bioRxiv preprint doi: https://doi.org/10.1101/268359; this version posted February 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Delineating feedback activity in the MAPK and AKT pathways using feedback-enabled Inference of Signaling Activity

Bram Thijssen[1,2], Katarzyna Jastrzebski[1], Roderick L. Beijersbergen[1,*], Lodewyk F.A. Wessels[1,2,3,*]

[1]Division of Molecular Carcinogenesis, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands.

[2]Oncode Institute, The Netherlands Cancer Institute, The Netherlands

[3]Faculty of EEMCS, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands.

*To whom correspondence should be addressed: r.beijersbergen@nki.nl and l.wessels@nki.nl

Abstract

An important aspect of cellular signaling networks is the existence of feedback mechanisms. However, due to the complexity of signaling networks, as well as the presence of multiple interrelated feedback events, it can be difficult to identify which signaling routes are active in any particular context. We have previously shown that Inference of Signaling Activity (ISA) can be a useful method to study steady-state oncogenic signaling across different cell lines and inhibitor treatments. However, ISA did not explicitly include feedback signaling events. Incorporating feedback will increase the complexity and computational cost of the model, and more data is likely to be needed to infer feedback activities. Here, we developed feedback-ISA (f-ISA), an extension of the ISA modeling approach which incorporates feedback signaling events. It also includes integrated batch correction in order to fit the models to multiple, independent datasets simultaneously. We find that the identifiability of feedback activities can be counter-intuitive, which shows the importance of analyzing the full, joint uncertainty in model parameters. By iteratively adapting the model and including multiple datasets, including both steady state and intervention data, we constructed a model that can explain a large part of the phosphorylation levels of several signaling molecules in the MAPK and AKT pathways, across many breast cancer cell lines and across various conditions. The resulting model delineates which routes in the signaling network are likely to be active in each cell line and condition, given all of the data. Additionally, such models can indicate whether datasets agree with each other, and identify which parts of the data cannot be explained, thereby highlighting gaps in the current knowledge. We conclude that this modeling approach can be useful to quantitatively understand how complex cellular signaling networks behave across different cell lines and conditions.

bioRxiv preprint doi: https://doi.org/10.1101/268359; this version posted February 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

The mitogen-activated protein kinase (MAPK) and AKT signaling pathways are two central regulatory mechanisms which are often deregulated in cancer cells. Many inhibitors which block key kinases in these signaling pathways have been developed as potential cancer therapies. These kinase inhibitors can be very potent anticancer drugs, but cancer cells are often intrinsically resistant or develop resistance over time [1]. As a result, patients have a highly variable response to kinase inhibitors, and this variability is also seen in cell lines [2–4]. To understand this variability in response, and develop effective and selective (combination) therapies, a detailed quantitative understanding of cellular regulatory networks would be highly useful.

Signaling networks are complex, with many interrelated signaling events. One of the important features of cellular signaling networks is the existence of feedback mechanisms. These feedback loops can, for example, provide robustness to a signaling network or modulate the sensitivity to external inputs [5]. To date, many details of numerous signaling networks have been discovered, including various feedback events. In the MAPK pathway, an important feedback mechanism is the inactivation of RAF by ERK [6,7]. In the AKT pathway, two feedback mechanisms involve the down-regulation of the insulin receptor substrate IRS1 [8], and the modulation of SIN1 activity [9,10], a component of the MTORC2 complex. Nevertheless, several aspects of these feedback loops remain unclear.

For instance, in the case of SIN1, there has been debate about the regulation and function of its phosphorylation sites, in particular T86 and T398. Liu *et al* [9] have shown that both S6K and AKT can phosphorylate SIN1, and that this phosphorylation suppresses MTORC2-mediated phosphorylation of AKT. Conversely, Humphrey *et al* [11] have shown that AKT phosphorylates SIN1 and that this stimulates MTORC2 instead. Yang *et al* [10] further investigated this in more cell lines and conditions, arguing that AKT rather than S6K is the major SIN1 kinase and that the feedback is positive, in line with Humphrey *et al*.

For IRS1, one aspect which is unclear is the phosphorylation of S312 (in human IRS1). Note that this site corresponds to mouse S307, while the human S307 constitutes yet another phosphorylation site on IRS1. Signaling by IRS1 is regulated by multiple phosphorylation events [12], where serine/threonine phosphorylation inhibits its activity while tyrosine phosphorylation activates it. For S312, it is known that JNK (among other kinases) phosphorylates this site [13], but it has also been shown that this phosphorylation is MTORC1-dependent [14]. S6K phosphorylates IRS1 on several other sites, including S307 [15] and S270 [16]. Given the MTORC1-dependence of IRS1 S312 phosphorylation, it is sometimes assumed that S312 phosphorylation is also S6K-dependent. Indeed the databases Uniprot [17]and PhosphoSitePlus [18] currently list S6K as kinase for IRS1 S312 phosphorylation (putatively in the case of PhosphoSitePlus), even though to the best of our knowledge there is no direct evidence for this. Rather, it has been reported as unlikely [15], and S312 is also not part of an S6K target motif. MTORC1 may also affect S312 phosphorylation through an effect on protein phosphatase PP2A [19] or indirectly through dependencies between phosphorylation sites. In addition to these details, it is unclear in which contexts feedback through IRS1 -- whether mediated by MTORC1, JNK, S6K or yet other regulators -- is important. For example, feedback to IRS1 is involved in insulin resistance [8], and mediates re-activation of the AKT pathway after rapamycin treatment [20], but seems not to be involved in re-activation of AKT after AKT-inhibitor treatment [21,22].

Computational models can be used to better understand these complex regulatory networks. Different modeling frameworks have been used to quantitatively study oncogenic cellular signaling pathways, including dynamic models [23,24] and steady state models [25–27]. Dynamic models can describe detailed kinetics of a system, but they are costly to simulate, especially when the full parameter uncertainty is analyzed [28–30]. The computational cost is further exacerbated when multiple datasets are included, resulting in many model conditions which have to be evaluated for each parameter value. Logic models allow significantly larger models to be evaluated [31,32], but it is more difficult to model quantitative differences between cell lines and conditions in such a framework.

bioRxiv preprint doi: https://doi.org/10.1101/268359; this version posted February 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

We have previously developed Inference of Signaling Activity (ISA), a steady state modeling approach to study signaling activities across cell lines along with different inhibitor treatments [27]. One major assumption in ISA is the absence of feedback signaling, which allowed fast model evaluations and hence relatively large signaling models. Models without feedback can give good fits to drug response cell viability data [27], arguing that feedback events are not crucial to describe differences in the relative viability between cell lines after 72 hour drug treatment. However, it is clear that feedback signaling events are a major component of cellular signaling, and are important to consider in many situations. We therefore set out to expand ISA with feedback signaling events, to explore whether the method could also be used to infer signaling activities in models that include feedback events.

Methods

Below, we outline the substantial changes we have made to ISA [27], resulting in feedback-ISA (f-ISA), an approach capable of modelling feedback mechanisms in signaling pathways across different cell lines and conditions.

Model structure

In ISA, the activities of signaling molecules are modeled by a continuous latent variable x_i that can take values between 0 and 1. This signaling activity is denoted by x_i where i indexes the signaling molecule. This activity is a function of the upstream signaling nodes, as well as a basal activity, the expression of the signaling molecule itself and any kinase inhibitors that may be present. In the original ISA, the activity was restricted between 0 and 1 using a clamping function. In order to accommodate feedback loops, we now change the activation function to a logistic function. Specifically, we calculate the signaling activity as

$$x_i = u_b e_i \frac{1}{1+\exp[-k(b_i + (\sum_{j \in \text{parents}_x(i)} u_a u_s a_{j,i} x_j) + (\sum_{j \in \text{parents}_m(i)} a_{j,i} m_j) - s)]} = \mathrm{h}_i(\mathbf{x}) \quad (1.1)$$

Here u_a, u_b and u_s are the kinase inhibitor effects (defined in more detail later), e_i is the expression of signaling molecule i, b_i is the basal activity of the signaling molecule i, $a_{j,i}$ is the strength of signaling from molecule j to molecule i, m_j is a binary variable denoting whether mutation j is present, while k and s are the constant steepness and inflection point of the logistic function. The logistic function is more expensive to compute than a clamping function, but in contrast to a clamping function, the logistic function is smooth (i.e. continuously differentiable), which simplifies the process of solving the systems of equations with feedbacks. To keep the number of parameters manageable, we fix both the steepness k and the inflection point s to a set value. The steepness is set to 9.19024 such that the activity is 0.01 and 0.99 when the total input is 0 and 1 respectively, and the inflection point is set to 0.5. An example of an activity calculation is shown in Figure 1B.

Without feedback events, the signaling activities can be calculated from the upstream molecules downwards. However, by including feedback events, the equations for the signaling activities become coupled, and as a result the activities have to be calculated by solving a system of nonlinear equations. We use the Newton-Raphson method for this; that is, we solve the equation

$$\mathbf{x} = \mathbf{h}(\mathbf{x}) \quad (1.2)$$

by iterating through

$$\mathbf{x}^{n+1} = \mathbf{x}^n - \mathbf{J}^{-1}(\mathbf{x}^n)\mathbf{f}(\mathbf{x}^n) \ , \quad (1.3)$$

where $\mathbf{x}^n$ represents the solution at the n-th iteration, and

bioRxiv preprint doi: https://doi.org/10.1101/268359; this version posted February 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

$$\mathbf{f}(\mathbf{x}) = \mathbf{h}(\mathbf{x}) - \mathbf{x}\,. \tag{1.4}$$

The Jacobian matrix **J** is given by

$$J_{i,j} = \frac{\partial \mathrm{f}_i}{\partial x_j} = \begin{cases} -1 & \text{if } i = j \\ 0 & \text{if } j \notin \text{parents}(i) \\ u_b u_a u_s e_i a_{j,i} \cdot \dfrac{kp_i}{(p_i+1)^2} & \text{if } j \in \text{parents}(i) \end{cases}, \tag{1.5}$$

with

$$p_i = \exp[k(b_i + (\sum_{j \in \text{parents}(i)} u_a u_s a_{j,i} x_j) + (\sum_{j \in \text{parents}_m(i)} a_{j,i} m_j) - s)]\,. \tag{1.6}$$

If the system of equations contains four or fewer signaling molecules, Equation 1.3 is most efficiently calculated by taking the inverse of the Jacobian directly. If the system contains more than four signaling molecules, it is more efficient to use LU-decomposition to solve Equation 1.3 instead. We stop the Newton-Raphson iteration when the last change in **x** is less than 10^{-5} in each direction.

Although it is possible to calculate the entire model with Equations 1.2 and 1.3, the LU-decomposition scales cubically in the number of signaling molecules. It would therefore be beneficial to decrease the system size as much as possible. The system of equations is not fully coupled however. Some signaling molecules are not involved in any feedback loop, but are only upstream or downstream of other molecules. The activities of these molecules can be calculated directly, without having to solve a system of equations. Additionally, some signaling molecules which are affected by a feedback loop, may not be affected by another loop. We can therefore decrease the size of the systems to be solved by decomposing the model into smaller systems composed of signaling molecules which are coupled by a feedback signaling loop. To identify which parts need to be solved as a system of equations, we use Tarjan's strong connectivity algorithm [33] as a preprocessing step. This algorithm identifies all the strongly connected components of a directed graph, as well as providing a topological ordering of the graph. To calculate the signaling activities, we then iterate over the topological ordering, and calculate the activities either directly (for connected components of size 1, i.e. individual signaling molecules that are not part of any feedback system) or by solving the corresponding system of equations (for connected components with size larger than 1). Figure 1C illustrates the decomposition of a model into individual molecules and systems.

Note that the signaling activity vector **x** is a function of the mutations, **m,** protein expression, **e**, and the presence of any inhibitors, *u*, (all of which can change between cell lines and conditions), as well as a number of parameters (which remain constant between cell lines and conditions). As such, the model is constrained to explain all observed signaling activities from the mutation and protein expression patterns. This is a fundamentally different approach from estimating signal strengths separately for each cell line and condition, as done for example by [24]. Rather than inferring cell-line specific signaling strengths, we attempt to infer signal strengths that can fit the data across all cell lines and conditions.

