Liebig's law of the minimum in the TGF-β pathway

Yuchao Li[1,#], Difan Deng[2], Chris Tina Höfer[2], Jihye Kim[3], Won Do Heo[3], Quanbin Xu[4], Xuedong Liu[4], Zhike Zi[1,2,5, #,*]

1 Max Planck Institute for Molecular Genetics, Otto Warburg Laboratory, 14195 Berlin, Germany

2 German Federal Institute for Risk Assessment, Department of Experimental Toxicology and ZEBET, 10589 Berlin, Germany

3 Department of Biological Sciences, KAIST Institute for the BioCentury, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, 305701 Republic of Korea

4 Department of Biochemistry, 3415 Colorado Ave, University of Colorado Boulder, Boulder, Colorado 80303-0596, United States

5 CAS Key Laboratory of Quantitative Engineering Biology, Shenzhen Institute of Synthetic Biology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

# These authors contributed equally to this work

* Corresponding author: Zhike Zi

Abstract

Cells use signaling pathways to sense and respond to their environments. The transforming growth factor-β (TGF-β) pathway produces context-specific responses. Here, we combined modeling and experimental analysis to study the dependence of the output of the TGF-β pathway on the abundance of signaling molecules in the pathway. We showed that the TGF-β pathway processes the variation of TGF-β receptor abundance using Liebig's law of the minimum, meaning that the output-modifying factor is the signaling protein that is most limited, to determine signaling responses across cell types and in single cells. We found that the abundance of either the type I (TGFBR1) or type II (TGFBR2) TGF-β receptor determined the responses of cancer cell lines, such that the receptor with relatively low abundance dictates the response. Furthermore, nuclear SMAD2 signaling correlated with the abundance of TGF-β receptor in single cells depending on the relative expression levels of TGFBR1 and TGFBR2. A similar control principle could govern the heterogeneity of signaling responses in other signaling pathways.

Keywords: systems biology; mathematical modeling; TGF-β pathway; cellular heterogeneity; synthetic biology

Introduction

Cells sense and respond to their environment through signaling pathways. These pathways convert extracellular factors into signaling activities, resulting in the responses such as regulation of gene expression and determination of cell fate. Heterogeneous signaling responses to the same environmental cue often occur in different cell types and in individual cells. Intrinsic noise and extrinsic noise contribute to cell-to-cell variability (Eldar & Elowitz, 2010; Eling *et al*, 2019; Losick & Desplan, 2008; Mitchell & Hoffmann, 2018; Niepel *et al*, 2009; Snijder & Pelkmans, 2011). Intrinsic noise reflects the stochastic fluctuations of biochemical reactions, whereas extrinsic noise originates from differences in protein abundance, cell cycle stage, and other parameters. Cellular protein abundances have a large dynamic range with several orders of magnitude (Buccitelli & Selbach, 2020). Variation in protein abundance affects the heterogeneity of signaling responses (Cohen-Saidon *et al*, 2009; Filippi *et al*, 2016; Gabor *et al*, 2021; Kramer *et al*, 2022; Lun *et al*, 2017; Strasen *et al*, 2018; Wu *et al*, 2021).

The transforming growth factor-β (TGF-β) pathway regulates many fundamental processes, including cell proliferation, differentiation, and migration (Shi & Massague, 2003). The effects of TGF-β depend on the cellular context (Massague, 2012; Zhang, 2018), providing a useful system to investigate the relationship between signaling protein abundance and signaling output. The core of the TGF-β pathway is composed of TGF-β receptors (TGFBR1, TGFBR2) that perceive the ligand TGF-β, receptor-regulated SMADs (R-SMADs: SMAD2, SMAD3) and common SMAD (Co-SMAD: SMAD4) effectors that mediate the transcriptional response, and inhibitory SMAD (I-SMAD: SMAD7). The abundance of the core components and feedback regulators could determine TGF-β responses by regulating SMAD signaling in the nucleus (Derynck & Budi, 2019; Massague, 2012). Heterogeneous SMAD signaling responses in single cells can arise from the variability in protein levels (Frick *et al*, 2017; Strasen *et al.*,

2018). For the bone morphogenetic protein (BMP) pathway, diverse response profiles depend on the combinations of BMP ligands with different receptor binding affinities (Antebi *et al*, 2017; Klumpe *et al*, 2022; Mattiazzi Usaj *et al*, 2021). Manipulation of the abundance of BMP receptor subunits reprograms response profiles for different BMP ligand pairs (Antebi *et al.*, 2017). The same TGF-β ligand can initiate different signaling responses depending on the cellular context, but the underlying control principle remains unclear.

In this work, we examined how differences in signaling protein abundance dictates TGF-β signaling response profiles in different cell lines and in single cells using a combination of computational and experimental approaches. Computational modeling revealed that SMAD signaling responses can be determined by the abundance of the type I or type II TGF-β receptor. The type of TGF-β receptor present in relatively low numbers played a dominant role in shaping signaling responses, which resembles Liebig's law of the minimum (Liebig, 1840). Further experiments showed that different cell types have distinct response sensitivities to the variation of TGF-β receptors, depending on the relative expressions of TGFBR1 and TGFBR2. In addition, we found that the nuclear SMAD2 signaling strongly correlates with the expression of either of the TGF-β receptors in single cells, depending on the expression pattern of the receptors. Together, these results revealed an effect of the minimum control in the TGF-β pathway, which may be an important principle of control in signaling pathways with context-dependent outputs.

Results

TGF-β receptors display heterogeneous expression profiles across diverse cell types

Utilizing the Human Protein Atlas data (Karlsson *et al*, 2021), we first examined the transcript expression profiles of TGF-β receptors across various cell types. Among 1055 cell lines, both type I (TGFBR1) and type II (TGFBR2) TGF-β receptors displayed large variations (Figure 1A). The TGFBR1-to-TGFBR2 expression ratios were also heterogeneous in these cell lines (Figure 1B). Numerous cancer cell lines had imbalanced expression levels of TGFBR1 and TGFBR2. For instance, RH-30, TE-4 and NCI-H524 considerably higher levels of TGFBR1 compared to TGFBR2, while HepG2, RT-4, and KP-4 had elevated TGFBR2 expression relative to TGFBR1.

A minimal model predicts Liebig's law of the minimum in the TGF-β pathway

The heterogeneous expression profile of TGF-β receptors inspired us to analyze how different types of cells to process the variations of TGF-β receptors. To address this question, we first developed a minimal mathematical model that describes the key aspects of the TGF-β pathway with two reactions (Figure 2A, Supplementary Information). The first reaction represents the binding of the TGF-β ligand (L) to type I (R_1) and type II (R_2) TGF-β receptors, which forms the ligand-receptor complex (LRC). The second reaction describes the phosphorylation and dephosphorylation of the SMAD protein. As seen in the mathematical models of the BMP pathway (Antebi *et al.*, 2017; Su *et al*, 2022), the minimal model assumes that the phosphorylation of the SMAD protein is proportional to the activated ligand-receptor complex and the total amount of the species in the model remains constant (conservation of mass).

Using this simple model, we investigated the impact of TGF-β receptor variations on steady-state TGF-β signaling activity (S_a), which has an analytic solution (Supplementary

Information). To quantitively characterize the signal response, we analyzed the signaling activity (S_a) with respect to a 2-dimensional matrix of TGFBR1 and TGFBR2 numbers spanning 5 magnitudes. The TGF-β signaling activity profile revealed a notable pattern: signaling activity was primarily dictated by the receptor with lower abundance (Figure 2B). When TGFBR1 and TGFBR2 were equally expressed, TGF-β signaling activity was proportional to the receptor numbers (the dashed line in Figure 2B). Interestingly, when the amounts of TGFBR1 and TGFBR2 were imbalanced (denoted as stars in Figure 2B), TGF-β signaling is sensitive to changes of the TGF-β receptor with lower abundance, while remaining robust to changes of the TGF-β receptor with higher abundance. This observation aligns with Liebig's law of the minimum, which states that the growth rate of a species is determined by the scarcest resource (Liebig, 1840). It is worth noting that similar response profiles were predicted using different parameter values in the minimal model (Figure S1).

An extended model further demonstrates that TGF-β signaling is governed by the type of TGF-β receptor in lower abundance

For simplicity, the above-mentioned minimal model made several assumptions by neglecting specific features of the TGF-β pathway such as receptor internalization and recycling, nuclear import and export of SMAD protein, and the feedback regulations. Consequently, it remains unclear whether the predicted TGF-β signaling response profiles holds true in a more intricate TGF-β pathway when accounting for additional signaling processes and regulations. To tackle this issue, we developed an extended mathematical model that characterizes TGF-β signaling processes with greater complexity, including ligand-receptor interaction, receptor endocytosis, SMAD2 nucleocytoplasmic shuttling, and negative feedback regulation (Figure 3A). Building upon our previous model (Zi *et al*, 2011), the extended model employs mass action-based

ordinary differential equations to describe the dynamics of the TGF-β pathway (Supplementary Information).

To parameterize the extended model, we conducted quantitative immunoblotting experiments using HaCaT cells. These experiments allowed us to measure the absolute abundance of SMAD2 and TGFBR2 (Figure S2), the rate of TGFBR2 internalization (Figure S3), and the half-life of SMAD2, TGFBR1, and TGFBR2. Using these experimental data, we determined the initial conditions and some model parameter values (Supplementary Information). To constrain the remaining nine unknown parameters, we fitted the model to 102 data points obtained from HaCaT cells exposed to varying concentrations of TGF-β. Notably, the extended model simulations accurately reproduced the experimental data, encompassing the dynamics of TGF-β in the medium, overall SMAD2 phosphorylation (P-SMAD2), cytoplasmic SMAD2 and P-SMAD2, as well as nuclear SMAD2 and P-SMAD2 (Figure S4). Profile likelihood analysis confirmed the practical identifiability of the estimated model parameters (Figure S5). Consequently, the extended model was rigorously refined with meticulous parameterization.

We next conducted simulations on the extended model to investigate the variations in signaling components among different types of cells. Specifically, we simultaneously varied the abundances of TGF-β receptors and SMAD2 in the extended model by randomly sampling from multi-dimensional spaces (Figure 3B). By repeating the model simulations and analyzing the SMAD2 signaling output for 100,000 variations of TGF-β receptors and SMAD2, we observed interesting findings. Our results showed that the abundance of TGF-β receptors, rather than SMAD2, played a predominant role in controlling the variations of relative P-SMAD2 (the amount of P-SMAD2 normalized to SMAD2 abundance) and normalized nuclear SMAD2 signaling (SMAD2 N2C fold change) (Figure 3C, Figure S6A).

To gain further insights into the specific patterns of SMAD signaling response resulting from different TGF-β receptor abundances, we plotted the two SMAD2 signaling outputs in relation

to the abundance of TGFBR1 and TGFBR2 (Figure 3D, Figure S6B). Remarkably, the extended model predicted similar response patterns for relative P-SMAD2 and normalized nuclear SMAD signaling, reinforcing the idea that the TGF-β receptor with lower abundance acts as a key determinant of SMAD signaling response. Taken together, our findings from both the simplified and extended models support the notion that cells utilize a minority rule to process the variations in TGFBR1 and TGFBR2, ultimately influencing SMAD signaling output.

SMAD signaling is sensitive to perturbations of TGF-β receptor in lower abundance, but is robust to changes of TGF-β receptor in higher abundance

To experimentally test how SMAD signaling is changed with respect to the perturbations of TGF-β receptors in different cell types, we selected three cell lines with different ratios of TGFBR1 and TGFBR2: HaCaT cells with similar amounts of TGFBR1 and TGFBR2, RH-30 cells with more TGFBR1 than TGFBR2, and HepG2 cells with more TGFBR2 than TGFBR1 (Figure 3E). The relative abundances of TGFBR1 and TGFBR2 in these cell lines were confirmed by immunoblotting experiments (Figure S7).

If Liebig's law of the minimum holds for the TGF-β pathway, our model simulations suggested that HaCaT, RH-30, and HepG2 cells will display distinct sensitivities in response to variations in TGFBR1 and TGFBR2 (Figure 3D). Notably, RH-30 cells demonstrate significantly lower expression levels of TGFBR2 compared to TGFBR1. Consequently, the computational model simulations predicted that SMAD2 signaling in RH-30 cells would be sensitive to alterations in TGFBR2, but not TGFBR1. On the other hand, HepG2 cells display a higher abundance of TGFBR2 relative to TGFBR1, leading the model to predict that SMAD2 signaling in these cells would be sensitive to the reduction of TGFBR1, rather than TGFBR2. Finally, given that

HaCaT cells express comparable quantities of both receptors, the model predicts that SMAD2 signaling in these cells would exhibit sensitivity to the reduction of either receptor.

To test these predictions, we compared the profiles of SMAD2 signaling to changes of TGFBR1 and TGFBR2 abundance in HaCaT, RH-30, and HepG2 cells. Using different concentrations of small interfering RNA (siRNA), we knocked down individual TGF-β receptors to varying amounts in these cell lines. Previous studies showed that SMAD2 phosphorylation peaked around 1 hour after addition of TGF-β to the media (Strasen *et al.*, 2018; Zi *et al.*, 2011). Therefore, we measured SMAD2 phosphorylation and TGF-β receptor amounts in the siRNA knockdown samples relative to the non-targeting control (ntc) siRNA samples at 1 hour after TGF-β stimulation (Figure 4). The experimental data validated each prediction of the model: in RH-30 cells, SMAD2 signaling is sensitive to the reduction of TGFBR2, but not TGFBR1. Conversely, in HepG2 cells, it displayed a notable dependence on the reduction of TGFBR1, but not TGFBR2. Lastly, in HaCaT cells, SMAD2 signaling demonstrated comparable sensitivity to knockdown of either receptor (Figure 4).

These results suggested that the type of TGF-β receptor with the lower abundance serves as a limiting factor for TGF-β signaling to SMAD2. Furthermore, the data indicated that manipulation of the lower abundance TGF-β receptor might be an effective way to influence TGF-β signaling in different cellular contexts.

TGF-β signaling correlates with the low abundance TGF-β receptor in single cells

Because individual cells can have different amounts of TGF-β signaling proteins, we explored the TGF-β signaling responses in single cells by performing single-cell model simulations (Figure 5A). The single-cell models had the same molecular interactions and kinetic parameter values as the population model, but signaling protein concentrations (TGFBR1, TGFBR2, SMAD2) were varied according to log-normal distributions. The variations of TGF-β receptors

were achieved by setting the production rate as a constant with a value randomly generated from log-normal distributions. The mean value of each log-normal distribution is the same as the corresponding value in the population model. To mimic the variation of the strength of negative feedback regulation (NFR), the parameter (k_NFR) for the constant representing ligand-induced receptor degradation was sampled from a log-normal distribution. A coefficient of variation (CV) of 0.1 was applied for the variations of signaling proteins and NFR strength for simplicity. The single-cell model simulations predicted that the fold change in the SMAD2 N2C response at 1 h is strongly positively correlated with the amount of TGF-β receptors and weakly negatively correlated with SMAD2 and the strength of NFR (Figure 5B). However, the predicted response at 8 h was less positively correlated with the amount of TGF-β receptors and was more strongly negatively correlated with SMAD2 and NFR strength compared with the 1 h response (Figure 5C). These results suggested that heterogeneous amounts of TGF-β receptors mainly contribute to the variations of short-term SMAD signaling in single cells, whereas NFR strength and SMAD2 abundance contributes to determining long-term SMAD signaling dynamics.

