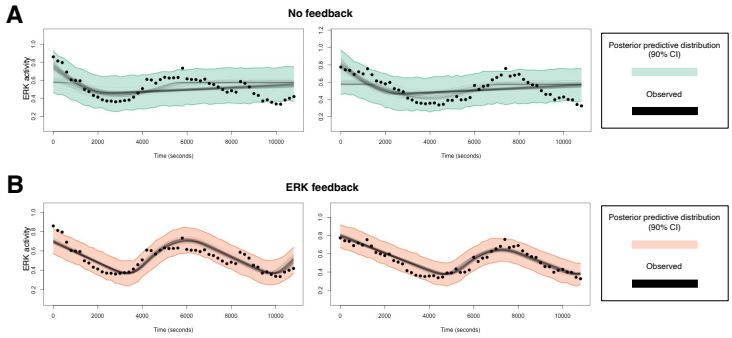

Figure 4.4: Posterior predictive distribution (ppd) of ERK protein activity for 9 KRAS mutant cells reported by Ponsioen et al. [20]. (A) No feedback model ppd (only two cells are shown) (B) ERK feedback model (only two cells are shown).

Figure 4.5 shows that the ERK feedback model also predicts the oscillatory behavior for MEK protein activity. RAS protein activity is quite stable.

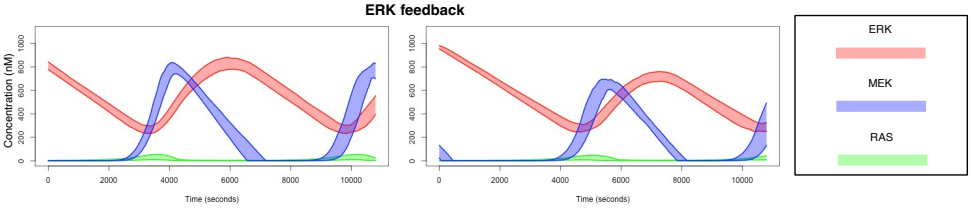

Figure 4.5: Posterior predictive distribution (ppd) of ERK, MEK and RAS protein activity for 9 KRAS mutant cells reported by Ponsioen et al. [20] (only two cells are shown).

These results prompt the question of why the other models performed so poorly compared to the models with ERK, ERK + DUSP, and ERK + DUSP + SPRY feedback inhibition. The model without any feedback inhibition was expected to have a low marginal likelihood as we normally expect oscillations in the activity of ERK. Sustained oscillations cannot mathematically be described by a model without any feedback mechanism [32]. For the other models, it seems that the ERK feedback mechanism is necessary to accurately fit ERK protein activity. Additionally, the SPRY feedback inhibition decreases the ability of the model to describe the ERK dynamics. This could indicate that the GRB - SOS complex is not required to accurately describe the MAPK dynamics.

Figure 4.3B shows the model posterior for kI_{dusp} and kI_{spry} for the model ERK + DUSP + SPRY model, which indicates that both of these kinetic constants are above 10^2. Figure 4.3C shows the simulated values of SPRY and DUSP over time, both of which have a maximum concentration of around 3 nM. Together, these results indicate that the effect of the DUSP and SPRY inhibition are not significant. This can be shown by calculating the inhibitive effect of the DUSP and SPRY feedback mechanism (see table 4.1):

$$\begin{aligned} \frac{1}{1+\frac{SPRY}{kI_{spry}}} &= \frac{1}{1+\frac{3}{100}} = 0.97 \simeq 1 \\ \frac{1}{1+\frac{DUSP}{kI_{dusp}}} &= \frac{1}{1+\frac{3}{100}} = 0.97 \simeq 1 \end{aligned} \tag{4.2}$$

Thus, it can be concluded that the effect of DUSP and SPRY is negligible. The same is true for the effect of DUSP in the model with ERK + DUSP. This result indicates that the model with just the ERK feedback mechanism is the most accurate description of the MAPK dynamics given these KRAS mutant cells.

4.2.1. Conclusion

This section compared the different versions of the MAPK pathway model to each other. It was found that only direct feedback of ERK on the activation of RAS is necessary to describe the ERK protein activity of KRAS mutant cells. This is in line with the seminal work by Kholodenko et al. [32]. However, this model is even more minimal as it doesn't contain the triple phosphorylation states of ERK and MEK.

4.3. KRAS and wildtype signaling do not show a clear difference in RAS activity

To confirm the biological relevance of this model a control experiment was performed. This experiment should indicate if the parameter estimates of the model are reliable. In this experiment, it was tested whether the model inferred the gain-of-function of RAS due to KRAS gene mutation without any a priori information on this phenomenon.

For this comparison, the model was fitted to ERK protein activity time series of nine wildtype cells presented by Ponsioen et al. [20]. Figure 4.6 presents the model fit, which shows that we can also accurately fit the ERK protein activity of wildtype cells.