Kinase inhibitors in the model can be of two types: they can either inhibit the activity of their target, or they can inhibit the activation of the target. The type of the inhibitor is derived from the literature. For example, the AKT inhibitor GSK690693 is an ATP-competitive inhibitor which inhibits the activity of AKT, while the AKT inhibitor MK2206 is an allosteric inhibitor which inhibits the activation of AKT. For activity-inhibiting drugs we set u_a=u in equation 1.2 (for the specific signaling link that is inhibited), and for activation-inhibiting drugs we set u_b=u (for the

bioRxiv preprint doi: https://doi.org/10.1101/268359; this version posted February 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

specific target that is inhibited). Note that u_a affects the activity of the children of the target, whereas u_b affects the activity of the target itself. The type of the inhibitor is specified in the SBML annotation (described later). Regardless of whether u_a or u_b is used, the inhibitory effect is calculated as

$$u = \begin{cases} q & \text{if a single concentration is used} \\ q + \dfrac{1-q}{10^{k_u(c-c_{IC50}))+1}} & \text{if more than one concentration is used} \\ 1 & \text{if the inhibitor is not present} \end{cases} \quad (1.7)$$

where q is the maximal inhibitory effect (or the inhibitory effect at the particular concentration used if there is only one concentration), k_u is the steepness of the inhibitory curve, c is the concentration and c_{IC50} is the 50% inhibitory concentration. In contrast to the activation function in equation 1.1, the steepness and inflection point of the kinase inhibition curves (i.e., k_u and c_{IC50}) are included as free parameters to be inferred. Finally an inhibitor may increase the susceptibility of its target to *incoming* signals, a process which has been named "inhibitor hijacking" [21]. The variable u_s is included in Equation 1.1 to reflect this effect. This parameter is only used if the drug is specified to alter the susceptibility of its target, and we use it specifically to allow ATP-competitive AKT inhibitors to alter the susceptibility of AKT phosphorylation by MTORC2.

Likelihood function

Jastrzebski *et al* [27] used protein phosphorylation data as well as untreated proliferation rates and relative viability upon kinase inhibitor treatment to infer the signaling activities. However, since f-ISA is computationally more expensive than the original ISA framework, we only use protein phosphorylation data for the inference here. Including cell viability data, in particular dose response curves, would result in too many model conditions as well as additional parameters to be estimated. Mutation data, copy number data and total protein expression data are still used as before [27], by directly setting the corresponding variable (m_i or e_i) to the observed value. The structure of a small part of a model is shown in template notation in Figure 1D.

The likelihood of an observed protein phosphorylation measurement is defined as

$$P(y_{i,j,k,l} \mid \theta) = t(y_{i,j,k,l} \mid \mu = z_{i,j,k}(\theta), \sigma = \sigma_{i,k}, \nu = 3) \ , \quad (1.8)$$

where Θ is the vector containing all model parameters, $y_{i,j,k,l}$ is the measurement data for observed variable i, cell line j, dataset k and replicate l, $z_{i,j,k}$ is the modeled variable (defined further below in Equation 1.9) and $\sigma_{i,k}$ is the variance of observed variable y_i in dataset k. The variance $\sigma_{i,k}$ for observed variable i is shared by all cell lines and biological replicates, but is specific for each dataset k.

The modeled variable z is defined as

$$z_{i,j,k}(\theta) = g_{i,k} + d_{i,k} x_{i,j}(\theta) \ , \quad (1.9)$$

where $g_{i,k}$ is the background signal generated by aspecific binding of the antibody and $d_{i,k}$ is a scaling factor to account for differences between datasets (recall that k indexes the dataset). If a particular phosphorylation is only measured by one dataset, then $d_{i,k}$ is set to 1, and if the phosphorylation is measured by more than one dataset, then $d_{i,k}$ is included as a free parameter to be inferred, for every dataset. $g_{i,k}$ is always included as free parameter to be inferred for every epitope and dataset, regardless of whether the epitope is uniquely measured by only one dataset. Finally, we assume independence between the epitopes, datasets, cell lines and replicates, and the total likelihood is calculated as a product of the likelihoods of the individual data points.

bioRxiv preprint doi: https://doi.org/10.1101/268359; this version posted February 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Although the likelihood function includes a scaling variable to perform batch correction, it is useful to have the measurements on approximately the same scale between datasets before running the inference. Therefore, as a preprocessing step, the measured phosphorylation levels are reduced to [0,1] by dividing the measurements of each epitope by the maximum value observed for that epitope in any cell line or condition in that dataset, if that value is larger than 1. With the RPPA quantification procedure that was used (SuperCurve [34]), maximum observed values below 1 indicate that for that epitope, all conditions in that dataset have low phosphorylation levels, relative to the set of standard lysates. We therefore do not *increase* the signal to 1, to prevent amplifying experimental noise. With all measurement values in the range [0,1], the prior for the scaling parameters $d_{i,k}$ can then be set to a small uniform distribution, between 0 and 2, which reduces the parameter space that has to be searched.

Parameter priors

The model contains several different types of parameters that are inferred from the data. Table 1 lists these parameters, along with the prior distribution that is used for each of them.

The strength parameter for reactions with a known influence sign is inferred on a logarithmic scale. This is done to give equal prior weight to weak interactions (strength 0.01-0.1), interactions with intermediate strength (0.1-1) and strong amplifying interactions (1-10). If this range from 0-10 was not log-transformed and a uniform distribution would still be used, most of the prior weight would be on amplifying signals, and all signaling activities would be heavily saturated *a priori*, especially in a signaling cascade. When the sign of the influence is not known, that is, when the signal could be either activating or inhibiting, then the prior is set on a regular scale from -2 to 2.

For drug inhibition, the drug can be specified to either use a single drug concentration, giving only a single parameter for the inhibition at that concentration, or to use multiple drug concentrations, giving three parameters: the maximum inhibition, the 50% inhibitory concentration (IC_{50}) and the steepness. The prior for the 50% inhibitory concentration is centered on the concentration at which it inhibits its target for 50%, as determined by *in vitro* inhibition experiments described in the literature.

Table 1 Model parameters that are inferred from the data.

Parameter	Symbol	Prior
base_[*molecule*]	b_i	uniform(a=0,b=1)
strength_[*molecule*]_[*molecule*]	$s_{j,i}$	uniform(a=-2,b=1) – $\log_{10}$ scale
maxinhib_[*drug*]_[*molecule*]	q	uniform(a=0,b=1)
[drug]_[molecule]_susceptibility	u_s	uniform(a=-2,b=1) – $\log_{10}$ scale
ic50_[*drug*]_[*molecule*]	c_{IC50}	normal(μ=x,σ=2) – $\log_{10}$ scale
logsteepness_[*drug*]_[*molecule*]	k_u	uniform(a=-1,b=1) – $\log_{10}$ scale
base_measurement_[*dataset*]_[*epitope*]	$g_{i,l}$	uniform(a=0,b=1)
scale_measurement_[*dataset*]_[*epitope*]	$d_{i,l}$	uniform(a=0,b=2)
variance_measurement_[*dataset*]_[*epitope*]	$\sigma_{i,l}$	exponential(λ=5)

Model structure specification

The signaling graph of the model is constructed in CellDesigner [35], version 4.4. Using this tool, the model is specified in Systems Biology Graphical Notation as an Activity Flow diagram [36]. That is, rather than specifying precise molecule states and reactions, only the abstract influences between signaling molecules are specified, which fits with the modeling paradigm of ISA. The model is stored as an SBML file with CellDesigner extension annotation. This extension annotation allows specification of the types of the signaling molecules (i.e. whether a node is a signaling molecule, mutation or drug) as well as specification of the signaling influences and the types of drug inhibition.

Inference

bioRxiv preprint doi: https://doi.org/10.1101/268359; this version posted February 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

To infer the posterior distribution of the parameters, we use the BCM software package [37]. To incorporate f-ISA in BCM, we developed the tool `sbmlpdinf`, which can read the activity flow diagram described in the SBML/CellDesigner file, as well as two XML files specifying the likelihood and the prior. All data is stored in a single NetCDF4 file. The likelihood file specifies which measurement in the data file corresponds to which node in the model, as well as all the model conditions, such as the presence of a particular kinase inhibitor. The model simulation code was written in C++ and made thread-safe for use in the parallelized inference algorithms of BCM. The largest model presented here, with 18 signaling molecules and two feedback systems, could be evaluated at approximately 1.1 million model evaluations per second using 18 threads on an Intel Xeon E5-2697 v4 processor.

We sampled the posterior distribution using feedback-optimized, parallel tempered MCMC [38] with automated parameter blocking [39]. The marginal likelihood was calculated using thermodynamic integration [40]. The temperature schedule was optimized twice, which also served as burn-in period. After this optimization, 1,000 samples were generated from the posterior (after subsampling), with the amount of subsampling chosen such that the autocorrelation was negligible and at least 100 roundtrips from prior to posterior were performed. Specifically, when one dataset is included in the inference we use 24 parallel chains and run the inference with a subsampling of 1 in 2,000 and probability of choosing a temperature swap move of 0.9. Each MCMC move constitutes updating every parameter or parameter block once. When more than one dataset is included we increased the number of chains to 36 and use a subsampling of 1 in 4,000. Sample traces are provided in Supplementary File 1. Despite the high performance implementation, inference with a single dataset and a medium-sized model takes several hours, and the largest inference presented here, which included 108 model conditions and 139 free parameters, required approximately 48 hours to run with the aforementioned processor.

Quantifying reduction in uncertainty

To quantify the reduction in uncertainty, we used the Occam factor introduced by MacKay [41], which quantifies the change in volume of the parameter space that is accessible from the posterior compared to the prior. In other words, it measures which fraction of the prior parameter space is consistent with the data. MacKay used a Gaussian approximation to calculate the Occam factor; but since we evaluated the marginal likelihood as part of the inference here, we can calculate the Occam factor directly by

$$\log \text{Occam factor} = \log\ P(y \mid M) - \log P(y \mid M, \theta_{MAP}) \tag{1.10}$$

where $P(y|M)$ is the marginal likelihood of the data given the model and $P(y|M,\theta_{MAP})$ is the likelihood of the data at the maximum a posteriori value of the parameters.

Cell lines

All cell lines used have been described, including their culture conditions, in Supplementary Table 1 of Jastrzebski *et al* [27].

Measurement of on-treatment phosphorylation

Cell lines were seeded in 60 mm dishes (BT549 at $4x10^5$ cells/dish; HCC1954 at $8x10^5$ cells/dish; MCF7 at $4x10^5$ cells/dish; MM231 at $8x10^5$ cells/dish; MM453 at $1x10^6$ cells/dish; MM468 at $1x10^6$ cells/dish; SKBR3 at $6x10^5$ cells/dish; T47D at $8x10^5$ cells/dish). Following 24 h of incubation, all but the exponentially growing cells were serum starved in unsupplemented base medium containing penicillin/streptomycin (Gibco) for a further 24 h. Cells were then treated with either DMSO vehicle control or one of the following three inhibitors - PD0325901 at 50 nM, GDC0941 at 10 μM or AZD8055 at 1 μM – for 30 min, after which stimulation with 10 ng/ml EGF, where indicated, was carried out for a further 20 min. Cells were then placed on ice, washed with ice-cold PBS, and lysed in 150 μl/dish of RIPA buffer (20 mM Tris-HCl, pH 8, 150 mM NaCl, 1% NP40, 0.5% sodium deoxycholate, 0.1% SDS) supplemented with cOmplete protease and phosSTOP phosphatase inhibitor cocktails (Roche). Lysates were cleared by centrifugation at 4°C and 20,800 x g, and protein concentration determined using the Pierce BCA protein

bioRxiv preprint doi: https://doi.org/10.1101/268359; this version posted February 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

assay (Thermo Fisher Scientific). The supernatant was normalized to 1 µg/µl with RIPA buffer and supplemented with SDS sample buffer to a final concentration of 62.5 mM Tris-HCl pH 6.8, 10% glycerol, 2% SDS, 2.5% (v/v) 2-mercaptoethanol. Samples were further assayed at the MD Anderson Cancer Center RPPA Core Facility, as outlined previously [27].

IRS1 disruption experiment

Cell lines were seeded in 6-well plates (BT549 at $4x10^5$ cells/well; HCC1954 at $2x10^5$ cells/well), and following a 48 h incubation, treated with either DMSO vehicle control or 5 µM NT157 for a further 24 h. They were then treated for a further 1 h with either DMSO vehicle control, 1 µM GSK-690693 or 5 µM PF-4708671. Cells were then lysed and protein concentration determined as outlined above. Twenty µg of protein were supplemented with Novex® LDS Sample Buffer and Sample Reducing Agent, heated at 70°C for 10 min and separated on 4-12% gradient gels (Thermo Fisher Scientific). Separated proteins were transferred onto Immobilon-P PVDF membranes (Merck Millipore) using a Trans-Blot® system (Bio-Rad). Blocking was performed in TBS supplemented with 0.1% Tween and 3% BSA (TBS-TB) for 1 h at room temperature, followed by overnight immunoblotting at 4°C with the following primary antibodies: IRS1 (MERCK/Millipore 06-248); pIRS1-S312 (Cell Signaling Technology 2381); AKT (Cell Signaling Technology 9272); pAKT-S473 (Cell Signaling Technology 4060); SIN1 (Cell Signaling Technology 12860); pSIN1-T86 (Cell Signaling Technology 14716); PRAS40 (Cell Signaling Technology 2610); pPRAS40-T246 (Cell Signaling Technology 2997); S6 (Cell Signaling Technology 2217); pS6 (S235/236) (Cell Signaling Technology 2211); Vinculin (Sigma V9131). Membranes were then washed with TBS supplemented with 0.1% Tween (TBS-T) and probed with secondary goat anti-mouse or anti-rabbit HRP-conjugated antibodies (Bio-Rad) diluted in TBS-TB for 2 h at room temperature. Finally, membranes were washed in TBS-T, an ECL reaction was carried out using the Clarity™ Western ECL Substrate (Bio-Rad) and the signal detected using a ChemiDoc Touch instrument (Bio-Rad). Quantification of the exposures was performed using ImageQuant software (Bio-Rad).

