学位論文

Extraction of simple relation in growth factor-specific multiple-input and multiple-output system in cell-fate decision by backward elimination PLS regression

（細胞運命決定機構多入力多出力経路における

backward elimination PLS regression法を用いた解析）

平成24年　12月　博士(理学)申請

東京大学大学院理学系研究科

生物化学専攻

秋元　勇輝

1. ABSTRACT

Cells use common signaling molecules for the selective control of downstream gene expression and cell-fate decisions. The relationship between signaling molecules and downstream gene expression and cellular phenotypes is a multiple-input and multiple-output (MIMO) system and is difficult to understand due to its complexity. For example, it has been reported that, in PC12 cells, different types of growth factors activate MAP kinases (MAPKs) including ERK, JNK, and p38, and CREB, for selective protein expression of immediate early genes (IEGs) such as c-FOS, c-JUN, EGR1, JUNB and FOSB, leading to cell differentiation, proliferation and cell death; however, how multiple-inputs such as MAPKs and CREB regulate multiple-outputs such as expression of the IEGs and cellular phenotypes remains unclear. To address this issue, I employed a statistical method called partial least squares (PLS) regression, which involves a reduction of the dimensionality of the inputs and outputs into latent variables and a linear regression between these latent variables. I measured 1,200 data points for MAPKs and CREB as the inputs and 1,900 data points for IEGs and cellular phenotypes as the outputs, and I constructed the PLS model from these data. The PLS model highlighted the complexity of the MIMO system and growth factor-specific input-output relationships of cell-fate decisions in PC12 cells. Furthermore, to reduce the complexity, I applied a backward elimination method to the PLS regression, in which 60 input

variables were reduced to 5 variables, including the phosphorylation of ERK at 10 min, CREB at 5 min and 60 min, AKT at 5 min and JNK at 30 min. The simple PLS model with only 5 input variables demonstrated a predictive ability comparable to that of the full PLS model. The 5 input variables effectively extracted the growth factor-specific simple relationships within the MIMO system in cell-fate decisions in PC12 cells.

2. TABLE OF CONTENTS

3. INTRODUCTION

3.1 Multi-input and multi-output (MIMO) system

Cells use common signaling molecules to selectively control downstream gene expression and cell-fate decisions. The relationship between signaling molecules and gene expression or cellular phenotypes was previously thought to be a one-to-one correlation. However, recent studies have revealed that signaling molecules and downstream gene expression levels and cellular phenotypes are mutually connected, and their relationship appears to be a multiple-input and multiple-output (MIMO) system [1-6].

For example, PC12 cells, an adrenal chromaffin cell line, have been shown to undergo cell differentiation, proliferation and death in response to various growth factors (Fig. 1) [7-11]. Nerve growth factor (NGF) and pituitary adenylate cyclase-activating polypeptide (PACAP) induce differentiation and neurite extension, epidermal growth factor (EGF) induces cell proliferation, and the protein synthesis inhibitor anisomycin induces cell death [9-18]. These growth factors use common signaling pathways. NGF induces differentiation via the receptor-tyrosine kinase, TrkA, which causes a sustained activation of downstream signaling pathways, including both the ERK and AKT pathways [9,10,19]. PACAP activates the G protein type receptor

PAC1, which phosphorylates CREB through cAMP-dependent protein kinase A (PKA) activation, leading to cell differentiation [10,20,21]. EGF induces cell proliferation by activating the tyrosine kinase receptor EGFR, which transiently activates the ERK and AKT pathways [9,15,22,23]. Anisomycin activates mitogen-activated protein kinase (MAPK) cascades, such as JNK and p38, as well as caspases, including Caspase 3, which leads to cell death. Moreover, signaling molecules transmit information downstream via the protein expression of immediate early genes (IEGs), including c-Fos, c-Jun, EGR1, FosB and JunB [24,25]. Thus, a wide range of growth factors encode information into specific temporal patterns and combinations of the multiple-inputs, such as MAPKs and CREB, that are further decoded by the multiple-outputs, such as expression of IEGs to exert biological functions in PC12 cells. However, the essential and simple relationship in the MIMO system remains to be elucidated.

3.2 Partial least squares (PLS) regression method

The statistical analysis called partial least square (PLS) regression was originally developed for econometrics and chemometrics [26]. The PLS regression has been applied to the biological MIMO system, signaling molecules and cellular phenotypes such as apoptotic signaling pathways [1-3,27-30]. The application of PLS regressions to the MIMO system involve reducing the dimensionality of the inputs and outputs into latent variables, which are selectively weighted linear combinations of the inputs and outputs. A linear regression is then performed between the latent variables of the inputs and the outputs. Because the latent variables explain the characteristics of the data using a smaller number of latent variables than the number of original variables, those latent variables are called principal components. This method can relate multiple signaling molecules to multiple downstream functions based on heterogeneous multivariate signaling in response to various growth factors. The principal components in the PLS model consist of linear combinations of all variables. Because the number of variables is not reduced and complexity still remains, the result of the PLS regression is difficult to intuitively understand. To facilitate a better understanding of the MIMO system, a method for further reducing the number of variables is required.

3.3 Introduction of this study

In this study, I employed PLS regression and analyzed the complex relationship between the phosphorylation of signaling molecules and the expression of IEGs and cellular phenotypes in PC12 cells in response to various growth factors. The PLS model highlighted the complex characteristics of the MIMO system and growth factor-specific input-output relationships of cell-fate decisions in PC12 cells. Furthermore, to reduce the number of input variables in the PLS model, I applied a backward elimination method to the PLS regression model and obtained a simple PLS model with 5 input variables. The simple PLS model with only 5 input variables demonstrated a predictive ability comparable to that of the full PLS model with 60 variables. The 5 input variables effectively highlight the simple relationships within the MIMO system and growth factor-specific input-output relationships of cell-fate decisions in PC12 cells. the simple relationships can be intuitively understood and easily observed by visual inspection.

4 MATERIALS AND METHODS

4.1 Antibodies

Mouse anti-phospho-ERK1/2 (Thr 202/Tyr 204) monoclonal antibody (mAb) (#9106), rabbit anti-phospho-CREB (Thr 133) mAb (#9198), rabbit anti-phospho-JNK (Thr183/Tyr185) mAb (#4668), rabbit anti-EGR1 mAb (#4154), rabbit anti-c-JUN mAb (#9165), rabbit anti-c-FOS mAb (#2250), rabbit anti-JUNB mAb (#3753), rabbit anti-FOSB mAb (#2251), and rabbit anti-cleaved Caspase 3 mAb (#9664) were purchased from Cell Signaling Technology (Beverly, MA). Rabbit anti-phospho p38 mAb (#v1211) was purchased from Promega (Madison, WI).

4.2 Cell culture and treatments

PC12 cells (kindly provided by Masato Nakafuku, Cincinnati Children's Hospital Medical Center, Ohio) were cultured at 37°C under 5% CO_2 in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum and 5% horse serum (Invitrogen, Carlsbad, CA). Cells were stimulated using recombinant mouse β-NGF (R&D Systems, Minneapolis, MN), EGF (Roche, Mannheim, Germany), PACAP (Sigma, Zwijndrecht, The Netherlands), or anisomycin (EMD Biosciences, Inc., San Diego, CA) as previously described [31]. I used a low dose of anisomycin (50 nM)

to activate p38 and JNK without inhibiting translation. For the QIC assays, cells were seeded at a density of 10^4 cells per well in 96-well poly-L-lysine–coated glass-bottomed plates (Thermo Fisher Scientific, Pittsburgh, PA) and then starved in DMEM containing 25 mM HEPES and 0.1% bovine serum albumin for approximately 18 h before stimulation. Cells seeded in 96-well microplates were stimulated by replacing the starvation medium with the medium containing the stimulant using a liquid handling system (Biomek® NX Span-8, Beckman Coulter, Fullerton, CA) with an integrated heater-shaker (Variomag®, Daytona Beach, FL) and robotic incubator (STX-40, Liconic, Mauren, Liechtenstein). All of the cells within a plate were fixed simultaneously to prevent their exposure to formaldehyde vapor during the treatment.

4.3 QIC (Quantitative Image Cytometry)

QIC was performed as previously described [32]. Briefly, after growth factor stimulation, the cells were fixed, washed with phosphate-buffered saline (PBS), and permeabilized with blocking buffer (0.1% Triton X-100, 10% fetal bovine serum in PBS). The cells were then washed and incubated for 2 h with primary antibodies diluted in Can Get Signal immunostain Solution A (Toyobo, Osaka, Japan). The cells were washed three times and then incubated for 1 h with secondary antibodies. After

immunostaining, the cells were treated for nucleus and cytoplasm staining by incubating with Hoechst 33342 (Invitrogen, Carlsbad, CA) and CellMask Deep Red stain (Invitrogen Carlsbad, CA), respectively. The images of the stained cells were acquired using a CellWoRx (Thermo Fisher Scientific, Pittsburgh, PA) automated microscope with a ×10 objective. For QIC analyses, I acquired two different fields for each well and obtained 1238 ± 356 (mean ± SD) cells for each well. All liquid handling for the 96-well microplates was performed using a Biomek® NX Span-8 liquid handling system (Beckman Coulter, Fullerton, CA).

4.4 Quantitative analysis of the neurite length

PC12 cells (0.5×10^4 cells/well) were fixed using a 10% formalin solution (Wako, Osaka, Japan) for 10 minutes. Cells were washed with phosphate-buffered saline (PBS), incubated with 1 μg/ml Hoechst 33342 solution (Life Technologies, Carlsbad, CA) and 1 μg/ml CellMask (Life Technologies) in PBS for 1 hour at room temperature and then washed with PBS. Images were captured using a CellWoRx microscope (Thermo Fisher Scientific, Rockford, IL). Using the CellMask signal as the neuronal cell image and the Hoechst signal as the nuclear image, the lengths of the neurites were measured with the

NeuroTracer NIH ImageJ plug-in [33]. The length of the neurites of cells under each stimulation condition was represented as the averaged neurite lengths of cells.

4.5 Cell viability assay (mitochondrial respiratory chain activity)

Cell viability was determined by measuring mitochondrial reduction of the MTS dye [3-(4,5-dimethythiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium] reagent into a soluble formazan product (Promega, Madison, WI) for the quantification of the respiratory chain activity of the mitochondria. PC12 cells were plated on Poly-L-lysine (PLL)-coated 96-well plates. After incubation, cells were treated with MTS solution (1 mg/ml), and the intracellular soluble formazan produced by the cellular reduction of the MTS was determined by recording the absorbance of each 96-well plate using a Mithras LB940 microplate reader (Berthold Japan, Tokyo, Japan) at a wavelength of 490 nm.

4.6 Cell death assay (Activity of Caspase 3)

Cell death was determined by measuring of activation of Caspase 3 as cleaved Caspase 3 using western blot assays. Cell lysates were subjected to standard sodium dodecyl

sulfate polyacrylamide gel electrophoresis (SDS-PAGE). After fractionation by SDS-PAGE and transfer to nitrocellulose membranes, the blots were incubated with antibodies directed at Cleaved Caspase 3 (1:1000 dilution; Cell Signaling Technology, Danvers, MA, #9664) or pan ERK1/2 (1:2000 dilution; Cell Signaling Technology, #9102) followed by incubation with horseradish peroxidase-conjugated rabbit IgG (GE Healthcare, Buckinghamshire, England). Chemiluminescence was detected using Immobilon Western (Millipore, Billerica, MA). The resulting image was captured with a luminescent image analyzer LAS-4000 (Fujifilm, Tokyo, Japan). The signal intensity was quantified using Phoretix 1D software (TotalLab Ltd, Newcastle upon Tyne, UK).