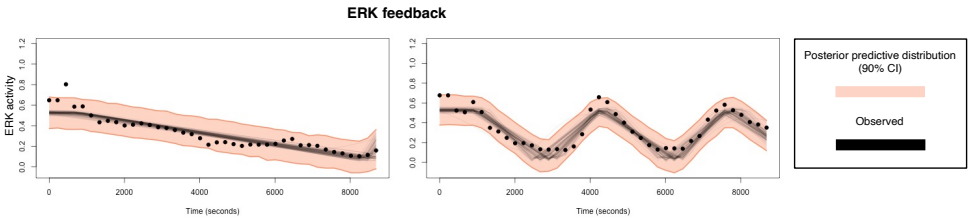

Figure 4.6: Posterior predictive distribution (ppd) of ERK protein activity for 9 wildtype mutant cells reported by Ponsioen et al. [20]. Only the model with ERK feedback was used.

Figure 4.7 presents the posterior samples of RAS activation and RAS deactivation for wildtype and KRAS cells. These samples indicate that the deactivation of RAS is increased in KRAS mutant cells. We would expect due to the gain-of-function in RAS that this would be reversed. This result indicates that the kinetic parameter estimates are less accurate than hoped. The result could be due to the model being fitted to only one observed protein activity. A more accurate estimate might be attained when the kinetic parameters are inferred from more than one dataset or a dataset with multiple protein activity measurements.

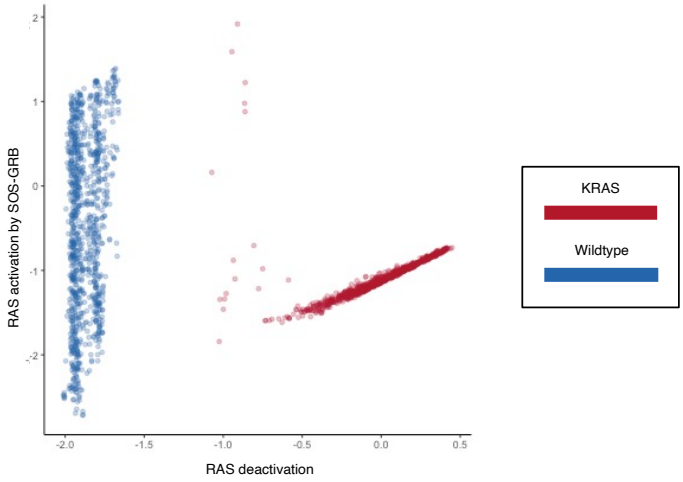

Figure 4.7: Posterior samples of RAS activation and deactivation. The posterior samples for the wildtype and KRAS fit are presented. These samples were both attained by fitting 9 wildtype and 9 KRAS mutant cells reported by Ponsioen et al. [20] (only two cells are shown). Only the model with ERK feedback was used.

4.3.1. Conclusion

This section explored the ability of the proposed model and method to infer kinetic parameter estimates. It was found that these estimates were less accurate than hoped. The gain-of-function was not inferred from the protein activity data. This could be explained by the fact that this model was fitted to just one dataset with only one protein activity observation.

4.4. No clear kinetic differences between unperturbed and perturbed KRAS mutant cells

The main goal of this research is to uncover the mechanistic difference between unperturbed KRAS mutant cells and drug perturbed KRAS mutant cells. Specifically, it is of interest why the combination of a MEK and HER inhibitor induces resistance.

To make this comparison the different versions of the model were fitted to ERK protein activity time series data of nine KRAS mutant cells treated with trametinib (MEKi) and cetuximab (HERi) presented by Ponsioen et al. [20]. The concentration of trametinib was assumed to be 100 nM [40]. 18 cells were simulated in total.

In the full model we can observe the same phenomenon that was seen in unperturbed cells. Figure 4.8 shows that the inhibitory effect of DUSP and SPRY is negligible following the same rationale used for unperturbed cells.

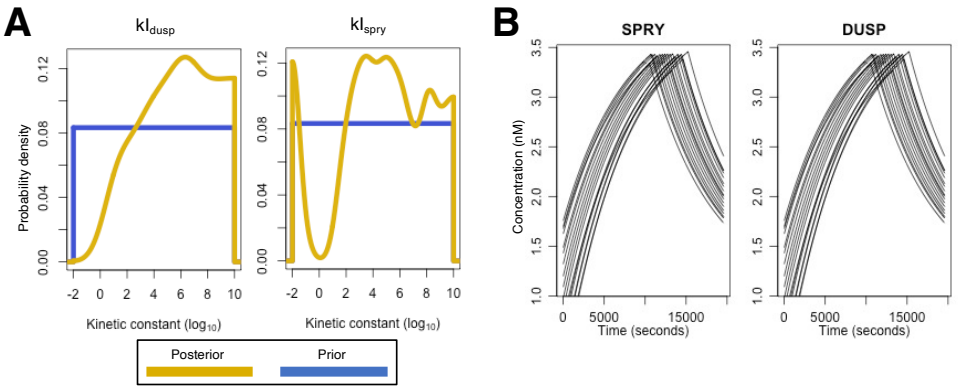

Figure 4.8: DUSP and SPRY concentration and kinetics fitted to ERK protein activity for 9 KRAS mutant cells treated with trametinib (100 nM) and cetuximab (3.289 nM) reported by Ponsioen et al. [20]. The full MAPK model was used with all three feedback inhibitions. (A) Marginal posterior distribution of DUSP and SPRY inhibition for the full model version. (B) Simulated DUSP and SPRY concentrations for the full model version.

