RESEARCH ARTICLE

DREISS: Using State-Space Models to Infer the Dynamics of Gene Expression Driven by External and Internal Regulatory Networks

Daifeng Wang1,2, Fei He3, Sergei Maslov3,4, Mark Gerstein5,6,7*

1 Department of Biomedical Informatics, Stony Brook University, Stony Brook, New York, United States of America, 2 Stony Brook Cancer Center, Stony Brook Medicine, Stony Brook, New York, United States of America, 3 Biology Department, Brookhaven National Laboratory, Upton, New York, United States of America, 4 Department of Bioengineering and Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States of America, 5 Program in Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut, United States of America, 6 Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut, United States of America, 7 Department of Computer Science, Yale University, New Haven, Connecticut, United States of America

* mark@gersteinlab.org

Abstract

Gene expression is controlled by the combinatorial effects of regulatory factors from different biological subsystems such as general transcription factors (TFs), cellular growth factors and microRNAs. A subsystem’s gene expression may be controlled by its internal regulatory factors, exclusively, or by external subsystems, or by both. It is thus useful to distinguish the degree to which a subsystem is regulated internally or externally—e.g., how non-conserved, species-specific TFs affect the expression of conserved, cross-species genes during evolution. We developed a computational method (DREISS, dreiss.gersteinlab.org) for analyzing the Dynamics of gene expression driven by Regulatory networks, both External and Internal based on State Space models. Given a subsystem, the “state” and “control” in the model refer to its own (internal) and another subsystem’s (external) gene expression levels. The state at a given time is determined by the state and control at a previous time. Because typical time-series data do not have enough samples to fully estimate the model’s parameters, DREISS uses dimensionality reduction, and identifies canonical temporal expression trajectories (e.g., degradation, growth and oscillation) representing the regulatory effects emanating from various subsystems. To demonstrate capabilities of DREISS, we study the regulatory effects of evolutionarily conserved vs. divergent TFs across distant species. In particular, we applied DREISS to the time-series gene expression datasets of C. elegans and D. melanogaster during their embryonic development. We analyzed the expression dynamics of the conserved, orthologous genes (orthologs), seeing the degree to which these can be accounted for by orthologous (internal) versus species-specific (external) TFs. We found that between two species, the orthologs have matched, internally driven expression patterns but very different externally driven ones. This is particularly true for genes with evolutionarily ancient functions (e.g. the ribosomal proteins), in contrast to those with more recently evolved functions (e.g., cell-cell communication). This suggests that
despite striking morphological differences, some fundamental embryonic-developmental processes are still controlled by ancient regulatory systems.

Author Summary
The dynamics of a biological system can be controlled by its own internal mechanisms and external perturbations. To gain intuition on this, we may draw a comparison with a mass hanging from a spring. The mass will move naturally by itself but its dynamics is also affected by one's pulling it. That is, the dynamics of the mass is governed by the effect of the external perturbations superimposed on the internal mechanism of the spring (i.e. Hooke's law). Similarly, given a group of genes, their temporal gene expression dynamics can be controlled by both transcription factors inside the group and external regulatory factors. Therefore, it is useful to identify the expression dynamics that are exclusively controlled by internal or external factors and compare them across various systems. While state-space models have been widely used to decouple the internal and external effects in physical systems, such as the mass and spring, typical biological systems do not have enough time samples to infer all the model's parameters, and applications of state-space models were not very effective in these instances. Hence, we developed a general-purpose computational method by integrating state-space models and dimensionality reduction to identify temporal gene expression patterns driven by internal and external regulatory networks. We applied our method to the embryonic developmental datasets in the worm and fly (and also in a human cancer context). We successfully identified the temporal expression dynamics of cross-species conserved genes that were driven by conserved and species-specific regulatory networks.

This is a PLOS Computational Biology Methods paper.

Introduction
Gene regulatory networks systematically control the gene expression dynamics. These networks are highly modular, and consist of various sub-networks. Each sub-network contains a number of regulatory factors representing a subsystem that drives specific gene regulatory functions [1,2]. The subsystems interact with one another, and work together to carry out the entire gene regulatory function. For example, the gene expression in embryogenesis is controlled by the combinatorial effects of various regulatory subsystems composed of complex evolutionary regulatory networks [3]. These regulatory subsystems drive very diverse developmental programs, from the highly conserved (e.g. DNA replication) to the species-specific (e.g. body segmentation). As such the orthologous genes that are evolutionarily conserved genes across species can therefore be regulated by both orthologous and species-specific transcription factors (TFs) [4]. The orthologous TFs form an "internal" regulatory network, while the species-specific TFs form an "external" one. Unfortunately, existing experimental gene expression data cannot decouple the expression components that are driven by the different subsystems. Thus, computational methods are required to assess the contribution from each factor or subsystem from the gene expression data. In this study, we propose a novel computational method,
DREISS—dynamics of gene expression driven by external and internal regulatory networks based on state space model. Using DREISS, we are able to identify temporal gene expression dynamic patterns for evolutionarily conserved genes during embryonic development, as driven by conserved and species-specific regulatory subsystems. These results advance our current understanding of gene regulatory networks during evolution, as well as the differentiation during development.

