


Provided by the author(s) and University College Dublin Library in accordance with publisher policies. Please cite the published version when available.

Title	On the personalised modelling of cancer signalling
Author(s)	Fey, Dirk; Kuehn, Axel; Kholodenko, Boris N.
Publication date	2016-10-12
Publication information	IFAC-PapersOnLine, 49 (26): 312-317
Conference details	6th IFAC Conference on Foundations of Systems Biology in Engineering (FOSBE 2016), Magdeburg, Germany, 9-12 October 2016
Publisher	Elsevier
Item record/more information	<a href="http://hdl.handle.net/10197/8317">http://hdl.handle.net/10197/8317</a>
Publisher's version (DOI)	<a href="http://dx.doi.org/10.1016/j.ifacol.2016.12.145">http://dx.doi.org/10.1016/j.ifacol.2016.12.145</a>

Downloaded 2017-11-03T17:51:21Z

The UCD community has made this article openly available. Please share how this access benefits you. Your story matters! (@ucd\_oa) 

Some rights reserved. For more information, please see the item record link above.



## On the personalised modelling of cancer signalling<sup>\*</sup>

Dirk Fey<sup>\*</sup> Axel Kuehn<sup>\*\*</sup> Boris N. Kholodenko<sup>\*</sup>

<sup>\*</sup> *Systems Biology Ireland, School of Medicine, University College Dublin, Ireland (e-mails: [dirk.fey@ucd.ie](mailto:dirk.fey@ucd.ie), [boris.kholodenko@ucd.ie](mailto:boris.kholodenko@ucd.ie)).*

<sup>\*\*</sup> *Cancer Bioinformatics and systems biology, Institute of Cancer Research of Montpellier, France (e-mail: [axel.kuehn@inserm.fr](mailto:axel.kuehn@inserm.fr))*

**Abstract:** Dynamic modelling has long been used to understand fundamental principles of cell signalling and its dysregulation in cancer. More recently, these models have also been used to understand the individual risks of cancer patients, and predict their survival probabilities. However, the current methodologies for integrating tumour data and generating patient-specific simulations suffer from the lack of general applicability; they only work for cell signalling models in which only posttranslational protein modifications are considered, so that the total protein concentrations are conserved. Here, we present novel, generally applicable method. The method is based on a simple theoretical framework for modelling gene-regulation, and the indirect estimation of patient-specific parameters from tumour data. Because our method does not require time-invariance of the total-protein concentrations, it can be applied to models of any nature, including the many cancer signalling models involving gene-regulation.

© 2016, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.

**Keywords:** Systems biology, ordinary differential equations, parameter estimation.

### 1. INTRODUCTION

The dynamic states of cancer cell signalling are intimately linked to cancer cell fate decisions and clinical outcomes (Kolch et al., 2015; Fey et al., 2012). Recently, multiple studies have shown that patient-specific differences in the dynamic behaviour of these signalling networks underlie individual pathogenetic changes and disease manifestation (Fey et al., 2015; Flanagan et al., 2015; Lindner et al., 2013; Murphy et al., 2013). For example, a dynamic model of the JNK pathway could predict the survival probabilities of cancer patients in neuroblastoma, a common childhood cancer (Fey et al., 2015). These predictions were based on personalised simulations for over 700 patients, and revealed that a high amplitude, switch-like JNK activation was associated with neuroblastoma cell death, and better patient survival.

To generate the patient-specific simulations, the signalling model was personalised by adjusting the total protein concentration of each model component to the measured values from the patient's tumour sample. This approach works for purely posttranslational signalling models in which the total protein concentration of each gene is conserved (constant), and the only dynamic changes arise from the regulation of protein-protein interactions and enzyme activities, for example by phosphorylation.

In fact, all the personalised models mentioned are based on this simple principle of directly using the measured mRNA or protein concentrations as static parameters in the model (Fey et al., 2015; Flanagan et al., 2015; Lindner et al., 2013; Murphy et al., 2013). Thus, all these personalised models are based on the assumption that the

total protein concentrations of the modelled genes are constant for all genes. Because only then do these measured concentrations constitute static parameters in the model that can readily be adjusted using the measured tumour data. Unfortunately, this assumption is often violated and prevents the application of this approach to systems in which the protein levels themselves are regulated.