External data

In addition to the data we generated here, we included a part of the dataset provided by Korkola *et al* [42] in the inference. They performed RPPA measurements of 15 breast cancer cell lines, treated with either lapatinib (250 nM), GSK690693 (250 nM), a combination of the two, or a DMSO control, at 8 time points from 30 minutes to 72 hours. Since we focus here on fast-acting post-translational feedback here, we selected the 1-hour time point from their data, which most closely matched the 50-minute treatment of our on-treatment phosphorylation measurements. We used only the 9 cell lines which overlap with our cell line panel, since we have mutation and copy number data available for these cell lines. To have the measurements on the same scale, we reverse-log-transformed the data and divided by the maximum value for each epitope as described above.

Results

Constructing a model of steady state signaling with feedback loops

To better understand the signaling activities in the MAPK and AKT regulatory networks, we followed the ISA modeling approach [27] to construct a simplified model of these pathways. We based our starting model (see Figure 1A) on the simplified, joint drug-model of [27]. The starting model includes three receptor tyrosine kinases (RTKs), a simplified representation of the MAPK and AKT pathways, and 10 important oncogenic mutations which are observed in breast cancer cells. In addition, we now added four known feedback signaling events.

Briefly, in ISA, the steady state activity of a signaling molecule is modeled as a latent variable with a continuous value between 0 and 1. The signaling activity is a function of a basal activity, the expression of the signaling molecule itself, the activities of upstream signaling molecules, the effect of mutations, and any kinase inhibitors that are present. Importantly, the parameters of the model are shared across all cell lines and conditions. That is, the parameter values are the same for all cell lines in all conditions. For example, while each cell line can have a

bioRxiv preprint doi: https://doi.org/10.1101/268359; this version posted February 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

different amount of AKT activity in a particular condition, a given amount of AKT signal always gives rise to the same amount of input signal to MTORC1.

To be able to include feedback events in ISA, we make two main changes to the modeling approach. First, to constrain the activity of the signaling molecules between 0 and 1, we use a logistic function rather than a clamping function (see Figure 1B). Although this is computationally more expensive, it makes the activation function smooth, which greatly simplifies solving the systems of equations when feedbacks are present. Second, to calculate the activity of the molecules which are part of a feedback system, we use Newton-Raphson root-finding rather than calculating the activities in one pass from upstream to downstream, as was done before. Since the computation of one Newton-Raphson step scales cubically with the number of signaling molecules, a significant speed-up can be obtained if we restrict the root-finding to the part of the model that contains feedback. We therefore first identify the strongly connected components in the model, using Tarjan's strong connectivity algorithm [33], illustrated in Figure 1D. This algorithm also provides a topological ordering, which is needed to calculate the signaling activities in the decomposed equations in the correct order. The system of equations corresponding to each strongly connected component is then solved separately, with the benefit that these systems are typically much smaller than the complete model. All equations and the methodology for solving them are described in detail in the Methods section.

Feedback activity from ERK to RAF is partially identifiable from pre-treatment, non-intervention data

Having a methodology to infer steady state signaling activities in models with feedback, we first wondered whether feedback loops are already identifiable using non-intervention data. To test this, we fitted the starting model (Figure 1A) to the phosphorylation data of thirty, untreated breast cancer cell lines, grown under normal culturing conditions; i.e. the phosphorylation data of Jastrzebski *et al* [27]. We first included only the phosphorylation data of the downstream signaling kinases (Figure 2A). To our surprise, these non-intervention data already suggest that the feedback loop from ERK to RAF is likely to be active (Figure 2B). A possible explanation for the identifiability of feedback from non-intervention data may be that there are many inputs into the MAPK pathway; in the model we included three different RTKs, as well as mutations in KRAS or BRAF. These five inputs together could lead to over-activation of the MAPK pathway, so to prevent this, the inputs need to be restrained in some way. It is easier to accomplish this by a negative feedback loop, since this requires only one parameter to be given a high value, rather than by each input to the MAPK pathway being weak, which requires five parameters to have a low value. In such situations, the Bayesian inference follows the principle of Occam's razor by preferring a simple explanation over a more complex one.

To test whether it is indeed the inputs to the MAPK pathway that determine the high values for ERK-to-RAF feedback activity, we artificially forced the inputs to be weak by restricting the prior for the strength parameters of these inputs (Figure 2C). In this case, we indeed see that the feedback becomes weaker as well, indicating that the negative feedback is used to balance the inputs.

Nevertheless, the posterior probability distribution for the strength of the ERK-to-RAF feedback loop has non-zero probability for low values as well. This means that a weak or very weak feedback loop is also consistent with the data. A model entirely excluding this negative feedback loop is less likely to represent the data, but the difference is small (the log Bayes factor is 1.0 in favor of the model including the feedback loop). In summary, given the pre-treatment, non-intervention data of the intracellular kinases, this feedback loop is likely to be active, but based on the single dataset used so far we cannot rule out that it the feedback from ERK to RAF is inactive.

Adding data can increase the uncertainty in individual parameters

We next tested whether the addition of more data helps in identifying the feedback loop. We first added data for the surface receptor tyrosine kinases, which provide information on the inputs to the pathways (Figure 3A). Interestingly, if we include phosphorylation and protein expression data of EGFR and ERBB2, the feedback activity from ERK to RAF in fact becomes less identifiable (Figure 3B). With these data, the model can infer that the activities of EGFR and ERBB2 are low in the majority of cell lines, and this constrains at least some inputs to the

bioRxiv preprint doi: https://doi.org/10.1101/268359; this version posted February 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

MAPK pathway to low strengths. Given this, a strong feedback is no longer required to balance the activation of the MAPK pathway. Since the feedback is not required, but still feasible, the activity of the feedback has become less certain.

To see whether the RTK data does provide some information overall, we tested whether the joint uncertainty in all parameters does decrease. This can be quantified using Occam's factors [41], which measure the reduction in the parameter space that is accessible from the posterior compared to the prior. The model with RTK data has a log Occam factor of -127.6, compared to -97.4 without the RTK data. The posterior space collapses more when the RTK data is included than when it is excluded, meaning that the uncertainty in all parameters together is reduced by the inclusion of the RTK data. Together, this shows that, while the total uncertainty is reduced by adding data, the uncertainty of individual parameters can increase.

Incorporating on-treatment measurements confirms ERK to RAF feedback activity

Intervention data should be more informative for identifying feedback loops. We therefore performed on-treatment measurements, using the MEK-inhibitor PD0325901, the PI3K inhibitor GDC0941 and the dual MTORC1/2 inhibitor AZD8055, in a smaller panel of cell lines (Figure 4A). Cells were either allowed to grow exponentially for 48 hours, or were starved for 24 hours, followed by 30 minute pre-treatment with inhibitors, before stimulation with EGF to increase signaling activity. The EGF stimulation is included in the model by setting EGF activity to 1 for the EGF-stimulated conditions and to 0 otherwise.

This on-treatment dataset was generated separately, at a different time, from the 30-cell line pre-treatment dataset. Although the RPPA measurements include a set of standard control lysates, the spot intensities are quantified separately and there are likely to be differences between the two data batches. Furthermore, we cannot always rely on a sufficient number of overlapping measurement conditions to align multiple datasets. We therefore incorporated batch correction directly into the inference, such that the differences between batches are automatically accounted for, and balanced against the most likely signaling strengths. This is accomplished by adding an offset and a scale parameter for each epitope in each dataset (see Figure 1C and the Methods section for details).

Focusing on the MEK inhibitor, we see that using the on-treatment data, feedback activity from ERK to RAF is clearly identifiable (Figure 4B, first column). This shows that on-treatment data is indeed more informative for inferring feedback activity, as expected. The result of the feedback is also clearly visible in the data (Figure 4A) given that MEK phosphorylation greatly increases after treatment with a MEK inhibitor. Although the increase in MEK phosphorylation upon MEK inhibitor treatment could also be a direct effect of the inhibitor rather than a feedback event (discussed in more detail for ATP-competitive AKT inhibitors later), PD0325901 is an allosteric inhibitor [43,44], and the increased phosphorylation of MEK is rather more likely to be through ERK to RAF feedback signaling [45].

When considering several other parameters, we see that both the pre-treatment and on-treatment measurement provide useful information (Figure 4B). The pre-treatment measurements were collected for a single condition for more cell lines, while the on-treatment measurements included more conditions for fewer cell lines. Combining the two datasets provides a broad coverage over cell lines as well as intervention measurements in a smaller panel to help identify signal strengths. The on-treatment data not only helps to identify feedback activity from ERK to RAF, but also, for example, the strength of AKT to MTORC1 signaling. Conversely, the broader pre-treatment data helps identify, for example, the effect of PIK3R1 mutations and FGFR2 amplifications, neither of which were found in the smaller cell line panel used for generating the on-treatment data.

Negative feedback is likely to be active in the AKT pathway

One of the main advantages of f-ISA lies in identifying which signaling activities and strengths are most strongly supported by the data. This advantage becomes most apparent when there are multiple interrelated feedback events, combined with information from multiple datasets. In this case it is impossible to manually keep track of all the

bioRxiv preprint doi: https://doi.org/10.1101/268359; this version posted February 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

constraints, and computational modeling is necessary to obtain a quantitative understanding of the regulatory network. To illustrate this, we next focus on the AKT pathway, where several different feedback events exist, further complicated by conflicting reports in the literature. As a first approximation, and as can be seen in Figure 1A, we incorporated the two hypotheses regarding feedback through SIN1, described in the introduction, through a negative link between MTORC1 and MTORC2 (representing inhibitory phosphorylation of SIN1 by S6K), and a positive link between AKT and MTORC2 (representing stimulating phosphorylation of SIN1 by AKT). The feedback through IRS1 is included as a negative link between MTORC1 and PI3K.

Using the two datasets described so far (see Figure 4A and 5A), the model identifies that there is likely to be a negative feedback in the AKT pathway (Figure 5B). The positive feedback through SIN1 is identified to be weak (Figure 5B, top-left panel), while the negative feedback through SIN1 is likely to be active (Figure 5B, top-right panel). This is partially driven by the same effect as in the MAPK pathway, that is, there are again many inputs into the AKT pathway, which are most easily balanced by a negative feedback loop. This effect can also be seen by a very strong correlation between the strength of signaling from PI3K to MTORC2 and the negative feedback from MTORC1 to MTORC2 (Figure 5B, bottom-right panel). However, apart from this effect, the model also infers that the negative feedback from MTORC1 to MTORC2 is used to reduce AKT-pathway activity in cell lines with high MAPK pathway activity, including MDA-MB-231 and ZR-75-30.

The negative feedback in the AKT pathway could be achieved in several different ways, either through IRS1 or through SIN1. To further resolve which of these proteins mediates the feedback, we extended the model to explicitly include IRS1 as signaling molecule (Figure 5C), and we incorporated information from the dataset of Korkola *et al* [42], who performed time-course RPPA measurements with a different panel of cell lines upon inhibitor treatment, including the AKT inhibitor GSK690693. We selected the 1 hour time point from their dataset as it is closest to the 50 minute inhibitor treatment (30 minute inhibitor pre-treatment plus 20 minute subsequent EGF stimulation) in our intervention dataset. To accommodate this dataset, we also added GSK690693 as inhibitor to the model.

A complication when using ATP-competitive AKT inhibitors like GSK690693 is that these inhibitors can directly alter the susceptibility of AKT phosphorylation by PDK1 and MTORC2, independently of AKT kinase activity [21]. This effect causes an increased AKT phosphorylation upon AKT inhibitor treatment, independent of any feedback events. We incorporated this effect in the model by allowing a kinase inhibitor to alter the strength of incoming signals to the targeted protein (see Methods section; note that this effect cannot be seen in the SBGN schematics but is included in the SBML model annotation).

The model can fit several aspects of the Korkola data, including the expected reduction in S6 phosphorylation and increased AKT phosphorylation upon AKT inhibitor treatment (Figure 5D), without compromising the fit of the other two datasets (Supplementary File 1). Furthermore, we see that the increased AKT phosphorylation upon AKT inhibitor treatment is indeed most likely explained by an increase in AKT phosphorylation susceptibility (Figure 5E, right-most panel). Despite this, a negative feedback in the AKT pathway is still likely to be present (Figure 5E, left-most panel). This, however, is unlikely to be through IRS1 S312 phosphorylation (Figure 5E, 4^{th} and 5^{th} panel). Consistent with this, the measured IRS1 S312 phosphorylation levels are not consistently changed upon AKT inhibitor treatment in this dataset (Figure 5D, bottom panel). Finally, the presence of a positive feedback in the AKT pathway, acting via SIN1, is still deemed unlikely in this model with IRS1 included as signaling molecule (Figure 5E, 2^{nd} panel).