Developmental gene regulatory networks control gene expression during the developmental processes. These particular regulatory networks have evolved, making it difficult to understand their regulatory mechanisms at the system level. Hence, one typically compares developmental gene expression across species to infer biological activities of developmental gene regulatory networks. For example, embryogenesis provides a platform to study the evolution of gene expression between different species. Recent work has showed that significant biological insight can be gained by cross-species comparisons of the expression profiles during embryogenesis for worms [5], flies [6], frogs [7] and several other vertebrates [8]. It was found that the orthologous genes have minimal temporal expression divergence during the phylotypic stage, a middle phase during the embryonic development across species within the same phylum. These patterns are often characterized as "hourglass" [9]. In addition, the conserved hourglass patterns were observed even within a single species while comparing the developmental gene expression data across distant species, such as worm and fly [10]; i.e., the expression divergence among evolutionarily conserved genes become minimal during the phylotypic stage in both worm and fly. However, much less is known about how the orthologous genes in each species eventually contribute to their species-specific phenotypes due to the lack of appropriate computational approaches. Thus, we aim to use DREISS to discover the components of the orthologous gene expression during embryonic development driven by species-specific transcription factors.

The state-space model has been widely used in engineering [11], and also in biology for the analysis of gene expression dynamics [12–14]. It models the dynamical system output as a function of both the current internal system state and the external input signal. A well-known example in engineering is the vehicle cruise control system where the system state can be the vehicle's speed. Based on the road conditions, the cruise control requires various fuel amounts in order to keep the desired speed level. In biology, we can look at the transcription factors and microRNAs as internal and respectively external regulatory factors of the protein-coding genes expression (See more internal-external examples in S1 Table). Similarly, the state-space model can be applied for studying the expression of orthologous genes at different developmental stages using information regarding their expression (internal) and species-specific regulatory factors (external) at the current known developmental stage. Unlike earlier studies that calculate the expression correlation between individual genes, the state-space model predicts the temporal causal relationships at the system level; i.e., the state at a time is determined by the state and external input at the previous time. The earlier work applied the state-space model to study the gene expression dynamics focusing on small-scale systems, and did not explore the analytic dynamic characteristics of the inferred state-space models. The complex and large-scale biological datasets, especially temporal gene expression data, are very noisy, and high dimensional (i.e., the number of genes is much greater than the number of time samples), thereby preventing an accurate estimation of the state-space model’s parameters. The dimensionality reduction techniques have thus been used to project high-dimensional genes to low-dimensional meta-genes (i.e., the selected features representing de-noised and systematic expression patterns [1,15,16]) as well as the principal dynamic patterns for those meta-genes [17,18]. Using DREISS, we are able to apply the dimensionality reduction to the gene expression data, and develop an effective state-space model for their meta-genes, and then identify a group of
canonical temporal expression trajectories representing the dynamic patterns driven by the effective conserved and species-specific meta-gene regulatory networks according to the model's analytic characteristics. These dynamic patterns reveal temporal gene expression components that are controlled by conserved or species-specific GRNs.

DREISS is a general-purpose tool and can be used to study the gene regulatory effects from any different subsystems for a given group of genes. As an illustration, we applied DREISS to the gene expression data during embryonic development for two model organisms, worm (*Caenorhabditis elegans*) and fly (*Drosophila melanogaster*). In both species, we were able to identify the expression patterns of worm-fly orthologs driven by the conserved regulatory network consisting of the worm-fly orthologous TFs (i.e., the conserved regulatory subsystems between two species), as well as the worm/fly-specific regulatory network consisting of non-orthologous TFs (i.e., the species-specific regulatory subsystem). Our results reveal that, in addition to executing conserved developmental functions between worm and fly, their orthologous genes are also regulated by species-specific TFs to involve in species-specific developmental processes. In summary, DREISS provides a framework to analyze both distantly and closely related species allowing for a better understanding of the gene regulatory mechanisms during development.

**Methods**

DREISS consists of five major steps as detailed in Fig 1

**Step A**: DREISS models temporal gene expression dynamics using state-space models in control theory. In this step, we need to define the internal and external groups of genes and input their time-series gene expression data that we are interested to study. We assume that the time-series gene expression data fits a state-space module. In the state-space model, the “state” refers to the expressions for a large group of genes of interest, such as the worm-fly orthologous genes investigated here. The “control” refers to any other group of genes that contribute to the gene expression of the “state”, such as the species-specific TFs contributed to control orthologous gene expression.

**Step B**: Due to the limited number of temporal samples in gene expression experiments, we do not have enough data to accurately estimate the parameters of the state-space models that capture interactions among hundreds of genes. Therefore, DREISS projects high-dimensional gene expression space to lower-dimensional meta-gene expression spaces using dimensionality reduction techniques.