Using single-cell model simulations, we investigated single-cell responses in cells with imbalanced amounts of TGFBR1 and TGFBR2. The simulations were performed like those shown in Figure 5, but the mean amount of TGFBR1 or TGFBR2 was increased to 5-fold of the corresponding value in the population model. The simulations predicted that normalized nuclear SMAD2 signaling in single cells is highly correlated with the TGF-β receptor that is less abundant (Figure 6). These results resemble Liebig's law of the minimum at a single-cell level. Our model predictions predicted that SMAD2 signaling is correlated with both TGFBR1 and TGFBR2 when they are present in equal amounts or with constant ratios in single cells (the dashed line in Figure 3D). Therefore, we tested this prediction in the single-cell model by setting the same protein abundance values for TGFBR1 and TGFBR2 in each single cell

simulation. As expected, the model predicted that SMAD2 signaling response is highly correlated with the number of TGF-β receptors in single cells (Figure 7A).

To test these model predictions in single cells, we used the optogenetic TGF-β receptor system (optoTGFBRs) (Li *et al*, 2018). In this system, SMAD2 signaling is activated with blue light through the reversible dimerization of the N-terminal end of CIB1 (CIBN) and the PHR domain of Cryptochrome2, which are fused to chimeric TGFBR1 and TGFBR2, respectively, and fluorescent SMAD2 N2C is detected with the reporter iRFP-SMAD2 (Figure 7B). We used HeLa cells stably expressing optoTGFBR1, optoTGFBR2-tdTomato, and iRFP-SMAD2. These cells are called optoTGFBRs-HeLa. To ensure similar amounts of each receptor, a viral 2A peptide sequence (P2A) was included between the coding sequences of optoTGFBR1 and optoTGFBR2-tdTomato to induce self-cleavage during translation enabling co-expression of these two proteins (Ahier & Jarriault, 2014; Li *et al.*, 2018) (Figure 7C).

With this system, we measured SMAD2 N2C fold changes in individual optoTGFBRs-HeLa cells at different time points after blue light stimulation. As predicted by the model, the amplitude of iRFP-SMAD2 signaling increased with the signal from optoTGFBR2-tdTomato, indicating the abundance of optogenetic TGFBR2 receptor, in individual cells (Figure 7D-E). Furthermore, the SMAD2 N2C fold change responses at 20 min and 1 h were highly correlated with the amount of optoTGFBR2-tdTomato (Figure 7F-G). We observed the same correlation when this system was stably expressed in HaCaT cells (optoTGFBRs-HaCaT) (Figure S8). Together, these results showed that variable but matched amounts of TGF-β receptors in individual cells produced heterogeneous TGF-β signaling responses at the single-cell level.

Discussion

Signaling proteins are present in different quantities across cell lines and within individual cells. The variability of signaling protein abundance affects the transmission of molecular

signal in a nuanced and sophisticated manner. In this study, we investigated the relationship between TGF-β signaling responses, in particular SMAD2 activity, and the variation of protein abundance. The TGF-β pathway uses two types of receptors (TGFBR1 and TGFBR2) to propagate the signal. Our analyses showed that heterogeneity in SMAD2 signaling follows specific patterns according to the variations in TGF-β receptor abundance. We found that different cell lines show distinct sensitivities to perturbation of receptor abundance depending on the relative amounts of TGFBR1 and TGFBR2. When the amounts of TGFBR1 and TGFBR2 are imbalanced, as occurs in some cell lines (such as RH-30 and HepG2 cells studied here), nuclear SMAD2 signaling is sensitive to variation of low abundance receptor but is robust to variation of the high abundance receptor (Figure 4). This feature is consistent with Liebig's law of the minimum (Liebig, 1840). Similar to the rate-limiting concept in enzymology, our work indicated that the low abundance TGF-β receptor may serve as a rate-limiting factor that dictates the TGF-β signaling response in different cellular contexts. It will be of great interest to see whether the principle of minimum control governs the heterogeneity of signaling responses in other signaling pathways.

The responses of TGF-β signaling can display diverse patterns in different time scales at the population and single-cell levels (Strasen *et al.*, 2018; Zi *et al.*, 2011). Multiple studies report that SMAD signaling in the short-term (here, considered within 1 hour) senses the amount of TGF-β through the activation of receptors (Frick *et al.*, 2017; Nunns & Goentoro, 2018; Schmierer *et al*, 2008; Zi *et al.*, 2011). In this work, we found that short-term SMAD signaling is highly correlated with the number of TGF-β receptors in single cells, when each receptor is present at an equivalent amount using an engineered optogenetic system (Figure 7). In the long-term (8 hours or longer), downstream TGF-β signaling patterns diverge as negative feedback regulators alter the availability of TGF-β ligand and receptors (Strasen *et al.*, 2018; Zi *et al.*, 2011). Our computational analysis showed that long-term SMAD signaling negatively

correlated with the strength of NFR (Figure 5C). This is consistent with a study showing that knocking down SMAD7, a negative feedback regulator of TGF-β signaling, reduces the heterogeneity of long-term SMAD2 dynamics in single cells (Strasen *et al.*, 2018). However, a direct quantification of the relationships between TGF-β signaling responses and feedback proteins is missing at the single-cell level. Single-cell protein profiling approaches could provide key information about signaling relationships that produce heterogeneous responses (Gut *et al*, 2018; Kramer *et al.*, 2022; Lun *et al.*, 2017). Advances in this area should help to clarify the dependence of signaling responses on protein abundance in spaces with higher dimensionality.

TGF-β signaling plays an important role in cancer progression. Differential expression of the genes encoding TGF-β receptors occurs frequently in human cancers (Dzieran *et al*, 2013; Jakowlew *et al*, 1997; Korkut *et al*, 2018; Venkatasubbarao *et al*, 2000). Many pharmacological compounds have been designed to target TGFBR1 and TGFBR2 (Colak & Ten Dijke, 2017). Our finding indicates the importance of considering the relative abundance of TGF-β receptors for drug targeting in the TGF-β pathway. Selective inhibition of TGFBR2 could be more effective than targeting TGFBR1 in tumors with a high TGFBR1-to-TGFBR2 abundance or gene expression ratio and vice versa. We envision that the principle of minimum control could help us to understand why specific signaling proteins function differently in diverse cellular contexts. Determining the relationships between signaling outputs and the abundance of signaling components will be useful for manipulating pathway activities in therapeutic applications.

Materials and Methods

Cell culture and the constructs of stable cell lines

HaCaT, HepG2, optoTGFBRs-HeLa, and optoTGFBRs-HaCaT cells were cultured in DMEM medium supplemented with 10% fetal bovine serum, 100 units/mL penicillin and 100 μg/mL streptomycin, and 2 mM L-glutamine at 37 °C and 5% CO_2. RH-30 cells were cultured in RPMI 1640 medium with the same supplements as those in the complete medium for culturing other cells. The optoTGFBRs-HaCaT cell line was established following the same procedure as for the development of the optoTGFBRs-HeLa cell line (Li *et al.*, 2018).

Cell lysate preparation and immunoblotting

Cell lysates were prepared with RIPA buffer (Cell Signaling, #9806), supplemented with 1 mM PMSF, 1 mM NaF, protease inhibitor cocktail (Roche, #04 693 132 001), and phosphatase inhibitor cocktail (Roche, #04 906 845 001). SDS–PAGE gels were transferred onto nitrocellulose membranes.

We used the following primary antibodies against: Phospho-SMAD2 (Ser465/467) (Cell Signaling, #3108), SMAD2 (Cell Signaling, #3103), β-Actin (Cell Signaling, #3700), TGFBR1 (Santa Cruz, #sc-398), TGFBR2 (abcam, #ab184948), and integrin β1 (Cell Signaling, #9699). All Western blot membranes were incubated with primary antibodies overnight at 4 °C. Secondary antibodies used for Western blotting were anti-rabbit IgG (H+L) DyLight 800 (Cell Signaling, #5151) and anti-mouse IgG (H+L) DyLight 680 (Cell Signaling, #5470). Western blot images were acquired and quantified using Odyssey CLx Imaging System (LI-COR Biosciences, #9140).

siRNA-mediated knockdown

Cells were seeded in 6-well plates at 50%-70% confluency and incubated overnight. On day 2, cells were transfected with different concentrations of SMARTpool siRNAs (final concentrations: 0.08-100 nM) using DharmaFECT1 reagents following the manufacturer's protocol. Non-targeting controls corresponded to the samples transfected with non-targeting control siRNA (final concentration: 20 nM). Each SMARTpool siRNA product contains four types of individual siRNAs targeting a single gene. After two days of transfection, cells were cultured with standard cell culture medium and where indicated the cells were stimulated with 100 pM (2.5 ng/mL) TGF-β1 for 1 h.

The following siRNA reagents were purchased from Horizon Discovery Group: TGFBR1 siRNA (catalog number: L-003929-00-0010), TGFBR2 siRNA (catalog number: L-003930-00-0010), and non-targeting control siRNA (catalog number: D-001810-10-05).

Live-cell imaging and photoactivation

Live-cell imaging experiments were performed using a Zeiss LSM 710 NLO 2-photon/confocal laser scanning microscope, which was equipped with an incubator to maintain the samples at 37 °C and with 5% CO_2. optoTGFBRs-HeLa and optoTGFBRs-HaCaT cells were activated with blue laser light at 488 nm (laser power:12.4 μW). The light intensity was fine-tuned by changing the pixel dwell time (Li & Zi, 2022).

Image analysis and quantification

Live-cell imaging data were quantified using ImageJ. We first quantified the SMAD2 N2C response by calculating the ratio of nuclear to cytoplasmic iRFP-SMAD2 signal. The SMAD2 N2C fold change response in each cell was quantified by normalizing SMAD2 N2C to the corresponding SMAD2 N2C signal at 0 h in the same cell. The optoTGFBR2-tdTomato amount was quantified as the tdTomato fluorescence intensity in the cytoplasmic area after subtracting

the background signal measured in the areas without cells. Multiple areas were quantified, and the average values were used for calculations.

Fitting of P-SMAD2 dose response curve to TGFBR1 or TGFBR2 knockdown

To quantify how P-SMAD2 responded to siRNA knockdown perturbations of TGFBR1 and TGFBR2, we fitted a Hill function to the dose response curves (Figure 4). The amount of P-SMAD2 and the amount of TGFBR1 (or TGFBR2) in each perturbation were first normalized to those in the non-targeting control (ntc) samples. Then the dose response curve was fitted with a 3 parameter Hill function using least squares fitting in MATLAB.

Mathematical modeling

Model simulations were performed using MATLAB. Parameter estimation and identifiability analyses were performed with the parallel parameter estimation tool SBML-PET-MPI (Zi, 2011). The details about model development, assumptions, parameter estimation, the system of ordinary differential equations and kinetic parameter values are described in the Supplementary Information.

Data availability

All data is included in the manuscript and/or supporting information.

Acknowledgements

This work was supported by BMBF e:Bio SyBioT project (031A309) to ZZ. The authors thank Dr. Ping Wei for critical reading of this manuscript and Dr. Nancy R. Gough (BioSerendipity, LLC) for editorial support.

Author contribution

YL, DD and ZZ: experiments and data analysis; ZZ: mathematical modeling; CTH contributed to the setup of live-cell imaging experiments; JK and WDH contributed to constructing the optoTGFBRs plasmid; QX and XL provided insights on the data analysis. ZZ wrote the manuscript with contributions from all the authors. ZZ conceived the study and supervised the research.

Conflict of interest

X.L. is a cofounder of OnKure Therapeutics, Inc and Vesicle Therapeutics, Inc and owns equity in both companies. None of these companies has relationships or competing interests to this study. All other authors declare that they have no conflict of interests.

References

Ahier A, Jarriault S (2014) Simultaneous expression of multiple proteins under a single promoter in Caenorhabditis elegans via a versatile 2A-based toolkit. *Genetics* 196: 605-613

Antebi YE, Linton JM, Klumpe H, Bintu B, Gong M, Su C, McCardell R, Elowitz MB (2017) Combinatorial Signal Perception in the BMP Pathway. *Cell* 170: 1184-1196 e1124

Buccitelli C, Selbach M (2020) mRNAs, proteins and the emerging principles of gene expression control. *Nat Rev Genet* 21: 630-644

Cohen-Saidon C, Cohen AA, Sigal A, Liron Y, Alon U (2009) Dynamics and variability of ERK2 response to EGF in individual living cells. *Mol Cell* 36: 885-893

Colak S, Ten Dijke P (2017) Targeting TGF-beta Signaling in Cancer. *Trends Cancer* 3: 56-71

Derynck R, Budi EH (2019) Specificity, versatility, and control of TGF-beta family signaling. *Sci Signal* 12

Dzieran J, Fabian J, Feng T, Coulouarn C, Ilkavets I, Kyselova A, Breuhahn K, Dooley S, Meindl-Beinker NM (2013) Comparative analysis of TGF-beta/Smad signaling dependent cytostasis in human hepatocellular carcinoma cell lines. *PLoS One* 8: e72252

Eldar A, Elowitz MB (2010) Functional roles for noise in genetic circuits. *Nature* 467: 167-173

Eling N, Morgan MD, Marioni JC (2019) Challenges in measuring and understanding biological noise. *Nat Rev Genet* 20: 536-548

Filippi S, Barnes CP, Kirk PD, Kudo T, Kunida K, McMahon SS, Tsuchiya T, Wada T, Kuroda S, Stumpf MP (2016) Robustness of MEK-ERK Dynamics and Origins of Cell-to-Cell Variability in MAPK Signaling. *Cell Rep* 15: 2524-2535

Frick CL, Yarka C, Nunns H, Goentoro L (2017) Sensing relative signal in the Tgf-beta/Smad pathway. *Proc Natl Acad Sci U S A* 114: E2975-E2982

Gabor A, Tognetti M, Driessen A, Tanevski J, Guo B, Cao W, Shen H, Yu T, Chung V, Single Cell Signaling in Breast Cancer DCm *et al* (2021) Cell-to-cell and type-to-type heterogeneity of signaling networks: insights from the crowd. *Mol Syst Biol* 17: e10402