Figure 4.9 shows that the model without a feedback mechanism cannot describe the transient ERK protein activity in one of the cells. However, Figure 4.9B shows that the model with ERK inhibition is able to describe this behavior. This result again confirms that the ERK feedback model is sufficient to describe ERK protein activity, also under perturbation.

Thus, from these results, it can be concluded the ERK model is the most relevant model to use for KRAS mutant cells treated with a MEK and HER inhibitor.

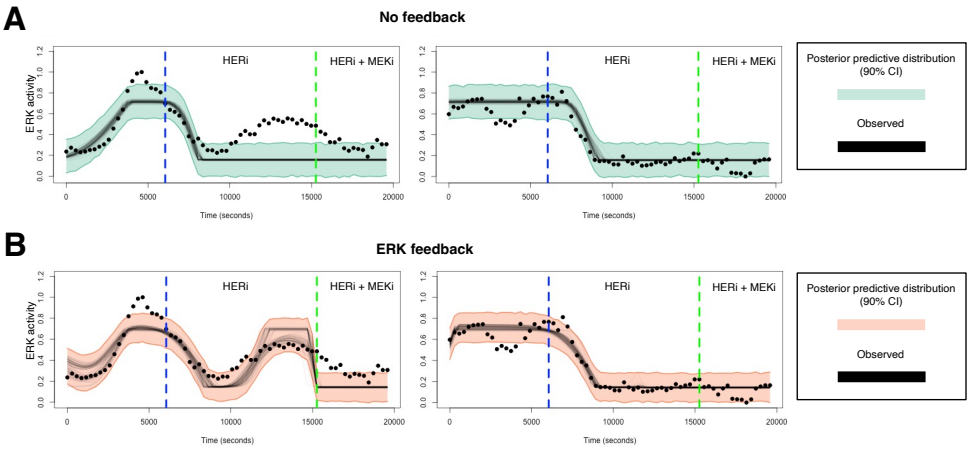

Figure 4.9: Posterior predictive distribution (ppd) of ERK protein activity for 9 KRAS mutant cells treated with trametinib (100 nM) and cetuximab (3.289 nM) reported by Ponsioen et al. [20]. (A) No feedback model ppd (only two cells are shown) (B) ERK feedback model (only two cells are shown).

Figure 4.10 shows that the model doesn't predict oscillatory behavior for MEK protein activity. This is in contrast to the unperturbed KRAS mutant cells. This is probably due to the inhibition of the activity of MEK and its target ERK. As ERK protein activity doesn't oscillate as much, due to the perturbation, it also doesn't affect MEK as much due to the negative feedback on the activation of RAS. This result is in concordance with the expectation.

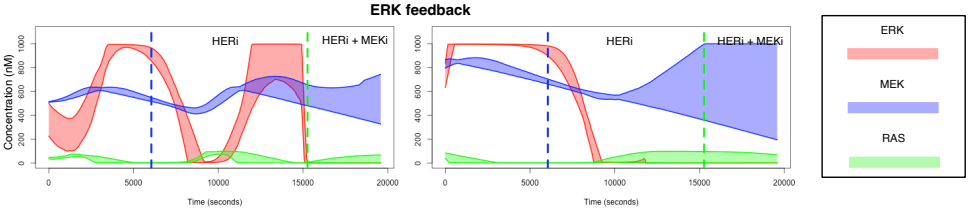

Figure 4.10: Posterior predictive distribution (ppd) of ERK, MEK and RAS protein activity for 9 KRAS mutant cells treated with trametinib (100 nM) and cetuximab (3.289 nM) reported by Ponsioen et al. [20] (only two cells are shown). Only the model with ERK feedback was used

Figure 4.11A shows the most important non-variant kinetic variable posterior distributions of the Ponsioen model fits for unperturbed and perturbed KRAS mutant cells. It is clear that for all of these parameters the activation and deactivation change on treatment. However, the effects balance each other out. If the activation is increased then the deactivation is increased. There is no clear change in kinetics shown in these model fits between perturbed and unperturbed cells.

Figure 4.11B shows the inferred variance term for the breakdown of ERK. Here we do see that there is a large variance in the breakdown of ERK for untreated cells. This could indicate that under perturbed conditions the cells are more homogeneous in their behavior.

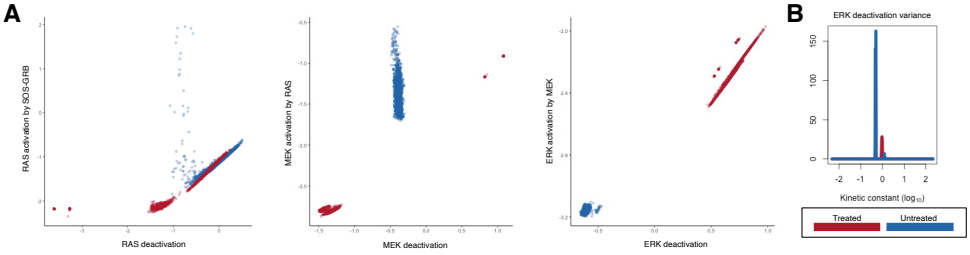

Figure 4.11: Posterior samples and distributions for the perturbed and unperturbed KRAS mutant cells are presented. These samples were both attained by fitting 9 unperturbed and 9 perturbed KRAS mutant cells reported by Ponsioen et al. [20]. Only the model with ERK feedback was used. (A) Posterior samples of RAS activation and deactivation. (B) Posterior samples of RAS activation and deactivation. (C) Posterior samples of RAS activation and deactivation. (D) Posterior distribution of ERK variance.