4.7 Cell cycle assay (cell cycle S-phase fraction)

Cell cycle S-phase fraction was determined by the incorporation of 5-*ethynyl*-2′-deoxyuridine (EdU) using the Click-iTEdU Cell Proliferation AssayKit (Invitrogen). The PC12 cells were incubated with 10 μM EdU for 1 hour before fixation, permeabilization, and EdU staining, which were performed according to the kit manufacturer's protocol. The proportion of nucleated cells incorporating EdU was determined by fluorescence microscopy using a CellWoRx microscope (Thermo Fisher Scientific), and the fraction of cells in the S-phase was measured using MATLAB

software (MathWorks).

4.8 Partial Least Squares Regression method

The partial least squares regression method used in this study was described in a previous publication [1,34]. Partial least squares regression is a predictive two-block regression method based on estimated latent variables and is applied for the simultaneous analysis of two data sets. The purpose of PLS regression is to build a linear model that enables the prediction of outputs from inputs. In this study, PLS regression analysis was performed using the MATLAB (Mathworks) software suite. Data were normalized by mean centering and variance scaling the different measurements.

Let $\mathbf{X}$ be the (20×60) input data matrix for PLS modeling. The i-th $(1\le i\le 20)$ row vector of $\mathbf{X}$ is the input vector $\mathbf{x}_i^t$ where $\cdot^t$ denotes the transpose of a vector or matrix $\cdot$. The input vector $\mathbf{x}_i$ consists of 60 metric variables which are time course points of MAPKs and CREB. I used 20 doses of stimuli to obtain 20 samples as input vectors, hence, i corresponds to the attribute of stimulation. Let $\mathbf{Y}$ be the (20×95) output matrix. The j-th $(1\le j\le 95)$ column vector $\mathbf{y}_j$ of $\mathbf{Y}$ is the output vector of

which each variable correspond to the attribute of stimulation, and j corresponds to the attribute of time course point of the IEGs or phenotype.

The PLS model can be understood as two steps regression model developed simultaneously. The first step can be considered as consisting of the development of outer relations ($\mathbf{X}$ and $\mathbf{Y}$ metric individually). These data matrix were decomposed in latent variables plus a residue metrics. The sub-matrices can be represented as the product of the scores and the loadings which can be re-grouped in independent matrices for the $\mathbf{X}$ and $\mathbf{Y}$ metrics as follows:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^t + \mathbf{E}$$

$$\mathbf{Y} = \mathbf{U}\mathbf{Q}^t + \mathbf{F}$$

where $\mathbf{T}$ and $\mathbf{U}$ are the scores, and $\mathbf{P}$ and $\mathbf{Q}$ are the loadings, for the $\mathbf{X}$ and $\mathbf{Y}$ metrics, respectively. The matrices $\mathbf{E}$ and $\mathbf{F}$ correspond to the residues associated with the PLS modeling. The second step is a linear inner relation linking between $\mathbf{T}$ and $\mathbf{U}$,

$$\mathbf{U} = \mathbf{T}\mathbf{D} + \mathbf{H}$$

where $\mathbf{D}$ is the diagonal matrix and $\mathbf{H}$ denotes the residual matrix.

Eventually, PLS regression is yielded by

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \boldsymbol{\varepsilon}$$

where $\mathbf{B}$ is the matrix of regression coefficients

$$\mathbf{B} = \mathbf{X}^t\mathbf{U}\left(\mathbf{T}^t\mathbf{X}\mathbf{X}^t\mathbf{U}\right)^{-1}\mathbf{T}^t\mathbf{Y}$$

and $\boldsymbol{\varepsilon}$ is the residual matrix.

The optimum number of components were determined by minimizing MSE of Leave-one-out cross-validation as LOOCV MSE[35],

$$\text{LOOCV MSE} = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\sqrt{\left(\hat{y}_{ij}^{\setminus i} - y_{ij}\right)^2}$$

where y_{ij} is the (i, j) variables of the ($N \times M$) matrix $\mathbf{Y}$ and $\hat{y}_{ij}^{\setminus i}$ is the prediction for y_{ij} by PLS model which was trained by the data set removed i-th sample.

4.9 Backward elimination PLS regression

I applied a backward elimination variable selection method [36] for PLS because backward elimination can improve the accuracy of a PLS model. The backward elimination PLS regression began with the full PLS model with the input vector of *M* variables. I define the LOOCV MSE removing k-th variable from input vector as

$$\text{LOOCV MSE } \setminus k = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\sqrt{\left(\hat{y}_{ij}^{\setminus ik} - y_{ij}\right)^2}$$

$\hat{y}_{ij}^{\setminus ik}$ is the prediction for y_{ij} by PLS model which was trained by the data set removed i-th sample and k-th variable of input vector. In the n-th step of procedure, the elimination k^*-th

variable of input vector was determined by minimizing LOOCV MSE $\backslash k$ for $1 \le k \le M-n-1$

$$k^* = \arg\min_k \text{LOOCV MSE } \backslash k$$

then, k^*-th variable was eliminated from the input vector, and redefined the new input vector of which k^*-th variable was eliminated for $(n+1)$-th step. I iterated this procedure until the 4 variables remained, which is the same number as the principal components.

4.10 Variable Importance in Projection (VIP)

We calculated the *Variable Importance in Projection* (VIP) of [41] to summarize each variable contribution to the model. VIP describes which ***X*** variables characterize the ***X*** block well and which variables correlate with ***Y***. VIP values summarize the overall contribution of each input variable to the PLS model, summed over all components and weight~~ed~~ according to the ***Y*** variation accounted for by each component. VIP is calculated as follows:

$$VIP_k = \left\{ \sum_{h=1}^{m} \sum_{j=1}^{M} R^2(\mathbf{y}_j, \mathbf{t}_h) w_{hk}^2 \Big/ \sum_{h=1}^{m} \sum_{j=1}^{M} R^2(\mathbf{y}_j, \mathbf{t}_h) \right\}^{1/2}$$

,

for each *k*-th input variable *k=1,…, p,* where $R^2(a,b)$ stands for the squared correlation between items in vector ***a*** and ***b***, $\mathbf{t}_h$ is the *h*-th column vector of the score

matrix ***T***, m is the number of principle components, and w_{hk} is the (h,k) element of the weight matrix ***W***.

5 RESULTS

5.1 The multiple-input and multiple-output system in PC12 cells

I stimulated PC12 cells with various doses of NGF, PACAP, EGF, and anisomysin and measured time series data of the phosphorylation of signaling molecules, including ERK (pERK), CREB (pCREB), JNK (pJNK), AKT (pAKT) and p38 (pp38) (Figs. 2A and 3), protein expression levels of immediate early genes (IEGs), including c-FOS, c-JUN, EGR1, JUNB and FOSB (Fig. 2B, Fig. 3), and cellular phenotypes, including neurite lengths, cell viability (respiratory chain activity of mitochondria), cell cycle (S-phase fraction) and cell death (Caspase3 activity) (Figs. 2B, Fig 3). Among many asssays for detection of apoptosis, we chose caspase3 activity because of the availability for high-throughput assay. NGF, PACAP, EGF, and anisomycin induced distinct temporal patterns and combinations of phosphorylation of signaling molecules, IEGs and cellular phenotypes. I did not observe obvious cell proliferation by EGF stimulation under the conditions. Here, I regarded the phosphorylation of signaling molecules as the inputs and the IEGs and cellular phenotypes as the outputs.

5.2 Construction of the PLS model

I applied PLS regression to infer the MIMO system underlying cell-fate decisions in PC12 cells (Fig. 4) [37]. PLS regression is a regression method for use with the MIMO system that involves reducing the dimensionality of the inputs and outputs into latent variables, which are selectively weighted linear combinations of the inputs and the outputs, denoted as principal components. A linear regression is then performed between the principal components of the inputs and the principal components of the outputs(see Materials and Methods). The principal component was determined to maximize the capture of the covariance between the input latent variable and the output latent variable, and the principal components were orthogonal to one another. Thus, the PLS regression predicts multiple output variables from multiple input variables.

The input data set consisted of 20x60 matrices of phosphorylation of signaling molecules at 12 time points (60 variables) that involved 5 doses of 4 growth factors (Fig. 4A, Tables 1, see Materials and Methods). The output dataset consisted of 20x95 matrices of the protein expression of 5 IEGs with 12 time points and cellular phenotypes of neurite lengths, cell viability and cell death at 9 time points and cell cycle at 8 time points (95 variables) that involved 5 doses of 4 growth factors (Fig. 4A, Tables 1, see Materials and Methods). I used the LOOCV MSE (leave-one-out cross validation

mean squared error) [35] as the estimated prediction error to optimize the number of model dimensions (Fig. 4B) and determined that the LOOCV MSE was minimized with four principal components. The first principal component captured approximately 45% of the total variance, the first and second principal components captured 67% of the total variance, and the first to fourth principal components captured approximately 85% of the total variance (Fig. 4C). The Pearson correlation coefficient between the measured outputs and the predicted outputs in the four principal components was 0.94 (Fig. 4D).

PLS regression characterizes the input-output system using "loadings", which are the vector projections of the unit direction vector of the principal component on each variable, and "scores", which are the projections of sample points on the principal component direction. In short, loadings represent the contribution of each variable to the principal component, and scores represent growth factor-specificity in the principal component.

In the input loadings of the first principal component, pERK, pCREB and pAKT were positive, whereas pJNK and pp38 were negative (Fig. 5A), indicating their opposing contributions to the first principal component. In the input scores of the first principal component, NGF, PACAP and EGF were positive, whereas anisomycin was

negative (Fig. 5B), indicating that anisomycin was inversely correlated to the other growth factors in the first principal component. In the output loadings of the first principal component of the outputs, the neurite length, cell viability, and all IEGs were positive, whereas cell death and cell cycle were negative (Fig. 5C). In the output scores, NGF and PACAP were positive, whereas anisomycin was negative (Fig. 5D), indicating that anisomycin was inversely correlated with the other growth factors in the first principal component. These results indicate that the first PLS component divided the data into cell survival/differentiation and cell death. In the input loadings of the second principal component, pCREB and pJNK, late pAKT, and pp38 were positive, whereas pERK and early pAKT were negative. In the input scores, PACAP and anisomycin were positive, whereas NGF and EGF were negative. In the output loadings of the second principal component, cell death, cell cycle and c-FOS, JUNB and FOSB were positive, whereas the neurite length, cell viability, c-JUN and EGR1 were negative. In the output scores, PACAP and anisomycin were positive, whereas NGF and EGF were positive. These observations indicate that the second principal component divided the data into receptor-type tyrosine kinase (NGF and EGF) and other receptor types (PACAP and anisomycin). The third principal component divided the data into higher and lower doses of stimuli, and the fourth principal component divided the data into EGF and

others.

The first and second principal components captured approximately 67% of the variance (Fig. 4C), and I plotted both the loadings and scores on these two principal components (Fig. 5E-H). In the first quadrant of the loadings, pCREB and late pAKT in the inputs were correlated with c-FOS, JUNB and FOSB in the outputs (Fig. 5E, G). In the second quadrant of the loadings, pJNK and pp38 in the input were correlated with cell death and cell cycles in the outputs (Fig. 5E, G). In the fourth quadrant, pERK and early pAKT in the inputs were correlated with the neurite length, cell viability, EGR1 and c-JUN in the outputs. In the scores, the first, second, and forth quadrant involved PACAP, anisomycin and NGF, respectively (Fig. 5F, H), indicating that these quadrants represent growth factor-specific input-output relationships. Thus, the loadings and scores of the first and second principal components highlight characteristics of the MIMO system and growth factor-specific input-output relationships of cell-fate decisions in PC12 cells, respectively.