Regulation of gene expression on the level of mRNA and total protein is quite common (Kolch et al., 2015; Nakakuki et al., 2010). Examples that are particularly relevant for cancer are regulation of the cell-cycle or the DNA-damage response. In these systems, both mRNA and total protein concentrations are regulated, and changing over time. Moreover, these dynamic changes are critical aspects of the systems behaviour. So far, the assumption of time-invariant total protein levels has hampered personalised modelling of these important processes. Thus, there is an urgent need for novel model-personalisation methodologies that are generally applicable.

Here we propose such a methodology.

In contrast to the current approach, the here proposed methodology does not require that the total-protein concentrations are constant. Instead, we propose a theoretical framework in which the patient-specific mRNA and total protein concentrations arise naturally from patient-specific mRNA synthesis parameters. Rather than time-invariant parameters, the mRNA and total protein concentrations are now dynamic states described by ordinary differential equations. This framework allows us to personalise models of any nature, involving signalling, gene-regulation or both, by indirectly estimating the patient-specific synthesis parameters.

<sup>\*</sup> Supported by EU FP7 grant "SynSignal" (No. 613879).

The current manuscript is organised as follows. First, we will present a simple, yet generally applicable theoretical framework for the personalised modelling of gene expression. Second, we show how the patient-specific parameters can be estimated from expression data measured in the patients' tumour samples. Third, we illustrate the methodology by testing it on simulated data from a p53 DNA-damage response model.

## 2. A MODELLING FRAMEWORK FOR PERSONALISED GENE EXPRESSION

Although often considered separately, the posttranslational and gene-regulatory levels of cell signalling are intrinsically linked (Fig. 1). On the one hand, signalling on the protein level affects the gene-regulatory level, for instance by triggering mRNA synthesis through the activation of transcription factors. On the other hand, the gene-regulatory level affects the signalling level by controlling the expression of the proteins that mediate and process the signals.

### 2.1 Combined modelling of gene-regulation and signalling

Mathematically, we can think of signalling as a system consisting of two layers, linked by feedback (Fig. 1). Each layer consists of a set of ordinary differential equations. The signalling layer has the concentrations of all post-translational protein forms as states, and takes the vector of mRNA concentrations as input. Vice versa, the gene-regulatory layer has the mRNA concentrations as states and takes the vector of protein concentrations as input.

Formally, we have the following system

$$\frac{d}{dt}y = f(y, p) + B \operatorname{diag}(k_{\text{transl}})x \quad (1a)$$

$$\frac{d}{dt}x = \rho + g(y, k) - \operatorname{diag}(k_{\text{rdeg}})x, \quad (1b)$$

where  $x \in \mathbb{R}^m$  and  $y \in \mathbb{R}^n$  are vectors of mRNA and protein concentrations, respectively.  $f(y, p)$  and  $g(y, k)$  are functions describing posttranslational modifications and gene-regulatory effects, respectively. Note that  $f(y, p)$  should contain a degradation term, i.e.  $f(y, p) = \tilde{f}(y, p) - \operatorname{diag}(k_{\text{pdeg}})y$ .  $B \in \mathbb{R}^{n \times m}$  is a matrix containing  $m$  different unit vectors  $e_i \in \mathbb{R}^n$ , indicating the subset of unmodified proteins  $y_i$  that are translated (synthesised) from their corresponding mRNA templates.  $\operatorname{diag} k_{\text{transl}} \in \mathbb{R}^{m \times m}$  and  $\operatorname{diag} k_{\text{rdeg}} \in \mathbb{R}^{m \times m}$  are diagonal matrices containing the protein-translation and mRNA-degradation rate constants.  $\rho \in \mathbb{R}^m$  is a vector of patient-specific parameters.

*Remark on the notation.* In the following, and for simplicity of presentation, we often do not explicitly denote the dependency of  $f$  and  $g$  on the kinetic parameters  $p$  and  $k$ . That is, when referring to  $f$  and  $g$  we mean the functions  $f: y \mapsto f(y, p)$  and  $g: y \mapsto g(y, k)$ , respectively.

### 2.2 Rationale and assumptions behind the approach

The method is based on linking the measured patient-specific gene-expression differences on the systems level to parameter changes in the model on individual gene level. Because we want the patient-specific parameters to

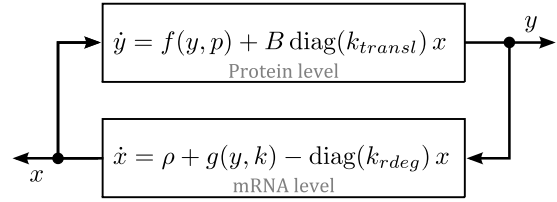


Fig. 1. The dynamic system used for combined modelling of posttranslational and gene-regulatory networks:  $y$  describes the protein level, and  $x$  the mRNA level.

be identifiable, we assume that the basal mRNA-synthesis rate for each gene varies between patients, but all other parameters do not.