Gut G, Herrmann MD, Pelkmans L (2018) Multiplexed protein maps link subcellular organization to cellular states. *Science* 361

Jakowlew SB, Moody TW, Mariano JM (1997) Transforming growth factor-beta receptors in human cancer cell lines: analysis of transcript, protein and proliferation. *Anticancer Res* 17: 1849-1860

Karlsson M, Zhang C, Mear L, Zhong W, Digre A, Katona B, Sjostedt E, Butler L, Odeberg J, Dusart P *et al* (2021) A single-cell type transcriptomics map of human tissues. *Sci Adv* 7

Klumpe HE, Langley MA, Linton JM, Su CJ, Antebi YE, Elowitz MB (2022) The context-dependent, combinatorial logic of BMP signaling. *Cell Syst* 13: 388-407 e310

Korkut A, Zaidi S, Kanchi RS, Rao S, Gough NR, Schultz A, Li X, Lorenzi PL, Berger AC, Robertson G *et al* (2018) A Pan-Cancer Analysis Reveals High-Frequency Genetic Alterations in Mediators of Signaling by the TGF-beta Superfamily. *Cell Syst* 7: 422-437 e427

Kramer BA, Sarabia Del Castillo J, Pelkmans L (2022) Multimodal perception links cellular state to decision-making in single cells. *Science* 377: 642-648

Li Y, Lee M, Kim N, Wu G, Deng D, Kim JM, Liu X, Heo WD, Zi Z (2018) Spatiotemporal Control of TGF-beta Signaling with Light. *ACS Synth Biol* 7: 443-451

Li Y, Zi Z (2022) Optogenetic Control of TGF-beta Signaling. *Methods Mol Biol* 2488: 113-124

Liebig JFv (1840) *Die organische Chemie in ihrer Anwendung auf Agricultur und Physiologie*. Friedrich Vieweg und Sohn Publishing Company, Braunschweig

Losick R, Desplan C (2008) Stochasticity and cell fate. *Science* 320: 65-68

Lun XK, Zanotelli VR, Wade JD, Schapiro D, Tognetti M, Dobberstein N, Bodenmiller B (2017) Influence of node abundance on signaling network state and dynamics analyzed by mass cytometry. *Nat Biotechnol* 35: 164-172

Massague J (2012) TGFbeta signalling in context. *Nat Rev Mol Cell Biol* 13: 616-630

Mattiazzi Usaj M, Yeung CHL, Friesen H, Boone C, Andrews BJ (2021) Single-cell image analysis to explore cell-to-cell heterogeneity in isogenic populations. *Cell Syst* 12: 608-621

Mitchell S, Hoffmann A (2018) Identifying Noise Sources governing cell-to-cell variability. *Curr Opin Syst Biol* 8: 39-45

Niepel M, Spencer SL, Sorger PK (2009) Non-genetic cell-to-cell variability and the consequences for pharmacology. *Curr Opin Chem Biol* 13: 556-561

Nunns H, Goentoro L (2018) Signaling pathways as linear transmitters. *Elife* 7

Schmierer B, Tournier AL, Bates PA, Hill CS (2008) Mathematical modeling identifies Smad nucleocytoplasmic shuttling as a dynamic signal-interpreting system. *Proc Natl Acad Sci U S A* 105: 6608-6613

Shi Y, Massague J (2003) Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell* 113: 685-700

Snijder B, Pelkmans L (2011) Origins of regulated cell-to-cell variability. *Nat Rev Mol Cell Biol* 12: 119-125

Strasen J, Sarma U, Jentsch M, Bohn S, Sheng C, Horbelt D, Knaus P, Legewie S, Loewer A (2018) Cell-specific responses to the cytokine TGFbeta are determined by variability in protein levels. *Mol Syst Biol* 14: e7733

Su CJ, Murugan A, Linton JM, Yeluri A, Bois J, Klumpe H, Langley MA, Antebi YE, Elowitz MB (2022) Ligand-receptor promiscuity enables cellular addressing. *Cell Syst* 13: 408-425 e412

Venkatasubbarao K, Ahmed MM, Mohiuddin M, Swiderski C, Lee E, Gower WR, Jr., Salhab KF, McGrath P, Strodel W, Freeman JW (2000) Differential expression of transforming growth factor beta receptors in human pancreatic adenocarcinoma. *Anticancer Res* 20: 43-51

Wu ZG, Park S, Lau CM, Zhong Y, Sheppard S, Sun JC, Das J, Altan-Bonnet G, Hsu KC (2021) Dynamic variability in SHP-1 abundance determines natural killer cell responsiveness. *Sci Signal* 14

Zhang YE (2018) Mechanistic insight into contextual TGF-beta signaling. *Curr Opin Cell Biol* 51: 1-7

Zi Z (2011) SBML-PET-MPI: a parallel parameter estimation tool for Systems Biology Markup Language based models. *Bioinformatics* 27: 1028-1029

Zi Z, Feng Z, Chapnick DA, Dahl M, Deng D, Klipp E, Moustakas A, Liu X (2011) Quantitative analysis of transient and sustained transforming growth factor-beta signaling dynamics. *Mol Syst Biol* 7: 492

Figures

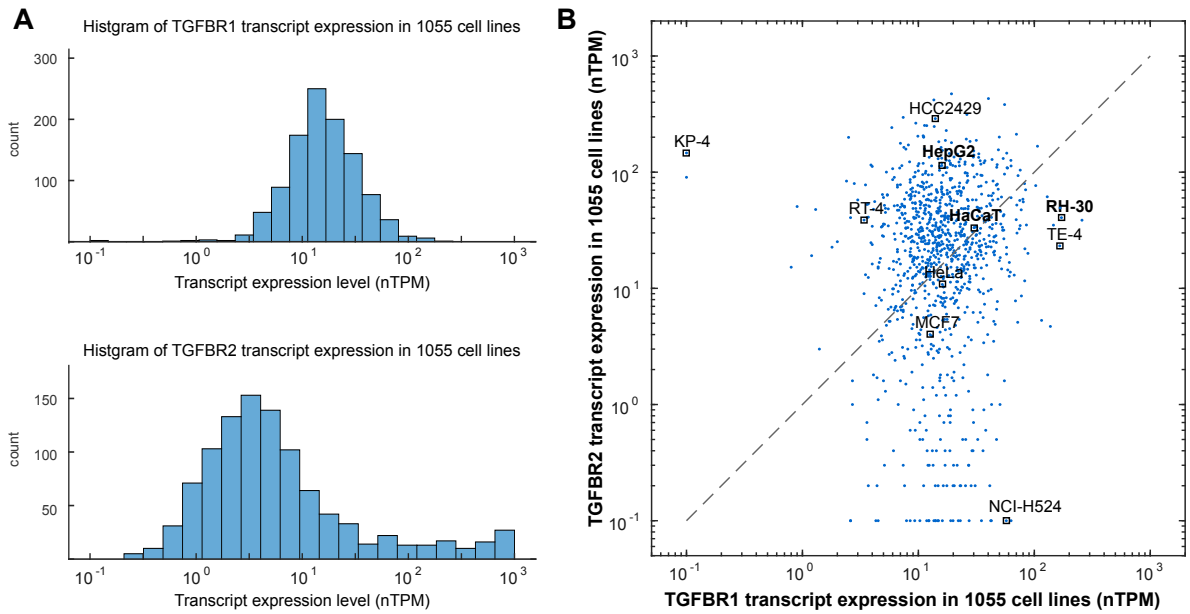

Figure 1: Transcript Expression Profile of TGFBR1 and TGFBR2 in 1055 Cell Lines

The transcript data for TGFBR1 and TGFBR2 were obtained from the Human Protein Atlas database version 22.0. The expression levels are represented in normalized transcripts per million (nTPM).

(A) Histogram depicting the distribution of TGF-β receptors expression across various cell lines.

(B) Imbalanced expression of TGFBR1 and TGFBR2 was observed in numerous cancer cell lines. Each dot in the plot represents the expression levels of TGFBR1 and TGFBR2 in one of the 1055 cell lines. Specific cell lines of interest are labeled as representatives.

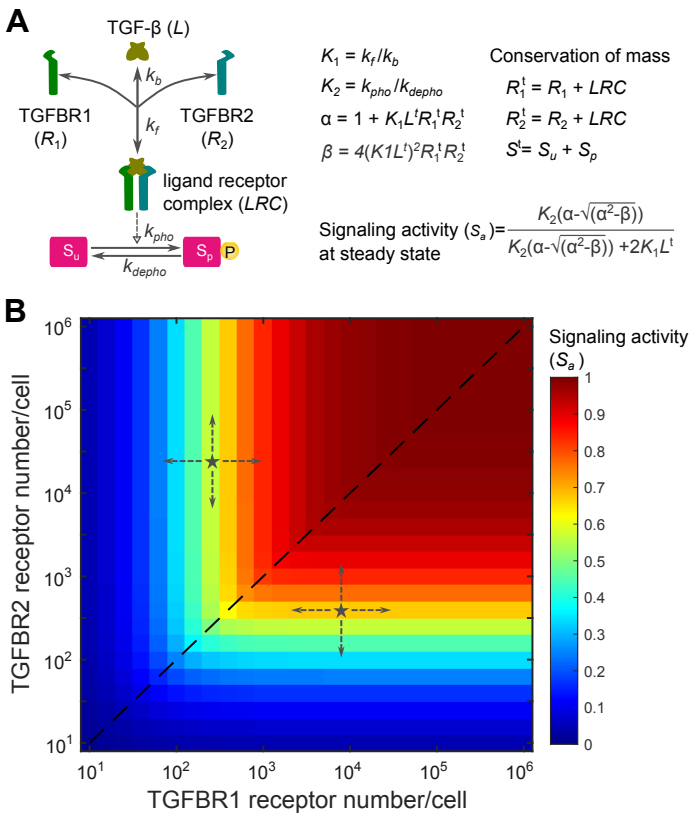

Figure 2: A minimal mathematical model predicts that TGF-β signaling activity is dictated by the type of TGF-β receptor in lower abundance

(A) A minimal model of the TGF-β pathway with two reactions including the TGF-β ligand (L), which interact with type I receptor (R_1) and type II receptor (R_2) with a binding affinity (K_1). The active ligand-receptor complex (LRC) phosphorylates the SMAD protein with an equilibrium coefficient (K_2). TGF-β signaling activity (S_a) at the steady state can be analytically calculated with equations (refer to Supplementary Information for model details and equation derivation).

(B) The response profile of TGF-β signaling activity (S_a) at the steady state is shown in relation to a 2-dimensional matrix of TGFBR1 and TGFBR2 expression numbers spanning five orders of magnitude. K_1=1000, K_2=10. TGF-β ligand concentration is set to be 0.1 nM.

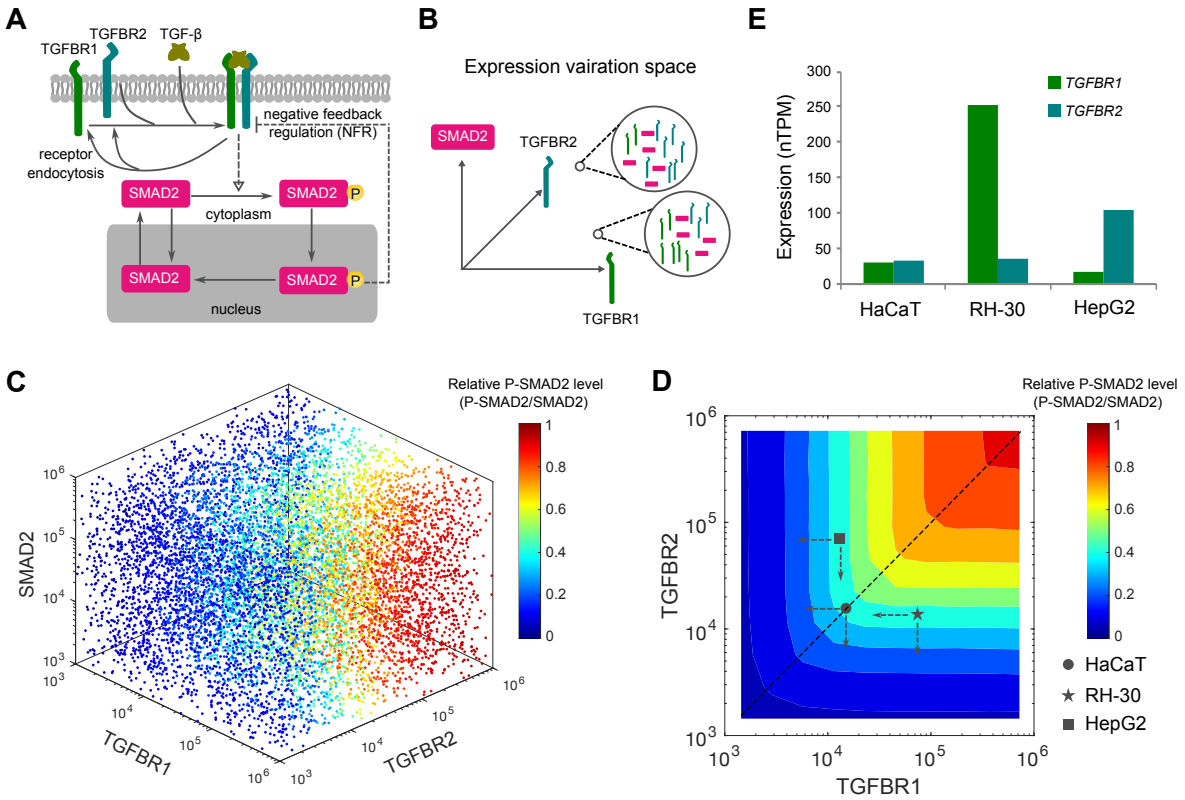

Figure 3. Computational modeling of heterogeneous signaling responses in the variation space of TGF-β receptors and SMAD2 expression

(A) Scheme of the extended TGF-β signaling model. The development of the extended model is described in the Supplementary Information.

(B) Heterogeneous expression of TGF-β receptors and SMAD2 is illustrated as points in 3 multi-dimensional spaces. The two magnified circles exemplify two possible expression profiles of TGFBR1, TGFBR2 and SMAD2.

(C) Model predictions for normalized P-SMAD2 responses in the expression space of TGFBR1, TGFBR2 and SMAD2. Each scatter point indicates a random set of protein expression levels in the space. The point colors indicate the normalized P-SMAD2 responses (P-SMAD2/SMAD2).

(D) Contour landscape of the normalized P-SMAD2 responses to the combinations of TGFBR1 and TGFBR2. The dashed line indicates relationship between the relative P-SMAD2 level and the receptors when equal amounts of each receptor are present. The abundances of TGFBR1 and TGFBR2 in HaCaT, RH-30 and HepG2 cell lines are mapped according to the relative expression data shown in Figure 3E and the measured HaCaT TGF-β receptor number shown in Figure S2B. The dashed arrows indicate the directions of receptor changes.