**Step C**: DREISS derives the effective state-space models for meta-genes so that model parameters can be estimated.

**Step D**: DREISS identifies the meta-gene expression dynamic patterns; i.e., canonical temporal expression trajectories driven by “state” (internal) and by “control” (external) based on the analytic solutions of the estimated models.

**Step E**: Finally, DREISS calculates the gene coefficients over canonical temporal expression trajectories based on linear transformations between genes and meta-genes. DREISS also allows us to compare the dynamic expression patterns of multiple datasets with samples taken at different times. We describe each DREISS step in detail as follows.

**State-space models for temporal gene expression dynamics**

A gene regulatory network is made up of various subsystems [1,2]. These subsystems work together to execute regulatory functions. Given a group of $N_1$ genes in a subsystem, defined as
the internal gene set, $\Omega$, their gene expression levels are not only controlled by internal interactions among $\Omega$, but also affected by the regulatory factors from other subsystems outside $\Omega$. We define an external gene set, $\Psi$ consisting of those external regulatory factors. For example, we consider the worm-fly orthologous genes as internal set $\Omega$. The worm-fly orthologous TFs from internal set $\Omega$ are the internal regulatory factors, and non-orthologous TFs such as worm- or fly-specific TFs are the external regulatory factors. Both the internal and external regulatory factors control gene expressions in dynamic ways (i.e., their regulatory signals at the current time will affect gene expressions at subsequent times). Thus, the regulatory mechanisms for gene expressions form a control system. In this study, we used a state-space model (defined by linear first-order difference equations, Fig 2A) to formulate temporal gene expression dynamics for internal set $\Omega$ (comprising $N_1$ genes) with external regulation from external

Fig 1. DREISS workflow. (A) DREISS models temporal gene expression dynamics using state-space models in control theory. The “state” refers to the expressions for a large group of genes of interest, such as the worm-fly orthologous genes investigated here. The “control” refers to any other group of genes that contribute to gene expressions of the “state”, such as the species-specific TF studied here. (B) it then projects high-dimensional gene expression space to lower-dimensional meta-gene expression spaces using dimensionality reduction techniques. (C) it derives the effective state-space models for meta-genes so that model parameters can be estimated. (D) it then identifies the meta-gene expression dynamic patterns; i.e., canonical temporal expression trajectories driven by “state” (internal) and by “control” (external) based on the analytic solutions to estimated models. (E) it finally calculates the coefficients of genes for the dynamic patterns of linear transformations between genes and meta-genes.

doi:10.1371/journal.pcbi.1005146.g001
set $\Psi$ (comprising $N_2$ genes) at time points $1, 2, \ldots, T$ as follows:

$$X_{t+1} = AX_t + BU_t$$

(1)

where the vector $X_t \in \mathbb{R}^{N_1 \times 1}$, the “state”, includes $N_1$ gene expression levels at time $t$ in $\Omega$, and the vector $U_t \in \mathbb{R}^{N_2 \times 1}$, the “input or control”, includes $N_2$ gene expression levels at time $t$ in $\Psi$. The system matrix $A \in \mathbb{R}^{N_1 \times N_1}$ captures internal causal interactions among genes in $\Omega$ (i.e., the $i^{th}, j^{th}$ element of $A, A_{ij}$ describes the contribution from the $j^{th}$ gene expression at time $t$ to the $i^{th}$ gene expression at the next time $t+1$), and the control matrix $B \in \mathbb{R}^{N_1 \times N_2}$ captures external causal regulations from the genes $\Psi$ to genes in $\Omega$ (i.e., the $i^{th}, j^{th}$ element of $B, B_{ij}$ describes the contribution from the $j^{th}$ gene expression in $\Psi$ at time $t$ to the $i^{th}$ gene expression in $\Omega$ at the next time $t+1$).