*Assumption 1.* The observed gene-expression differences in tumour samples are caused by patient-specific differences of the basal mRNA-synthesis rates.

Although some other parameters, most notably mRNA-degradation and protein-translation rate constants, might also vary between patients, we only estimate the patient-specific mRNA-synthesis rates. All other kinetic parameters not directly relating to gene expression – such as catalytic rate constants and binding affinities – are presumed not to vary between patients.

*Remark on violating Assumption 1.* Although it is a theoretical requirement for formulating the methodology, violating Assumption 1 has limited practical consequences. As we will see later in Sec. 4.3, accurate, albeit less precise simulation results can be obtained even when the rates of mRNA-degradation and protein-translation for each gene differ between patients.

## 3. ESTIMATING THE PATIENT-SPECIFIC PARAMETERS

Our aim is to estimate the patient specific parameter  $\rho$  from gene expression measurements of the patients' tumours. Such tumour data are usually obtained from tissue samples collected in biopsies or surgery, and would usually reflect the homeostatic, unperturbed state of the patient's tumour. Thus, these tumour data would correspond to the unstimulated steady-state of our model (1).

*Assumption 2.* Measured gene expression values from tumour samples reflect basal (unstimulated) steady-state values.

This is a fair assumption considering that most tumour data are collected prior to pharmacological treatment (de Gramont et al., 2015; Juty et al., 2015).

Put more formally, the goal is to estimate  $\rho$  from steady state measurements of either  $x$  or  $y$ , based on the model in (1) with given functions  $f, g$  and given kinetic parameters  $p, k, k_{\text{transl}}$  and  $k_{\text{rdeg}}$ .

For the steady-state we have

$$0 = f(y, p) + B \operatorname{diag}(k_{\text{transl}})x, \quad (2a)$$

$$0 = \rho + g(y, k) - \operatorname{diag}(k_{\text{rdeg}})x. \quad (2b)$$

The next step is to solve (2) for the patient-specific parameters  $\rho$ . Two scenarios can be distinguished.

### 3.1 Estimation based on mRNA measurements ( $x$ is measured)

In the first scenario, all mRNA concentrations are measured. The following theorem provides the main result.

**Theorem.** Assume that the function  $f : y \mapsto f(y, p)$  is injective. Then we can estimate the patient specific parameters  $\rho$  from steady state measurements of  $x$ .

**Proof** (by construction). Because  $f$  is injective, a left-inverse of  $f$  with  $f^{inv} \circ f = Id_y$  exists. Thus, in steady-state  $\frac{d}{dt}y = 0$ , it hold that

$$y = f^{inv}(B \text{diag}(k_{transl})x, p), \quad (3)$$

which means given steady-state measurements of  $x$ , we know the steady-state of  $y$ . Further, it follows from the steady-state condition  $\frac{d}{dt}x = 0$  of  $x$ , that

$$\rho = k_{rdeg}^T Ix - g \circ f^{inv}(B \text{diag}(k_{transl})x). \quad (4)$$

□

Equation (4) is the analytical solution to the problem in (2), and provides an explicit formula for estimating the patient-specific parameters from steady-state mRNA measurements.

### 3.2 Estimation based on protein measurements ( $y$ is measured)

In the second scenario, the protein concentrations are measured. The following theorem provides the main result.

**Theorem.** Assume non-zero translation rates for all genes, i.e. all elements of  $k_{transl}$  are greater than zero. Then we can estimate the patient-specific parameters from steady-state measurements of  $y$ .

**Proof.** By definition  $B$  is composed of a subset of the unit vectors, e.g.  $B = [e_i \ e_j \ \dots]$ . It holds that  $B^T B = I$ , which means we can solve (2a) for  $x$

$$x = -\text{diag}^{-1}(k_{transl})B^T f(y, p). \quad (5)$$

To see this, note that right-multiplying (2a) with  $B^T$  gives  $0 = B^T f(y) + B^T B \text{diag}(k_{transl})x$ . After substitution into (2b) we can solve for the patient-specific parameter

$$\rho = -k_{rdeg}^T \text{diag}^{-1}(k_{transl})B^T f(y, p) - g(y, k). \quad (6)$$

□

Equation (6) is the analytical solution to the problem in (2), and provides an explicit formula for estimating the patient-specific parameters from steady-state protein measurements.