4.4.1. Conclusion

This section explored the difference in signaling dynamics described by the ERK feedback model for perturbed and unperturbed KRAS mutant cells. It was found that the ERK model was the most relevant for describing the ERK protein activity under perturbation. This was also true for unperturbed cells. The predicted levels of MEK and RAS were in concordance with the expectation which confirmed that the prediction of unobserved protein activity is robust. The kinetic parameter analysis did not indicate any functional difference under perturbations. This could be due to the models being fitted to only one dataset and/or there only being one observed protein activity. It indicates that there is not enough information to infer these parameters.

4.5. Comparison to CyTOF data

A second confirmation experiment was performed to test if this model can predict ERK and MEK protein activity after drug treatment based only on baseline ERK activity in unperturbed cells.

The dataset used for comparison was presented by Brandt et al. [30]. KRAS colorectal cancer PDOs were treated with a MEK inhibitor (selumetinib 1000 nM). CyTOF was performed at baseline and after 3 hours of treatment. We are specifically interested in the measurements of ERK and MEK protein levels. The baseline level of ERK was used as input for the $ratio_{erk}$ variable.

Next, the posterior samples from the previously mentioned Ponsioen et al. ERK protein activity data in perturbed cells were used to simulate the ODE for 3h (see Methods). Figure 4.12B shows that the model does predict a similar result. There are some deviations, most notably the activity of ERK is more extremely inhibited in the simulation. Considering that protein activity measurements are from a different data modality, in PDOs from different patients, and gathered at different labs this result is positive. It was shown that the model can accurately predict protein activity over time.

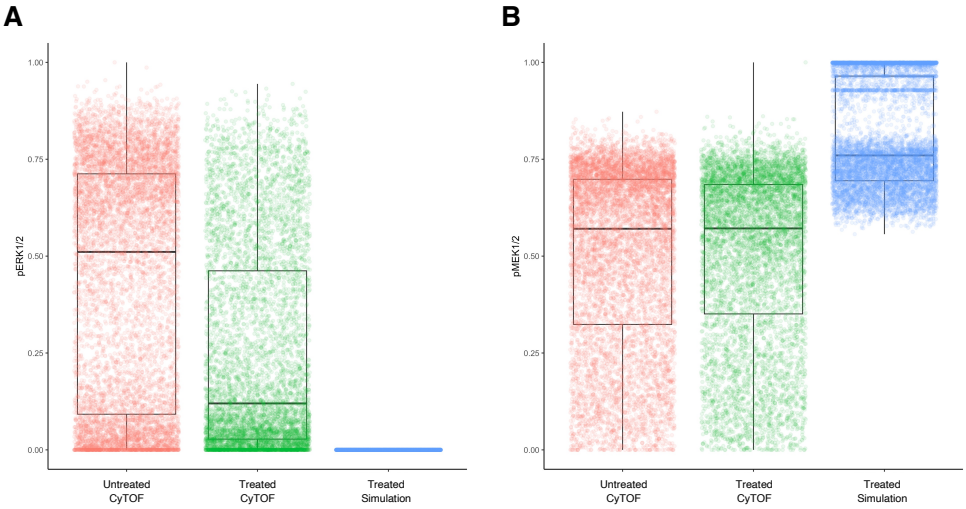

Figure 4.12: A: CyTOF measurement of ERK activity in untreated KRAS mutant cells. CyTOF measurement of ERK activity in KRAS mutant cells treated for 3h with selumetinib. Simulated ERK activity in KRAS mutant cells treated for 3h with selumetinib. For this simulation, the cells were initiated with the ERK activity measured with CyTOF in the untreated KRAS mutant cells B: CyTOF measurement of MEK activity in untreated KRAS mutant cells. CyTOF measurement of MEK activity in MEK mutant cells treated for 3h with selumetinib. Simulated MEK activity in KRAS mutant cells treated for 3h with selumetinib. For this simulation, the cells were initiated with the ERK activity measured with CyTOF in the untreated KRAS mutant cells.

4.5.1. Conclusion

This section explored the capabilities of this model to predict protein activity in an external dataset. It was found that the activity of ERK was predicted to be more extreme than it was in reality. However, the result still showed that this model can predict protein activity levels for different types of perturbations. This is useful as a confirmation of the biological relevance of this model.

5

Conclusion

KRAS mutants are very prevalent in many types of cancer [3, 4]. These mutations are important drivers of pancreatic, colorectal and non-small cell lung cancer. In vitro experiments have shown that the combination of a MEK and HER inhibitor is very potent in terms of its ability to kill KRAS mutant cells [11]. However, clinical results were not conformable to this notion [3–5]. This work aimed to investigate why some tumor cells remain viable after treatment, based on single-cell observations.