Testing the strength of negative feedback through IRS1 by modulating its expression

The above described inference indicated that while there is negative feedback in the AKT pathway, the feedback through IRS1 S312 phosphorylation is likely to be weak. However, there is uncertainty in the strength of IRS1 to PI3K signaling. To further resolve this, we experimentally disrupted the IRS1 feedback mechanism by pre-treating cells with the IGF1R inhibitor NT157, which has been shown to induce degradation of IRS1 [46,47], and measured

bioRxiv preprint doi: https://doi.org/10.1101/268359; this version posted February 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

how this affects phosphorylation upon AKT- or S6K inhibitor treatment (Figure 6A). To accommodate this data more precisely, it is also useful to expand the model by including S6K and SIN1 as separate nodes (the resulting model is shown Figure 6B). Since we now explicitly include SIN1, we model the unknown effect of SIN1 on MTORC2 by allowing this signal to be either positive or negative.

With some exceptions, the model can describe the IRS1-disruption experiment very well (Figure 6C), without compromising the fit to the other datasets (Supplementary File 1). The data shows a clear increase in AKT phosphorylation after AKT-inhibitor treatment, but not after S6K-inhibitor treatment. We also see that the increased AKT phosphorylation is reduced by pre-treatment with NT157. This is consistent with the report that AKT-inhibitor-induced AKT-phosphorylation is dependent on MTORC2 activity [21], and MTORC2 being dependent on IRS1 signaling. The S6K inhibitor, although it does reduce S6 S235/236 phosphorylation, does not change the phosphorylation levels of AKT S473, SIN1 T86 or IRS1 S312, suggesting it is not critically involved in mediating the feedback loop under these conditions.

We can then use the model to delineate all the feedback activities (Figure 6D). In contrast to the inference based on the previous datasets, the model can now infer that the signal from IRS1 to PI3K is strong, perhaps even a point of signal amplification in the pathway, given that the inferred strength is larger than 1 (Figure 6D, panel 7). Hence a relatively weak negative feedback to IRS1 can still affect the AKT pathway. Furthermore, there appears to be a weak, though nonzero, feedback to IRS1 (Figure 6E). The model cannot entirely resolve whether the feedback is mediated by MTORC1 only or also through S6K (Figure 6E), due to the uncertainty in how strongly the S6K inhibitor is inhibiting its target (Figure 6D, panel 8) - the more likely it is that PF4708671 is inhibiting S6K, the less likely S6K is to phosphorylate IRS1 on S312. Despite this uncertainty, an S6K-independent signal to IRS1-pS312 is approximately three times more likely than an S6K-dependent signal (Figure 6D and E).

The hyperphosphorylation of AKT upon GSK690693 treatment is still mainly explained by the inhibitor-induced change in phosphorylation susceptibility (Figure 6D, panel 4). It is also clear that SIN1 T86 phosphorylation is dependent on AKT (panel 1), providing further information for the debate on the regulation of this site. Furthermore, SIN1 T86 phosphorylation is unlikely to be mediated by S6K (panel 2), although there is again some amount of uncertainty caused by the poor and uncertain efficacy of the S6K inhibitor (panel 8). However, given the present data, feedback through SIN1 is predicted to have only a minor negative effect on MTORC2 activity, if any (panel 3).

Taking into account 1254 data points spread across 108 different conditions, the most likely feedback in the AKT pathway is a strong effect of AKT back to SIN1, which only weakly inhibits MTORC2, combined with a weak signal from MTORC1 to IRS1 S312 (but unlikely through S6K), which is however amplified by IRS1.

Using ISA to test the agreement between datasets

Individual datasets are often insufficient to constrain most of the signaling strengths. Figure 7A shows how much each parameter is constrained by each dataset, as well as all datasets together. We can see that in many cases, a parameter is constrained by a single dataset. In this way, each dataset provides complementary information, and thus all datasets together are able to constrain most of the parameters to some extent.

Since datasets often provide different measurements, it is not clear how they can be directly compared against each other. However, we can use f-ISA to test whether datasets agree with each other given a model. To explore this, we calculated the disagreement in each parameter between any pair of the four datasets. We calculate the disagreement by taking one minus the overlap coefficient between the posterior 90% confidence intervals (CI) of the parameter given each dataset, where the overlap coefficients is the size of the intersection divided by minimum size of either CI. A disagreement of 1 means that the posterior 90% CIs do not overlap, whereas a disagreement of 0 means that the CI of one posterior is completely contained within the other, or are even exactly the same. Figure 7B shows this disagreement between datasets. Reassuringly, for most of the parameters the disagreement is small, indicating that in most cases the posterior distribution is accurate between datasets (an example is shown in Figure 7C). Note that this

bioRxiv preprint doi: https://doi.org/10.1101/268359; this version posted February 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

also includes many cases where one or both of the datasets simply do not constrain a parameter at all, hence the model may not be able to give a precise prediction for a parameter based on a dataset, but the posterior is still accurate in that it quantifies the uncertainty correctly.

There are, however, also some parameters for which the datasets disagree. Most disagreement occurs between the pre-treatment and intervention datasets. For example, the strength of signaling from MEK to ERK is estimated differently by the pre-treatment and intervention datasets. Assuming that the measurements in both datasets can be trusted, this would indicate that the model is overestimating its certainty. Nevertheless, despite the disagreement in some parameters, the model can provide a good fit against all data simultaneously (Figure 7D), and the resulting joint posterior is the most likely compromise between the posteriors inferred by the datasets separately.

Identification of unexplained data points

Such models can also be used to identify unexpected parts of the data. Although the model can explain a large part of the data (Figure 7D), there are also data points which cannot be recapitulated by the model. For example, in the intervention dataset, EGF stimulation leads to a large increase in ERK phosphorylation in four of the eight cell lines, which the model cannot reproduce (see Figure 4A). The cell line specificity is partially explained by EGFR expression, but not completely, given that despite EGFR expression is included as input the model cannot explain the variability observed. In addition to the cell line specificity, there is only a very modest increase in MEK phosphorylation upon EGF stimulation, raising the question of how the signal from EGF to ERK is transmitted. One explanation for a difference in the changes of MEK and ERK phosphorylation could be that MEK amplifies the signal. However, the activation function allows for a signal amplification of up to 10-fold, and hence the constraints posed by all other data do not seem to allow this explanation. An explanation rooted in signal amplification also does not explain why this occurs only in some of the cell lines. Alternative explanations might be that there are additional feedback mechanisms or feed-forward signaling pathways which transmit the EGF signal downstream, that other phosphorylation sites are involved, or that the discrepancy is a result of the specific timing of starvation, inhibition and stimulation.

Another interesting part of the data that cannot be explained by the model, is the increased S6 phosphorylation in HCC1954 upon NT-157 pretreatment (Figure 6C). It has been shown that NT-157 treatment can induce MAPK pathway activation by inducing IGF1R-SHC complex formation [46], and the increased MAPK signaling could lead to elevated S6 phosphorylation [48,49]. These are links that could be included in the model, although it would then still be unclear why the increased S6 phosphorylation is specific for HCC1954 and does not occur in the BT474 cell line. Alternatively, other pathways may be activated as a result of the disruption of IRS1 signaling, such as a stress response, potentially leading to S6K activation. Indeed, following treatment with NT-157, we did observe a reduction in cell viability of the HCC1954 cell line, while the BT474 cells appeared unaffected by the treatment.

These two examples show that the computational model can highlight cases where our knowledge, as summarized in the model, is incomplete and unable to fully describe the data. To resolve these discrepancies, additional rounds of model extension, combined with further experimental measurements to constrain the parameters, can be performed. Alternatively, the identified discrepancies can help define precise questions for follow-up functional genetic screening.

Discussion

As our knowledge of biological signaling networks grows, it is increasingly difficult to fully understand the behavior of these networks. Computational modeling then provides a means to quantify the interactions as well as the contributions of different signaling molecules and genetic aberrations. We can also use these models to test whether our knowledge is sufficient to explain the data. As our models of biological signaling get more complex, we will also increasingly need to integrate multiple datasets to be able to identify the key parameters.

bioRxiv preprint doi: https://doi.org/10.1101/268359; this version posted February 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

In this work, we used four datasets to constrain the parameters in a signaling model. Each of the data sets provided complementary information, and the largest number of the parameters could only be constrained with all four datasets employed simultaneously. We also described several non-intuitive behaviors of the identifiability of parameters. We found that feedback loops can sometimes be partially identified from non-intervention data, including the negative feedback from ERK to RAF, as well as a feedback from MTORC2 to MTORC1 used to reduce AKT signaling activity in cell lines with high MAPK pathway activity. Additionally, for the feedback from ERK to RAF, we found that adding more data to the inference can lead to an increased uncertainty in individual parameters. This is not caused by inconsistency of the data but by shifts in which signaling activities are more likely to explain all data simultaneously. We find that the aggregate uncertainty in all parameters does decrease when more data is employed in the inference.

Various other modeling approaches have previously been used to understand these signaling pathways. Thobe *et al* used logical modeling of MTORC2 signaling, including hypotheses of the regulation through SIN1 [50]. However, they were not able to resolve a preference between models, and given such a non-quantitative approach all models were found to be in agreement with the data. We find that using a quantitative approach as presented here, it is possible to deduce a preference for one feedback loop over another. Dalle Pezze *et al* constructed ODE models of mTOR signaling [51], although they did not include the SIN1 feedbacks (these had not been reported yet). In their ODE model, they used fixed parameter values, although they did perform a sensitivity analysis around the optimum. Given that in our study there are 108 model simulations required to evaluate the model with respect to the four datasets, and the many parameter values that need to be considered to characterize the multidimensional parameter space, it is presently not feasible to use such detailed ODE models in combination with a full uncertainty analysis. Nevertheless, with a quasi-steady state approach we were able to quantify the uncertainty in the feedback mechanisms.

Alongside the abovementioned advantages, the f-ISA modeling framework posseses several disadvantages as well. First, the model is restricted to known mechanisms. It is possible that other links provide an equally good fit. For example, there could be other negative feedbacks ending in or upstream of MEK and AKT, or the involvement of other phosphorylation sites on IRS1. Our approach does not search for additional links that may fit the data better. Rather, the goal is to test whether a particular model can describe the data well. Unfortunately, combining extensive parameter uncertainty analysis with a search for network topology is computationally intractable at present. Second, the calculation of signaling with feedback events as done here is significantly slower than the original ISA modeling approach without feedback. It is therefore still impractical to infer signaling activities for large models with feedback while also including full dose response curves, especially with multiple drugs. For the dose response data presented in Jastrzebski *et al* [27], this would add another 300 model conditions per drug, as well as additional model parameters. It may be feasible to perform inference sequentially, thereby reducing the number of model evaluations required [52]. Alternatively, evaluating the gradient of the likelihood and using sampling algorithms that can leverage this gradient information may speed up the inference [53,54].

A third limitation is that f-ISA is currently only able to handle feedback events which occur on one timescale. We focused here on fast-acting post-translational feedback. This approach could also be used to study transcriptional regulation; in this case total protein or mRNA levels could be used as inference data instead of phosphorylation data. It is also possible to include both transcriptional and post-translational feedback in the same model, but this would assume that the feedbacks occur on approximately the same timescale, which is probably unreasonable. To disentangle feedback mechanisms acting over different timeframes, it would be necessary to add further extensions that can calculate steady states with feedback loops occurring on multiple timescales, or it may be more beneficial to switch to dynamic models in this case.

Despite these limitations, we believe f-ISA is a useful approach to delineate context-dependent activities in signaling networks with multiple feedback paths. By iteratively incorporating additional detail in the signaling networks, and

bioRxiv preprint doi: https://doi.org/10.1101/268359; this version posted February 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

additional measurements to constrain the parameters, it is possible to obtain an increasingly thorough, quantitative understanding of a regulatory signaling network.

Acknowledgments

We are grateful to Jordi Vidal Rodriguez for experimental support with the on-treatment, intervention RPPA experiment.