## 4. EXAMPLE

To illustrate the proposed methodology, we applied it to a simple toy model inspired by the p53 DNA damage response system, using simulated data for testing. On the gene-regulatory level, p53 induces the expression of MDM2, which in turn promotes the degradation of p53 (Fig. 2). On the posttranslational level, the phosphorylation of p53, for example induced by DNA damage, inhibits p53 degradation, thus stabilising p53 protein, resulting in increased p53 protein expression. Thus, this system

contains an interesting crosstalk between the posttranslational level describing p53 phosphorylation and the gene-regulatory level describing p53 expression. Indeed, this was our rationale behind choosing this model. We wanted to look at a system that contains a feedback loop involving both transcriptional and post-translational events, as we thought such a system would constitute the most difficult challenge for model personalisation.

### 4.1 Equations of the example system

Our example system, based on a core model of the p53 DNA-damage response, is a prototypical example for a signalling system that involves gene-regulation (Fig. 2). Note that there are many other gene-regulatory systems that follow similar regulatory patterns (Avraham and Yarden, 2011), for example NF- $\kappa$ B (Perkins, 2012).

The model equations are

$$\dot{y}_1 = -v_1(y_1) + v_2(y_3) - p_{-1}y_2y_1 + k_{t,1}x_1, \quad (7a)$$

$$\dot{y}_2 = -p_{-2}y_2 + k_{t,2}x_2, \quad (7b)$$

$$\dot{y}_3 = +v_1(y_1) - v_2(y_3) - p_{-3}y_3, \quad (7c)$$

$$\dot{x}_1 = \rho_1 - k_{rdeg,1}x_1, \quad (7d)$$

$$\dot{x}_2 = \rho_2 + k_1 H(y_1 + y_3) - k_{rdeg,2}x_2, \quad (7e)$$

where  $y_1, y_2, y_3$  and  $x_1, x_2$  denote the protein and mRNA concentrations of Y1, Y2, pY1 and X1, X2, respectively, and with parameters described below, and functions

$$v_1(y_1) = p_1 u(t) \frac{y_1}{y_1 + K_{m,1}}, \quad (7f)$$

$$v_2(y_3) = p_2 \frac{y_3}{y_3 + K_{m,2}}, \quad (7g)$$

describing the rates of Y1 phosphorylation and dephosphorylation, respectively, and the Hill function

$$H(y) = \frac{y^n}{K_h^n + y^n}, \quad (7h)$$

describing the Y1 mediated induction of X2 synthesis.  $u(t)$  denotes a possibly time-dependent input signal, which we can interpret as the activity of the upstream kinase catalysing the phosphorylation of Y1. Constants  $\rho_1$  and  $\rho_2$  are patient specific parameters that, for the purposes of our simulation tests, are randomly chosen from a lognormal distribution (denoted  $\ln \mathcal{N}$ ) with mean  $\mu = 1$  and standard deviation  $\sigma = 0.33$ . All other, nonrandom parameters are

$$k_{t,1} = 10, \quad k_{t,2} = 10 \quad (8a)$$

$$p_{-1} = 5, \quad p_{-2} = 5, \quad p_{-3} = 0.5, \quad (8b)$$

$$k_1 = 5, \quad k_{rdeg,1} = 1, \quad k_{rdeg,2} = 1, \quad (8c)$$

$$p_1 = 100, \quad K_{m,1} = 1, \quad K_h = 2, \quad (8d)$$

$$p_2 = 10, \quad K_{m,2} = 1, \quad n = 2. \quad (8e)$$

What makes this system particularly interesting is the link between the gene-expression and posttranslational levels, in which the phosphorylation of Y1 regulates the expression of the total Y1 protein concentration  $[Y1]_{total} = y_1 + y_3$ . This link is established by a condition on the parameter values  $p_{-1} \gg p_{-3}$  which makes the degradation of Y1 in the unphosphorylated form  $y_1$  much faster than the degradation of the phosphorylated form  $y_3$ .

Figure 3 shows the behaviour of the model with the nominal patient-specific parameters  $\rho_1 = 1$ ,  $\rho_2 = 1$  in response to a unit step input at  $t = 10$ , that is  $u = 0$  for  $t < 10$ , and  $u = 1$  for  $t \geq 10$ .