5.1. Contributions

Several research aims were set, and this section will evaluate how well each of these questions was answered:

1. **What is the minimal mitogen-activated protein kinase pathway model that can describe KRAS mutant ERK protein activity in a population of cells?**
   It was found that a model with just two phosphorylation states of MEK-ERK-RAS was enough to describe the ERK protein activity of a single KRAS mutant cell. A minimal model integrating the most important feedback inhibitions was constructed. Three feedback mechanisms were included, namely negative feedback directly on the activation of RAS by ERK, inhibition of ERK activation by DUSP protein, and feedback inhibition through SPRY protein in the upstream part of the MAPK pathway. However, it was found that only the direct ERK feedback mechanism was required to describe the MAPK pathway behavior. This was true for unperturbed and perturbed KRAS mutant cells. GRB-SOS was included to allow for the simulation of HERi perturbation. This novel minimal model was proposed to allow for the analysis of cell populations, as larger models would make this computationally intractable.

2. **Can we infer the gain-of-function in a population KRAS mutant cells?**
   It was found that the gain-of-function that was expected in KRAS mutant cells could not be detected with this model. This indicates that the kinetic estimates were less accurate than hoped.

3. **Are any alternative pathways activated under perturbed conditions?**
   The main research question was whether any alternative pathways are activated under perturbed conditions. Specifically, after treatment with a MEK and HER inhibitor. The analysis did not indicate any real kinetic difference under perturbations, as all changes functionally balanced each other out.

4. **Can the minimal MAPK pathway model predict protein activity in a population of perturbed KRAS mutant cells in an external validation dataset?**
   The model was found to be able to predict protein activity in an external dataset. The model was also able to infer changes in protein activity after a MEKi perturbation. For this purpose, a new method was implemented to simulate drug perturbations. These results indicate that the model is relevant in terms of its protein activity levels. Thus, if we compare the results under treatment and without treatment, we can say with more certainty that MEK protein activity is more stable under perturbation with a MEK and HER inhibitor. Additionally, this indicates a certain level of biological relevance of this model.

5.2. Limitations

The most striking limitation of this work is the uncertainty in the parameter estimations made using the proposed model. The gain-of-function in RAS activity could not be detected without prior information. Additionally, there were no clear pathway changes under perturbation. Which does not have to imply that these estimates are faulty. However, we would expect that there are small changes in the kinetic parameters that could explain the transient reactivation after treatment of some cells. This limitation is probably due to only having one observed longitudinal protein activity measurement. Acquiring more observed species is hard if this FRET based method is used due to the limited detection bandwidth that can be used for fluorescence measurement.

Another limitation of the method used in this work is the need to choose specific parameters that vary between cells. This feature also makes the method computationally tractable for larger populations of cells. However, it does require some very important assumptions to be made, which can heavily affect the inference results.

5.3. Future work recommendations

Given the limitations described in the previous section, it would be probably prudent to estimate kinetics based on multiple datasets, and datamodality types. In this work, we already presented a comparison to CyTOF data. Such datasets could also be employed for the process of parameter estimation. However, at this point, there are not enough of these datasets reported in the literature to accurately extend the presented analysis.

Possible improvements to the parameter inference method would be to vary all the kinetic parameters but make more assumptions about the shape of the posterior distributions, as was done by Loos et al. [26]. However, such approaches also have clear drawbacks. Another solution might be to find a more reproducible way to choose parameters to vary between cells. For example, by constructing a method to find which parameters vary the most between single-cell fits and choosing these to vary between cells. A similar approach would be to use a dimensionality reduction technique on these single cell posterior samples to find latent dimensions that can describe the variance between cells. These could then be employed to simulate heterogeneous cell populations.

references

[1] B. Alberts, A. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of The Cell. 6º ed.* Vol. cap 19. Cell junctions and the extracellular matrix. 2015, pp. 1035–1081. ISBN: 9788578110796. arXiv: arXiv:1011.1669v3.

[2] J. C. D. KM; "Classification of KRAS-activating mutations and the implications for therapeutic intervention". In: *Cancer discovery* (). URL: https://pubmed.ncbi.nlm.nih.gov/35373279/.

[3] S. C. F. A. Huijberts, R. M. J. M. van Geel, E. M. J. van Brummelen, F. L. Opdam, S. Marchetti, N. Steeghs, S. Pulleman, B. Thijssen, H. Rosing, K. Monkhorst, A. D. R. Huitema, J. H. Beijnen, R. Bernards, and J. H. M. Schellens. "Phase I study of lapatinib plus trametinib in patients with KRAS-mutant colorectal, non-small cell lung, and pancreatic cancer". In: *Cancer Chemotherapy and Pharmacology* 85.5 (Apr. 2020), pp. 917–930. DOI: 10.1007/s00280-020-04066-4. URL: https://doi.org/10.1007/s00280-020-04066-4.