(E) TGF-β receptor expression profiles for HaCaT, RH-30 and HepG2 cell lines. Bars indicate expression levels of each receptor (normalized transcripts per million, nTPM). The data is based on the Human Protein Atlas database version 22.0.

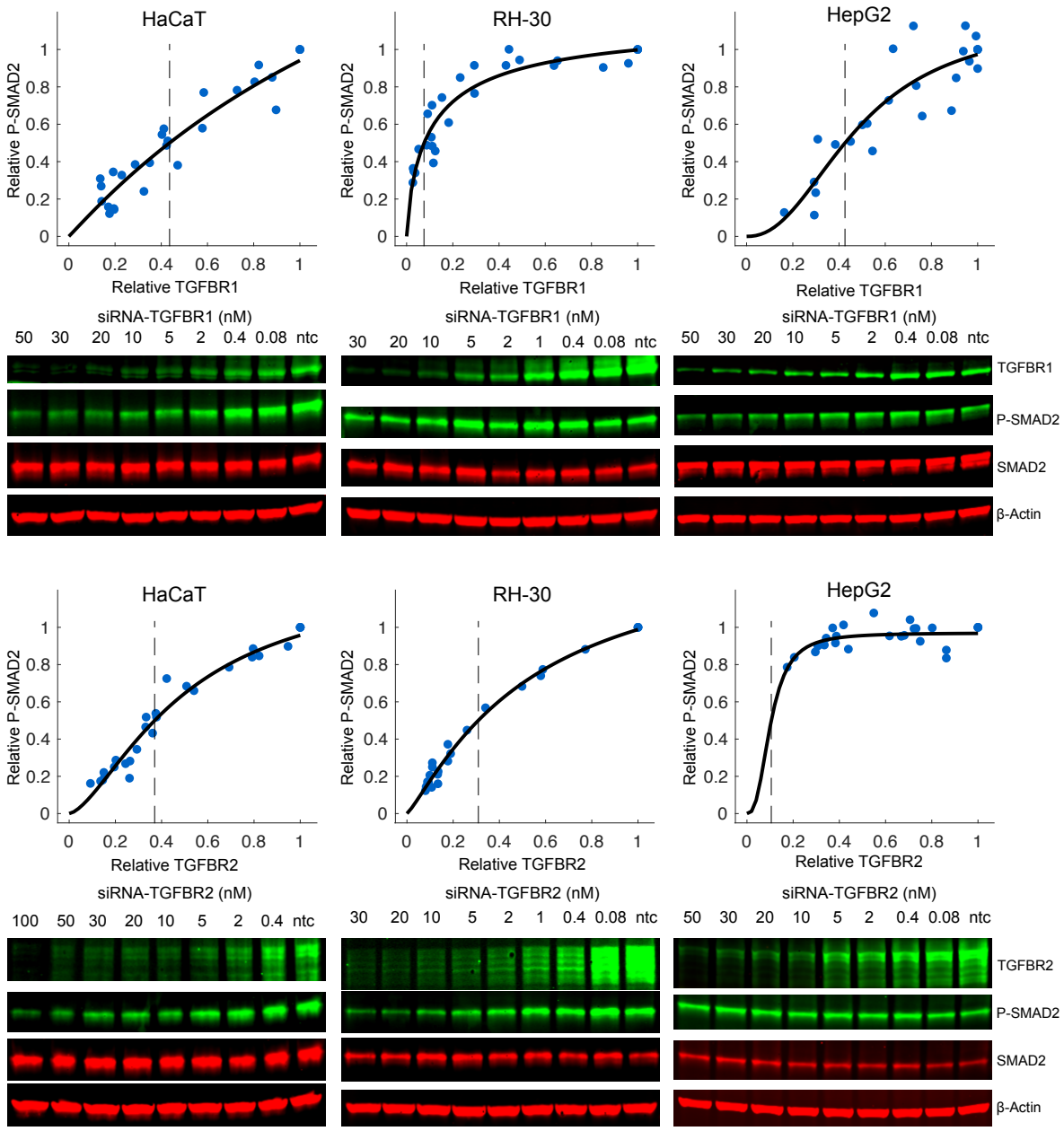

Figure 4. Different types of cells show distinct sensitivities to the knockdown of TGF-β receptors

Each type of TGF-β receptors was knocked down to different levels using various concentrations of siRNA. The relative P-SMAD2 responses and TGF-β receptor expression levels at 1 h after 100 pM TGF-β stimulation were measured and normalized to the non-targeting control (ntc) siRNA sample in the same experiment. Each plot contains the quantified data from three biological replicates. Immunoblotting results from one representative experiment are shown below each plot. The black curve represents the fitted Hill curve and the dashed lines indicate the fold change of TGF-β receptor leads to the half of the P-SMAD2 response in the ntc sample.

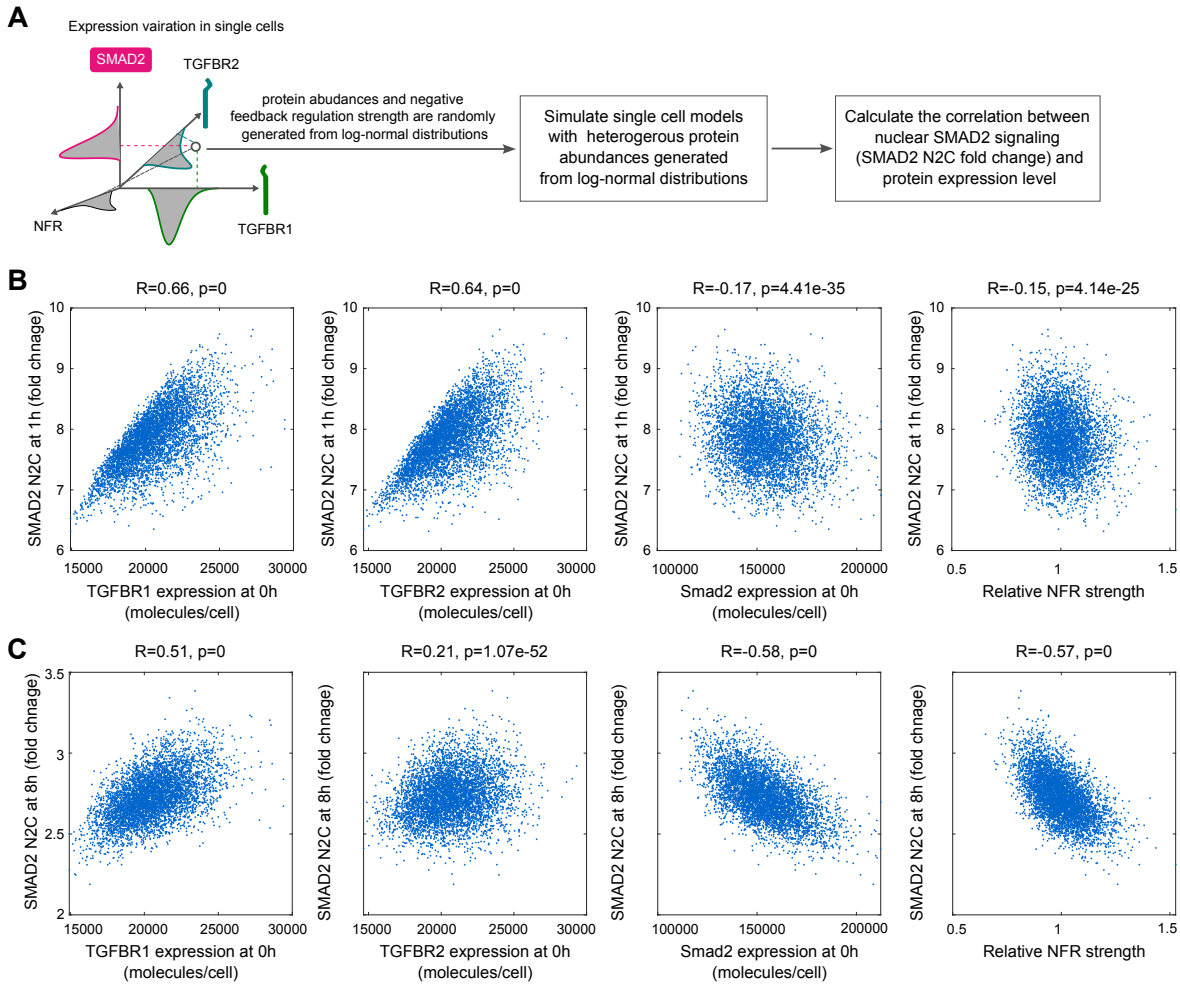

Figure 5. Model predictions for the correlation between the SMAD2 signaling responses and the expressions of TGF-β signaling proteins in single cells

(A) Illustration of model simulations for the SMAD2 signaling response to the variations of protein expressions in single cells.

(B) Model simulations for the correlation between SMAD2 N2C fold change response at 1 h after 100 pM TGF-β stimulation and the expression of TGFBR1, TGFBR2, SMAD2 at 0 h and negative feedback regulation (NFR) strength.

(C) Model simulations for the correlation between SMAD2 N2C fold change response at 8 h after 100 pM TGF-β stimulation and the expression of TGFBR1, TGFBR2, SMAD2 at 0h and negative feedback regulation (NFR) strength.

The correlations (R-values) in the plots were calculated with Pearson correlation coefficients.

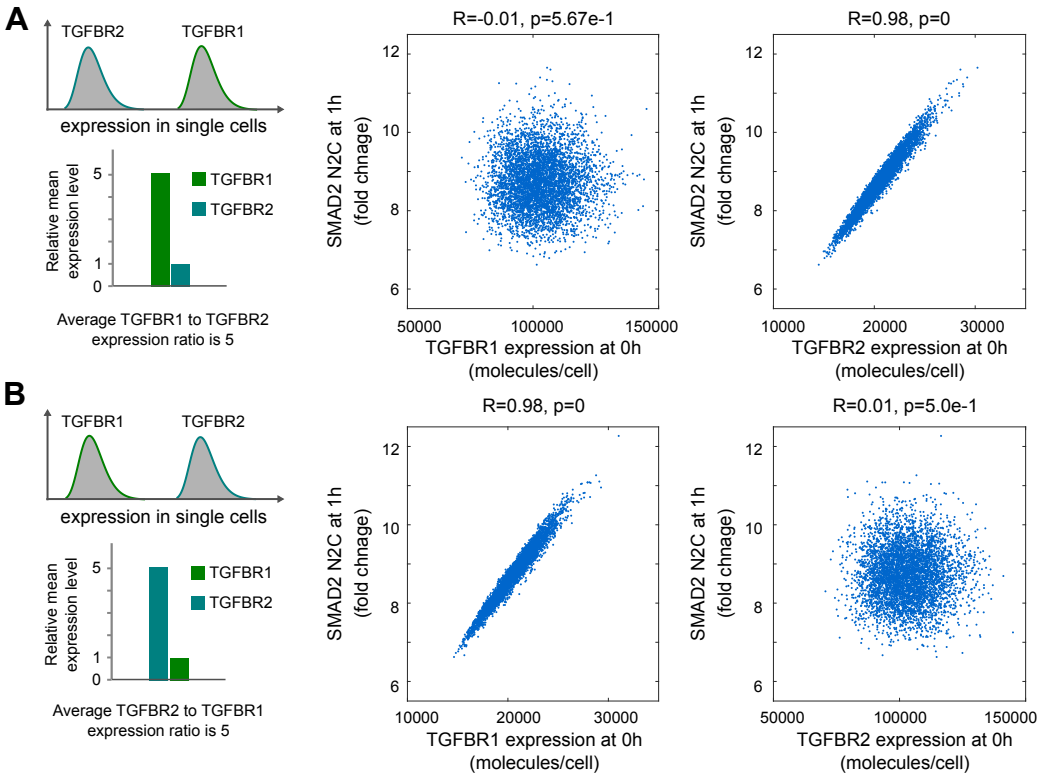

Figure 6. Model predictions for SMAD2 signaling responses with imbalanced expression of TGFBR1 and TGFBR2 in single cells

(A) Model predicts that SMAD2 N2C fold change is correlated with TGFBR2 when TGFBR1 is expressed at higher level than TGFBR2 in single cells (in this example, the average TGFBR1 to TGFBR2 expression ratio was set to 5).

(B) Model predicts that SMAD2 N2C fold change is correlated with TGFBR1 when TGFBR1 is expressed at lower level than TGFBR2 in single cells (in this example, the average TGFBR2 to TGFBR1 expression ratio was set to 5).

The correlations (R-values) in the plots were quantified with Pearson correlation coefficients.