5.3 Validation of the PLS model

I validated the PLS model using additional experimental data including inhibitors of signaling molecules. I perturbed the activity of signaling molecules by adding inhibitors and measured the inputs and cellular phenotypes. I used PD0325901 (MEK inhibitor), H89 (PKA inhibitor), SP600125 (JNK inhibitor), SB203580 (p38 inhibitor), and LY294002 (PI3K inhibitor) that are thought to inhibit pERK, pCREB, pJNK, pp38 and pAKT, respectively. Because of prominent effects of NGF and PACAP on neurite lengths and MTS, and of anisomycin on cell cycle and cell death, we chose these stimuli for validation by the inhibitor experiments.

Using the measured inputs in the presence of the inhibitors, the PLS model predicted the neurite length in the presence of the inhibitors (Fig. 6A, B). The predicted neurite length showed a high correlation (r>=0.7) with the measured neurite length in response to NGF (r=0.78) and PACAP (r=0.82). The predicted c-FOS, c-JUN and EGR1 expression levels showed high correlations with the measured data in response to NGF (r=0.93 for c-FOS, r=0.94 for c-JUN, r=0.81 for EGR1) (Fig. 6A, B). The predicted c-FOS FOSB and JUNB expression levels were highly correlated with the measured data in response to PACAP (r=0.91 for c-FOS, r=0.93 for FOSB, r=0.78 for JUNB) (Fig. 6B). However,

we observed a low correlation (r=<0.5) between JUNB and NGF (r=0.50) and between EGR1 and PACAP (r=-0.1) (Fig. 6A, B). This lack of correlation may have attributed to the low expression levels of the IEGs (Table S2) rather than a low predictive ability of the PLS model. Thus, the PLS model predicted data that correlated well with the inhibitor experimental data regarding the neurite length and protein expression levels of the IEGs.

There were no significant correlations regarding the cell viability (r=0.32 for PACAP), or cell death (r=0.16 for anisomycin) (Fig. 6C), indicating that the PLS model shows a low predictive ability for the cell cycle and cell death in the presence of the inhibitors. This low predictive ability may occur due to the existence of nonlinearity between the inputs and the outputs that cannot be modeled by PLS regression or because the inputs of the PLS model may overlook essential signaling molecules required for representing

5.4 Reduction of the PLS model by backward elimination PLS regression

Although PLS regression involves reducing the dimensionality of the inputs into principal components, the principal components continue to involve all of the inputs, making it difficult to intuitively understand how the combination of inputs correlates with the outputs. To facilitate an intuitive understanding of the relationship by reducing the number of variables, I employed a backward elimination variable selection process [36] in the PLS regression called the backward elimination PLS regression method. I constructed a set of single variable-eliminated PLS models and estimated the LOOCV MSE of each model. I eliminated the variable with the minimum LOOCV MSE. I then reconstructed each single variable-eliminated PLS model using the remainder of the variables. I iterated this step sequentially. As a result, as the model was reduced in order, the LOOCV MSE decreased and reached a minimum with 22 input variables (Fig. 7A-C, Table 2). The 22-input variable PLS model provided the best predictive accuracy of any PLS model. The eliminated inputs are also considered to be factors that decrease or do not affect the predictive accuracy of the model. As the input variables were sequentially eliminated, the error increased and reached a level similar to that of the full PLS model when 5 input variables remained. The last 5-input variable model was denoted as the simple PLS model (Fig. 7D). The input variables in the simple PLS model were pERK

at 10 min (pERK10), pCREB at 5 and 30 min (pCREB5 and pCREB60), pAKT at 5 min (pAKT5) and pJNK at 30 min (pJNK30). This result indicates that pERK10, pCREB5, pCREB60, pAKT5 and pJNK30 were the minimum set of the inputs that showed a comparable predictive accuracy as that of the entire data set of the outputs in the full PLS model with 60 input variables. We further validated the simple model by use of the inhibitor experiments in Figure 4 and found that the simple model similarly predicted the inhibitor experimental result as the full and best models (Fig. 8 and 9). The variables pERK10, pAKT5 and pJNK30 were considered to indicate the peak activity of pERK, pAKT and pJNK, respectively (Fig. 3). The variables pCREB5 and pCREB60 may correspond to pERK and PKA activity, respectively, because pCREB is regulated by pERK and PKA. I then plotted the loadings and scores of the 5 variables in the first and second principal component axes (Fig. 7E, F). I also reduced the number of output variables by selecting the single output variable with the maximum norm that is considered to indicate the maximum amount of information regarding the output variable. The selected output variables were cell death at 12 hour (Cell death12), cell cycle at 32 hour (Cell cycle32), JUNB at 270 min (JUNB270), FOSB at 180 min (FOSB180), c-FOS at 60 min (c-FOS60), neurite length at 48 hour (Neurite length48), cell viability at 48 hour (Cell viability48), c-JUN at 90 min (c-JUN90) and EGR1 at 90

min (EGR1-90). I plotted the loadings and scores of the selected single outputs for each variable (Fig. 7G, H) of the simple PLS model with 5 input and 9 output variables (Fig. 8A). The variable pJNK30 correlated with Cell death12 and pCREB5, and pCREB60 correlated with JUNB270, FOSB180 and c-FOS60. pERK10 and pAKT5 correlated with c-JUN90 and EGR1-90. Neurite length48 and Cell viability48 correlated with pCREB5 and pCREB60, and pERK10. Anisomycin, PACAP, NGF and EGF were plotted in the first, second, third and fourth quadrants, respectively. The results of the simple PLS model are consistent with those of the full PLS model, indicating that the simple PLS model effectively represents the relationships between the inputs and the outputs underlying cell-fate decisions in PC12 cells.

6 DISCUSSION

In this study, I employed PLS regression to describe the relationship between the phosphorylation of signaling molecules and the expression of IEGs and cellular phenotypes, which has been applied to similar biological data sets [1-3,27-30]. The loadings and scores of the PLS model highlight characteristics of the MIMO system and growth factor-specific input-output relationships of cell-fate decisions, respectively, in PC12 cells.

One of the technical highlights of this study is the model reduction via backward elimination PLS regression. I found that a reduction of the number of input variables provided better predictive ability, and the 22-variable model showed the best predictive ability. I further reduced the number of variables, and found that the simple PLS model with 5 input variables showed a comparable predictive ability to that of the full PLS model with 60 input variables. I also reduced the number of outputs by selecting the output with the maximum norm, which is considered to encode the maximum information of each output. Such reduction methods help provide an intuitive understanding of the complex relationships in the MIMO system and can be widely applied to any signaling and cellular phenotypes. The variable importance in the projection (VIP) scores is indicative of the importance of each variable in the projection

used in a PLS model and are often used for variable selection [1]. I calculated the VIP scores of all variables in the full PLS model and found that the 5 variables in the simple PLS model were included in the top 20 variables selected by VIP (Table 3). Furthermore, we compared the simple PLS model with the VIPs-PLS model with 5 input variables selected from the highest VIP score (Table 4). The simple PLS model with 5 input variables showed lower LOOCV MSE and higher correlation than the VIPs-PLS models with 5 input variables. indicating that the prediction ability of the simple model is higher than the VIPs-PLS model. These results support the importance of these 5 variables in the simple PLS model.

The simple PLS model with 5 variables demonstrated a comparable predictive ability to that of the full PLS model. The simple PLS model with 5 input variables and 9 output variables (Fig. 10A) facilitates an intuitive understanding of the MIMO system and growth factor-specific input-output relationships of cell-fate decision in PC12 cells (Fig. 10B). The simple model showed the similar predictive ability to the full model against the inhibitor experiment (Fig. 8 and 9), indicating the simple model essentially captures the input-output relationship of the MIMO system. The selected input variables in the simple PLS model were pERK10, pCREB5, pCREB60, pAKT5 and pJNK30. Among the selected inputs, pERK10, pAKT5 and pJNK30 correspond to their peak activities

and are considered to encode the maximum information of the signaling molecules (Fig. 3). It has been reported that sustained ERK activity is required for cell differentiation in PC12 cells [9,38,39]; however, late pERK was not selected in the simple PLS model, although the model can capture the characteristics of the neurite length. This omission may occur because late pERK information is encoded by pCREB60 levels, a downstream molecule of ERK, in the simple PLS model. Two different time points of pCREB, pCREB5 and pCREB60, were selected, suggesting that pCREB5 and pCREB60 encode different information. The phosphorylation of CREB has been reported to be regulated by ERK and PKA [40]. NGF and EGF have been reported to induce the phosphorylation of CREB via ERK, and PACAP has been reported to induce the phosphorylation of PKA [10,40,41]. Therefore, pCREB5 and pCREB60 likely encode different information for growth factors and upstream molecules.

I found that the simple model shows the similar predictive ability for the inhibitor experiments to the full and best models, meaning that the selected 5 input variables in the simple model have high explanatory abilities to the outputs. I further examined the contribution of the 5 input variables in the simple model to the specific outputs. I made the best models for each output and compared to the selected input variables to those in the simple models (Table 5). The selected input variables in the best models for each

output included different sets of the input variables in the simple model, suggesting that the different sets of the input variables in the simple model specifically contributed to each output. For example, the cell cycle and cell death shared the same set of the input variables such as pERK10, pJNK30 and pAKT5, suggesting that these output may share the same upstream dependency. Each IEG had the different sets of the selected input variables except c-FOS and FOSB, indicating that the regulation of c-FOS and FOSB were similar, and the regulation of other IEGs were different.

PLS regression reduces dimensionality of the inputs and outputs into respective latent variables. In this study, I further reduced the PLS model by eliminating inputs variables from the original PLS model using the backward elimination method, and found that accuracy of the eliminated PLS model (best PLS model) was improved. These results demonstrate that the backward elimination PLS regression has two-fold advantage compared to the conventional PLS regression; reduction of input variables and improvement of accuracy of prediction. I further eliminated the input variables and identified 5 inputs as the minimum set of the inputs that characterized the MIMO system in PC12 cells. Thus, our data-driven statistical modeling method is widely useful to effectively extract simple relationships of the cellular MIMO system from large-scale data sets.