[4] R. M. J. M. van Geel, E. M. J. van Brummelen, F. A. L. M. Eskens, S. C. F. A. Huijberts, F. Y. F. L. de Vos, M. P. J. K. Lolkema, L. A. Devriese, F. L. Opdam, S. Marchetti, N. Steeghs, K. Monkhorst, B. Thijssen, H. Rosing, A. D. R. Huitema, J. H. Beijnen, R. Bernards, and J. H. M. Schellens. "Phase 1 study of the pan-HER inhibitor dacomitinib plus the MEK1/2 inhibitor PD-0325901 in patients with KRAS-mutation-positive colorectal, non-small-cell lung and pancreatic cancer". In: *British Journal of Cancer* 122.8 (Mar. 2020), pp. 1166–1174. DOI: 10.1038/s41416-020-0776-z. URL: https://doi.org/10.1038/s41416-020-0776-z.

[5] E. M. J. van Brummelen, S. Huijberts, C. van Herpen, I. Desar, F. Opdam, R. van Geel, S. Marchetti, N. Steeghs, K. Monkhorst, B. Thijssen, H. Rosing, A. Huitema, J. Beijnen, R. Bernards, and J. Schellens. "Phase I study of afatinib and selumetinib in patients with KRAS-mutated colorectal, non-small cell lung, and pancreatic cancer". en. In: *Oncologist* 26.4 (Apr. 2021), 290–e545.

[6] L. Huang, Z. Guo, F. Wang, and L. Fu. "KRAS mutation: from undruggable to druggable in cancer". In: *Signal transduction and targeted therapy* 6.1 (2021), p. 386.

[7] S. B. Kemp, N. Cheng, N. Markosyan, R. Sor, I.-K. Kim, J. Hallin, J. Shoush, L. Quinones, N. V. Brown, J. B. Bassett, *et al.* "Efficacy of a small-molecule inhibitor of KrasG12D in immunocompetent models of pancreatic cancer". In: *Cancer discovery* 13.2 (2023), pp. 298–311.

[8] D. Lake, S. A. Corrêa, and J. Müller. “Negative feedback regulation of the ERK1/2 MAPK pathway”. In: *Cellular and Molecular Life Sciences* 73.23 (2016), pp. 4397–4413.

[9] S. Sasagawa, Y.-i. Ozaki, K. Fujita, and S. Kuroda. “Prediction and validation of the distinct dynamics of transient and sustained ERK activation”. In: *Nature Cell Biology* 7.4 (Mar. 2005), pp. 365–373. DOI: 10.1038/ncb1233. URL: https://doi.org/10.1038/ncb1233.

[10] T. A. Ahmed, C. Adamopoulos, Z. Karoulia, X. Wu, R. Sachidanandam, S. A. Aaronson, and P. I. Poulikakos. “SHP2 Drives Adaptive Resistance to ERK Signaling Inhibition in Molecularly Defined Subsets of ERK-Dependent Tumors”. In: *Cell Reports* 26.1 (Jan. 2019), 65–78.e5. DOI: 10.1016/j.celrep.2018.12.013. URL: https://doi.org/10.1016/j.celrep.2018.12.013.

[11] C. Sun, S. Hobor, A. Bertotti, D. Zecchin, S. Huang, F. Galimi, F. Cottino, A. Prahallad, W. Grernrum, A. Tzani, A. Schlicker, L. F. Wessels, E. F. Smit, E. Thunnissen, P. Halonen, C. Lieftink, R. L. Beijersbergen, F. Di Nicolantonio, A. Bardelli, L. Trusolino, and R. Bernards. “Intrinsic Resistance to MEK Inhibition in KRAS Mutant Lung and Colon Cancer through Transcriptional Induction of ERBB3”. In: *Cell Reports* 7.1 (Apr. 2014), pp. 86–93. DOI: 10.1016/j.celrep.2014.02.045. URL: https://doi.org/10.1016/j.celrep.2014.02.045.

[12] G. Housman, S. Byler, S. Heerboth, K. Lapinska, M. Longacre, N. Snyder, and S. Sarkar. “Drug resistance in cancer: An overview”. In: *Cancers* 6.3 (2014), pp. 1769–1792. ISSN: 20726694. DOI: 10.3390/cancers6031769.

[13] P. M. Altrock, L. L. Liu, and F. Michor. “The mathematics of cancer: Integrating quantitative models”. In: *Nature Reviews Cancer* 15.12 (2015), pp. 730–745. ISSN: 14741768. DOI: 10.1038/nrc4029.

[14] I. Dagogo-Jack and A. T. Shaw. “Tumour heterogeneity and resistance to cancer therapies”. In: *Nature Reviews Clinical Oncology* 15.2 (2018), pp. 81–94. ISSN: 17594782. DOI: 10.1038/nrclinonc.2017.166. URL: http://dx.doi.org/10.1038/nrclinonc.2017.166.

[15] Y. Goyal, G. T. Busch, M. Pillai, J. Li, R. H. Boe, E. I. Grody, M. Chelvanambi, I. P. Dardani, B. Emert, N. Bodkin, J. Braun, D. Fingerman, A. Kaur, N. Jain, P. T. Ravindran, I. A. Mellis, K. Kiani, G. M. Alicea, M. E. Fane, S. S. Ahmed, H. Li, Y. Chen, C. Chai, J. Kaster, R. G. Witt, R. Lazcano, D. R. Ingram, S. B. Johnson, K. Wani, M. C. Dunagin, A. J. Lazar, A. T. Weeraratna, J. A. Wargo, M. Herlyn, and A. Raj. “Diverse clonal fates emerge upon drug treatment of homogeneous cancer cells”. In: *Nature* 620.7974 (Aug. 2023), pp. 651–659. ISSN: 1476-4687. DOI: 10.1038/s41586-023-06342-8. URL: https://doi.org/10.1038/s41586-023-06342-8.