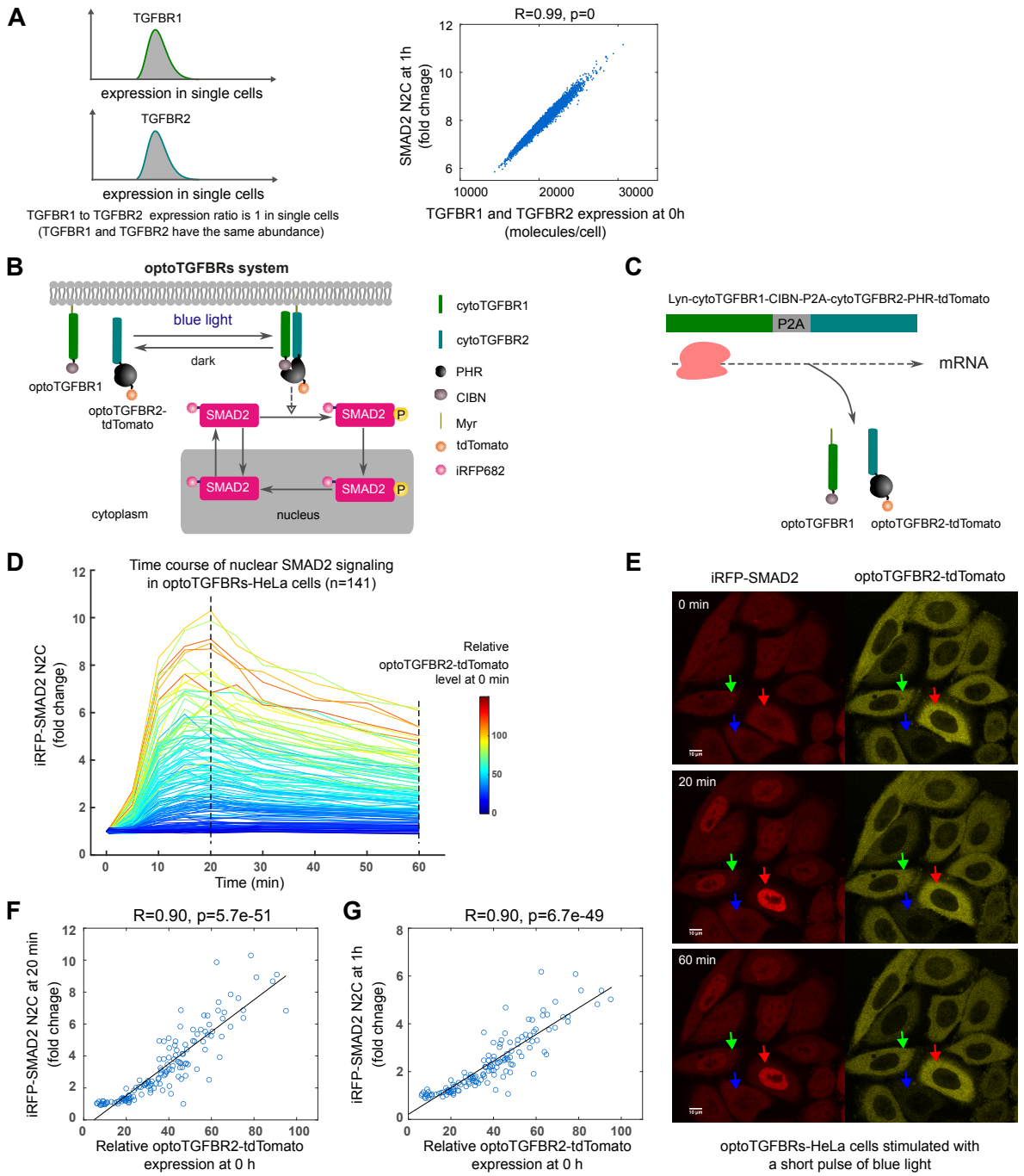

Figure 7. SMAD2 signaling responses are highly correlated with the amount of TGF-β receptors in single cells when TGFBR1 and TGFBR2 are expressed at a similar level

(A) Model prediction for the correlation between SMAD2 N2C fold change response at 1 h and the number of TGF-β receptors when TGFBR1 and TGFBR2 are expressed at a similar level in single cells. The protein abundance values for TGFBR1 and TGFBR2 were set to be the same in single cell simulations (TGFBR1 to TGFBR2 expression ratio is 1).

(B) Schematic representation of the optoTGFBRs optogenetic system. The development of the optoTGFBRs system was described in a previous publication (Li *et al.*, 2018).

(C) Illustration for the co-expression of optoTGFBR1 and optoTGFBR2-tdTomato with 2A peptide element (P2A).
(D) Time course profiles of iRFP-SMAD2 N2C fold change responses in individual optoTGFBRs-HeLa cells (n=141) upon one pulse of blue light stimulation (488 nm, 12.4 μW).
(E) Representative live-cell images of SMAD2 responses to one pulse of blue light stimulation in single cells. The red, green and blue arrows indicate cells that express optoTGFBR2-tdTomato receptor at a high, intermediate and low level, respectively.
(F) The plot of iRFP-SMAD2 N2C fold change responses at 20 min versus the expression of optoTGFBR2-tdTomato before blue light stimulation (0 h) in optoTGFBRs-HeLa cells.
(G) The plot of iRFP-SMAD2 N2C fold change responses at 1 h versus the expression of optoTGFBR2-tdTomato before blue light stimulation (0 h) in optoTGFBRs-HeLa cells.
The correlations (R-values) in the plots were calculated with Pearson correlation coefficients.

Supplementary Information for

Liebig's law of the minimum in the TGF-β pathway

Yuchao Li, Difan Deng, Chris Tina Höfer, Jihye Kim, Won Do Heo, Quanbin Xu, Xuedong Liu, Zhike Zi

Corresponding author: Zhike Zi

Table of Contents

1. Supplementary Methods 1

1.1 The minimal mathematical model of the TGF-β pathway 1

1.2 The extended model of the TGF-β pathway 3

1.3 Parameter estimation for the extend model 4

1.4. Single cell model simulations 7

1.5 Measurement of TGF-β concentration in media 7

1.6 Cell surface protein biotinylation and internalization 8

2. Supplementary Tables 9

Table S1: Summary of the derived parameter values based on experimental data 9

Table S2: Statistical analysis of the estimated parameters for the extended model 10

Table S3: Initial conditions of the extended model 11

Table S4: Ordinary differential equations of the extended model 12

3. Supplementary Figures 13

Figure S1. The predicted response profile of TGF-β signaling activities (S_a) at steady state with different choice of parameter values in the minimal model. 13

Figure S2. Measurement of protein abundance in HaCaT cells 14

Figure S3. Measurement of TGFBR2 internalization rate constant in HaCaT cells 15

Figure S4: The comparison of model simulation results and experimental data 16

Figure S5. Profile likelihood analysis of estimated parameter values 17

Figure S6: Computational modeling of the ratio of nuclear-to-cytoplasmic SMAD2 in the variation space of TGF-β receptors and SMAD2. 18

Figure S7. TGFBR1 and TGFBR2 protein expression profiles in HaCaT, RH-30 and HepG2 cells 19

Figure S8. Quantification of SMAD2 N2C fold change responses in HaCaT cells expressing optoTGFBRs 20

4. Supplementary References 21

1. Supplementary Methods

1.1 The minimal mathematical model of the TGF-β pathway

To investigate the signaling activity profiles of the TGF-β pathway in response to variations in TGF-β receptors, we developed a minimal model that captures the essential components of the pathway through two reactions. The first reaction represents the binding of the TGF-β ligand (L) to type I (R_1) and type II (R_2) TGF-β receptors, forming the ligand-receptor complex (LRC). The second reaction describes the phosphorylation and dephosphorylation of the SMAD protein (S_u: unphosphorylated Smad, S_p: phosphorylated Smad).

Following the approach used in published BMP pathway models (Antebi *et al*, 2017; Su *et al*, 2022), we model the binding and unbinding of TGF-β ligand and its receptors as one-step reversible reaction with forward rate (k_f) and backward rate (k_b). We assumed that the phosphorylation of the SMAD protein is proportional to the activity of the ligand-receptor complex. In this minimal model, our focus is on the steady-state signaling activity of the TGF-β pathway. Additionally, we assumed the total amounts of receptors and SMAD proteins to remain constant (conservation of mass), following similar assumptions made in mathematical models of TGF-β and BMP pathway (Antebi *et al.*, 2017; Su *et al.*, 2022; Vilar *et al*, 2006). The two reactions can be summarized as:

$$L + R_1 + R_2 \underset{k_b}{\overset{k_f}{\rightleftharpoons}} LRC \tag{1}$$

$$S_u \underset{k_{depho}}{\overset{k_{pho}[LRC]}{\rightleftharpoons}} S_p \tag{2}$$

where S_u denotes unphosphorylated SMAD, S_p denotes phosphorylated SMAD.

Based on mass-action kinetics, we can formulate the following set of ordinary differential equations to describe the dynamics of these reactions:

$$\frac{d[R_1]}{dt} = k_b[LRC] - k_f[L][R_1][R_2] \tag{3}$$

$$\frac{d[R_2]}{dt} = k_b[LRC] - k_f[L][R_1][R_2] \tag{4}$$

$$\frac{d[LRC]}{dt} = k_f[L][R_1][R_2] - k_b[LRC] \tag{5}$$

$$\frac{d[pS]}{dt} = k_{pho}[S_u][LRC] - k_{depho}[S_p] \tag{6}$$

With the mass of conservation, we obtain the following constraints:

$$R_1^t = [R_1] + [LRC] \tag{7}$$

$$R_2^t = [R_2] + [LRC] \tag{8}$$

$$S^t = [S_u] + [S_p] \tag{9}$$

where the variables with superscript '*t*' denote the total of the respective species with mass conservation.

Furthermore, in typical experimental setups, the TGF-β ligand is prepared in a much larger volume of media compared to the volume of cells, resulting in a scenario where there are significantly more ligand molecules than receptors. In such cases, the concentration of the TGF-β ligand remains constant.

$$[L] = L^t \tag{10}$$

By plugging equation 9 to equation 6 and scaling it to S^t, equation 6 can be rewritten as:

$$\frac{dS_a}{dt} = k_{pho}(1 - S_a)[LRC] - k_{depho} \cdot S_a \tag{11}$$

where S_a represents the TGF-β signaling activity, which is the percentage of total SMAD proteins that are actively phosphorylated.

We next focused on the TGF-β signaling activity (S_a) at steady state after ligand stimulation, which can be solved based on the following algebraic equations:

$$k_b[LRC^{ss}] - k_f[L^{ss}][R_1^{ss}][R_2^{ss}] = 0 \tag{12}$$

$$k_{pho}(1 - S_a^{ss})[LRC^{ss}] - k_{depho} \cdot S_a^{ss} = 0 \tag{13}$$

Plugging equations 7-9 to equations 12 and 13, we can derive the steady state of the TGF-β signaling activity (S_a^{ss}) as the following:

$$LRC^{ss} = \frac{\left(1+\frac{k_f}{k_b}L^tR_1^tR_2^t\right) - \sqrt{\left(1+\frac{k_f}{k_b}L^tR_1^tR_2^t\right)^2 - 4\left(\frac{k_f}{k_b}L^t\right)^2 R_1^tR_2^t}}{2\frac{kf}{kb}L^t} \tag{14}$$

$$S_a^{ss} = \frac{k_{pho}[LRC^{ss}]}{k_{pho}[LRC^{ss}] + k_{depho}} \tag{15}$$

Defining the following parameters

$$K_1 = \frac{k_f}{k_b} \tag{16}$$

$$K_2 = \frac{k_{pho}}{k_{depho}} \tag{17}$$

$$\alpha = 1 + K_1 L^t R_1^t R_2^t \tag{18}$$

$$\beta = 4\left(K_1 L^t\right)^2 R_1^t R_2^t \tag{19}$$

Equation 14 and 15 can be rewritten as

$$[LRC^{ss}] = \frac{\alpha - \sqrt{\alpha^2 - \beta}}{2K_1 L^t} \tag{20}$$

$$S_a^{ss} = \frac{K_2[LRC^{ss}]}{K_2[LRC^{ss}]+1} = \frac{K_2\left(\alpha - \sqrt{\alpha^2-\beta}\right)}{K_2\left(\alpha - \sqrt{\alpha^2-\beta}\right)+2K_1 L^t} \tag{21}$$

With equation 18, 19 and 21, we can plot the response profiles of the TGF-β signaling activity at steady state (S_a^{ss}) with the combinations of different R1 and R2 values. The initial abundance (N_X, molecules per cell) and the initial concentration (X^0) of specie X can be exchanged with the following equation:

$$N_X = N_A V_{cell} X^0 \tag{22}$$

where N_A is Avogadro number (6.02×1023). V_{cell} is set to be 3.3 pL (3.3×10-12 L) according to previous reports (Schmierer et al, 2008). It's worth noting that the choice of V_{cell} value does not change the response profiles shown in Figure 2 and Figure S1 as cell volume is just a scaling factor for the conversion between specie concentration and abundance.

1.2 The extended model of the TGF-β pathway

We next developed an extended mathematical model based on our previously published TGF-β model (Zi *et al*, 2011). Compared to the minimal model, the extended model includes additional TGF-β signaling processes: (1) receptor production, degradation, endocytosis and activation; (2) SMAD2 nucleocytoplasmic shuttling, phosphorylation and dephosphorylation; (3) negative feedback on the degradation of ligand receptor complex; (4) reversible binding of TGF-β to the cell surface. In the extended model, we used a linear chain of 2 variables (z1, z2) with linear differential equations to model the time delay between nuclear SMAD2 activity (PSmad2n) and the negative feedback on the degradation of ligand-receptor complex.

The extended model consists of 13 species and 22 reactions, characterized by 20 kinetic parameters. Details regarding the initial conditions, parameter values, and equations can be found in Table S1-S4. To determine the initial condition of the extended model, we measured the abundance of SMAD2 and TGF-β receptors in the HaCaT cell line (as described in section 1.2). Additionally, we derived 11 kinetic parameters from experimental datasets, including our own and those from other studies. The remaining 9 unknown parameter values were estimated by fitting the model to 102 experimental data points obtained from HaCaT cells exposed to different TGF-β stimulation conditions. Notably, parameter identifiability analysis confirms the identifiability of these parameters (see section 1.3.5 for more details on the estimation of unknown parameters).

1.3 Parameter estimation for the extend model

1.3.1 The abundance of SMAD2 and TGF-β receptors

To determine the abundance of SMAD2 and TGF-β type II receptor (TGFBR2, referred to as T2R in the extended model), we performed quantitative immunoblotting on HaCaT cells. This technique allowed us to estimate the absolute protein levels based on a standard curve signal generated from recombinant protein standards. The detailed procedure is described in a published protocol (Deng & Zi, 2022). The measured abundances of SMAD2 and TGFBR2 proteins in HaCaT are 1.54e5 and 2.06e4 molecules/cell, respectively (Figure S2).

According to the Human Protein Atlas database, the expression of TGFBR1 RNA in HaCaT cells is similar to that of TGFBR2 (Figure 3E). In the extended model, we set the protein abundance of TGFBR1 to be the same as TGFBR2 (2.06e4 molecules/cell), which aligns with a previously reported value (Wakefield et al, 1987).

1.3.2 TGF-β receptor internalization and recycling rate constants

We measured the internalization rate of TGFBR2 using a reversible biotinylation method (Gabriel *et al*, 2009). We next fitted the relative change of surface receptor protein over time with an exponential decay function and successfully measured the internalization rate constant of TGFBR2: $k_i = 0.022\ min^{-1} = 1.32\ h^{-1}$ (Figure S3).

It is known that most of TGF-β receptors are internalized in endosomes and about 10% of TGF-β receptors locate at cell surface (Di Guglielmo *et al*, 2003). Here we set the recycling rate is 1/9 of the internalization rate, $k_r = k_i/9 = 0.147\ h^{-1}$, this will result in about 10% receptors at cell surface in our model.

1.3.3 TGF-β receptor production and degradation rate constants

We measured the half-lives of TGF-β receptors and SMAD2 proteins in HaCaT cells by cycloheximide chase assays (Kao *et al*, 2015). The measured half-lives for TGFBR1, TGFBR2 and SMAD2 are 6.6 h, 3.67h, >24 h, respectively. The SMAD2 protein is stable before and after TGF-β stimulation. It is reasonable to assume the total amount of SMAD2 protein is constant in the model.

By rescaling the distribution of internalized receptors in endosomes, we can derive the degradation rate constants for T1R: $k_{deg}^{T1R} = 0.117\ h^{-1}$ ($\frac{-ln0.5}{6.6h\times0.9}$), for T2R: $k_{deg}^{T2R} = 0.21\ h^{-1}$ ($\frac{-ln0.5}{3.67h\times0.9}$). Since TGF-β ligand receptor complexes have similar distribution and half-life as the T1R receptor at resting state (Di Guglielmo *et al.*, 2003), we set $k_{deg}^{LCR} = k_{deg}^{T1R} = 0.117\ h^{-1}$.