The rest terms $\sum_{k=1}^{l-2} A^k BU_{t-1-k} + BU_{t-1}$ captures the expression of $\Psi$ meta-gene expression levels at time $t+1$, the $i^{th}$ row of the matrix whose columns are right-singular vectors of the matrix $[X_1 X_2 \cdots X_T]$ in $\Omega$ by the singular value decomposition (SVD) [19]; the vector $U_t \in \mathbb{R}^{N_2 \times 1}$, the “meta-gene input or control”, includes $M_t$ ($<< N_1$ and $< T$) meta-gene expression levels; $i.e.,$ the first $M_x$ elements of the $f^{th}$ row of the matrix whose columns are right-singular vectors of the matrix $[X_1 X_2 \cdots X_T]$ in $\Omega$ by the singular value decomposition (SVD) [19]; the vector $U_t \in \mathbb{R}^{N_2 \times 1}$, the “meta-gene input or control”, includes $M_t$ ($<< N_2$ and $< T$) meta-gene expression levels; $i.e.,$ the first $M_x$ elements of the $f^{th}$ row of the matrix whose columns are right-singular vectors of the matrix SVD of matrix $[U_1 U_2 \cdots U_T]$ at time $t$ in $\Psi; W_x \in \mathbb{R}^{N_1 \times M_1}$ is the linear projection matrix of SVD from $M_1$ meta-gene expression space to $N_1$ gene expression space in $\Omega$, and $\perp$ is a pseudo-inverse operation; $i.e.,$ $W \perp W = I$, where $I$ is the identity matrix. (C) Effective state space model for meta-genes. The effective state-space model for meta-genes, Eq 5 is obtained by using linear projections $W_x$ and $U_t$ between genes and meta-genes from Eqs 1–4. The effective meta-gene system matrix $\tilde{A} = W_x^T A W_x \in \mathbb{R}^{N_1 \times N_1}$ captures internal causal interactions among meta-genes in $\Omega$ (i.e., the $i^{th}, j^{th}$ element of $\tilde{A}, \tilde{A}_{ij}$ describes the contribution from the $j^{th}$ meta-gene expression at time $t$ to the $i^{th}$ meta-gene expression at time $t+1$), and the effective control matrix $\tilde{B} = W_x^T B U_t \in \mathbb{R}^{N_1 \times N_2}$ captures external causal regulations from meta-genes in $\Psi$ to meta-genes in $\Omega$ (i.e., the $i^{th}, j^{th}$ element of $\tilde{B}, \tilde{B}_{ij}$ describes the contribution from the $j^{th}$ meta-gene expression in $\Psi$ at time $t$ to the $i^{th}$ meta-gene expression in $\Omega$ at the next time $t+1$). Eq 5 describes the effective state space model for the meta-genes in $\Omega$, whose expression dynamics are determined by $\tilde{A}$ and $\tilde{B}$. Because the meta-gene dimension, $M_t (M_x)$ is less than $T$, and much less than $N_1 (N_2)$, we can estimate $\tilde{A}$ and $\tilde{B}$. doi:10.1371/journal.pcbi.1005146.g002
expression vector of the gene components in Ω affected externally by the genes in Ψ. In particular, \( X^\text{ext}_t = BU_{t-1} \) represents the expression vector of gene components in Ω driven purely by the genes in Ψ since it only involves B and U, and \( X^\text{inter}_t = \sum_{k=1}^{t-2} A^k BU_{t-1-k} \) captures the expression vector of gene components in Ω driven by the interactions between internal and external groups for involving A, B and U. In this paper, we mainly focus on the purely internal dynamics. As for the external-related terms, we should emphasize that any subdivision of the rest of the terms \( \sum_{k=1}^{t-2} A^k BU_{t-1-k} + BU_{t-1} \) is completely arbitrary. That is, although we subdivided it into a purely external term and an interaction term here, one could subdivide it multiple ways. That is, given a particular type of subdivision, each of the subdivided terms sums up a group of terms from \( A^k BU_{t-1-k} \), \( k = 0, 1, 2, \ldots, t-2 \). For example, one can look at \( \sum_{k=1}^{t-2} A^k BU_{t-1-k} + (ABU_{t-2} + BU_{t-1}) \), where \( ABU_{t-2} + BU_{t-1} \) shows the contribution from the inputs up to two time steps back to \( X_t \).

### Dimensionality reduction from genes to meta-genes

The temporal gene expression experiments normally have limited time samples (for example, there may only be a dozen time points), which are far less than the time samples needed to estimate the large matrices A and B when internal and external groups, Ω and Ψ are composed of hundreds or thousands of genes. One way to deal with lack of time samples is dimensionality reduction. Thus, we project high dimensional temporal gene expressions to much lower dimensional meta-gene expression levels using a dimensionality reduction technique (Fig 2B). Those meta-gene expression levels should capture original gene expression patterns, such as the ones having the greatest degree of co-variation. We calculate the meta-gene expression levels as follows:

\[
\tilde{X}_t = W^* X_t; \quad \tilde{U}_t = W^*_U U_t
\]

where \( \tilde{X}_t \in \mathbb{R}^{M_t \times 1} \), the “meta-gene state” at time \( t \), includes \( M_t < N_t \) and \( < T \) meta-gene expression levels; i.e., the first \( M_t \) elements of the \( t \)th row of the matrix whose columns are right-singular vectors of the matrix \([X_1 X_2 \cdots X_T]\) in Ω by the singular value decomposition (SVD) [19]; the vector \( \tilde{U}_t \in \mathbb{R}^{M_t \times 1} \), the “meta-gene input or control” at time \( t \), includes \( M_2 < N_2 \) and \( < T \) meta-gene expression levels; i.e., the first \( M_t \) elements of the \( t \)th row of the matrix whose columns are right-singular vectors from SVD of the matrix \([U_1 U_2 \cdots U_T]\) in Ψ; \( W_X \in \mathbb{R}^{N_t \times M_t} \) is the linear projection matrix of SVD from \( M_t \) meta-gene expression space to \( N_t \) gene expression space in Ω, \( W_U \in \mathbb{R}^{N_t \times M_t} \) is the linear projection matrix of SVD from \( M_t \) meta-gene expression space to \( N_t \) gene expression space in Ψ, and \((\cdot)^+\) is a pseudo-inverse operation; i.e., \( WW^+ = I \), where I is the identity matrix.