7 FIGURES

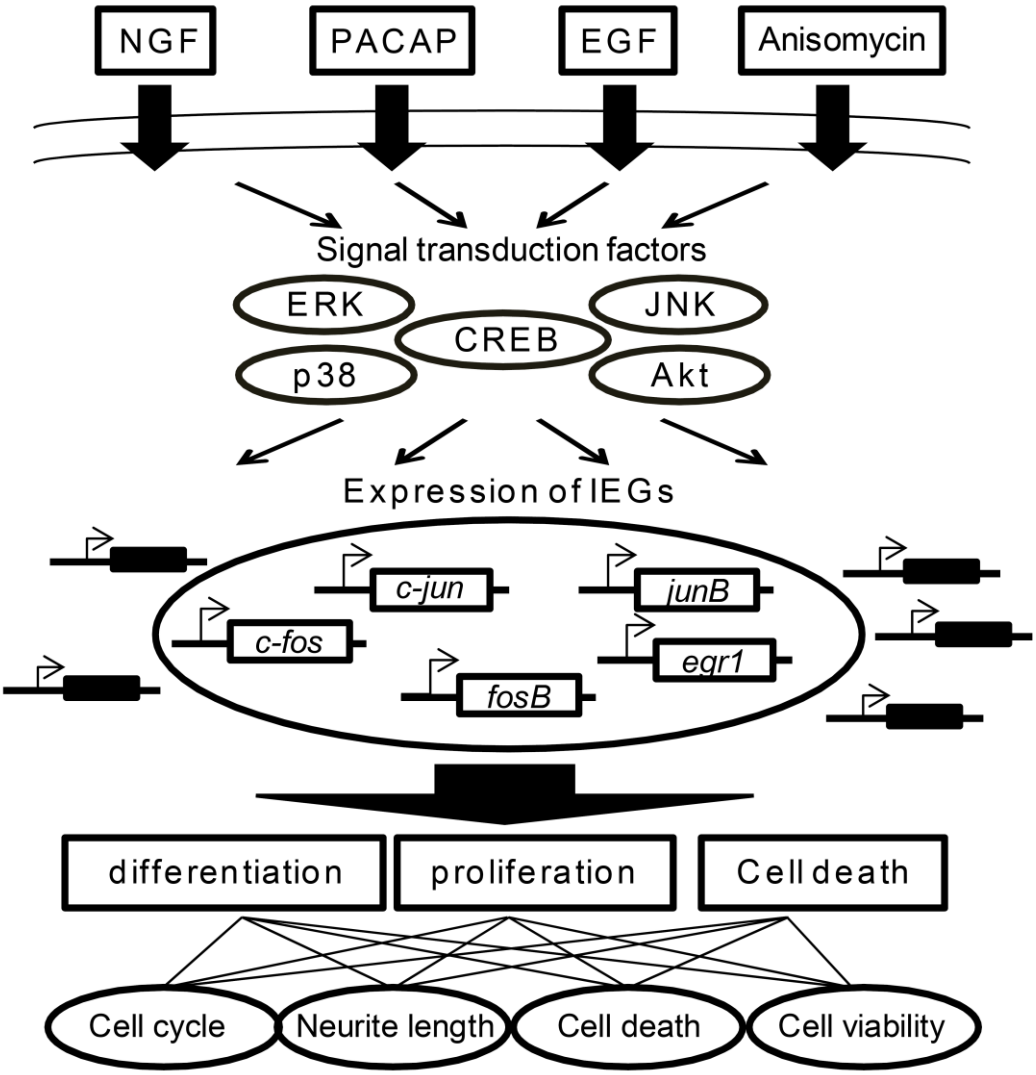


Figure 1. Multi-input and multi-output system in PC12 cell. PC12 cells exhibit several characteristic phenotypes in response to various external stimuli. These various stimuli are redundantly mediated by a few common signal transduction factors such as Akt, CREB and MAPKs, eventually increase expression levels of some proteins known as Immediate Early Genes (IEGs). These stimuli finally cause different types of cellular phenotype; cell differentiation, proliferation and cell death are caused by NGF and PACAP, by EGF and by anisomycin respectively. Each phenotype observes one or more properties such as cellcycle, neurite length, cell death and cell viability.

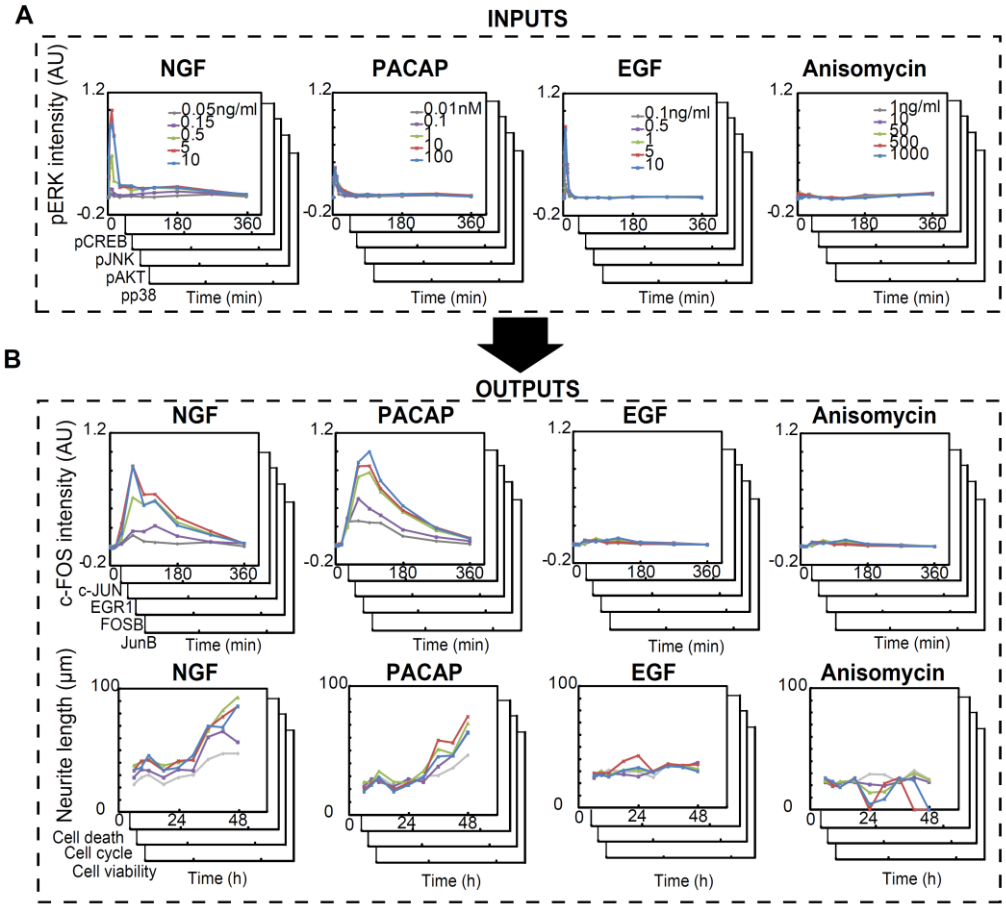


Figure 2. The inputs and outputs in the MIMO system for cell-fate decisions in PC12 cells. (**A**) The inputs consisted of the 12 time points of pERK, pCREB, pJNK, pAKT, and pp38 in response to 5 doses of 4 growth factors (Fig. 3). (**B**) The outputs consisted of the 12 time points for protein expression of c-FOS, c-JUN, EGR1, FOSB, and JUNB, and 9 time points for the neurite lengths, cell viability (respiratory chain activity of mitochondria), cell cycle (S-phase fraction) and cell death (Caspase3 activity) in response to 5 doses of 4 growth factors (Fig. 3). The doses of the growth factors are indicated by different colors (A and B).

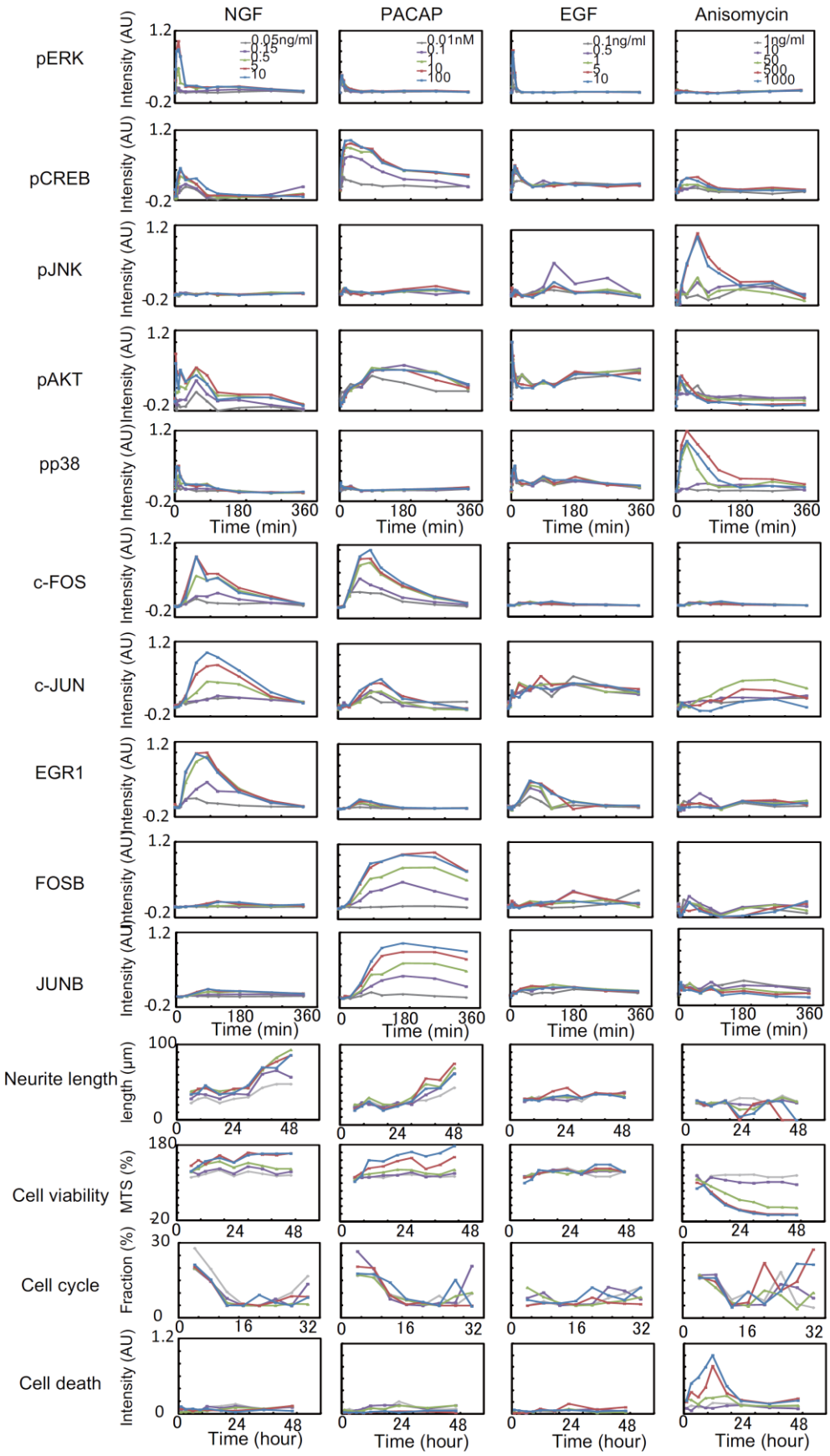

NGF
PACAP
EGF
Anisomycin
pERK
pCREB
pJNK
pAKT
pp38
c-FOS
c-JUN
EGR1
FOSB
JUNB
Neurite length
Cell viability
Cell cycle
Cell death
Intensity (AU)
length (μm)
MTS (%)
Fraction (%)
0.05ng/ml
0.15
0.5
1
5
10
0.01nM
0.1
1
10
100
0.1ng/ml
0.5
1
5
10
1ng/ml
10
50
500
1000
Time (min)
Time (hour)

Figure 3. Time courses of the inputs (phosphorylation of 5 signaling molecules) and theoutputs (5 IEGs and 4 cellular phenotypes) in response to NGF, EGF, PACAP and anisomycin for construction of the PLSmodel.