The initial conditions of TGF-β receptors are determined by the receptor production, internalization, recycling and degradation rate constants (Zi *et al.*, 2011; Zi & Klipp, 2007). At resting state, we derive the following equations for TGF-β receptors at cell surface ($T1R_s$, $T2R_s$) and in cell cytosol ($T1R_c$, $T2R_c$):

$$[T1R_s]_0 = \frac{k_{prod}^{T1R} \cdot kr}{k_{deg}^{T1R} \cdot ki} \tag{23}$$

$$[T1R_c]_0 = \frac{k_{prod}^{T1R}}{k_{deg}^{T1R}} \tag{24}$$

$$N_{T1R} = V_{cyt}([T1R_s]_0 + [T1R_c]_0)N_A \tag{25}$$

$$[T2R_s]_0 = \frac{k_{prod}^{T2R} \cdot kr}{k_{deg}^{T2R} \cdot ki} \tag{26}$$

$$[T2R_c]_0 = \frac{k_{prod}^{T2R}}{k_{deg}^{T2R}} \tag{27}$$

$$N_{T2R} = V_{cyt}([T2R_s]_0 + [T2R_c]_0)N_A \tag{28}$$

where N_A is Avogadro number (6.02×10^{23}), N_{T1R} and N_{T2R} are the abundances of TGFBR1 and TGFBR2 in molecules/cell, respectively.

The volume of cytoplasm (V_{cyt}) and nucleus (V_{nuc}) are set as 2.3 pL (2.3×10^{-12} L) and 1 pL (1×10^{-12} L), respectively according to previous reports (Schmierer *et al.*, 2008). Solving equation 1-6, we can derive the initial conditions for T1R and T2R at cell surface and in cytosol with the following equations:

$$[T1R_s]_0 = \frac{N_{T1R}}{V_{cyt}N_A} \cdot \frac{kr}{ki+kr} = 1.49 \times 10^{-9}\ M = 1.49\ nM \quad (29)$$

$$[T1R_c]_0 = \frac{N_{T1R}}{V_{cyt}N_A} \cdot \frac{ki}{ki+kr} = 13.4 \times 10^{-9}\ M = 13.4\ nM \quad (30)$$

$$[T2R_s]_0 = \frac{N_{T2R}}{V_{cyt}N_A} \cdot \frac{kr}{ki+kr} = 1.49 \times 10^{-9}\ M = 1.49\ nM \quad (31)$$

$$[T2R_c]_0 = \frac{N_{T2R}}{V_{cyt}N_A} \cdot \frac{ki}{ki+kr} = 13.4 \times 10^{-9}\ M = 13.4\ nM \quad (32)$$

Based on the equations 24, 30 and equations 27, 32, we derived the corresponding production rate constant for T1R ant T2R:

$$k_{prod}^{T1R} = k_{deg}^{T1R}[T1R_c]_0 = 0.117 \times 13.4 = 1.57\ nM/h \quad (33)$$

$$k_{prod}^{T2R} = k_{deg}^{T2R}[T2R_c]_0 = 0.21 \times 13.4 = 2.81\ nM/h \quad (34)$$

1.3.4 Initial conditions for cytosol and nuclear SMAD2

SMAD2 nuclear import and export rate constants has been experimentally measured (Schmierer & Hill, 2005; Schmierer *et al.*, 2008): $k_{imp}^{Smad2} = 0.156\ min^{-1} = 9.36\ h^{-1}$, $k_{exp}^{Smad2} = 0.763\ min^{-1} = 45.8\ h^{-1}$. We set the distribution of SMAD2 in cytoplasm and nucleus in steady state before TGF-β stimulation with the following equations

$$[Smad2_c]_0 = \frac{N_{Smad2}}{V_{cyt}N_A} \cdot \frac{k_{exp}^{Smad2}}{k_{imp}^{Smad2}+k_{exp}^{Smad2}} = 92.4\ nM \quad (35)$$

$$[Smad2_n]_0 = \frac{N_{Smad2}}{V_{nuc}N_A} \cdot \frac{k_{imp}^{Smad2}}{k_{imp}^{Smad2}+k_{exp}^{Smad2}} = 43.4\ nM \quad (36)$$

where N_A is Avogadro number (6.02×10^{23}), N_{SMAD2} is the abundances of SMAD2 in molecules/cell.

1.3.5 Estimation of the unknown parameters

We measured or derived most parameter values in the refined models based on experimental data. The detailed values for these parameters are shown in Table S1. In the refined model, there were still some parameters that cannot be measured or derived directly from experimental data. For estimation of these unknown parameters, we used a parallel parameter estimation tool SBML-PET-MPI (Zi, 2011), which includes a global optimization algorithm (SRES) using least squares to minimize the sum of squares of differences between model simulations and the corresponding experimental datasets. We estimated the values of 9 parameters in the model

with 102 experimental data points collected from HaCaT cells with different TGF-β stimulation conditions, which includes.

(1) Time course of medium TGF-β (TGFb_ex in the model) and SMAD2 phosphorylation (P-SMAD2) levels in response to 10 pM, 100 pM and 400pM TGF-β stimulations.

(2) Time course of cytosol and nuclear SMAD2 in response to 100 pM TGF-β stimulation.

(3) Relative P-SMAD2 level at 45 min in responses to different doses of TGF-β stimulations.

After parameter optimization, the refined model can reproduce various experimental data with different doses of TGF-β stimulations (Figure S4). To check the identifiability of the estimated parameters in the model, we performed profile likelihood analysis (Raue *et al*, 2009) and calculate 95% confidence intervals for the estimated parameters in the model. The profile likelihood analysis indicate that the estimated parameters are practically identifiable with the experimental datasets used for parameter estimation.

1.4. Single cell model simulations

To simulate the signaling responses in single cells, we use the same model as the population model except that the abundances of TGFBR1 (N_{T1R}), TGFBR2 (N_{T2R}) and SMAD2 (N_{Smad2}) are randomly generated from log-normal distributions with a mean of the corresponding reference value in the population model and a CV of 0.1. For each single cell model, we adjust the receptor production parameter values, while keep other kinetic parameter values fixed. The production rate constants for TGFBR1 (kprd_T1R) and TGFBR2 (kprd_T2R) are set according to equation 33 and equation 34, respectively. The initial conditions for surface and cytosol receptors can be derived with equation 29-32. The initial concentrations of cytosol SMAD2 and nuclear SMAD2 were calculated based on equation 35-36 accordingly. For the variation of negative feedback regulation (NFR), we randomly generate a value for k_NFR from a log-normal distribution with a mean of the corresponding reference value in the population model and a CV of 0.1. For each case, we performed 5000 single cell model simulations and the correlation analysis results were calculated with Pearson correlation coefficients.

1.5 Measurement of TGF-β concentration in media

Medium supernatants were collected from the culture of HaCaT cells at different time points after TGF-β1 (R&D Systems, Catalog Number: 240-B-002) stimulations. Medium TGF-β1 concentrations were measured using a TGF beta 1 ELISA kit (invitrogen, Catalog Number: 88-8350) following the manufacturer's instructions.

1.6 Cell surface protein biotinylation and internalization

The internalization rate of TGFBR2 was determined using a reversible biotinylation method (Gabriel *et al.*, 2009) with a cell surface protein isolation kit (Thermo Scientific Pierce, Catalog Number: 89881). Cell surface proteins were labeled with biotinylation reagent (sulfo-NHS-SS-biotin) at a low temperature (4 °C). A portion of the cells were shifted to permissive conditions for trafficking at 37 °C, allowing the biotinylated receptor proteins to internalize. Another set of cells was maintained at the low temperature as controls for total surface protein at time=0 and for stripping control. After different time of internalization, cells were moved back to low temperature to stop internalization and the residual surface biotin was stripped off with a reducing agent, which cleaves the disulfide-coupled biotin. The internalized receptor proteins with biotin were protected and they were not stripped. After cell lysis, biotinylated proteins were pulled down by streptavidin affinity chromatography according to the manufacturer's protocol and they were measured using quantitative immunoblotting. By subtracting the internalized receptor protein from the total surface protein control at time=0, the remaining surface receptor protein at different time points was determined. The dynamics of the remaining surface TGFBR2 were fitted with an exponential decay function to estimate the internalization rate constant.

2. Supplementary Tables

Table S1: Summary of the derived parameter values based on experimental data

Parameter	Value and Unit	Annotation	Reference
ki	1.32 h^{-1}	receptor internalization rate constant	this work
kr	0.147 h^{-1}	receptor recycling rate constant	this work
kprod_T1R	1.57 nM/h	TGFBR1 production rate constant	this work
kprod_T2R	2.81 nM/h	TGFBR2 production rate constant	this work
kdeg_T1R	0.117 h^{-1}	constitutive degradation rate constant for TGFBR1	this work
kdeg_T2R	0.21 h^{-1}	constitutive degradation rate constant for TGFBR2	this work
kdeg_LRC	0.117 h^{-1}	constitutive degradation rate constant for ligand receptor complex (LRC)	this work
kdeg1_TGFb	20.8 h^{-1}	constitutive degradation rate constant for internalized TGF-β	(Kaminska *et al*, 2005; Wakefield *et al*, 1990)
kimp_Smad2	9.36 h^{-1}	SMAD2 nuclear import rate constant	(Schmierer *et al.*, 2008; Zi *et al.*, 2011)
kexp_Smad2	45.8 h^{-1}	SMAD2 nuclear export rate constant	(Schmierer *et al.*, 2008; Zi *et al.*, 2011)
kimp_PSmad2	53.3 h^{-1}	P-SMAD2 nuclear import rate constant	(Schmierer *et al.*, 2008)
Vcyt	$2.3e^{-12}$ l	Cytoplasm volume	(Schmierer *et al.*, 2008)
Vnuc	$1e^{-12}$ l	Nucleus volume	(Schmierer *et al.*, 2008)
Vmed	$2e^{-9}$ l	Extracellular space volume	(Zi *et al.*, 2011)

Note: The derived parameters were rounded with 3 effective digits

Table S2: Statistical analysis of the estimated parameters for the extended model

Parameter (Unit)	Best fit	95% confidence interval	Annotation
ka_LRC ($nM^{-2}h^{-1}$)	6.57019e+02	(6.25195e+02; 6.89870e+02)	Activation rate constant for ligand-receptor complex
kdiss_LRC (h^{-1})	1.74990e+00	(1.68428e+00; 1.82099e+00)	Dissociation rate constant for ligand-receptor complex
k_NFR ($nM^{-1}h^{-1}$)	6.04124e-02	(5.89965e-02; 6.18755e-02)	Rate constant for negative feedback regulation
kdeg2_TGFb (h^{-1})	1.38784e-01	(1.25339e-01; 1.52663e-01)	Degradation of TGF-β through non-specific binding
kpho_Smad2 ($nM^{-1}h^{-1}$)	1.68603e+00	(1.64520e+00; 1.72687e+00)	SMAD2 phosphorylation rate constant
kdepho_Smad2 (h^{-1})	1.94430e+00	(1.89266e+00; 1.99747e+00)	SMAD2 dephosphorylation rate constant
kon_ns (h^{-1})	3.18848e-01	(3.09881e-01; 3.27816e-01)	Association rate constant for non-specific binding of TGF-β
koff_ns (h^{-1})	1.00995e+00	(9.73658e-01; 1.04783e+00)	Dissociation rate constant for non-specific binding of TGF-β
tau (h)	2.00973e+00	(1.95320e+00; 2.06782e+00)	Delay time between negative feedback regulation and P-Smad2n

Note: The estimated parameters were rounded with 6 effective digits.

Table S3: Initial conditions of the extended model

Species	Initial Conditions (Unit: nM)	Annotation
TGFb_ex	to be specified	TGF-β concentration in the medium
T1Rs	1.49	type I receptor at plasma membrane (cell surface)
T1Rc	13.4	type I receptor in early endosome
T2Rs	1.49	type II receptor at plasma membrane (cell surface)
T2Rc	13.4	type II receptor in early endosome
LRCs	0	LRC at plasma membrane (cell surface)
LRCc	0	LRC in early endosome
Smad2c	92.4	SMAD2 in cytoplasm
Smad2n	43.4	SMAD2 in nucleus
PSmad2c	0	phosphorylated SMAD2 monomer in cytoplasm
PSmad2n	0	phosphorylated SMAD2 monomer in nucleus
TGFb_c	0	TGF-β in early endosome
TGFb_ns	0	non-specific bound TGF-β
z1	0	linear chain variable for the delay between PSmad2n and negative feedback regulation (NFR)
z2	0	linear chain variable for the delay between PSmad2n and negative feedback regulation (NFR)

Table S4: Ordinary differential equations of the extended model

```
d[TGFb_ex]/dt = koff_ns*[TGFb_c] - kon_ns*[TGFb_ex] -
                Vcyt*kass_LRC*[TGFb_ex]*[T1Rs]*[T2Rs]/Vmed
d[T1Rs]/dt = kr*[T1Rc] - ki*[T1Rs] - kass_LRC*[TGFb_ex]*[T1Rs]*[T2Rs]
d[T1Rc]/dt = kprd_T1R - kdeg_T1R*[T1Rc] + kdiss_LRC*[LRCc] + ki*[T1Rs] - kr*[T1Rc]
d[T2Rs]/dt = kr*[T2Rc] - ki*[T2Rs] - kass_LRC*[TGFb_ex]*[T1Rs]*[T2Rs]
d[T2Rc]/dt = kprd_T2R - kdeg_T2R*[T2Rc] + kdiss_LRC*[LRCc] + ki*[T2Rs] - kr*[T2Rc]
d[LRCs]/dt = kass_LRC*[TGFb_ex]*[T1Rs]*[T2Rs] - k_NFR*[z2]*[LRCs] - ki*[LRCs]
d[LRCc]/dt = ki*[LRCs] - kdiss_LRC*[LRCc] - kdeg_LRC*[LRCc]
d[Smad2c]/dt = Vnuc*kexp_Smad2*[Smad2n]/Vcyt - kpho_Smad2*([LRCs]+[LRCc])*[Smad2c] -
               kimp_Smad2*[Smad2c]
d[Smad2n]/dt = Vcyt*kimp_Smad2*[Smad2c]/Vnuc + kdepho_Smad2*[PSmad2n] - kexp_Smad2*[Smad2n]
d[PSmad2c]/dt = kpho_Smad2*([LRCs]+[LRCc])*[Smad2c] - kimp_PSmad2*[PSmad2c]
d[PSmad2n]/dt = Vcyt*kimp_PSmad2*[PSmad2c]/Vnuc - kdepho_Smad2*[PSmad2n]
d[TGFb_c]/dt = kon_ns*[TGFb_ex] - koff_ns*[TGFb_c] - kdeg2_TGFb*[TGFb_c]
d[TGFb_ns]/dt = kdiss_LRC*[LRCc] - kdeg1_TGFb*[TGFb_ns]
d[z1]/dt = (2*([PSmad2n] - [z1]))/tau
d[z2]/dt = (2*([z1] - [z2]))/tau
```