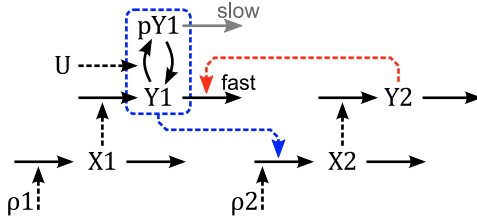


Fig. 2. The example system used to illustrate the proposed methodology: p53 protein (Y1) upregulates synthesis of MDM2 mRNA (X2) (blue), MDM2 protein (Y2) promotes p53 protein-degradation (red). X1 - p53 mRNA. Horizontal arrows - synthesis, degradation; curved arrows - phosphorylation, dephosphorylation; dashed arrows - regulatory influences.

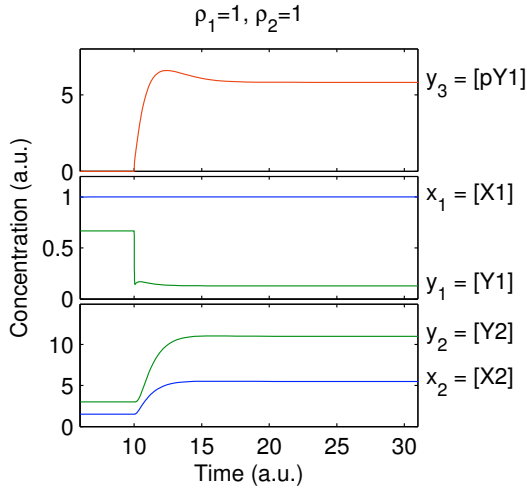


Fig. 3. Behaviour of the model depicted in Fig. 2 in response to a unit step input for the average patient, i.e.  $\rho_1 = 1, \rho_2 = 1$ . See (8) for all other parameters.

#### 4.2 Applying the proposed methodology

This section illustrates the application of the proposed methodology to the example system in Fig. 2.

According to the theoretical requirements of the methodology, we presume that the mRNA levels  $x = [x_1, x_2]^T$  are measured in the basal, unstimulated steady state, for which it holds that  $u = 0$ , and  $\dot{x}_1 = \dots = \dot{y}_3 = 0$ .

We can easily verify that the example system is of the form in (1), where  $y = [y_1, y_2, y_3]^T$ , and

$$f(y) = \begin{bmatrix} -v_1(y_1) + v_2(y_3) - p_{-1}y_2y_1 \\ -p_{-2}y_2 \\ v_1(y_1) - v_2(y_3) - p_{-3}y_3 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

and  $k_{transl} = [k_{t,1}, k_{t,2}]$ ,  $x = [x_1, x_2]^T$ .

Further, we can verify that  $f$  is injective for  $u = 0$ , and the inverse  $f^{inv}$  exists. To see this, we solve the steady state equations for 7a–c as follows. From (7b) we obtain

$$y_2 = \frac{k_{t,2}}{p_{-2}} x_2. \quad (9a)$$

From (7f) follows with  $u = 0$  that  $v_1 = 0$ , and therewith from (7c) that the only possible steady-state for pY1 is

$$y_3 = 0, \quad (9b)$$

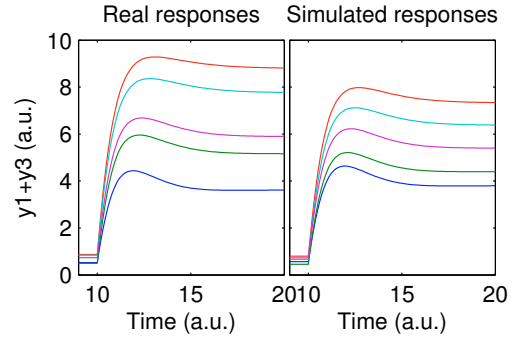


Fig. 4. Patient specific simulations of the model for the output  $[Y1]_{total} = y_1 + y_3$ . Left: Sample of the artificially generated data ( $\rho_1, \rho_2 \sim \ln \mathcal{N}$ ). Right: Corresponding simulated responses after estimating the patient-specific parameters from basal ( $u = 0$ ), noisy steady-state measurements as given in (10).

because  $v_1$  is the only term producing  $y_3$ . Finally, using  $y_3 = 0$  in (7g) gives  $v_2 = 0$ , which when substituted into (7a) gives  $0 = -p_{-1}y_2y_1 + k_{t,1}x_1$ . Substituting (9a) and solving for  $y_1$  yields

$$y_1 = \frac{k_{t,1} p_{-2} x_1}{k_{t,1} p_{-1} x_2}. \quad (9c)$$

This proves, by construction, that  $f$  in our example is invertible for  $u = 0$ . In fact, (9) is the explicit solution of the general form given in (3).