[16] S. M. Shaffer, M. C. Dunagin, S. R. Torborg, E. A. Torre, B. Emert, C. Krepler, M. Beqiri, K. Sproesser, P. A. Brafford, M. Xiao, E. Eggan, I. N. Anastopoulos, C. A. Vargas-Garcia, A. Singh, K. L. Nathanson, M. Herlyn, and A. Raj. "Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance". In: *Nature* 546 (7658 June 2017), pp. 431–435. ISSN: 14764687. DOI: 10.1038/nature22794.

[17] S. L. Spencer, S. Gaudet, J. G. Albeck, J. M. Burke, and P. K. Sorger. "Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis". In: *Nature* 459.7245 (2009), pp. 428–432. ISSN: 00280836. DOI: 10.1038/nature08012.

[18] G. Bunt and F. S. Wouters. "FRET from single to multiplexed signaling events". In: *Biophysical Reviews* 9.2 (Apr. 2017), pp. 119–129. ISSN: 1867-2469. DOI: 10.1007/s12551-017-0252-z. URL: https://doi.org/10.1007/s12551-017-0252-z.

[19] H. Sahoo. "Förster resonance energy transfer – A spectroscopic nanoruler: Principle and applications". In: *Journal of Photochemistry and Photobiology C: Photochemistry Reviews* 12.1 (2011), pp. 20–30. ISSN: 1389-5567. DOI: https://doi.org/10.1016/j.jphotochemrev.2011.05.001. URL: https://www.sciencedirect.com/science/article/pii/S138955671100

[20] B. Ponsioen, J. B. Post, J. R. B. des Amorie, D. Laskaris, R. L. van Ineveld, S. Kersten, A. Bertotti, F. Sassi, F. Sipieter, B. Cappe, S. Mertens, I. Verlaan-Klink, S. F. Boj, R. G. J. Vries, H. Rehmann, P. Vandenabeele, F. B. Riquet, L. Trusolino, J. L. Bos, and H. J. G. Snippert. "Quantifying single-cell ERK dynamics in colorectal cancer organoids reveals EGFR as an amplifier of oncogenic MAPK pathway signalling". In: *Nature Cell Biology* 23.4 (Apr. 2021), pp. 377–390. DOI: 10.1038/s41556-021-00654-5. URL: https://doi.org/10.1038/s41556-021-00654-5.

[21] A. Iyer, A. A. J. Hamers, and A. B. Pillai. "CyTOF® for the Masses". In: *Frontiers in Immunology* 13 (2022). ISSN: 1664-3224. DOI: 10.3389/fimmu.2022.815828. URL: https://www.frontiersin.org/articles/10.3389/fimmu.2022.815828.

[22] B. Thijssen, H. A. Segeren, Q. Liu, L. F. Wessels, and B. Westendorp. "Inferring single-cell protein levels and cell cycle behavior in heterogeneous cell populations". In: *bioRxiv* (2023), pp. 2023–08.

[23] S. Rogers and M. A. Girolami. *A First Course in Machine Learning*. Chapman and Hall / CRC machine learning and pattern recognition series. CRC Press, 2011, pp. I–XX, 1–285. ISBN: 978-1-43-982414-6.

[24] J. Hasenauer, S. Waldherr, M. Doszczak, N. Radde, P. Scheurich, and F. Allgöwer. "Identification of models of heterogeneous cell populations from population snapshot data". In: (2011). DOI: 10.1186/1471-2105-12-125. URL: http://www.biomedcentral.com/1471-2105/12/125.

[25] P. D. Dixit, E. Lyashenko, M. Niepel, and D. Vitkup. “Maximum Entropy Framework for Predictive Inference of Cell Population Heterogeneity and Responses in Signaling Networks”. In: *Cell Systems* 10 (2 Feb. 2020), 204–212.e8. ISSN: 2405-4712. DOI: 10.1016/J.CELS.2019.11.010.

[26] C. Loos, K. Moeller, F. Fröhlich, T. Hucho, and J. Hasenauer. “A Hierarchical, Data-Driven Approach to Modeling Single-Cell Populations Predicts Latent Causes of Cell-To-Cell Variability”. In: *Cell systems* 6 (5 May 2018), 593–603.e13. ISSN: 2405-4712. DOI: 10.1016/J.CELS.2018.04.008. URL: https://pubmed-ncbi-nlm-nih-gov.tudelft.idm.oclc.org/29730254/.

[27] L. Dharmarajan, H. M. Kaltenbach, F. Rudolf, and J. Stelling. “A Simple and Flexible Computational Framework for Inferring Sources of Heterogeneity from Single-Cell Dynamics”. In: *Cell Systems* 8 (1 Jan. 2019), 15–26.e11. ISSN: 24054720. DOI: 10.1016/J.CELS.2018.12.007.