### Estimation of effective state-space model for meta-gene expression dynamics

Next, we obtain the effective state-space model for meta-genes using linear projections \( W_X \) and \( W_U \) between genes and meta-genes as follows (Fig 2C). By replacing (1) using (3), we obtain that

\[
W_X \tilde{X}_{t+1} = AW_X \tilde{X}_t + BW_U \tilde{U}_t.
\]

After multiplying the pseudo-inverse of \( W_X \), \( W^*_X \in \mathbb{R}^{M_t \times N_t} \), s.t. \( W^*_X W_X = I \) where I is an identity matrix, at both sides of (4), we have that

\[
\tilde{X}_{t+1} = \underbrace{W^*_X AW_X}_{A} \tilde{X}_t + \underbrace{W^*_X BW_U}_{B} \tilde{U}_t = \tilde{A} \tilde{X}_t + \tilde{B} \tilde{U}_t
\]
Identification of internally and externally driven principal dynamic expression patterns of meta-genes (canonical temporal expression trajectories)

The analytic solution to a general first-order linear matrix difference equation \( [20] \), \( Q_{n+1} = CQ_n \), is \( Q_t = C^tQ_0 = (HE^{-1})^tQ_0 = HE^tQ_0 = HE^tS \), where the columns of the matrix \( H \) are eigenvectors of \( C \), the diagonal elements of the diagonal matrix \( E \) are eigenvalues of \( C \) such that \( CE = HE \), and the vector \( S = H^{-1}Q_0 \). Then, if we rewrite \( Q_t \) by a linear combination of the time exponential of eigenvalues of \( C \), we have that \( Q_t = HE^tS = \sum_{i=1}^{m_e} \alpha_i s_i H^i = \sum_{i=1}^{m_e} \alpha_i K_i \), where \( m_e \) is the total number of eigenvalues of \( C \), \( \alpha_i \) is the \( i^{th} \) eigenvalue of \( C \), \( s_i \) is the \( i^{th} \) element of \( S \), \( H^i \) is the \( i^{th} \) eigenvector of \( C \) (i.e., the \( i^{th} \) column of \( H \)), and \( K_i = s_i H^i \) is the coefficient vector of \( Q_0 \) over the \( i^{th} \) time exponential of \( \alpha_i \).

By Eq 5, the matrix \( \hat{A} \) determines the meta-gene states components whose expression dynamics are internally controlled by the meta-genes of \( \Omega \). As Eq 2, we define the expression of
the meta-gene components driven only internally by themselves in $\Omega$ at time $t$ as $\tilde{X}_t^{\text{INT}}$, an $M_1$-dimensional vector; i.e., their expression at two adjacent time points have

$$\tilde{X}_{t+1}^{\text{INT}} = \tilde{A}\tilde{X}_t^{\text{INT}} \in \mathbb{H}^{M_1 \times 1}.$$  

According to the above analytic solution, it can be a linear combination of $M_1$ dynamic patterns determined by the eigenvalues of the effective system matrix $\tilde{A}$ as follows:

$$\tilde{X}_t^{\text{INT}} = \sum_{p=1}^{M_1} \lambda_p \tilde{K}_p;$$  

i.e., the internally driven component of $i^{\text{th}}$ meta-gene's expression across all time points,

$$\begin{bmatrix}
\tilde{X}_1^{\text{INT}}(i) \\
\tilde{X}_2^{\text{INT}}(i) \\
\vdots \\
\tilde{X}_{M_1}^{\text{INT}}(i)
\end{bmatrix} = \sum_{p=1}^{M_1} \tilde{K}_p(i) \begin{bmatrix}
\lambda_1^p \\
\lambda_2^p \\
\vdots \\
\lambda_{M_1}^p
\end{bmatrix}$$  

(9)

where $\lambda_p$ and $\tilde{K}_p \in \mathbb{C}^{M_1 \times 1}$ are the $p^{\text{th}}$ eigenvalue of $\tilde{A}$ and its coefficient vector from the analytic solution, which determines the $p^{\text{th}}$ dynamic pattern driven by effective internal regulations, defined as the $p^{\text{th}}$ internal principal dynamic pattern (iPDP) $= [\lambda_1^p \lambda_2^p \ldots \lambda_{M_1}^p]$, in which $\lambda_p$ represents the $p^{\text{th}}$ power of $\lambda_p$, and $\tilde{K}(i)$ represents $i^{\text{th}}$ element of the vector $\tilde{K}$. $\mathbb{C}$ represents the complex number domain. If an eigenvalue $\lambda$ is complex when $\tilde{A}$ is asymmetric, then its conjugate $\bar{\lambda}$ is also an eigenvalue, so we sum its iPDP and its conjugate eigenvalue, $\bar{\lambda}$'s iPDP, as a unified iPDP with real elements equal to $[\lambda_1^p + \bar{\lambda}_1^p \lambda_2^p + \bar{\lambda}_2^p \ldots \lambda_{M_1}^p + \bar{\lambda}_{M_1}^p]$.