#### 4.3 Simulation results

Next, we generated some artificial data upon which our methodology can be tested. To that end, we simulated the model (7) with different patient-specific parameters  $\rho_1$  and  $\rho_2$  sampled from the lognormal distribution with mean  $\mu = 1$  and standard deviation  $\sigma = 0.33$ . The steady-state prior to stimulation (for  $u = 0$ ) was used to generate the artificial data by adding 10% measurement noise

$$x_{i,measured} = x_{i,simulated} \cdot \epsilon_i, \quad (10)$$

where  $\epsilon_i$  are a random numbers from a lognormal distribution with mean  $\mu = 1$  and standard deviation  $\sigma = 0.1$ .

To further test the method's robustness with respect to violating Assumption 1, we also generated a set of artificial data in which in addition to the mRNA-synthesis rates  $\rho_1, \rho_2 \sim \ln \mathcal{N}$ , the mRNA-degradation and protein-translation rates were also varied, i.e.  $k_{rdeg,1}, k_{rdeg,2} \sim \ln \mathcal{N}$  and  $k_{t,1}, k_{t,2} \sim \ln \mathcal{N}$ , respectively.

In the following sections, we used these artificial data to assess the accuracy and precision of both i) the personalised parameter estimates and ii) the dynamic responses arising from these estimates. To analyse the dynamic responses we used an unit step input, and focussed on the total Y1 protein concentration  $[Y1]_{total} = y_1 + y_3$  and the phosphorylated Y1 concentration  $[pY1] = y_3$  as outputs.

**Correct model, perfect measurements** First, we confirmed that the method is accurate when the conditions are perfect. That is when no assumptions are violated (only  $\rho_1, \rho_2$  are patient-specific), and no measurement noise is present. In this ideal scenario, the methodology recovers the patient-specific parameters perfectly (Tab. 1).

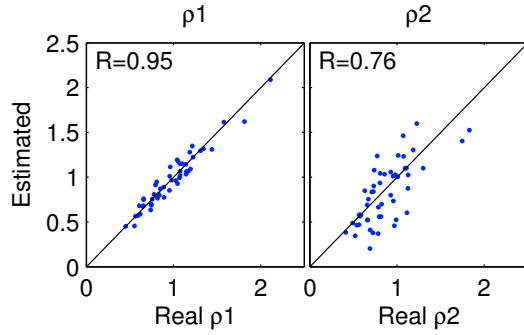


Fig. 5. Correlation between the real and estimated patient-specific parameter values. Test 2: Correct model, noisy measurements.

Correlation between real and simulated responses

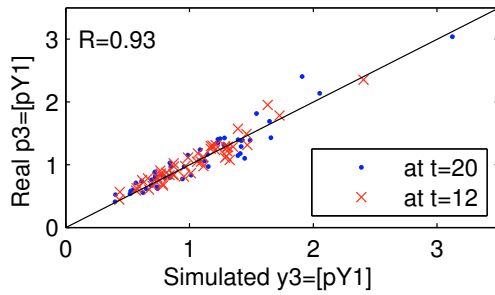


Fig. 6. Correlation between the real and estimated dynamic responses for the system output  $[pY1] = y_3$ . Test 2: Correct model, noisy measurements.

*Correct model, noisy measurements* A more realistic scenario is when the patient data are corrupted with measurement noise. In this case, we observed a good correlation between the real and estimated patient-specific parameters with about 90% and 53% of the variability explained for  $\rho_1$  and  $\rho_2$ , respectively (Fig. 5). Accordingly, the simulated protein levels in response to a unit step input also correlated well with the real response (Fig. 6).

*Incomplete model, perfect measurements* Next we assessed what happens when Assumption 1 is violated, and the real system contains additional patient-specific parameters, but no measurement noise. Specifically, we used variations in the mRNA-degradation constants  $k_{rdeg,1}, k_{rdeg,1}$  and the protein-translation parameters  $k_{t,1}, k_{t,1}$  in addition to  $\rho_1, \rho_2$  for generating the patient-specific data. Basically, this means that the model used for re-engineering the patient-specific differences (1) is incomplete: Because our methodology keeps the degradation and translation parameters constant, all patient specific differences will be mapped onto  $\rho_1$  and  $\rho_2$ . Naturally, this impacted on the patient-specific parameter estimates (Fig. 7), which are now less correlated with the real values (20 – 50% of variation explained). In contrast, the simulated dynamic responses correlated well with the real responses. About 38% of the variability in the steady state response was explained by the simulations (Fig. 8).