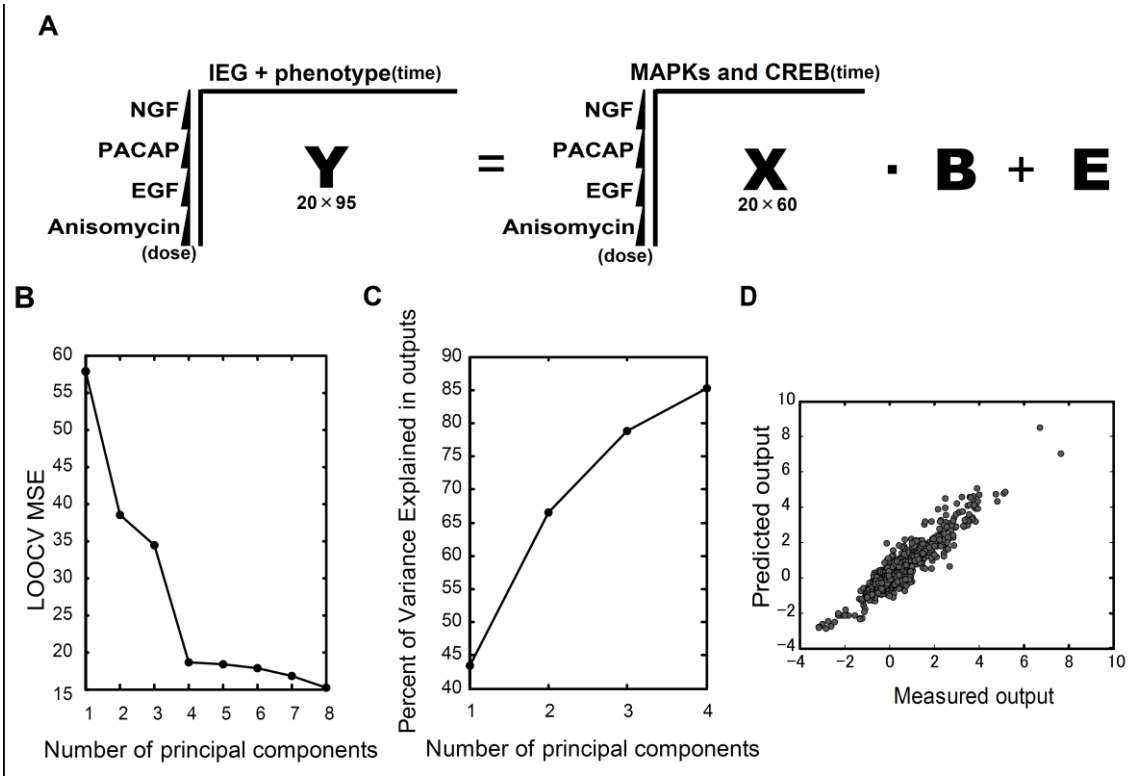


Figure 4. Construction of the PLS model (A) Construction of the PLS model. Inputs metrics **X** (20×60) regressed against the output metrics **Y** (20×95). Each column and row in **X** correspond with time course points of MAPKs and CREB, and the doses of stimuli, respectively. Each column and row in **Y** correspond with time course points of the IEGs and phenotypes, and with doses of stimuli, respectively. **B** is the coefficient metrics and **E** is the residue metrics of the PLS model. (**B**) LOOCV MSE (leave-one-out cross validation mean square error) as a function of the number of principal components. (**C**) The cumulative contribution percentage of the principal components. (**D**) Correlation plots between the measured and predicted outputs. The Pearson correlation coefficient, *r*, was 0.94. Each dot represents a single time point for one of the outputs.

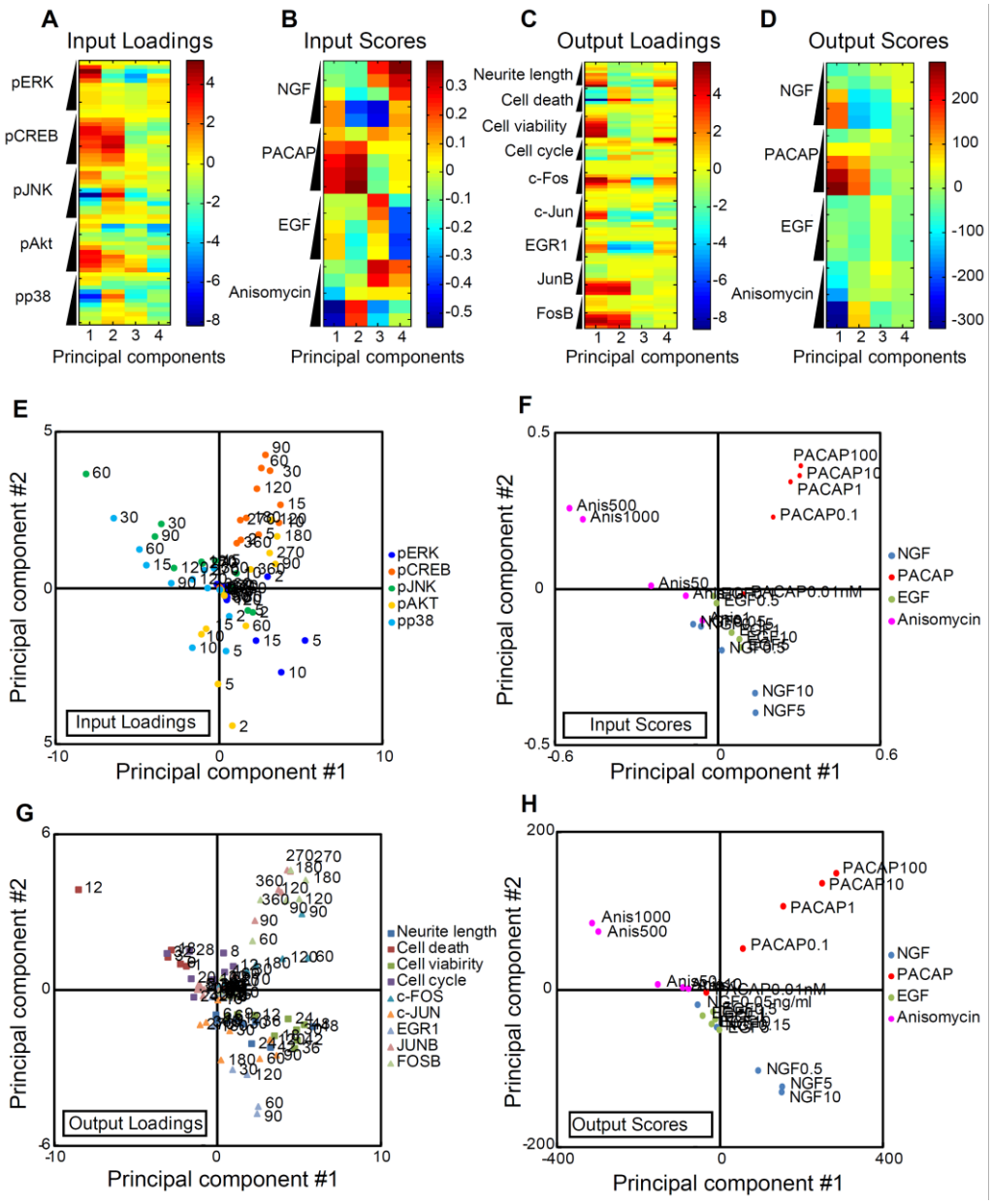


Figure 5. Loadings and scores of the principal components of the inputs and outputs. (**A**) Input loadings. A wedge indicates the temporal evolution of the indicated molecules from 0 to 360 min (Table 1). (**B**) Input scores. A wedge indicates the doses of the stimulant. (**C**) Output loadings. A wedge indicates the temporal evolution (Table 1). (**D**) Output scores. A wedge indicates the doses of the indicated molecules (Fig. 3). The red and blue colors indicate positive and negative values, respectively (**A-D**). Scatter plots of input loadings (**E**), input scores (**F**), output loadings (**G**) and output scores (**H**) of the first and second principal components. The colors correspond to the latent variables (**E**, **G**) and growth factors (**F**, **H**). The numbers indicate the time (minute for

pERK, pCREB, pJNK, pAKT, pp38, c-FOS, c-JUN, EGR1, JUNB and FOSB and hour for neurite lengths, cell viability, cell cycle and cell death).

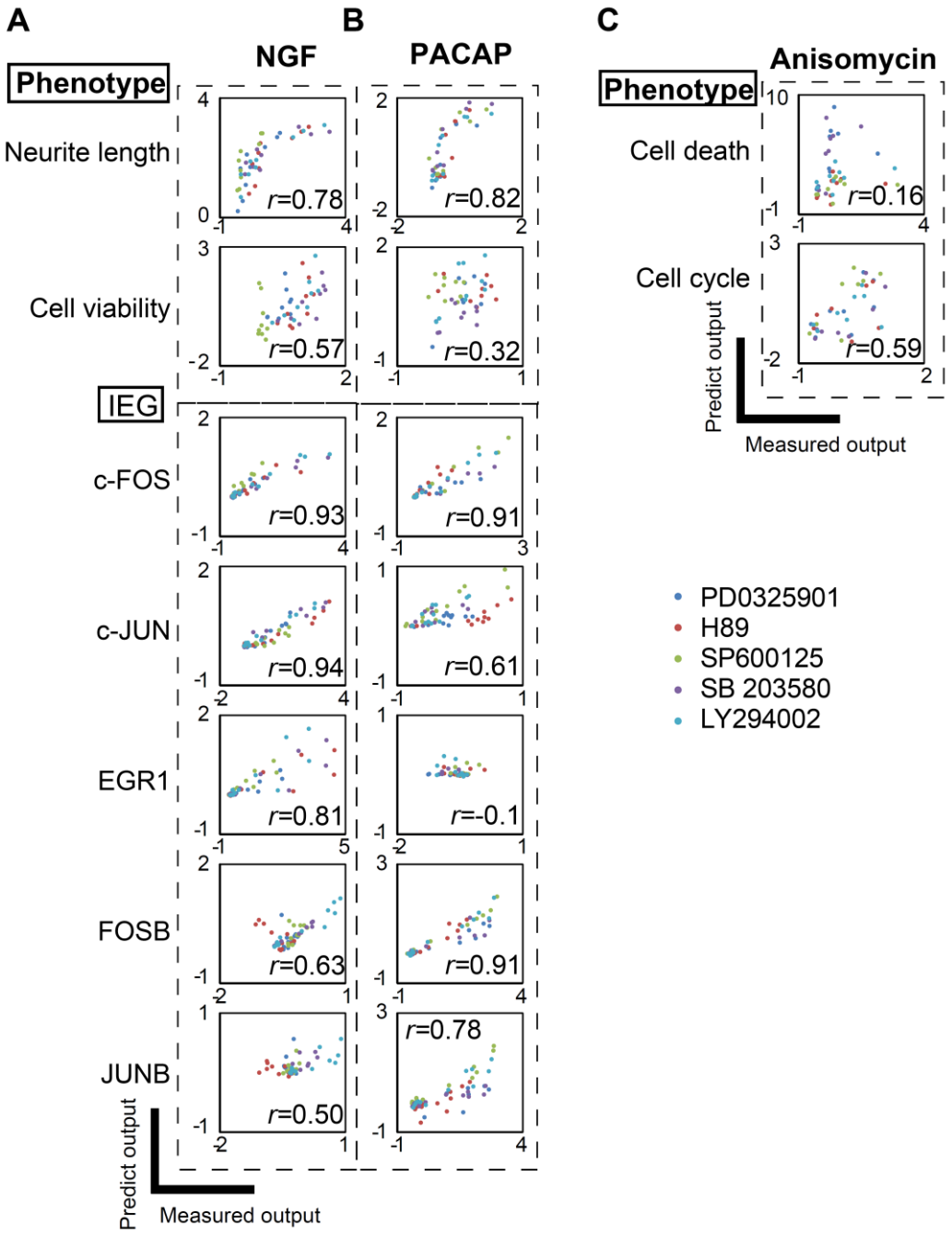


Figure 6. Validation of the PLS model using inhibitor experiments. Correlation plots between the measured outputs and predicted outputs with NGF (**A**), PACAP (**B**) and anisomycin (**C**). The correlation coefficient, *r*, is indicated in each plot. Each dot represents a single time point. The data sets with PD0325901 (MEK inhibitor), H89 (PKA inhibitor), LY294002 (PI3K inhibitor), SB203580 (p38 inhibitor), SP600125 (JNK inhibitor) are indicated by the various colors.