3. Supplementary Figures

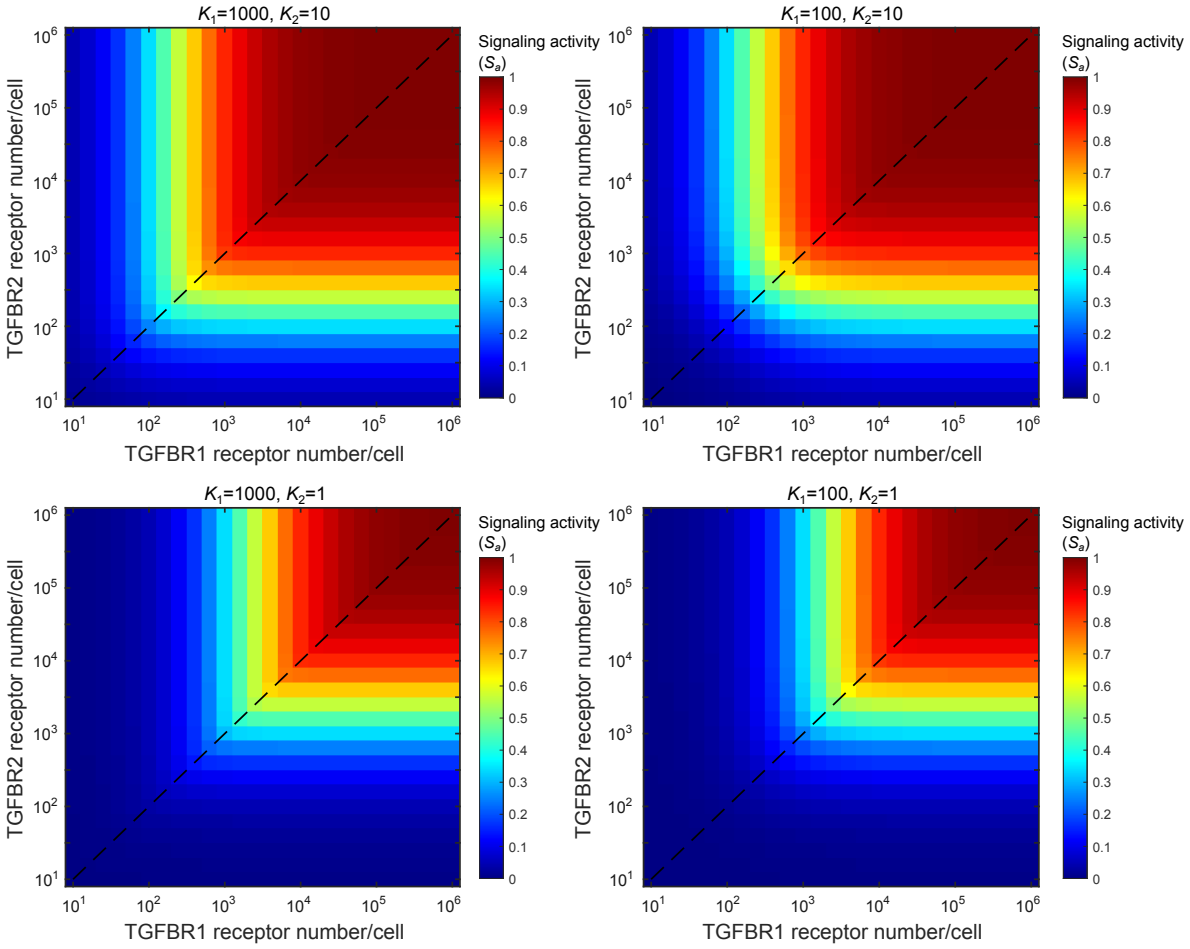

Figure S1. The predicted response profile of TGF-β signaling activities (S_a) at steady state with different choice of parameter values in the minimal model.

The ligand concentration L is set to be 0.1 nM (100 pM) in all the simulations.

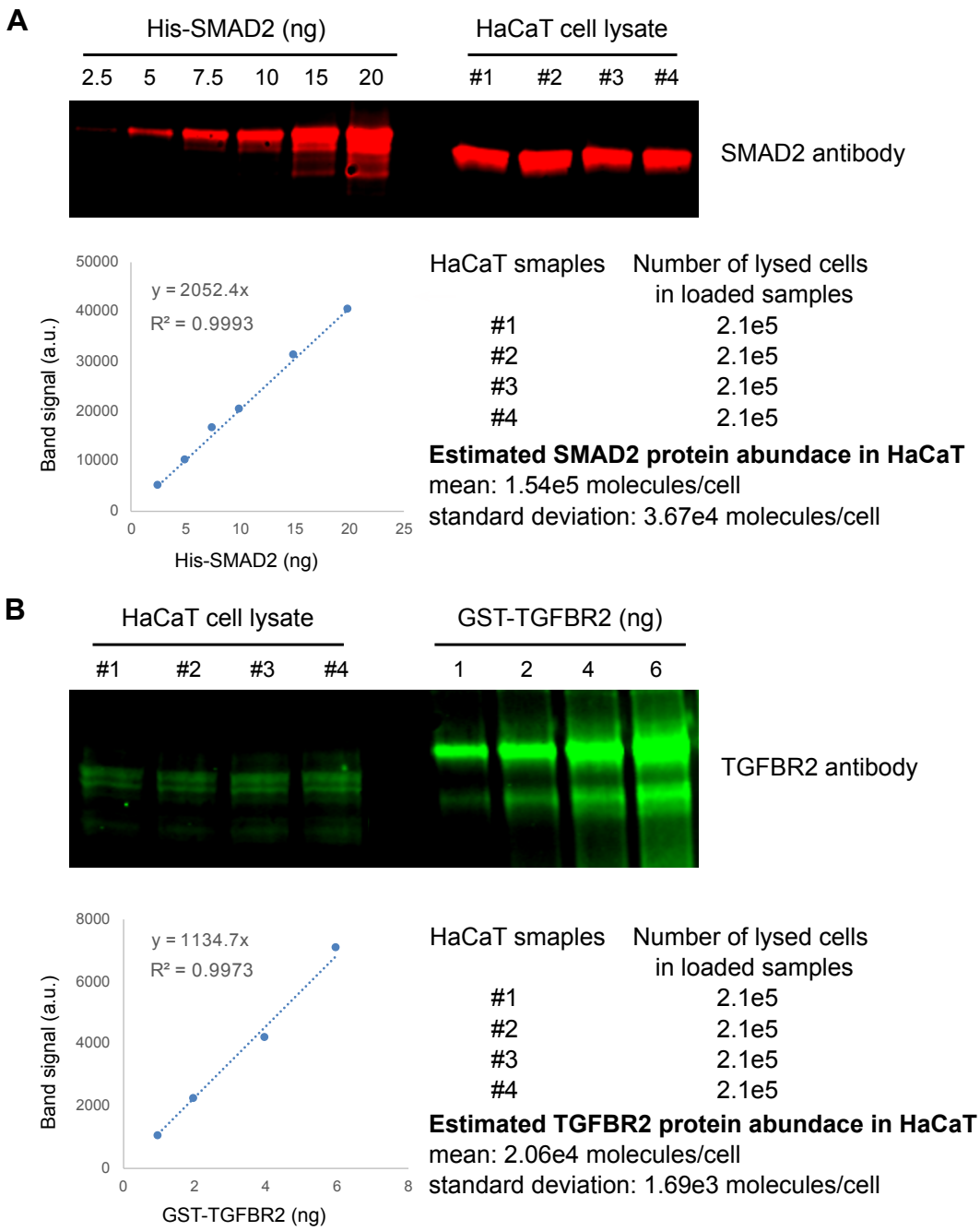

Figure S2. Measurement of protein abundance in HaCaT cells

Protein abundances were determined by quantitative immunoblotting with a series of recombinant protein standard.

(A) Estimation of SMAD2 abundance in HaCaT cells.

(B) Estimation of TGFBR2 abundance in HaCaT cells.

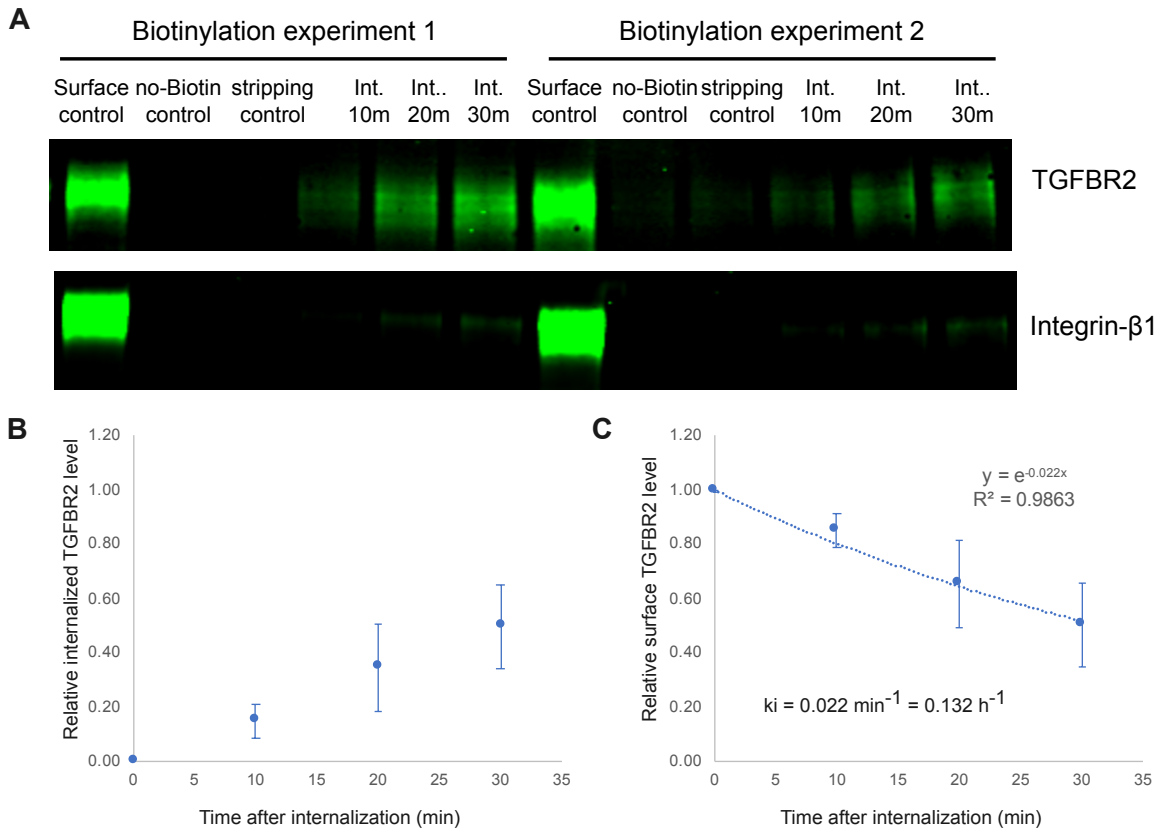

Figure S3. Measurement of TGFBR2 internalization rate constant in HaCaT cells

(A) Immunoblotting of the biotinylated protein samples. The biotin labeling procedure is described in the Supplementary Methods 1.6 section. Integrin-β1 is a cell surface protein marker.

(B) The kinetics of the internalized TGFBR2 were quantified from Western blots. The mean values and standard deviations from 4 experiments are shown.

(C) The kinetics of the remaining surface TGFBR2 after internalization. The mean values and standard deviations from 4 experiments are shown. The data points were fitted to an exponential decay function and the internalization rate constant (ki) was estimated.

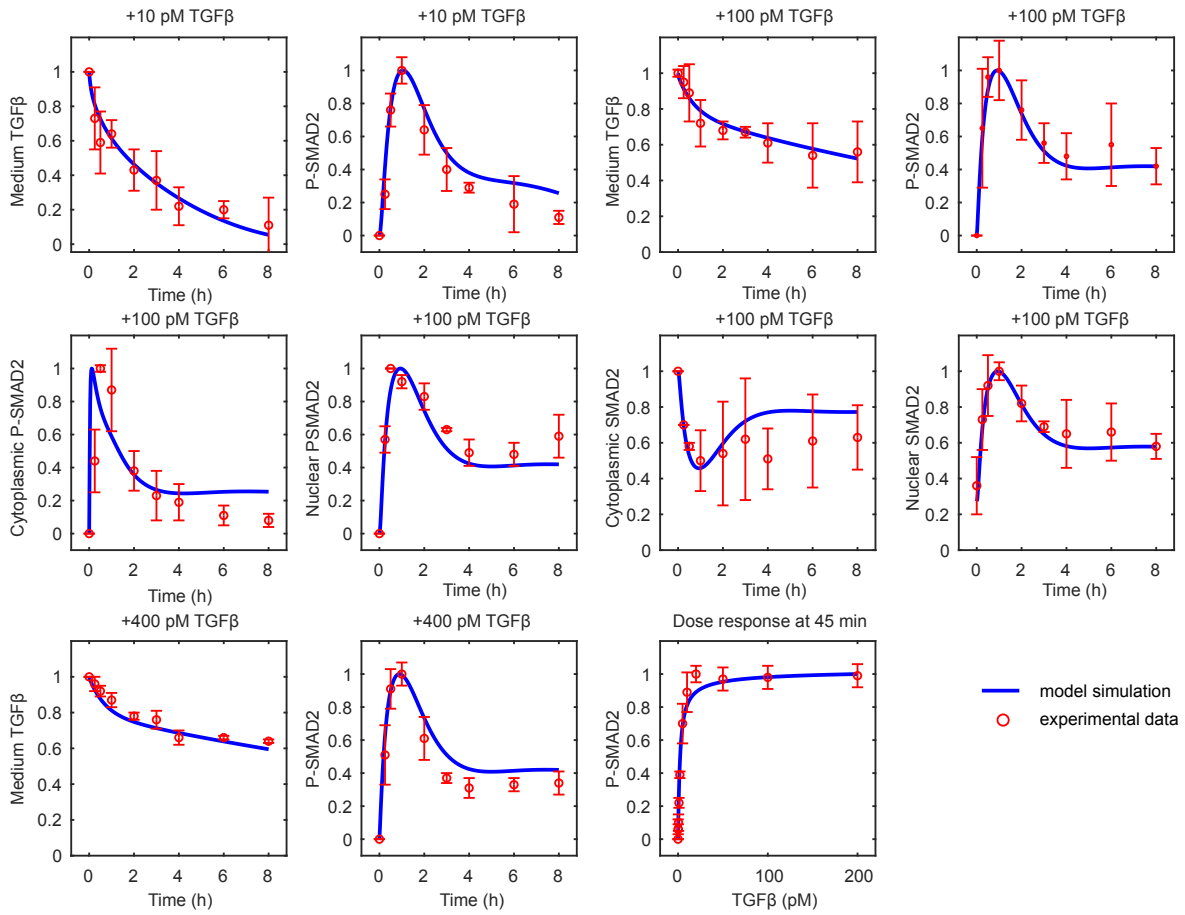

Figure S4: The comparison of model simulation results and experimental data

For comparison, the model simulation results and experimental data are normalized their maximum values in the corresponding data sets. The mean values and standard deviations of the normalized data from 3-5 experiments are shown.