The fact that the response predictions of the output were more precise than some parameter estimates, is explained as follows. The algorithm maps all patient-specific differences onto  $\rho_1$  and  $\rho_2$ , even the differences arising from

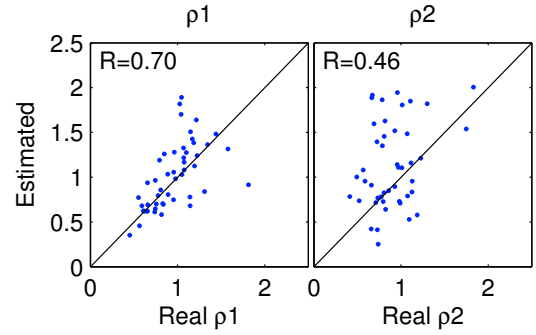


Fig. 7. Correlation between the real and estimated patient-specific parameter values. Test 3: Incomplete model, perfect measurements (see main text, and Tab. 1).

Correlation between real and simulated responses

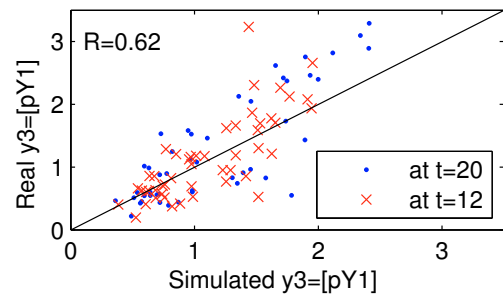


Fig. 8. Correlation between the real and estimated dynamic responses for the systems output  $[pY1] = y_3$ . Test 3: Incomplete model, perfect measurements.

$k_{rdeg}$  and  $k_t$ . Because they include these differences, the patient-specific parameter estimates are more variable. In contrast, the estimated mRNA concentrations are precise, because (4) guarantees that the modelled mRNA values match the measured values  $x_{model} = x_{measured}$ . This means that for predicting the dynamic responses, the precision of the state-estimates is more important than the precision of the parameter-estimates.

*Incomplete model, noisy measurements* Finally, we also analysed the worst case scenario, in which both the model and the measurements contain errors. The results were not markedly different from the ones obtained for noise-free measurements (Fig. 9). Despite the fact that the patient-specific estimates were not very precise (especially for  $\rho_2$  with a correlation coefficient of only  $R = 0.42$ ), the output predictions were precise, with the simulated pY1 responses explaining 48% of the variability present in the real responses (Fig. 9).

## 5. CONCLUSIONS

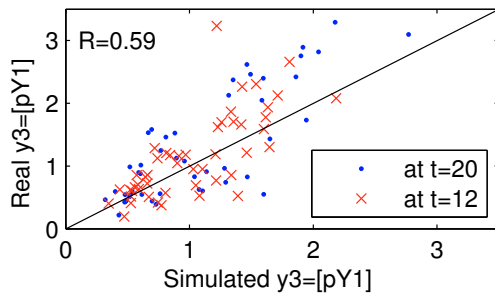
This paper proposes a generally applicable methodology for personalising models of cancer signalling. The advantage of using such personalised dynamic models over static biomarkers, has already been demonstrated in several case studies (Fey et al., 2015; Flanagan et al., 2015; Murphy et al., 2013; Lindner et al., 2013). In contrast to previous approaches, the here proposed methodology is not restricted to models in which the total protein concentrations are time-invariant parameters.



Table 1. Overview of the performed simulation tests ( $n = 1,000$  patients, 95%CI for  $R < 0.1$ ).

Test no.	Description	Simulation test setup, errors were included for			Test results ( $R^2$ values)			
		$k_{rdeg}$	$k_{transl}$	$x_{measured}$	Parameters		Responses	
					$\rho_1$	$\rho_2$	$[Y1]_{total}$	$[pY1]$
Test 1	Correct model, perfect measurement	-	-	-	1	1	1	1
Test 2	Correct model, noisy measurement	-	-	✓	0.90	0.86	0.87	0.87
Test 3	Incomplete model, perfect measurement	✓	✓	-	0.49	0.21	0.38	0.38
Test 4	Incomplete model, noisy measurement	✓	✓	✓	0.46	0.18	0.35	0.35

Correlation between real and simulated responses

Fig. 9. Correlation between the real and estimated dynamic responses for the systems output  $[pY1] = y_3$ . Test 4: Incomplete model, noisy measurements.