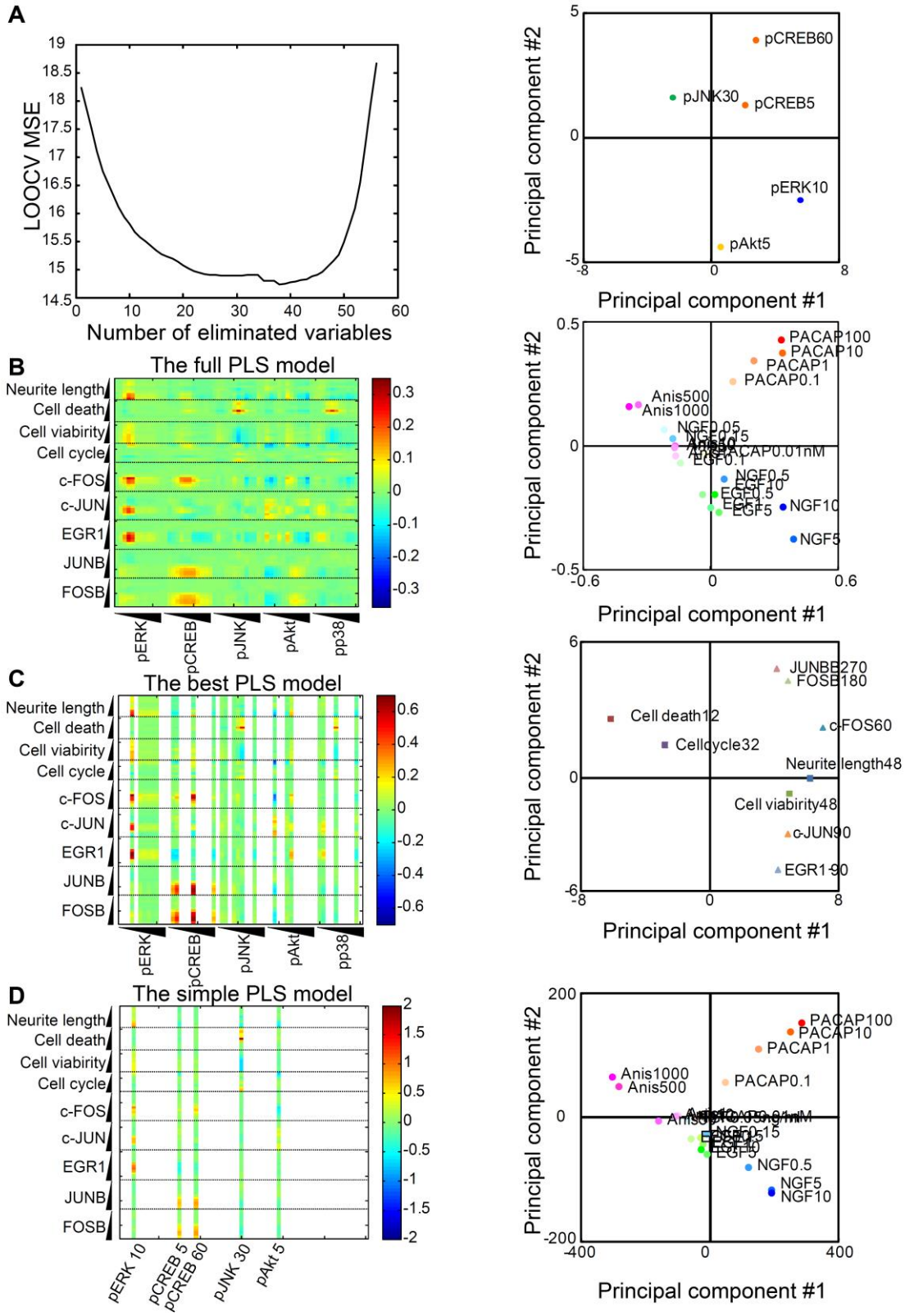


Figure 7. Reduction of the PLS model by backward elimination PLS regression. (**A**) MSE of LOOCV as a function of number of the eliminated variables via the

backward elimination PLS regression. Coefficient matrix of the full PLS model with 60 input variables (**B**), the best PLS model with 22 input variables (**C**) and the simple PLS model with 5 input variables (**D**). The red and blue colors indicate positive and negative values, respectively. As the number of the variables reduced, the contribution of remained variables relatively increased, and as a result, magnitude of the regression coefficient increased. The scatter plots of the input loadings (**E**), input scores (**F**), output loadings (**G**) and output scores (**H**) of the first and second principal components of the simple PLS model. The colors correspond to the latent variables (**E**, **G**) and growth factors (**F**, **H**).

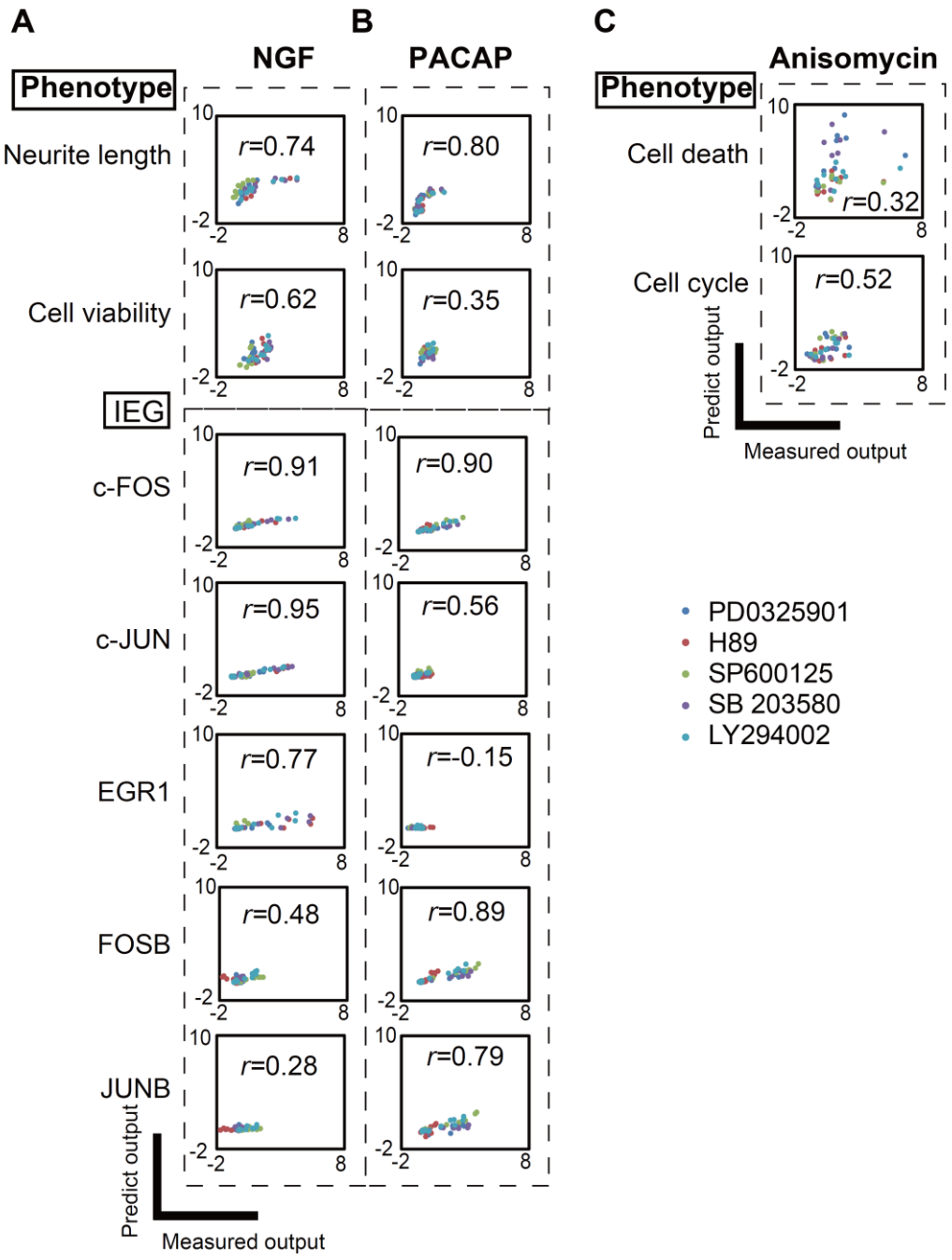


Figure 8. Validation of the simple PLS model using inhibitor experiments.
Correlation plots between the measured outputs and predicted outputs using same experiment as Figure 4 with NGF (**A**), PACAP (**B**) and anisomycin (**C**).

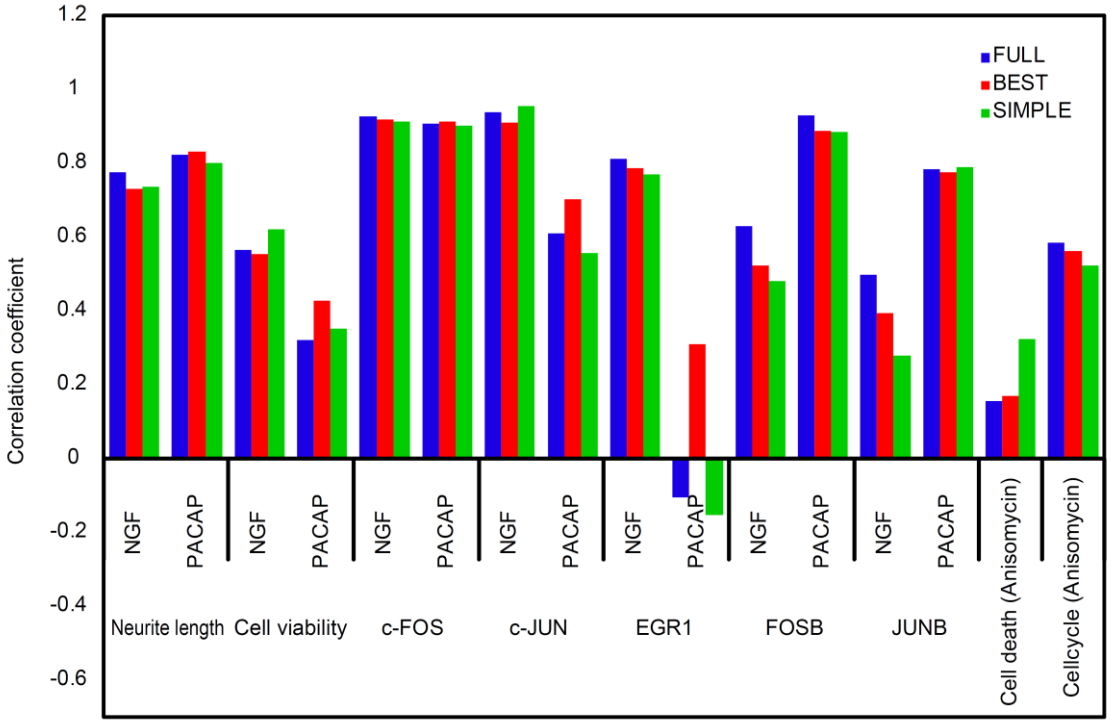


Figure 9. The correlation coefficient of the full, best and simple PLS model between the measured outputs and predicted outputs using the inhibitor experiments. The high correlations were found between NGF and neurite length (r=0.73),between PACAP and neurite length(r=0.80), between NGF and c-Fos (r=0.91), between PACAP and c-FOS(r=0.90), between NGF and c-JUN(r=0.95), between NGF and EGR1(r=0.77), between PACAP and FOSB (r=0.89) and between PACAP and JUNB (r=0.79); and the low correlations were found between PACAP and MTS (r=0.35), between PACAP and EGR1(r=-0.15), between NGF and FOSB (r=0.48), between NGF and JUNB (r=0.28), between anisomycin and cell death (r=0.32).