The internal principal dynamic patterns (iPDPS) represent canonical temporal expression trajectories, which can be either increasing, or damped oscillation and so on depending on iPDP’s eigenvalues (Fig 3). The iPDPs can be ordered by sorting their eigenvalues.

Also by Eqs 2 and 5, the expression of the meta-gene states components driven purely by the external group $\Psi$ at time $t$ is defined as $\tilde{X}_t^{\text{EXT}}$, an $M_1$-dimensional vector, and its expression dynamics is determined by the equation $\tilde{X}_t^{\text{EXT}} = A\tilde{X}_t^{\text{EXT}} \in \mathbb{H}^{M_1 \times 1}$; i.e., the externally driven components of meta-gene states at two adjacent time points. In particular, the externally driven component of $i^{\text{th}}$ internal meta-gene’s expression across time points:

$$\begin{bmatrix}
\tilde{X}_2^{\text{EXT}}(i) \\
\tilde{X}_3^{\text{EXT}}(i) \\
\vdots \\
\tilde{X}_{M_1}^{\text{EXT}}(i)
\end{bmatrix} = \sum_{q=1}^{M_1} \tilde{B}_{i q} \begin{bmatrix}
\tilde{U}_1(q) \\
\tilde{U}_2(q) \\
\vdots \\
\tilde{U}_{M_1(q)}(q)
\end{bmatrix}$$  

(10)

where $\tilde{X}_i^{\text{EXT}}(i)$ and $\tilde{U}_i(q)$ are $i^{\text{th}}$ and $q^{\text{th}}$ elements of $\tilde{X}_t^{\text{EXT}}$ and $\tilde{U}_i$, respectively with $t = 1, 2, \ldots, T$.

| PDP | Real | Complex (radius) |
|-----|------|-----------------|
| eigenvalue | >1 | =1 | <1 & >0 | <0 & > -1 | = -1 | < -1 | > 1 | =1 | < 1 |
| Canonical temporalexpression trajectory (initial) | Increasing (F) | De-creasing (D) | Vibrating early (VE) | Vibrating (V) | Vibrating late (VL) | Under-damped Oscillation (O) | Damped Oscillation (DO) |

**Fig 3. Classification of canonical temporal expression trajectories for iPDP eigenvalue types.** The internal principal dynamic patterns (iPDPS) represent canonical temporal expression trajectories, which can be either increasing, or damped oscillation and so on depending on iPDP’s eigenvalues (The bottom row).

doi:10.1371/journal.pcbi.1005146.g003
the vector \( \begin{bmatrix} \bar{U}_1(q) & \bar{U}_2(q) & \ldots & \bar{U}_{T-1}(q) \end{bmatrix} \) is defined as \( q \)th external principal dynamic pattern (ePDP), and \( B_{t,q} \) is the element of \( B \) at \( t \)th row and \( q \)th column, which is also the coefficient of the externally driven component of \( q \)th internal meta-gene’s expression over \( q \)th ePDP. Based on Eq 2, the expression of the meta-gene components driven by the interactions between internal and external meta-genes is given by \( X^\text{INT} = \sum_{k=1}^{T-2} A^k \bar{B} \bar{U}_{t-1} \). In this paper, we focus on the purely driven internal patterns (i.e., iPDPs) and compare them across different biological systems.

**Identification of gene coefficients of principal expression dynamic patterns**

Because genes and meta-genes have linear relationships in terms of their expression levels as described in Eq 2, the components of gene expression levels in \( \Omega \) driven by internal regulations, \( X^\text{INT} \in \mathbb{R}^{N_t \times 1} \) can be also expressed as linear combinations of \( M \) iPDPs:

\[
X^\text{INT}_t = W^\text{INT}_t \tilde{X}^\text{INT} = \sum_{p=1}^{M} \lambda^p X^\text{INT}_p = \sum_{p=1}^{M} \lambda^p C^p \; ; \text{i.e.,}
\]

the internally driven component of \( t \)th gene’s expression across all time points,

\[
\begin{bmatrix} X^\text{INT}_1(i) & X^\text{INT}_2(i) & \ldots & X^\text{INT}_T(i) \end{bmatrix} = \sum_{p=1}^{M} C^p [ \lambda^1 \; \lambda^2 \; \ldots \; \lambda^T ]^{\text{iPDP}}
\]