[28] B. Thijssen, T. M. H. Dijkstra, T. Heskes, *et al.* “BCM: toolkit for Bayesian analysis of Computational Models using samplers”. In: *BMC Systems Biology* 10.1 (2016), p. 100. DOI: 10.1186/s12918-016-0339-3. URL: https://doi.org/10.1186/s12918-016-0339-3.

[29] R. E. Kass and A. E. Raftery. “Bayes factors”. In: *Journal of the American Statistical Association* 90.430 (1995), pp. 773–795. DOI: 10.1080/01621459.1995.10476572. URL: https://doi.org/10.1080/01621459.1995.10476572.

[30] R. Brandt, T. Sell, M. Lüthen, F. Uhlitz, B. Klinger, P. Riemer, C. Giesecke-Thiel, S. Schulze, I. A. El-Shimy, D. Kunkel, B. Fauler, T. Mielke, N. Mages, B. G. Herrmann, C. Sers, N. Blüthgen, and M. Morkel. “Cell type-dependent differential activation of ERK by oncogenic KRAS in colon cancer and intestinal epithelium”. In: *Nature Communications* 10.1 (July 2019). DOI: 10.1038/s41467-019-10954-y. URL: https://doi.org/10.1038/s41467-019-10954-y.

[31] S. C. Bendall, E. F. Simonds, P. Qiu, E.-A. D. Amir, P. O. Krutzik, R. Finck, R. V. Bruggner, R. Melamed, A. Trejo, O. I. Ornatsky, R. S. Balderas, S. K. Plevritis, K. Sachs, D. Pe’er, S. D. Tanner, and G. P. Nolan. “Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum”. en. In: *Science* 332.6030 (May 2011), pp. 687–696.

[32] B. N. Kholodenko. “Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades”. In: *European Journal of Biochemistry* 267.6 (Mar. 2000), pp. 1583–1588. DOI: 10.1046/j.1432-1327.2000.01197.x. URL: https://doi.org/10.1046/j.1432-1327.2000.01197.x.

[33] H. Ryu, M. Chung, M. Dobrzyński, D. Fey, Y. Blum, S. S. Lee, M. Peter, B. N. Kholodenko, N. L. Jeon, and O. Pertz. “Frequency modulation of ERK activation dynamics rewires cell fate”. In: *Molecular systems biology* 11.11 (2015), p. 838.

[34] L. Gerosa, C. Chidley, F. Fröhlich, G. Sanchez, S. K. Lim, J. Muhlich, J. Y. Chen, S. Vallabhaneni, G. J. Baker, D. Schapiro, M. I. Atanasova, L. A. Chylek, T. Shi, L. Yi, C. D. Nicora, A. Claas, T. S. Ng, R. H. Kohler, D. A. Lauffenburger, R. Weissleder, M. A. Miller, W. J. Qian, H. S. Wiley, and P. K. Sorger. “Receptor-Driven ERK Pulses Reconfigure MAPK Signaling and Enable Persistence of Drug-Adapted BRAF-Mutant Melanoma Cells”. In: *Cell Systems* 11 (5 Nov. 2020), 478–494.e9. ISSN: 24054720. DOI: 10.1016/j.cels.2020.10.002.

[35] B. Schoeberl, C. Eichler-Jonsson, E. D. Gilles, and G. Müller. “Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors”. In: *Nature biotechnology* 20.4 (2002), pp. 370–375.

[36] B. N. Kholodenko, O. V. Demin, G. Moehren, and J. B. Hoek. “Quantification of short term signaling by the epidermal growth factor receptor”. In: *Journal of Biological Chemistry* 274.42 (1999), pp. 30169–30181.

[37] J. F. Doody, Y. Wang, S. N. Patel, C. Joynes, S. P. Lee, J. Gerlak, R. L. Rolser, Y. Li, P. Steiner, R. Bassi, *et al.* “Inhibitory activity of cetuximab on epidermal growth factor receptor mutations in non–small cell lung cancers”. In: *Molecular cancer therapeutics* 6.10 (2007), pp. 2642–2651.

[38] R. Thota, D. B. Johnson, and J. A. Sosman. “Trametinib in the treatment of melanoma”. en. In: *Expert Opin. Biol. Ther.* 15.5 (May 2015), pp. 735–747.

[39] T. C. Yeh, V. Marsh, B. A. Bernat, J. Ballard, H. Colwell, R. J. Evans, J. Parry, D. Smith, B. J. Brandhuber, S. Gross, A. Marlow, B. Hurley, J. Lyssikatos, P. A. Lee, J. D. Winkler, K. Koch, and E. Wallace. “Biological characterization of ARRY-142886 (AZD6244), a potent, highly selective mitogen-activated protein kinase kinase 1/2 inhibitor”. en. In: *Clin. Cancer Res.* 13.5 (Mar. 2007), pp. 1576–1583.

[40] M. Koyama, M. Kitazawa, S. Nakamura, T. Matsumura, S. Miyazaki, Y. Miyagawa, F. Muranaka, S. Tokumaru, M. Okumura, Y. Yamamoto, *et al.* “Low-dose trametinib and Bcl-xL antagonist have a specific antitumor effect in KRAS-mutated colorectal cancer cells”. In: *International Journal of Oncology* 57.5 (2020), pp. 1179–1191.