A

Neurite length48	=	0.95	pERK10	- 0.21	pCREB5	+ 0.17	pCREB60	- 0.54	pJNK30	- 0.64	pAKT5
Cell viability48	=	0.44	pERK10	+ 0.17	pCREB5	+ 0.19	pCREB60	- 0.67	pJNK30	- 0.13	pAKT5
Cell death12	=	- 0.20	pERK10	- 0.21	pCREB5	+ 0.06	pCREB60	+ 2.05	pJNK30	+ 0.24	pAKT5
Cell cycle32	=	- 0.98	pERK10	- 0.19	pCREB5	- 0.02	pCREB60	+ 0.69	pJNK30	- 0.07	pAKT5
c - FOS60	=	0.97	pERK10	+ 0.07	pCREB5	+ 0.55	pCREB60	- 0.06	pJNK30	- 0.57	pAKT5
c - JUN90	=	0.69	pERK10	+ 0.09	pCREB5	+ 0.06	pCREB60	- 0.20	pJNK30	+ 0.18	pAKT5
EGR1 - 90	=	1.07	pERK10	- 0.44	pCREB5	- 0.28	pCREB60	- 0.03	pJNK30	- 0.07	pAKT5
JUNB270	=	- 0.04	pERK10	+ 0.67	pCREB5	+ 0.81	pCREB60	- 0.28	pJNK30	- 0.24	pAKT5
FOSB180	=	- 0.01	pERK10	+ 0.76	pCREB5	+ 0.83	pCREB60	- 0.38	pJNK30	- 0.14	pAKT5

B

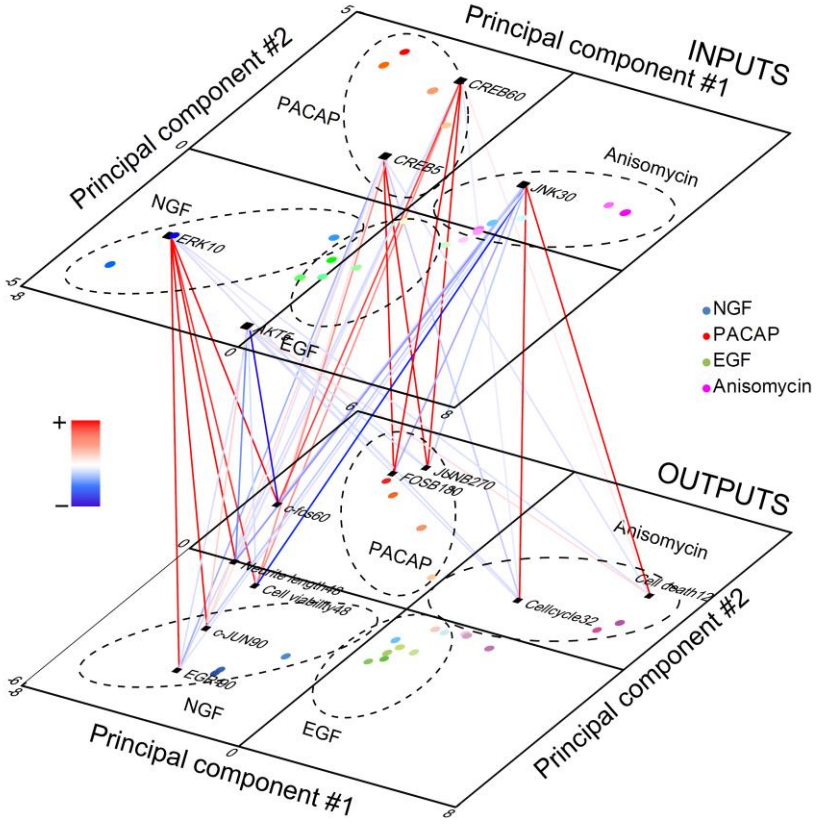


Figure 10. The simple relationships in the MIMO system in cell fate decision in PC12 cells. (A) The simple PLS model with 5 input and 9 output variables. (B) Loadings (squares) and scores (circles) of the first and second principal components in the simple PLS model with 5 input and 9 output variables were bi-plotted in input and output layers. Lines across the layers are coefficients of the matrix of the simple PLS model whose values are indicated by the color bar. The colors of circles indicate the growth factors.

8. TABLES

Table 1. Summary of the input (1,200 points) and output (1,900 points) data shown in Figs. 2 and 3.

	Molecule/Phenotype	Time points
Input	pERK	0, 2, 5, 10, 15, 30, 60, 90, 120, 180, 270, 360 (min)
	pCREB	
	pAKT	
	pJNK	
	pp38	
Output	c-FOS	0, 2, 5, 10, 15, 30, 60, 90, 120, 180, 270, 360 (min)
	FOSB	
	c-JUN	
	EGR1	
	JUNB	
	Neurite length	6, 9, 12, 18, 24, 30, 36, 42, 48 (hour)
	Cell viability	6, 9, 12, 18, 24, 30, 36, 42, 48 (hour)
	Cell death	1, 3, 6, 9, 12, 18, 24, 36, 48 (hour)
	Cell cycle	4, 8, 12, 16, 20, 24, 28, 32 (hour)

Table 2. The 22 input variables in the best model and five input variables in the simple model.

	2	5	10	15	30	60	90	120	180	270	360 (min)
pERK											
pCREB											
pJNK											
pAKT											
pp38											

Both the gray and black cells are the variables for the best PLS model, whereas the black cells alone are the variables for the simple PLS model.

Table 3. The VIP scores of the input variables.

	Variable	VIP
1	pJNK60	2.38
2	pERK10	2.17
3	pp38-30	1.82
4	pERK5	1.72
5	pERK15	1.69
6	pCREB30	1.56
7	pAkt2	1.52
8	pCREB60	1.52
9	pCREB90	1.49
10	pCREB15	1.48
11	pp38-60	1.35
12	pAkt5	1.32
13	pAkt360	1.25
14	pp38-15	1.25
15	pCREB10	1.23
16	pAkt180	1.16
17	pJNK90	1.16
18	pJNK30	1.16
19	pCREB120	1.13
20	pAkt270	1.08
21	pJNK120	1.07
22	pAkt120	1.03
23	pAkt90	0.96
24	pp38-10	0.93
25	pCREB5	0.92
26	pAkt60	0.92
27	pp38-90	0.89
28	pCREB180	0.84
29	pERK2	0.83
30	pp38-5	0.80
31	pAkt10	0.79
32	pCREB270	0.75
33	pp38-180	0.67
34	pCREB2	0.64
35	pJNK2	0.62
36	pAkt15	0.60
37	pCREB360	0.56
38	pp38-120	0.53
39	pJNK5	0.53
40	pp38-270	0.53
41	pJNK180	0.48
42	pERK30	0.45
43	pJNK270	0.45
44	pAkt30	0.39
45	pERK120	0.38
46	pp38-2	0.36
47	pERK90	0.35
48	pERK60	0.35
49	pERK180	0.35
50	pp38-360	0.33
51	pJNK10	0.31
52	pJNK360	0.29
53	pJNK15	0.27
54	pERK270	0.20
55	pJNK0	0.12
56	pERK360	0.09
57	pAkt0	0.05
58	pCREB0	0.02
59	pERK0	0.02
60	pp38-0	0.01

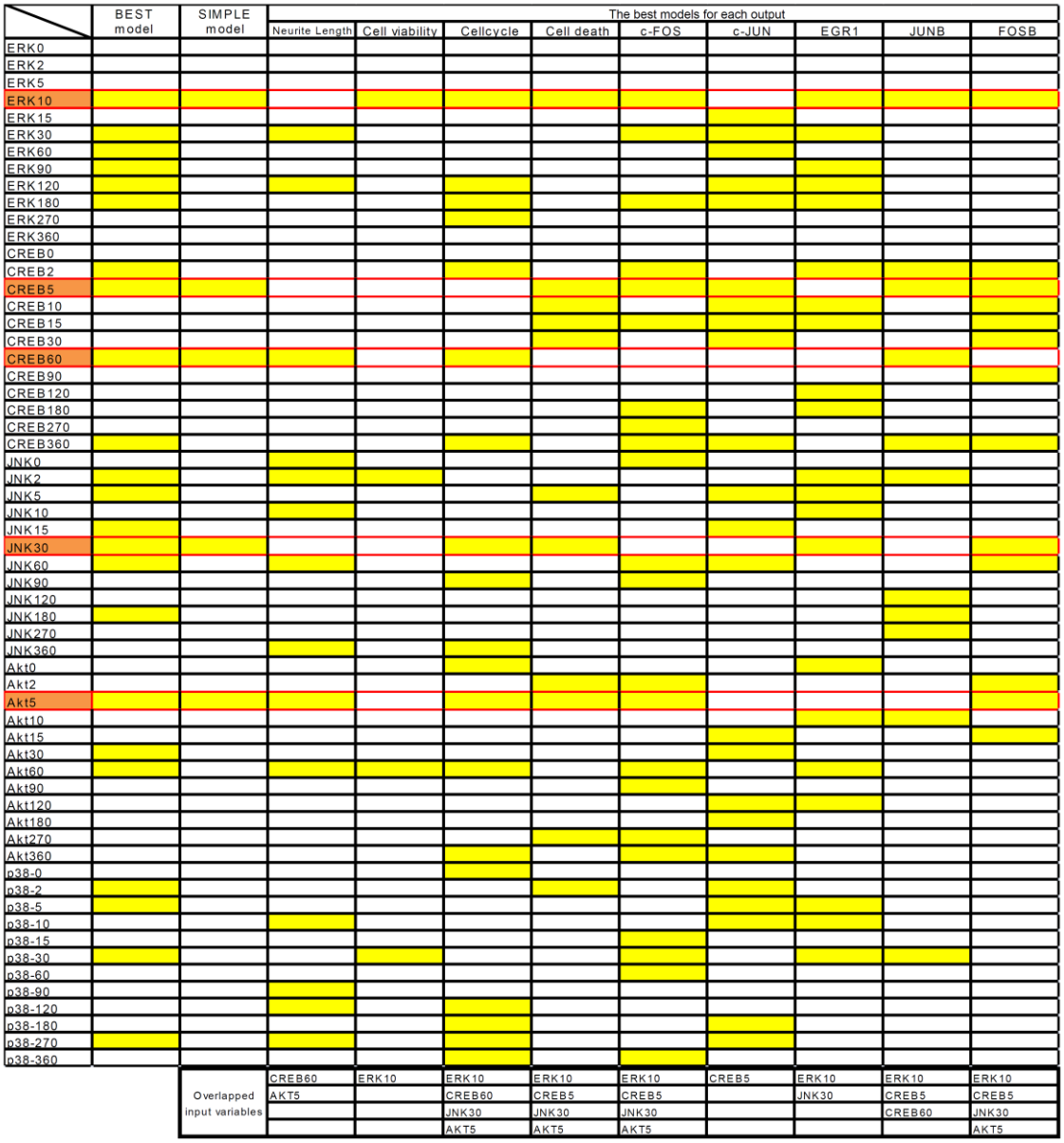
Table 4. The correlation coefficients and LOOCV MSE in the backward elimination PLS models and the VIPs-PLS model with 5 input variables.