where \( C^p = W^\text{INT}_t \tilde{K}^p \in \mathbb{C}^{M \times 1} \) represents the gene coefficient vector for \( p \)th iPDP. Similarly, the gene expression components driven by external genes in \( \Psi \), \( X^\text{EXT} \in \mathbb{R}^{N_t \times 1} \) can be also expressed as linear combinations of \( M \) ePDPs:

\[
X^\text{EXT}_t = W^\text{EXT}_t \tilde{X}^\text{EXT} = W^\text{EXT}_t \bar{B} \bar{U}_t = D \bar{U}_t \; ; \text{i.e.,}
\]

the externally driven component of \( t \)th gene’s expression across all time points,

\[
\begin{bmatrix} X^\text{EXT}_1(i) & X^\text{EXT}_2(i) & \ldots & X^\text{EXT}_T(i) \end{bmatrix} = \sum_{q=1}^{M} D^q [ \bar{U}_1(q) \; \bar{U}_2(q) \; \ldots \; \bar{U}_{T-1}(q) ]^{\text{ePDP}}
\]

where \( X^\text{EXT}_t(i) \) is \( i \)th element of \( X^\text{EXT}_t \) with \( t = 1, 2, \ldots, T \), and \( D_{t,q} \) is the element of \( D = W^\text{EXT}_t \bar{B} \) at \( t \)th row and \( q \)th column, which is also the coefficient of the externally driven component of \( i \)th gene’s expression over \( q \)th ePDP.

**Results**

Gene expression data during embryogenesis provide important information about the dynamics of genomic functions throughout the developmental process, from the conserved functions such as DNA replication to the species-specific functions such as body segmentation, but hardly reveal any data regarding the evolutionary gene regulatory subsystems that drive those developmental functions [3]. Thus, in order to understand the relationships between those subsystems and their driving genomic functions, we apply DREISS to worm and fly gene expression datasets during embryogenesis in modENCODE and we are able to identify various developmental genomic functions of worm-fly orthologous gene pairs driven by two different evolutionary regulatory subsystems, conserved (worm-fly orthologous TFs) and non-
conserved (worm/fly specific TFs). As model organisms for developmental biology, both worm and fly have been used previously to study embryogenesis.

Applications to worm and fly embryonic developmental data in modENCODE: Orthologous genes, transcription factors and gene expression datasets

DREISS enables us to compare expression dynamic patterns between two or more temporal gene expression datasets even though they have different numbers of samples, as well as differences in the times at which those samples were collected. For example, we can apply DREISS to two different datasets of the same group of genes, and identify both the common (similar) and the specific (different) dynamic patterns driven by internal regulations captured by the eigenvalues of the effective system matrices between the two datasets.

In this paper, we apply DREISS to 3,153 one-to-one orthologous genes between worm (*Caenorhabditis elegans*) and fly (*Drosophila melanogaster*) as internal group, Ω to study their expression dynamics during embryonic development [10]. We refer to species-specific TFs as external regulations; i.e., external group Ψ. We found that worm-fly orthologs have similar internal dynamic patterns, which may be mainly driven by conserved TFs, but have very different external dynamic patterns driven by species-specific TFs between worm and fly embryonic developmental stages. The data is summarized as follows.

We define internal group Ω as 3,153 one-to-one orthologous genes between worm and fly during embryonic development, and external group Ψ as all the species-specific TFs (509 worm-specific TFs, 442 fly-specific TFs) [21,22]. We used their temporal gene expression levels (as measured by the RPKM values in RNA-seq) during embryonic development from the mod-ENCODE project [10]. The worm embryonic development dataset includes \( T = 25 \) time stages at 0, 0.5, 1, 1.5, . . . , 12 hours, and the fly dataset includes \( T = 12 \) time stages at 0, 2, 4, . . . , 22 hours, but \( t = 1,2,...,25 \) for worm and \( t = 1,2,...,12 \) for fly are used in this paper, representing the relative time points for the entire embryonic development processes. Because \( M_1 + M_2 < T \) in Eq 8, we choose \( M_1 = M_2 = 5 \) meta-genes for fly (\( T = 12 \)), and find that five meta-genes of Ω and five meta-genes of Ψ capture ~98% of the co-variation of orthologous gene expressions and fly-specific TF gene expressions, respectively. In order to compare worm and fly, we also choose \( M_1 = M_2 = 5 \) meta-genes for worm, which capture ~98% of the co-variation of orthologous gene expressions and worm-specific TF gene expressions.