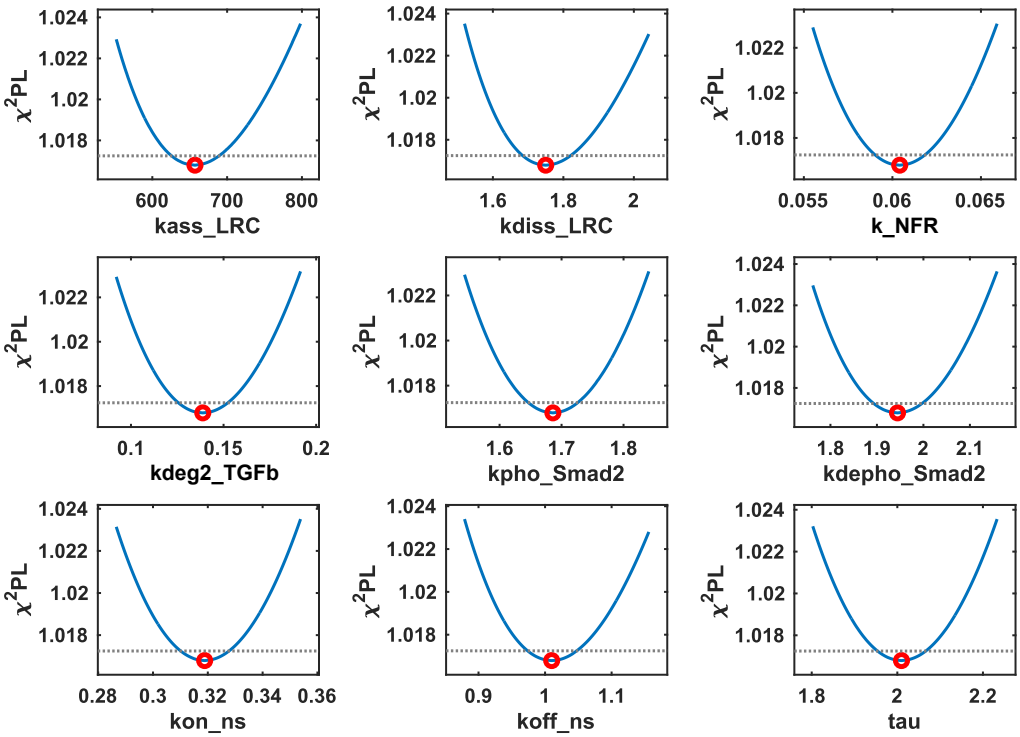

Figure S5. Profile likelihood analysis of estimated parameter values

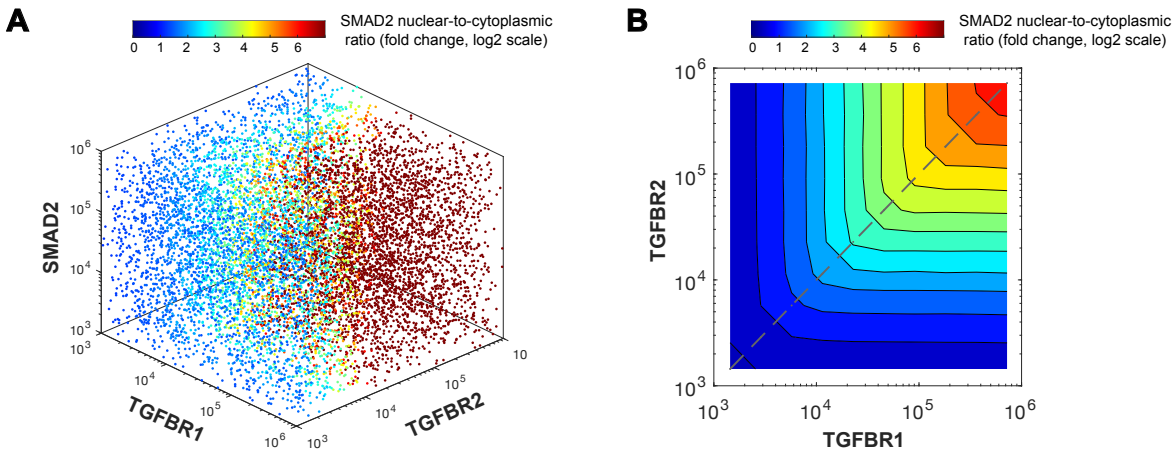

Figure S6: Computational modeling of the ratio of nuclear-to-cytoplasmic SMAD2 in the variation space of TGF-β receptors and SMAD2.

(A) Model predictions for the fold change in the ratio of nuclear-to-cytoplasmic SMAD2 (N2C) in the space of different amounts of TGFBR1, TGFBR2, and SMAD2 . Each point indicates a random set of protein amounts in the space. Colors indicate the fold change in the SMAD2 N2C ratio.

(B) Contour landscape of the fold change in SMAD2 N2C according to combinations of different amounts of TGFBR1 and TGFBR2. The dashed line indicates relationship between the fold change in SMAD2 N2C and the receptors when equal amounts of each receptor are present.

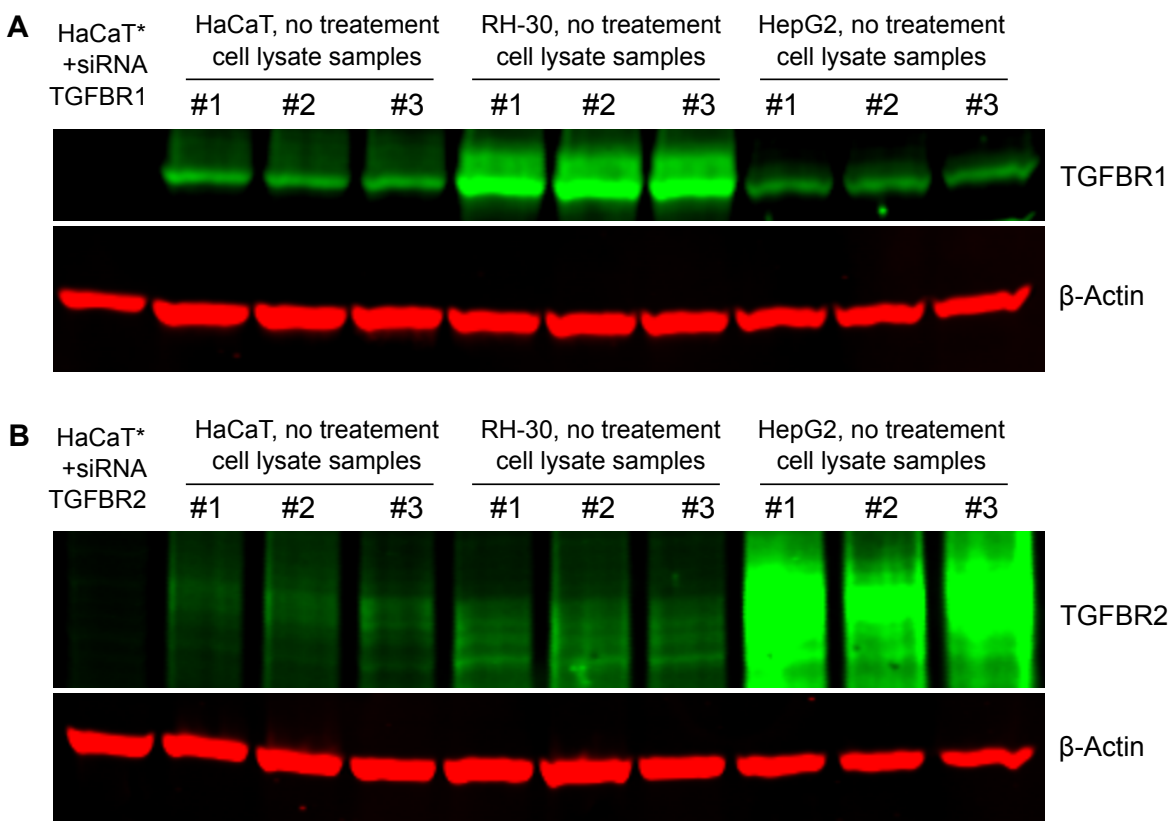

Figure S7. TGFBR1 and TGFBR2 protein expression profiles in HaCaT, RH-30 and HepG2 cells

Western blot for measuring relative abundance of TGFBR1 (A) and TGFBR2 (B) in HaCaT, RH-30 and HepG2 cells. The lysate samples from cells (#1, #2, #3) under normal growth conditions (no treatment) were loaded for each cell line. The lysate from HaCaT cells transfected with 100 nM TGFBR1 or TGFBR2 siRNA was loaded as a negative control (the first lane on the left) for the validation of the corresponding antibody.

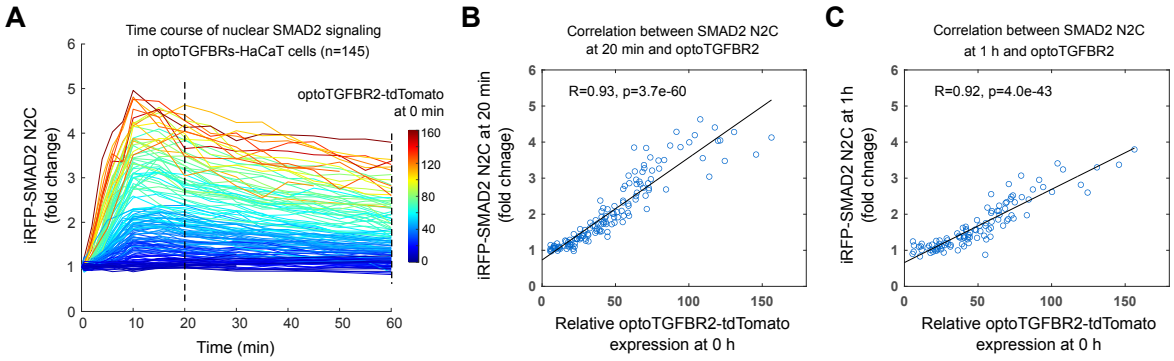

Figure S8. Quantification of SMAD2 N2C fold change responses in HaCaT cells expressing optoTGFBRs

(A) Time-course profiles of iRFP-SMAD2 N2C fold change responses in individual optoTGFBRs-HaCaT cells after one pulse of blue light stimulation (488 nm, 12.4 μW).

(B) The plot of iRFP-SMAD2 N2C fold change responses at 20 min versus the amount of optoTGFBR2-tdTomato before blue light stimulation (0 h) in optoTGFBRs-HaCaT cells.

(C) The plot of iRFP-SMAD2 N2C fold change responses at 1 h versus the amount of optoTGFBR2-tdTomato before blue light stimulation (0 h) in optoTGFBRs-HaCaT cells.

4. Supplementary References

Antebi YE, Linton JM, Klumpe H, Bintu B, Gong M, Su C, McCardell R, Elowitz MB (2017) Combinatorial Signal Perception in the BMP Pathway. *Cell* 170: 1184-1196 e1124

Deng D, Zi Z (2022) Absolute Quantification of TGF-beta Signaling Proteins Using Quantitative Western Blot. *Methods Mol Biol* 2488: 1-12

Di Guglielmo GM, Le Roy C, Goodfellow AF, Wrana JL (2003) Distinct endocytic pathways regulate TGF-beta receptor signalling and turnover. *Nat Cell Biol* 5: 410-421

Gabriel L, Stevens Z, Melikian H (2009) Measuring plasma membrane protein endocytic rates by reversible biotinylation. *J Vis Exp*

Kaminska B, Wesolowska A, Danilkiewicz M (2005) TGF beta signalling and its role in tumour pathogenesis. *Acta Biochim Pol* 52: 329-337

Kao SH, Wang WL, Chen CY, Chang YL, Wu YY, Wang YT, Wang SP, Nesvizhskii AI, Chen YJ, Hong TM *et al* (2015) Analysis of Protein Stability by the Cycloheximide Chase Assay. *Bio Protoc* 5

Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmuller U, Timmer J (2009) Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* 25: 1923-1929

Schmierer B, Hill CS (2005) Kinetic analysis of Smad nucleocytoplasmic shuttling reveals a mechanism for transforming growth factor beta-dependent nuclear accumulation of Smads. *Mol Cell Biol* 25: 9845-9858

Schmierer B, Tournier AL, Bates PA, Hill CS (2008) Mathematical modeling identifies Smad nucleocytoplasmic shuttling as a dynamic signal-interpreting system. *Proc Natl Acad Sci U S A* 105: 6608-6613

Su CJ, Murugan A, Linton JM, Yeluri A, Bois J, Klumpe H, Langley MA, Antebi YE, Elowitz MB (2022) Ligand-receptor promiscuity enables cellular addressing. *Cell Syst* 13: 408-425 e412

Vilar JM, Jansen R, Sander C (2006) Signal processing in the TGF-beta superfamily ligand-receptor network. *PLoS Comput Biol* 2: e3

Wakefield LM, Smith DM, Masui T, Harris CC, Sporn MB (1987) Distribution and modulation of the cellular receptor for transforming growth factor-beta. *J Cell Biol* 105: 965-975

Wakefield LM, Winokur TS, Hollands RS, Christopherson K, Levinson AD, Sporn MB (1990) Recombinant latent transforming growth factor beta 1 has a longer plasma half-life in rats than active transforming growth factor beta 1, and a different tissue distribution. *J Clin Invest* 86: 1976-1984

Zi Z (2011) SBML-PET-MPI: a parallel parameter estimation tool for Systems Biology Markup Language based models. *Bioinformatics* 27: 1028-1029

Zi Z, Feng Z, Chapnick DA, Dahl M, Deng D, Klipp E, Moustakas A, Liu X (2011) Quantitative analysis of transient and sustained transforming growth factor-beta signaling dynamics. *Mol Syst Biol* 7: 492

Zi Z, Klipp E (2007) Constraint-based modeling and kinetic analysis of the Smad dependent TGF-beta signaling pathway. *PLoS One* 2: e936