PLS model	*r**	LOOCV MSE
Full	0.90	18.64
Best	0.92	14.79
Simple	0.90	17.92
VIP_13variables	0.90	18.57
VIP_5variables	0.81	47.09

* Correlation coefficient between the mesured outputs and predicted outputs

Table 5. The best models for each output.

The high correlations were found between NGF and neurite length (r=0.73),between PACAP and neurite length(r=0.80), between NGF and c-Fos (r=0.91), between PACAP and c-FOS(r=0.90), between NGF and c-JUN(r=0.95), between NGF and EGR1(r=0.77), between PACAP and FOSB (r=0.89) and between PACAP and JUNB (r=0.79); and the low correlations were found between PACAP and MTS (r=0.35), between PACAP and EGR1(r=-0.15), between NGF and FOSB (r=0.48), between NGF and JUNB (r=0.28), between anisomycin and cell death (r=0.32).



8. REFERENCES

1. Janes KA, Albeck JG, Gaudet S, Sorger PK, Lauffenburger DA, et al. (2005) A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis. Science 310: 1646-1653.

2. Janes KA, Reinhardt HC, Yaffe MB (2008) Cytokine-induced signaling networks prioritize dynamic range over signal strength. Cell 135: 343-354.

3. Tentner AR, Lee MJ, Ostheimer GJ, Samson LD, Lauffenburger DA, et al. (2012) Combined experimental and computational analysis of DNA damage signaling reveals context-dependent roles for Erk in apoptosis and G1/S arrest after genotoxic stress. Mol Syst Biol 8: 568.

4. Li Y, Agarwal P, Rajagopalan D (2008) A global pathway crosstalk network. Bioinformatics 24: 1442-1447.

5. Citri A, Yarden Y (2006) EGF-ERBB signalling: towards the systems level. Nat Rev Mol Cell Biol 7: 505-516.

6. Oda K, Matsuoka Y, Funahashi A, Kitano H (2005) A comprehensive pathway map of epidermal growth factor receptor signaling. Mol Syst Biol 1: 2005 0010.

7. Ferrari G, Greene LA (1994) Proliferative inhibition by dominant-negative Ras rescues naive and neuronally differentiated PC12 cells from apoptotic death. EMBO J 13: 5922-5928.

8. Padmanabhan J, Park DS, Greene LA, Shelanski ML (1999) Role of cell cycle regulatory proteins in cerebellar granule neuron apoptosis. J Neurosci 19: 8747-8756.

9. Marshall CJ (1995) Specificity of receptor tyrosine kinase signaling: transient versus sustained extracellular signal-regulated kinase activation. Cell 80: 179-185.

10. Vaudry D, Stork PJ, Lazarovici P, Eiden LE (2002) Signaling pathways for PC12 cell differentiation: making the right connections. Science 296: 1648-1649.

11. Torocsik B, Szeberenyi J (2000) Anisomycin affects both pro- and antiapoptotic mechanisms in PC12 cells. Biochem Biophys Res Commun 278: 550-556.

12. Burstein DE, Blumberg PM, Greene LA (1982) Nerve growth factor-induced neuronal differentiation of PC12 pheochromocytoma cells: lack of inhibition by a tumor promoter. Brain Res 247: 115-119.

13. Lazarovici P, Jiang H, Fink D, Jr. (1998) The 38-amino-acid form of pituitary adenylate cyclase-activating polypeptide induces neurite outgrowth in PC12 cells that is dependent on protein kinase C and extracellular signal-regulated kinase but not on protein kinase A, nerve growth factor receptor tyrosine kinase, p21(ras) G protein, and pp60(c-src) cytoplasmic tyrosine kinase. Mol Pharmacol 54: 547-558.

14. Huff K, End D, Guroff G (1981) Nerve growth factor-induced alteration in the response of PC12 pheochromocytoma cells to epidermal growth factor. J Cell Biol 88: 189-198.

15. Gotoh Y, Nishida E, Yamashita T, Hoshi M, Kawakami M, et al. (1990) Microtubule-associated-protein (MAP) kinase activated by nerve growth factor and epidermal growth factor in PC12 cells. Identity with the mitogen-activated MAP kinase of fibroblastic cells. Eur J Biochem 193: 661-669.

16. Qui MS, Green SH (1992) PC12 cell neuronal differentiation is associated with prolonged p21ras activity and consequent prolonged ERK activity. Neuron 9: 705-717.

17. Traverse S, Gomez N, Paterson H, Marshall C, Cohen P (1992) Sustained activation of the mitogen-activated protein (MAP) kinase cascade may be required for differentiation of PC12 cells. Comparison of the effects of nerve growth factor and epidermal growth factor. Biochem J 288 (Pt 2): 351-355.

18. Torocsik B, Szeberenyi J (2000) Anisomycin uses multiple mechanisms to stimulate mitogen-activated protein kinases and gene expression and to inhibit neuronal differentiation in PC12 phaeochromocytoma cells. Eur J Neurosci 12: 527-532.

19. Segal RA, Greenberg ME (1996) Intracellular signaling pathways activated by neurotrophic factors. Annu Rev Neurosci 19: 463-489.

20. Deutsch PJ, Sun Y (1992) The 38-amino acid form of pituitary adenylate cyclase-activating polypeptide stimulates dual signaling cascades in PC12 cells and promotes neurite outgrowth. J Biol Chem 267: 5108-5113.

21. Richter-Landsberg C, Jastorff B (1986) The role of cAMP in nerve growth factor-promoted neurite outgrowth in PC12 cells. J Cell Biol 102: 821-829.

22. York RD, Molliver DC, Grewal SS, Stenberg PE, McCleskey EW, et al. (2000) Role of phosphoinositide 3-kinase and endocytosis in nerve growth factor-induced extracellular signal-regulated kinase activation via Ras and Rap1. Mol Cell Biol 20: 8069-8083.

23. Santos SD, Verveer PJ, Bastiaens PI (2007) Growth factor-induced MAPK network topology shapes Erk response determining PC-12 cell fate. Nat Cell Biol 9: 324-330.

24. Eriksson M, Taskinen M, Leppa S (2007) Mitogen activated protein kinase-dependent activation of c-Jun and c-Fos is required for neuronal differentiation but not for growth and stress response in PC12 cells. J Cell Physiol 210: 538-548.

25. Hazzalin CA, Mahadevan LC (2002) MAPK-regulated transcription: a continuously variable gene switch? Nat Rev Mol Cell Biol 3: 30-40.

26. H. Wold K-GJE (1982) Soft modelling, the basic design and some extensions: North-Holland Publishing Company, Amsterdam. pp. 1–54 p.

27. Janes KA, Albeck JG, Peng LX, Sorger PK, Lauffenburger DA, et al. (2003) A high-throughput quantitative multiplex kinase assay for monitoring information flow in signaling networks: application to sepsis-apoptosis. Mol Cell Proteomics 2: 463-473.

28. Lau KS, Juchheim AM, Cavaliere KR, Philips SR, Lauffenburger DA, et al. (2011) In vivo systems analysis identifies spatial and temporal aspects of the modulation of TNF-alpha-induced apoptosis and proliferation by MAPKs. Sci Signal 4: ra16.

29. Miller-Jensen K, Janes KA, Brugge JS, Lauffenburger DA (2007) Common effector processing mediates cell-specific responses to stimuli. Nature 448: 604-608.

30. Janes KA, Kelly JR, Gaudet S, Albeck JG, Sorger PK, et al. (2004) Cue-signal-response analysis of TNF-induced apoptosis by partial least squares regression of dynamic multivariate data. J Comput Biol 11: 544-561.

31. Sasagawa S, Ozaki Y, Fujita K, Kuroda S (2005) Prediction and validation of the distinct dynamics of transient and sustained ERK activation. Nat Cell Biol 7: 365-373.

32. Ozaki Y, Uda S, Saito TH, Chung J, Kubota H, et al. (2010) A quantitative image cytometry technique for time series or population analyses of signaling networks. PLoS One 5: e9955.

33. Pool M, Thiemann J, Bar-Or A, Fournier AE (2008) NeuriteTracer: a novel ImageJ plugin for automated quantification of neurite outgrowth. J Neurosci Methods 168: 134-139.

34. Janes KA, Yaffe MB (2006) Data-driven modelling of signal-transduction networks. Nat Rev Mol Cell Biol 7: 820-828.

35. Allen DM (1974) The Relationship between Variable Selection and Data Agumentation and a Method for Prediction. Technometrics Vol. 16, No. 1: 125-127.

36. Efroymson M. Stepwise regression—a backward and forward look; 1966; Florham Park, New Jersey

37. Geladi P, Kowalski BR (1986) Partial least-squares regression: a tutorial. Anal Chim Acta 185: 1-17.

38. Cowley S, Paterson H, Kemp P, Marshall CJ (1994) Activation of MAP kinase kinase is necessary and sufficient for PC12 differentiation and for transformation of NIH 3T3 cells. Cell 77: 841-852.

39. Kao S, Jaiswal RK, Kolch W, Landreth GE (2001) Identification of the mechanisms regulating the differential activation of the mapk cascade by epidermal growth factor and nerve growth factor in PC12 cells. J Biol Chem 276: 18169-18177.

40. Shaywitz AJ, Greenberg ME (1999) CREB: a stimulus-induced transcription factor activated by a diverse array of extracellular signals. Annu Rev Biochem 68: 821-861.

41. Watanabe K, Akimoto Y, Yugi K, Uda S, Chung J, et al. (2012) Latent process genes for cell differentiation are common decoders of neurite extension length. J Cell Sci 125: 2198-2211.

9. ACKNOWLEDGEMENTS

I am deeply grateful to Prof. Shinya Kuroda for teaching the principles of scientific research. I also want to thank our laboratory members for their support and Miharu Sato and Risa Kunihiro for their technical assistance with the experiments.

This work was supported by The Creation of Fundamental Technologies for Understanding and Control of Biosystem Dynamics, CREST, from the Japan Science and Technology (JST); by a KAKENHI Scientific Research grant (A) (#21240025) from the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT); and by a Human Frontier Science Project (HFSP) grant (RGP0061/2011).

I was supported by a Grant-in-Aid for the Global COE Program “Deciphering Biosphere from Genome Big Bang” from the Ministry of Education, Culture, Sports, Science and Technology of Japan.