References

1. Groenendijk FH, Bernards R. Drug resistance to targeted therapies: Deja vu all over again. Mol Oncol. 2014;8: 1067–1083.
2. Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW, et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature. 2012;483: 570–5.
3. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin A a, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature. 2012;483: 603–7.
4. Seashore-Ludlow B, Rees MG, Cheah JH, Cokol M, Price E V., Coletti ME, et al. Harnessing Connectivity in a Large-Scale Small-Molecule Sensitivity Dataset. Cancer Discov. 2015;
5. Stelling J, Sauer U, Szallasi Z, Doyle FJ, Doyle J. Robustness of cellular functions. Cell. 2004;118: 675–685.
6. Dougherty MK, Müller J, Ritt D a, Zhou M, Zhou XZ, Copeland TD, et al. Regulation of Raf-1 by direct feedback phosphorylation. Mol Cell. 2005;17: 215–24.
7. Fritsche-Guenther R, Witzel F, Sieber A, Herr R, Schmidt N, Braun S, et al. Strong negative feedback from Erk to Raf confers robustness to MAPK signalling. Mol Syst Biol. 2011;7: 489.
8. Shah OJ, Wang Z, Hunter T. Inappropriate activation of the TSC/Rheb/mTOR/S6K cassette induces IRS1/2 depletion, insulin resistance, and cell survival deficiencies. Curr Biol. 2004;14: 1650–6.
9. Liu P, Gan W, Inuzuka H, Lazorchak AS, Gao D, Arojo O, et al. Sin1 phosphorylation impairs mTORC2 complex integrity and inhibits downstream Akt signalling to suppress tumorigenesis. Nat Cell Biol. 2013;15: 1340–50.
10. Yang G, Murashige DS, Humphrey SJ, James DE. A Positive Feedback Loop between Akt and mTORC2 via SIN1 Phosphorylation. Cell Rep. 2015;12: 937–943.
11. Humphrey SJ, Yang G, Yang P, Fazakerley DJ, Stöckli J, Yang JY, et al. Dynamic adipocyte phosphoproteome reveals that akt directly regulates mTORC2. Cell Metab. 2013;17: 1009–1020.
12. Copps KD, White MF. Regulation of insulin sensitivity by serine/threonine phosphorylation of insulin receptor substrate proteins IRS1 and IRS2. Diabetologia. 2012;55: 2565–2582.
13. Lee YH, Giraud J, Davis RJ, White MF. c-Jun N-terminal kinase (JNK) mediates feedback inhibition of the insulin signaling cascade. J Biol Chem. 2003;278: 2896–2902.
14. Carlson CJ, White MF, Rondinone CM. Mammalian target of rapamycin regulates IRS-1 serine 307 phosphorylation. Biochem Biophys Res Commun. 2004;316: 533–539.
15. Shah OJ, Hunter T. Turnover of the Active Fraction of IRS1 Involves Raptor-mTOR- and S6K1-Dependent Serine Phosphorylation in Cell Culture Models of Tuberous Sclerosis. Mol Cell Biol. 2006;26: 6425–6434.
16. Zhang J, Gao Z, Yin J, Quon MJ, Ye J. S6K directly phosphorylates IRS-1 on Ser-270 to promote insulin resistance in response to TNF-α signaling through IKK2. J Biol Chem. 2008;283: 35375–35382.
17. The UniProt Consortium. UniProt: The universal protein knowledgebase. Nucleic Acids Res. 2016;45: D158–D169.
18. Hornbeck P V., Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E. PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations. Nucleic Acids Res. 2015;43: D512–D520.
19. Hartley D, Cooper GM. Role of mTOR in the degradation of IRS-1: Regulation of PP2A activity. J Cell Biochem. 2002;85: 304–314.
20. Wan X, Harkavy B, Shen N, Grohar P, Helman LJ. Rapamycin induces feedback activation of Akt signaling through an IGF-1R-dependent mechanism. Oncogene. 2007;26: 1932–1940.
21. Okuzumi T, Fiedler D, Zhang C, Gray DC, Aizenstein B, Hoffman R, et al. Inhibitor hijacking of Akt activation. Nat Chem Biol. 2009;5: 484–493.
22. Han EKH, Leverson JD, McGonigal T, Shah OJ, Woods KW, Hunter T, et al. Akt inhibitor A-443654 induces rapid Akt Ser-473 phosphorylation independent of mTORC1 inhibition. Oncogene. 2007;26: 5655–5661.
23. Kholodenko BN. Cell-signalling dynamics in time and space. Nat Rev Mol Cell Biol. 2006;7: 165–76.

bioRxiv preprint doi: https://doi.org/10.1101/268359; this version posted February 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

24. Eduati F, Doldàn-Martelli V, Klinger B, Cokelaer T, Sieber A, Kogera F, et al. Drug resistance mechanisms in colorectal cancer dissected with cell type-specific dynamic logic models. Cancer Res. 2017; canres.0078.2017.
25. Kholodenko BN, Kiyatkin A, Bruggeman FJ, Sontag E, Westerhoff H V, Hoek JB. Untangling the wires: A strategy to trace functional interactions in signaling and gene networks. Proc Natl Acad Sci. 2002;99: 12841–12846.
26. Stelniec-Klotz I, Legewie S, Tchernitsa O, Witzel F, Klinger B, Sers C, et al. Reverse engineering a hierarchical regulatory network downstream of oncogenic KRAS. Mol Syst Biol. 2012;8: 601.
27. Jastrzebski K, Thijssen B, Kluin RJC, Lint K de, Majewski IJ, Beijersbergen RL, et al. Integrative modeling identifies key determinants of inhibitor sensitivity in breast cancer cell lines. In Preparation.
28. Xu T-R, Vyshemirsky V, Gormand A, von Kriegsheim A, Girolami M, Baillie GS, et al. Inferring signaling pathway topologies from multiple perturbation measurements of specific biochemical species. Sci Signal. 2010;3: ra20.
29. Thijssen B, Dijkstra TMH, Heskes T, Wessels LFA. Bayesian data integration for quantifying the contribution of diverse measurements to parameter estimates. Bioinformatics. 2017;
30. Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MPH. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. J R Soc Interface. 2009;6: 187–202.
31. Saez-Rodriguez J, Alexopoulos L, Zhang M, Morris MK, Lauffenburger DA, Sorger PK. Comparing signaling networks between normal and transformed hepatocytes using discrete logical models. Cancer Res. 2011;71: 5400–5411.
32. Schlatter R, Schmich K, Avalos Vizcarra I, Scheurich P, Sauter T, Borner C, et al. ON/OFF and beyond--a boolean model of apoptosis. PLoS Comput Biol. 2009;5: e1000595.
33. Tarjan R. Depth-First Search and Linear Graph Algorithms. SIAM J Comput. 1972;1: 146–160.
34. Coombes KR, Neeley S, Joy C, Hu J, Baggerly K, Roebuck P. SuperCurve: RPPA Analysis Package. 2017.
35. Funahashi BA, Matsuoka Y, Jouraku A, Morohashi M, Kikuchi N, Kitano H. CellDesigner 3.5: A Versatile Modeling Tool for Biochemical Networks. Proc IEEE. 2008;96: 1254–1265.
36. Le Novère N, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, et al. The Systems Biology Graphical Notation. Nat Biotechnol. 2009;27: 735–742.
37. Thijssen B, Dijkstra TMH, Heskes T, Wessels LFA. BCM: toolkit for Bayesian analysis of Computational Models using samplers. BMC Syst Biol. 2016;10: 100.
38. Katzgraber HG, Trebst S, Huse DA, Troyer M. Feedback-optimized parallel tempering Monte Carlo. J Stat Mech Theory Exp. 2006;2006: P03018.
39. Turek D, de Valpine P, Paciorek CJ, Anderson-Bergman C. Automated Parameter Blocking for Efficient Markov-Chain Monte Carlo Sampling. Bayesian Anal. 2017;12: 465–490.
40. Gelman A, Meng X-L. Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. Stat Sci. 1998;13: 163–185.
41. MacKay DJC. Model Comparison and Occam's Razor. Information Theory, Inference, and Learning Algorithms. 2003. pp. 343–354.
42. Korkola JE, Collisson EA, Heiser L, Oates C, Bayani N, Itani S, et al. Decoupling of the PI3K Pathway via Mutation Necessitates Combinatorial Treatment in HER2+ Breast Cancer. PLoS One. 2015;10: e0133219.
43. Ohren JF, Chen H, Pavlovsky A, Whitehead C, Zhang E, Kuffa P, et al. Structures of human MAP kinase kinase 1 (MEK1) and MEK2 describe novel noncompetitive kinase inhibition. Nat Struct Mol Biol. 2004;11: 1192–1197.
44. Barrett SD, Bridges AJ, Dudley DT, Saltiel AR, Fergus JH, Flamme CM, et al. The discovery of the benzhydroxamate MEK inhibitors CI-1040 and PD 0325901. Bioorganic Med Chem Lett. 2008;18: 6501–6504.
45. Lito P, Saborowski A, Yue J, Solomon M, Joseph E, Gadal S, et al. Disruption of CRAF-Mediated MEK Activation Is Required for Effective MEK Inhibition in KRAS Mutant Tumors. Cancer Cell. 2014;25: 697–710.
46. Reuveni H, Flashner-Abramson E, Steiner L, Makedonski K, Song R, Shir A, et al. Therapeutic destruction of insulin receptor substrates for cancer treatment. Cancer Res. 2013;73: 4383–4394.
47. Ibuki N, Ghaffari M, Reuveni H, Pandey M, Fazli L, Azuma H, et al. The Tyrphostin NT157 Suppresses Insulin Receptor Substrates and Augments Therapeutic Response of Prostate Cancer. Mol Cancer Ther. 2014;13: 2827–2839.
48. Pende M, Um SH, Mieulet V, Goss VL, Mestan J, Mueller M, et al. S6K1-/-/S6K2-/- Mice Exibit Perinatal Lethality and Rapamycin-Sensitive 5'-Terminal Oligopyrimidine mRNA Translation and Reveal a Mitogen-

bioRxiv preprint doi: https://doi.org/10.1101/268359; this version posted February 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Activated Protein Kinase-Dependent S6 Kinase Pathway. Mol Cell Biol. 2004;24: 3112–3124.
49. Roux PP, Shahbazian D, Vu H, Holz MK, Cohen MS, Taunton J, et al. RAS/ERK signaling promotes site-specific ribosomal protein S6 phosphorylation via RSK and stimulates cap-dependent translation. J Biol Chem. 2007;282: 14056–64.
50. Thobe K, Sers C, Siebert H. Unraveling the regulation of mTORC2 using logical modeling. Cell Commun Signal. 2017;15: 6.
51. Dalle Pezze P, Sonntag AG, Thien A, Prentzell MT, Godel M, Fischer S, et al. A Dynamic Network Model of mTOR Signaling Reveals TSC-Independent mTORC2 Regulation. Sci Signal. 2012;5: ra25-ra25.
52. Thijssen B, Wessels LFA. Approximating multivariate posterior distribution functions from Monte Carlo samples for sequential Bayesian inference. arXiv Prepr. 2017; arXiv:1712.04200.
53. Duane S, Kennedy ADD, Pendleton BJ, Roweth D. Hybrid Monte Carlo. Phys Lett B. 1987;55: 2774–2777.
54. Hoffman MD, Gelman A. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. J Mach Learn Res. 2014;15: 1593–1623.

Figure Legends

Figure 1: Signaling model and system decomposition. (A) The signaling model of MAPK and AKT signaling in breast cancer constructed for the purpose of these studies. It is based on a reduced, joint-drug model of Jastrzebski *et al* [27]. The model is depicted in Activity Flow format of the Systems Biology Graphical Notation. Yellow boxes indicate genetic events including mutations and copy number aberrations, green nodes are the signaling molecules, and purple nodes are the kinase inhibitors. (B) The activation function used to calculate the activity of a signaling molecule from its input. The red lines indicate the constraints which were taken to determine the fixed steepness and inflection point of the logistic function. The blue lines give an example; in this case the signaling molecule has two upstream inputs (x_1 and x_2), which are multiplied by the respective signal strengths (s_1 and s_2) and summed together with a basal activity (b) to give a final signaling activity of approximately 0.7. (C) A small part of the model shown in template notation, illustrating the latent variable structure and the integrated batch correction. The text names indicate the names in the code while the symbols in brackets correspond to the equations in the Methods section. Each dataset (indexed by k) has multiple cell lines or conditions (indexed by j). The signaling activities are unique for each condition. The likelihood of the data depends on two batch correction variables (base and scale) as well as the variance; these variables are specific for each dataset. The signaling strengths and base activities are global parameters, which are shared by all datasets and all conditions. (D) Decomposition of the model into smaller modules that have to be solved as systems of equations (the red boxes; one system of 3 nodes and one system of 6 nodes), and a topological ordering (the red numbers) giving the order in which to calculate the nodes and systems.

Figure 2: Identification of feedback activity from ERK to RAF from pre-treatment, intracellular signaling measurements. (A) Data and posterior predictive for three of the six epitopes. Black dots indicate the measurement data and the shaded blue area is the 90% confidence interval of the posterior predictive distribution. (B) Posterior probability density for the strength of ERK->RAF feedback signal inferred from the data. (C) Restricting the prior for the inputs to the MAPK pathway to low values results in a weaker RAF->ERK feedback.