The mathematical framework underlying the methodology is based on two central biological principles: 1) mRNA serves as template for protein synthesis; 2) individual, cell-type specific differences arise from changes in gene expression.

Elaborating on the latter, a basic assumption in our approach is that the individual gene expression patterns arise from patient-specific differences in the basal gene expression rate parameters. Importantly, our approach is robust with respect to violating this assumption: The methodology can still be applied when other, unmodelled patient-specific differences are present.

Another, relatively weak assumption in framework is that all non-patient-specific parameters in the model are known. These parameters would usually be provided by the model, ideally as a result from constructing and calibrating the model using data from cell-culture experiments and parameter estimation. Literature reported values are also often used in these models, although care has to be taken for adopting parameter values out of context. In any case, measurements of mRNA and protein half-lives do exist for many genes (Schwanhäusser et al., 2011).

The validity of i) the approach and ii) the personalised model could be tested experimentally by measuring i) the synthesis and degradation parameters for example in pulse-chase experiments (Larance and Lamond, 2015), and ii) the dynamic responses in time-course experiments. This would i) test the assumption of patient-specific synthesis parameters, and ii) (in-)validate the predicted dynamic responses. Both experiments would require live cancer-cells from patients. Isolating live cancer-cells from tumour-samples can be technically challenging but the resulting cells are easy to assay. Tumour explants (little pieces of live tumour tissue) are easier to keep alive in the lab, but also contain non-cancerous cells and not all biochemical assays can be applied.

In conclusion, our methodology makes the personalised modelling of dynamic cell-signalling possible. In conjunction with tumour data (TCGA database, Tomczak et al., 2015)), it should be used to test the predictive and prognostic utility of the many signalling models constructed to date (BioModels database, Juty et al., 2015).

## REFERENCES

- Avraham, R. Yarden, Y. (2011). Feedback regulation of egfr signalling: decision making by early and delayed loops. *Nat Rev Mol Cell Biol*, 12(2), 104–117.
- de Gramont, A., Watson, S., et al. (2015). Pragmatic issues in biomarker evaluation for targeted therapies in cancer. *Nat Rev Clin Oncol*, 12(4), 197–212.
- Fey, D., Croucher, D.R., Kolch, W., Kholodenko, B.N. (2012). Crosstalk and signaling switches in mitogen-activated protein kinase cascades. *Front Physiol*, 3, 355.
- Fey, D., Halasz, M., et al. (2015). Signaling pathway models as biomarkers: Patient-specific simulations of jnk activity predict the survival of neuroblastoma patients. *Sci Signal*, 8(408), ra130.
- Flanagan, L., Lindner, A.U., et al. (2015). Bcl2 protein signalling determines acute responses to neoadjuvant chemoradiotherapy in rectal cancer. *J Mol Med (Berl)*, 93(3), 315–326.
- Juty, N., Ali, R., et al. (2015). Biomodels: Content, features, functionality, and use. *CPT Pharmacometrics Syst Pharmacol*, 4(2), e3.
- Kolch, W., Halasz, M., Granovskaya, M., Kholodenko, B.N. (2015). The dynamic control of signal transduction in cancer cells. *Nat Rev Cancer*, 15(9), 515–527.
- Larance, M. Lamond, A.I. (2015). Multidimensional proteomics for cell biology. *Nat Rev Mol Cell Biol*, 16(5), 269–280.
- Lindner, A.U., Concannon, C.G., et al. (2013). Systems analysis of bcl2 protein family interactions establishes a model to predict responses to chemotherapy. *Cancer Res*, 73(2), 519–528.
- Murphy, A.C., Weyhenmeyer, B., et al. (2013). Activation of executioner caspases is a predictor of progression-free survival in glioblastoma patients: a systems medicine approach. *Cell Death Dis*, 4, e629.
- Nakakuki, T., Birtwistle, M.R., et al. (2010). Ligand-specific c-fos expression emerges from the spatiotemporal control of erbb network dynamics. *Cell*, 141(5), 884–896.
- Perkins, N.D. (2012). The diverse and complex roles of nf- $\kappa$ b subunits in cancer. *Nat Rev Cancer*, 12(2), 121–132.
- Schwanhäusser, B., Busse, D., et al. (2011). Global quantification of mammalian gene expression control. *Nature*, 473(7347), 337–342.
- Tomczak, K., Czerwińska, P., Wiznerowicz, M. (2015). The cancer genome atlas (tcga). *Contemp Oncol (Pozn)*, 19(1A), A68–A77.