Figure 3: Addition of RTK phosphorylation data increases the uncertainty in ERK to RAF signaling. (A) Data and posterior predictive of the two epitopes that were added. Note that there is cross-reactivity between the antibodies against EGFR p-Y1068 and ERBB2 p-Y1248, hence both measurements are a linear combination of the EGFR and ERBB2 signaling estimates. (B) Posterior probability densities of the ERK->RAF feedback loop, and the signaling estimates of EGFR and ERBB2, with and without the RTK data. For the signaling estimates, the shaded gray areas indicate the 90% confidence interval.

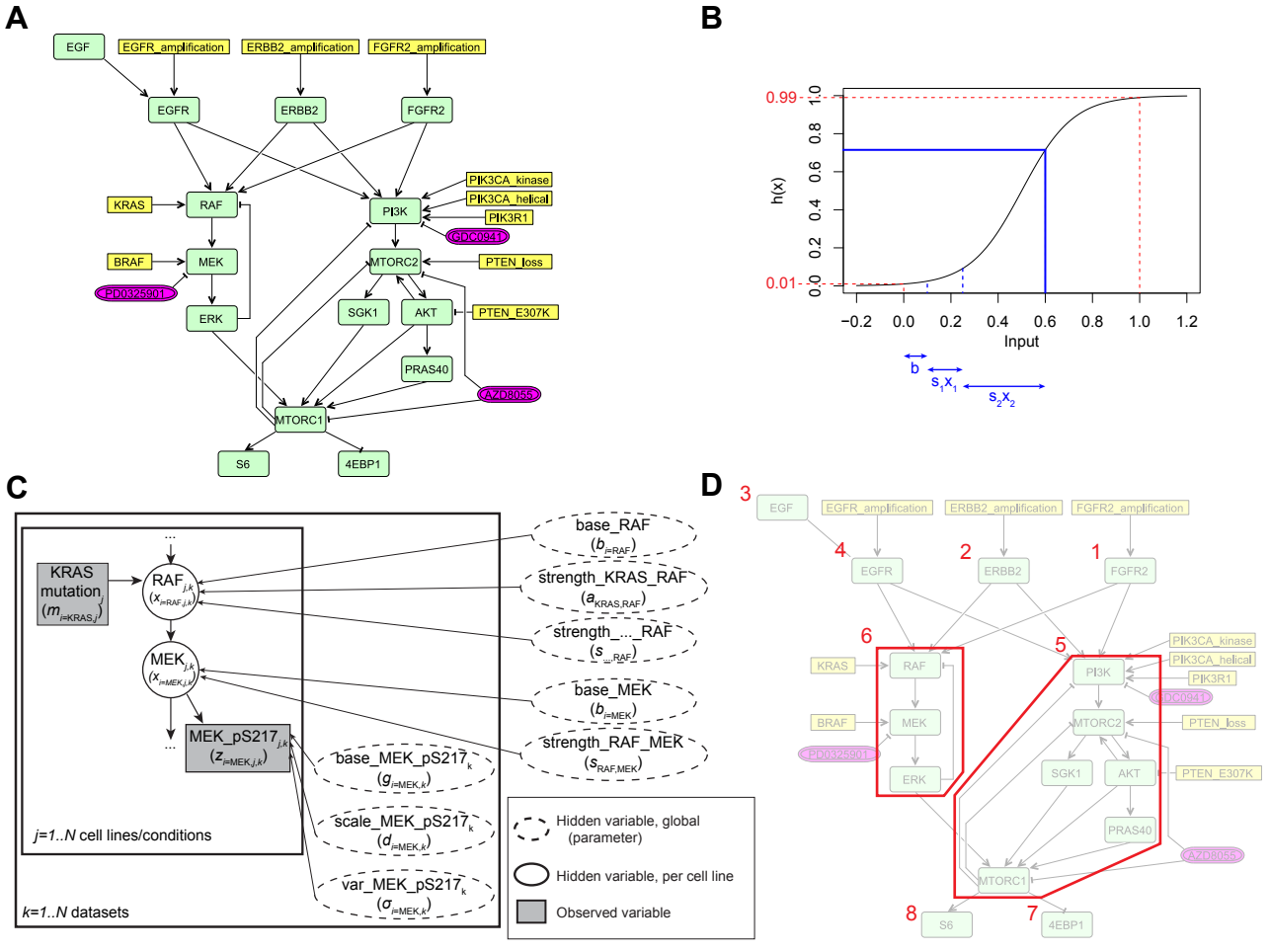
Figure 4: On-treatment data constrains the feedback from ERK to RAF. (A) Data and posterior predictive for three of the eight epitope added to the inference. The posterior predictive is obtained from inference with the pre-treatment and intervention datasets combined. (B) Comparison of densities of the ERK->RAF feedback loop and three other parameters with either dataset alone or together.

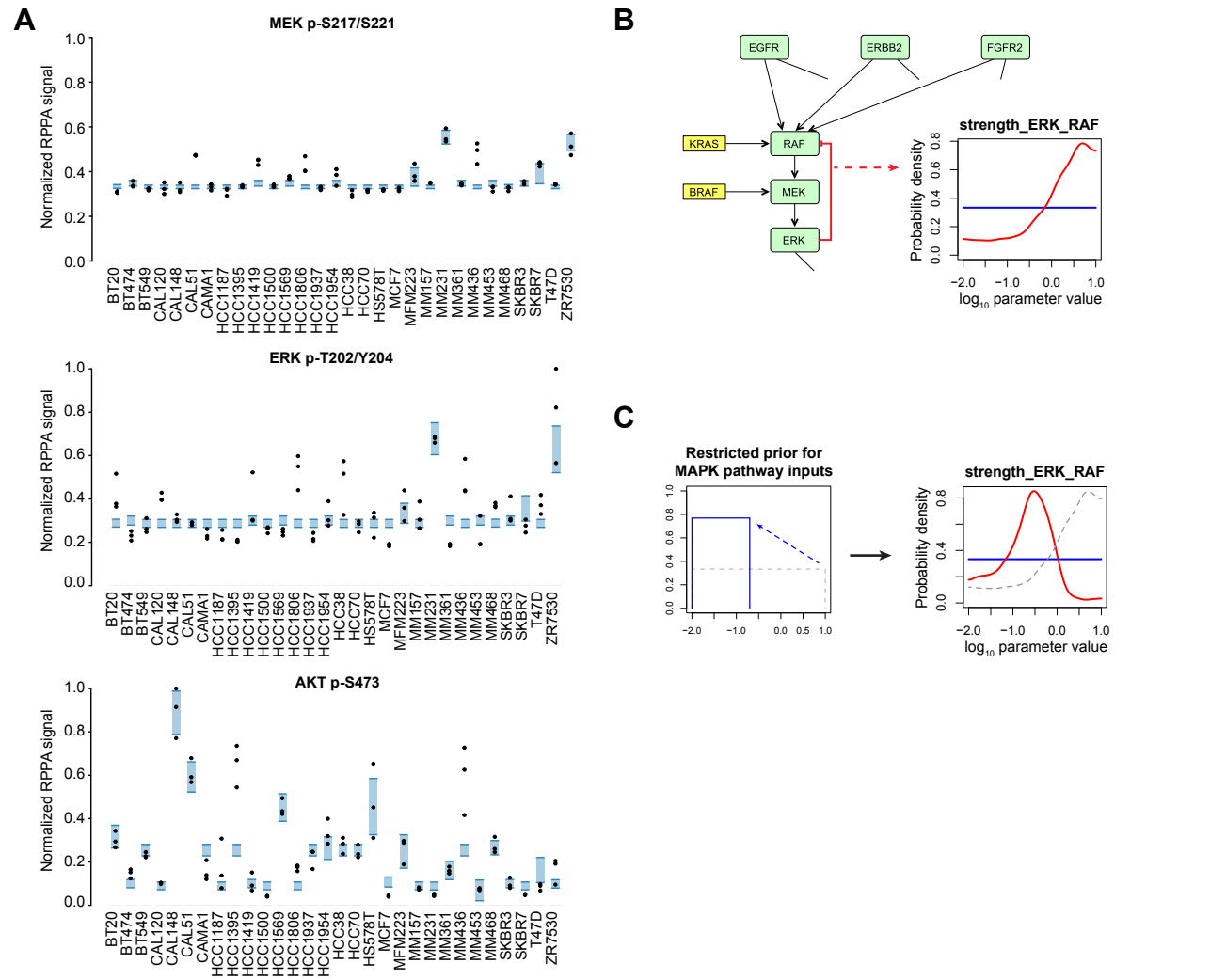
bioRxiv preprint doi: https://doi.org/10.1101/268359; this version posted February 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

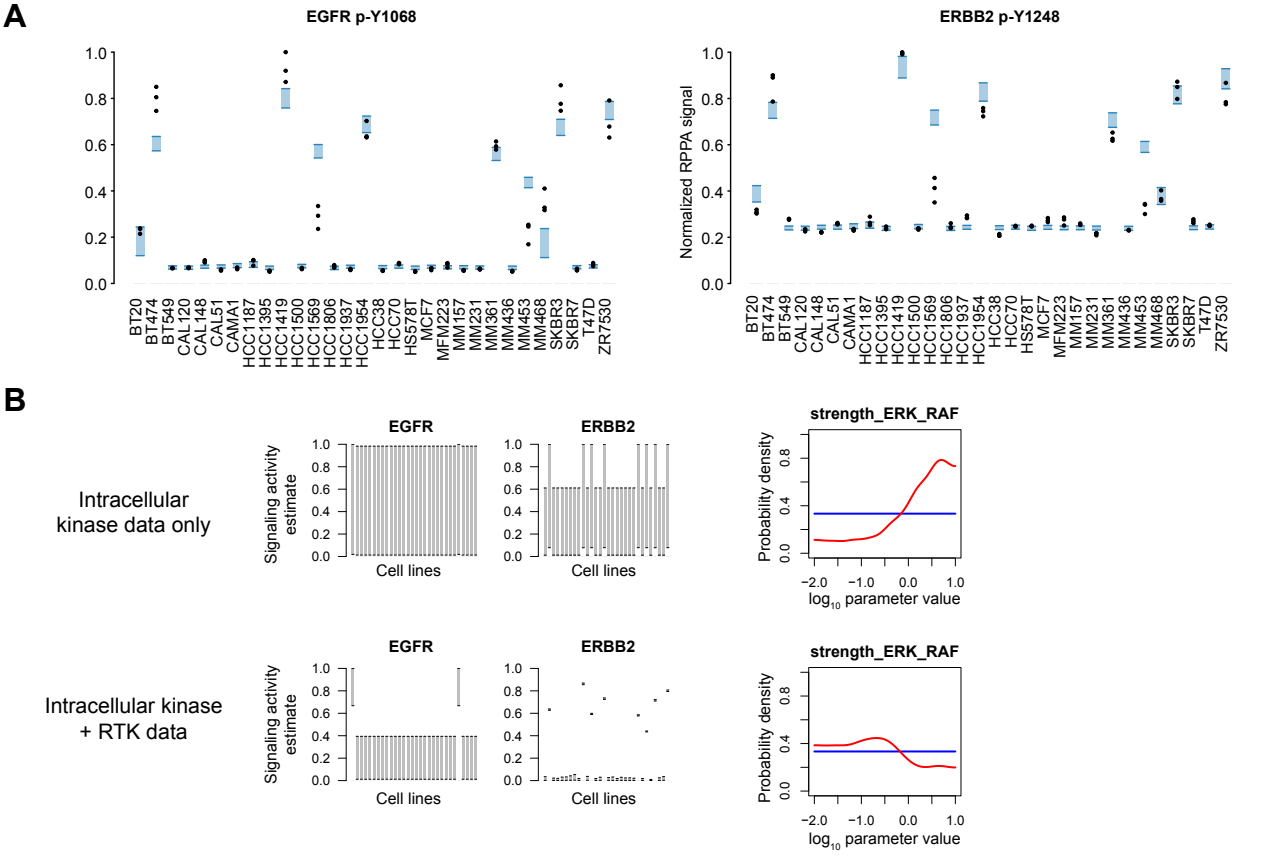
Figure 5: Feedback in the AKT pathway is likely to be negative. (A) Data and posterior predictive of AKT p-S473 in the pre-treatment dataset in the inference with the pre-treatment and intervention datasets combined; c.f. Figure 2A where the same data is shown but the intervention dataset was not included. (B) Feedback signaling strengths in the AKT pathway. The bottom right panel shows a scatter plot of the samples of the PI3K->MTORC2 and MTORC1->MTORC2 signal strengths. (C) Model with GSK-690693 as an additional inhibitor, and IRS1 explicitly included as a signaling molecule. The links for which the posterior is shown in E are highlighted in red. (D) Data and posterior predictive for three of the epitopes used for the inference; the data is the 1-hr time point of the dataset of Korkola *et al* [42]. The previous two datasets are included in the inference as well. (E) Posterior distribution of the feedback signaling events in the AKT pathway, given the three datasets together.

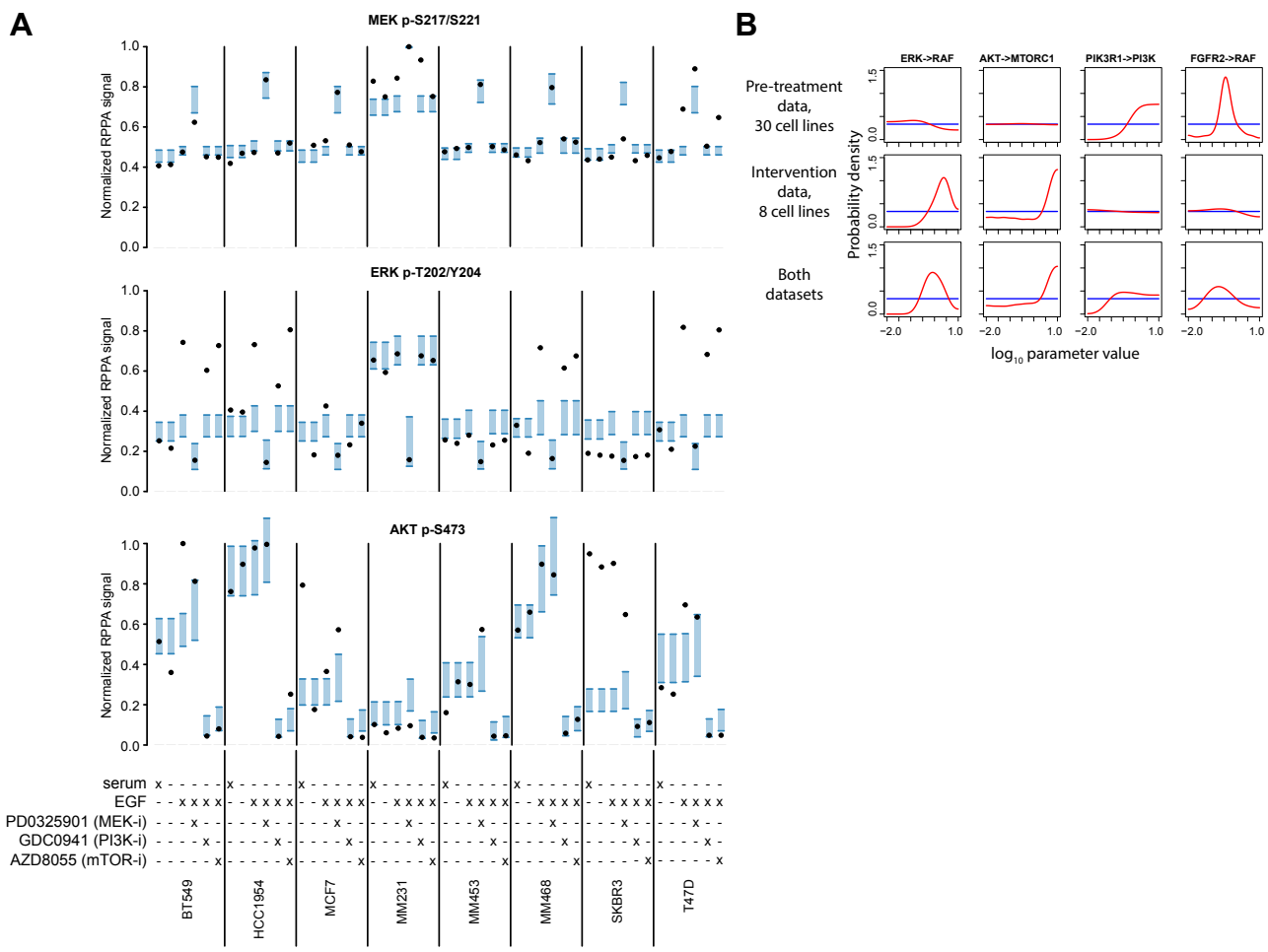
Figure 6: Resolving the feedback through IRS1. (A) BT474 and HCC1954 cell lines treated with the IRS1/2 inhibitor, NT-157 (NT), alone or in combination with either the AKT inhibitor, GSK690693 (GSK), or the S6K inhibitor, PF4708671 (PF), were analyzed by immunoblotting for expression and phosphorylation of key molecules in the AKT pathway. Control samples were treated with DMSO alone. (B) Model with S6K, SIN1 and PF4708671 added. The double, dashed arrows from SIN1 and MTORC2 indicate that the sign of this link is unknown. The links for which the posterior is shown in D are highlighted in red. (C) Quantification of the western blot signals shown in A, along with the posterior predictive of the fitted model. The previous three datasets are included in the inference as well. (D) Posterior distribution of the feedback signaling strengths in the AKT pathway, given the four datasets together. (E) Scatter plot of the posterior samples for the MTORC1->IRS1 and S6K->IRS1 signal strengths, with a 2D-KDE of the bivariate marginal posterior shown in contours.

Figure 7: Comparing the model fit and estimates across datasets. (A) Reduction in uncertainty for each of the 63 signaling parameters, given each dataset separately and together. Each dot indicates one parameter. Grey lines connect the same parameters between datasets. Some parameters are constrained by multiple datasets, while other parameters are uniquely constrained by one dataset. (B) Disagreement in parameter estimates between datasets; each bar represents on of the 63 signaling parameters. Disagreement is calculated as one minus the overlap coefficient, that is, $1 - IS / \min(CI_1, CI_2)$, where CI_i is size of the posterior 90% confidence interval given each dataset, and IS is the size of the intersection of the two CIs. Detailed figures with labels are included as Supplementary Figure 1. (C) Examples of two parameters for which the pretreatment dataset and the intervention dataset either agree (top panel) or disagree (bottom panel). (D) Scatter plot of all fitted data, represented by the median of the posterior predictive distribution, against the observed values.

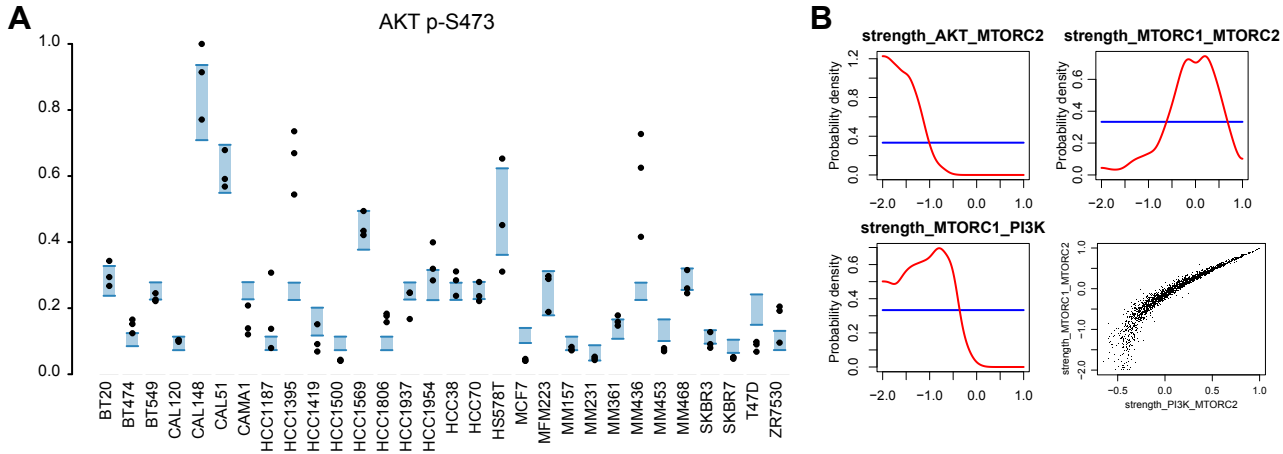

A
EGF
EGFR_amplification
ERBB2_amplification
FGFR2_amplification
EGFR
ERBB2
FGFR2
KRAS
RAF
PI3K
PIK3CA_kinase
PIK3CA_helical
PIK3R1
GDC0941
BRAF
MEK
MTORC2
PTEN_loss
PD0325901
ERK
SGK1
AKT
PTEN_E307K
PRAS40
AZD8055
MTORC1
S6
4EBP1
B
0.99
0.01
h(x)
0.0
0.2
0.4
0.6
0.8
1.0
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Input
b
s1x1
s2x2
C
KRAS mutation (mi=KRAS,j)
RAF j,k (xi=RAF,j,k)
MEK j,k (xi=MEK,j,k)
MEK_pS217 j,k (zi=MEK,j,k)
base_RAF (bi=RAF)
strength_KRAS_RAF (aKRAS,RAF)
strength_..._RAF (s..,RAF)
base_MEK (bi=MEK)
strength_RAF_MEK (sRAF,MEK)
base_MEK_pS217 k (gi=MEK,k)
scale_MEK_pS217 k (di=MEK,k)
var_MEK_pS217 k (σi=MEK,k)
j=1..N cell lines/conditions
k=1..N datasets
Hidden variable, global (parameter)
Hidden variable, per cell line
Observed variable
D
3
4
2
1
6
5
8
7


A
MEK p-S217/S221
ERK p-T202/Y204
AKT p-S473
Normalized RPPA signal
0.0
0.2
0.4
0.6
0.8
1.0
BT20
BT474
BT549
CAL120
CAL148
CAL51
CAMA1
HCC1187
HCC1395
HCC1419
HCC1500
HCC1569
HCC1806
HCC1937
HCC1954
HCC38
HCC70
HS578T
MCF7
MFM223
MM157
MM231
MM361
MM436
MM453
MM468
SKBR3
SKBR7
T47D
ZR7530
B
EGFR
ERBB2
FGFR2
KRAS
RAF
BRAF
MEK
ERK
strength_ERK_RAF
Probability density
log10 parameter value
C
Restricted prior for
MAPK pathway inputs
strength_ERK_RAF
Probability density
log10 parameter value

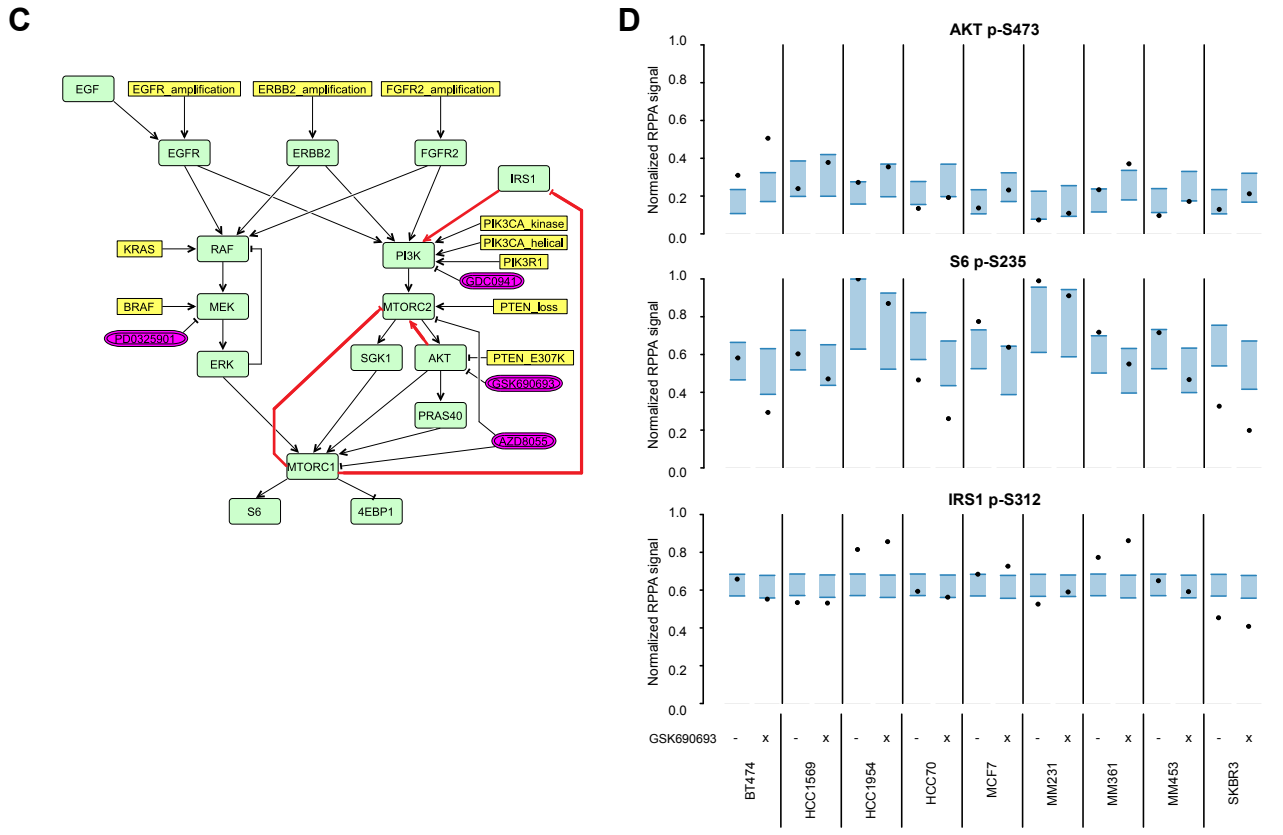

A
EGFR p-Y1068
ERBB2 p-Y1248
Normalized RPPA signal
1.0
0.8
0.6
0.4
0.2
0.0
BT20
BT474
BT549
CAL120
CAL148
CAL51
CAMA1
HCC1187
HCC1395
HCC1419
HCC1500
HCC1569
HCC1806
HCC1937
HCC1954
HCC38
HCC70
HS578T
MCF7
MFM223
MM157
MM231
MM361
MM436
MM453
MM468
SKBR3
SKBR7
T47D
ZR7530
B
Intracellular kinase data only
Intracellular kinase + RTK data
EGFR
ERBB2
Signaling activity estimate
Cell lines
strength_ERK_RAF
Probability density
log10 parameter value
−2.0
−1.0
0.0
1.0

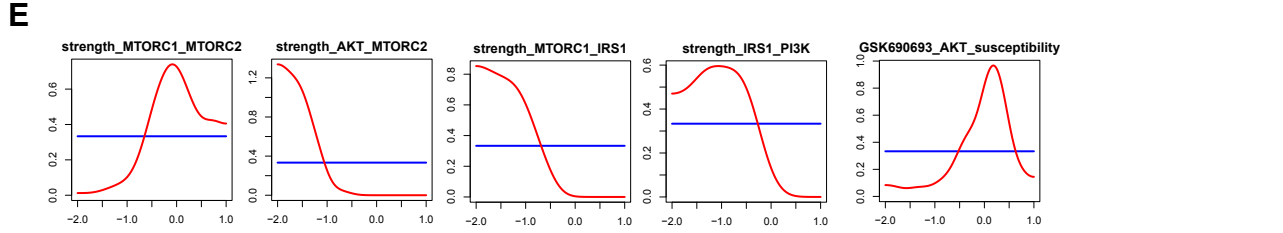

A
MEK p-S217/S221
ERK p-T202/Y204
AKT p-S473
Normalized RPPA signal
1.0
0.8
0.6
0.4
0.2
0.0
serum
EGF
PD0325901 (MEK-i)
GDC0941 (PI3K-i)
AZD8055 (mTOR-i)
BT549
HCC1954
MCF7
MM231
MM453
MM468
SKBR3
T47D
B
Pre-treatment data, 30 cell lines
Intervention data, 8 cell lines
Both datasets
ERK->RAF
AKT->MTORC1
PIK3R1->PI3K
FGFR2->RAF
Probability density
1.5
0.0
−2.0
1.0
log10 parameter value

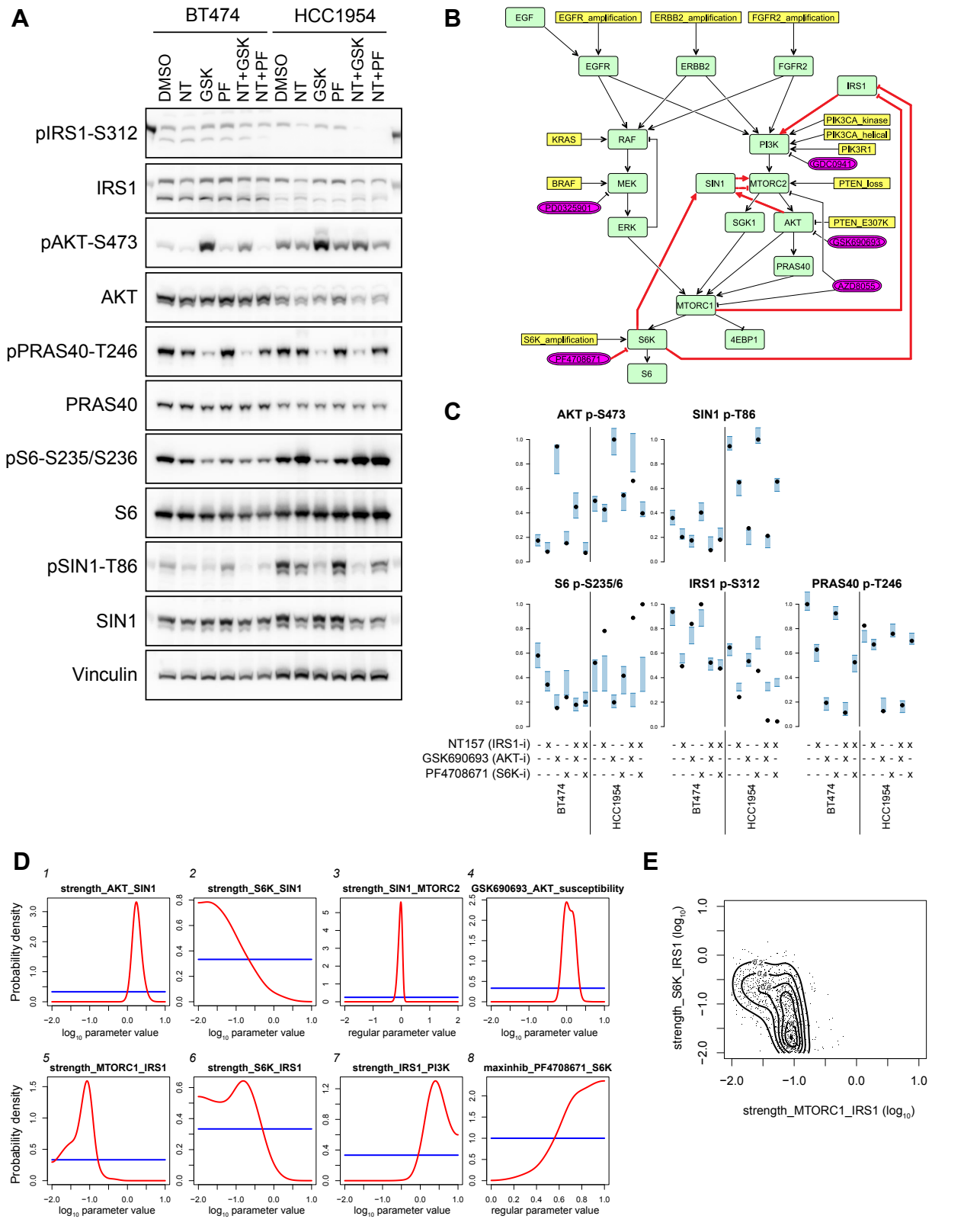
Pre-treatment + intervention data

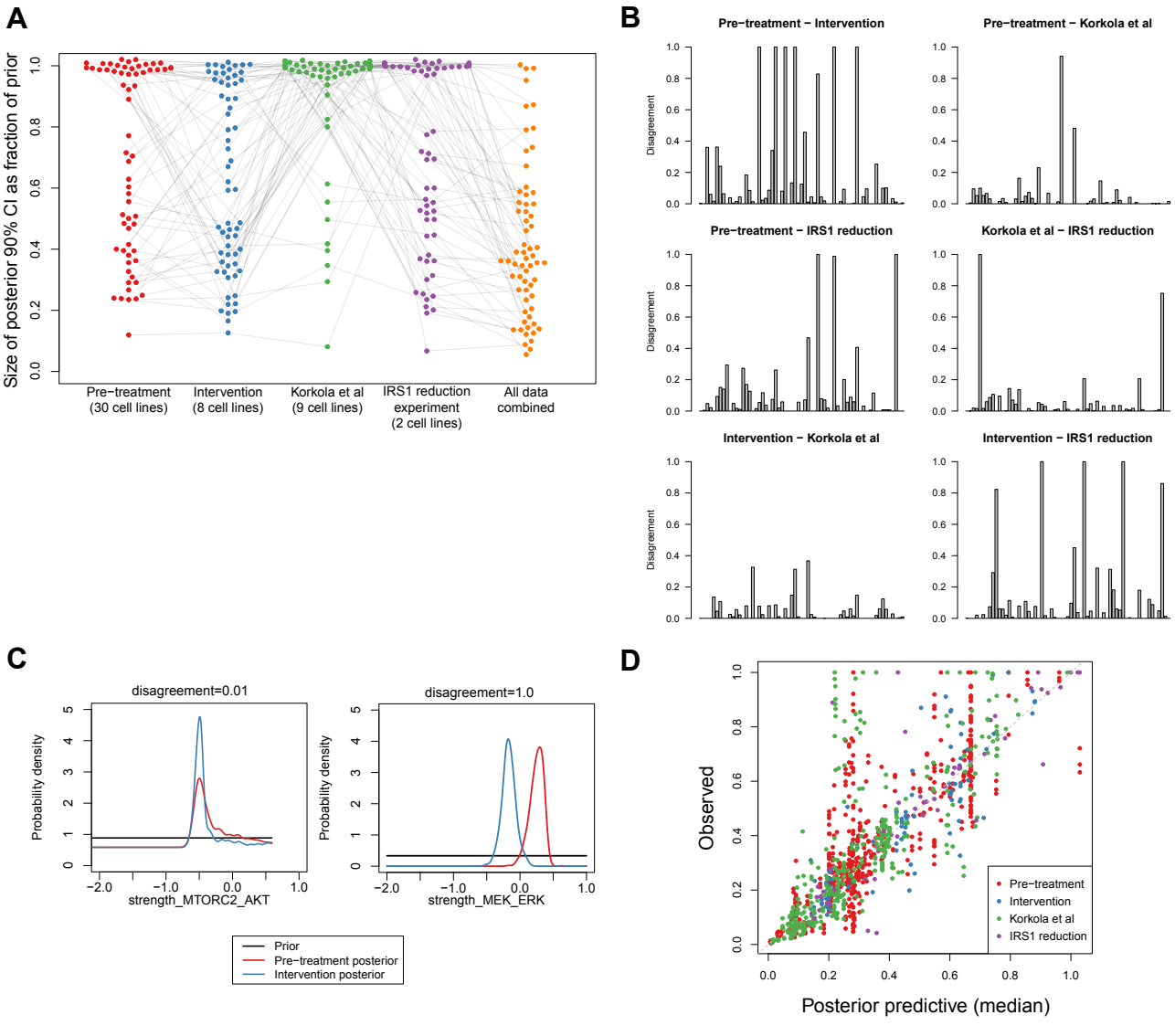


Pre-treatment + intervention + Korkola *et al*. data






A
BT474
HCC1954
DMSO
NT
GSK
PF
NT+GSK
NT+PF
pIRS1-S312
IRS1
pAKT-S473
AKT
pPRAS40-T246
PRAS40
pS6-S235/S236
S6
pSIN1-T86
SIN1
Vinculin
B
EGF
EGFR_amplification
ERBB2_amplification
FGFR2_amplification
EGFR
ERBB2
FGFR2
IRS1
PIK3CA_kinase
PIK3CA_helical
PIK3R1
KRAS
RAF
PI3K
GDC0941
BRAF
MEK
SIN1
MTORC2
PTEN_loss
PD0325901
ERK
SGK1
AKT
PTEN_E307K
GSK690693
PRAS40
AZD8055
MTORC1
S6K_amplification
S6K
4EBP1
PF4708671
S6
C
AKT p-S473
SIN1 p-T86
S6 p-S235/6
IRS1 p-S312
PRAS40 p-T246
NT157 (IRS1-i)
GSK690693 (AKT-i)
PF4708671 (S6K-i)
BT474
HCC1954
D
1 strength_AKT_SIN1
2 strength_S6K_SIN1
3 strength_SIN1_MTORC2
4 GSK690693_AKT_susceptibility
5 strength_MTORC1_IRS1
6 strength_S6K_IRS1
7 strength_IRS1_PI3K
8 maxinhib_PF4708671_S6K
Probability density
log10 parameter value
regular parameter value
E
strength_S6K_IRS1 (log10)
strength_MTORC1_IRS1 (log10)


A
Size of posterior 90% CI as fraction of prior
Pre−treatment (30 cell lines)
Intervention (8 cell lines)
Korkola et al (9 cell lines)
IRS1 reduction experiment (2 cell lines)
All data combined
B
Pre−treatment − Intervention
Pre−treatment − Korkola et al
Pre−treatment − IRS1 reduction
Korkola et al − IRS1 reduction
Intervention − Korkola et al
Intervention − IRS1 reduction
Disagreement
C
disagreement=0.01
disagreement=1.0
Probability density
strength_MTORC2_AKT
strength_MEK_ERK
Prior
Pre−treatment posterior
Intervention posterior
D
Observed
Posterior predictive (median)
Pre−treatment
Intervention
Korkola et al
IRS1 reduction