Meta-genes of worm-fly orthologous genes have similar internal, yet different external principal dynamic patterns during embryonic development

We find that the meta-gene canonical temporal expression trajectories driven by conserved regulatory networks (i.e., internal principal dynamic patterns, iPDPs) include four major patterns in both the worm and fly embryonic developmental process by order of eigenvalues: 1) a late highly varied pattern; 2) an early fast decaying pattern; 3) a slowly increasing pattern; and 4) an oscillating pattern (Fig 4A); i.e., the pattern of canonical trajectories (VL, D, I, O) in Fig 3. In contrast to the observed iPDP similarities, we find that worm and fly have very different external principal dynamic patterns (ePDPs) (Fig 4B); i.e., the expression dynamic patterns driven by species-specific TFs. The principal dynamic patterns driven by the worm-specific regulatory network; i.e., worm ePDPs, include a varied pattern that decreases until the middle stage and then increases, an increasing pattern, a varied pattern with a peak entering middle stage, a pattern that varies early and then increases during the embryonic development, and a cosine-like oscillating pattern with roughly two periods during the embryonic development. The
fly ePDPs, however, have a varied pattern with low expression at the early stage, a sine-like oscillating pattern with roughly two periods during the embryonic development, and a varied pattern that is like damped oscillation. In addition, we checked the sensitivity of iPDPs to small perturbations to internal/external regulatory networks by the leave-one-out method; i.e., we removed one gene in the internal/external group, ran DREISS, and obtained the ordered iPDP eigenvalues for the remaining genes. We repeated the leave-one-out method for all genes, and finally found the ranges in which iPDP eigenvalues vary shown as error bars in S1 Fig. We can see that the iPDP eigenvalues almost stay at the same values (small error bars) for both worm and fly, which implies that the principal dynamic patterns of worm-fly orthologous genes driven by their conserved regulatory network are robust to small changes.
The above results suggest that the conserved regulatory networks from orthologous meta-genes between worm and fly have similar effects to orthologous meta-genes, given their similar iPDPs (i.e., both have four patterns, as described above). The species-specific regulatory networks from species-specific meta-genes (i.e., worm-specific or fly specific TFs) have effects that differ from the orthologous meta-genes for their different ePDPs. In addition, the expression dynamic patterns driven by the interactions between internal orthologous genes and external species-specific TFs are also different between worm and fly (S2 Fig).

Orthologous genes have correlated coefficients between worm and fly for their matched internal principal dynamic patterns

In both worm and fly, we observe the similar four types of internally driven canonical temporal expression trajectories; i.e., four matched internal principal dynamic patterns (iPDPs) (Fig 4A). Thus, we are interested in seeing how individual orthologous genes relate to those dynamic patterns. We find that the worm-fly orthologous genes have correlated coefficients over each of the four iPDPs. Based on Eq 10, we can obtain the coefficients of orthologous genes for each iPDP. We find that their coefficients are significantly correlated between worm and fly iPDPs with a similar pattern (Fig 5): $r = 0.33$ ($p < 2.2e-16$) for the highly varied pattern at late embryonic development stages (first iPDP), $r = 0.66$ ($p < 2.2e-16$) for the fast decaying pattern at early embryonic development stages (second iPDP), $r = 0.67$ ($p < 2.2e-16$) for the slowly increasing pattern during embryonic development (third iPDP), and $r = 0.73$ ($p < 2.2e-16$) for the oscillation pattern during embryonic development (forth iPDP), where $r$ represents Spearman correlation of iPDP coefficients of 3,153 orthologous genes between worm and fly. This implies that, not only do the orthologous meta-genes have similar internal (conserved) regulatory effects (i.e., similar iPDPs), but the worm-fly orthologous genes also have similar internally-driven expression dynamics as resulted from their significantly correlated coefficients for iPDPs. The ePDPs between worm and fly generally do not show a high degree of matching similarity, but the worm ePDP No. 2, and the fly ePDPs No. 3 are roughly representing the growing patterns. We find that orthologous gene correlation coefficients between these ePDP patterns are very small (Spearman correlation $r = -0.22$ of the orthologous gene coefficients of worm ePDP No.2 and fly ePDP No. 3).

Ribosomal genes have significantly larger coefficients for the internal than external principal dynamic patterns, but signaling genes exhibit the opposite trend

The ribosome produces proteins, which is an ancient process and conserved across worm and fly, organisms separated by almost a billion years of evolution. The ribosomal genes are highly expressed during embryogenesis, since intensive cell division and migration require a large amount of proteins to be synthesized. We collected 195 ribosome-related genes based on the GO annotations. We ranked the coefficients of orthologous genes for each iPDP and ePDP in ascending order, and compared the rank values of iPDP and ePDP coefficients of ribosomal genes. We found that their average ranks of iPDP coefficients are significantly larger than ePDP ones in both worm ($t$-test $p < 2.2e-16$) and fly ($t$-test $p < 2.6e-13$) as shown in Fig 6. This means that the ribosomal gene expression is significantly more influenced by the conserved regulatory network than by the species-specific regulatory network, which is consistent with ribosomal genes having conserved functions during embryonic development.

The orthologous genes related to signal transduction for cell-cell communication (a significantly more recent evolutionary adaptation relative to the ribosome) exhibit the opposite trend. We found that 320 signaling genes from GO annotations have significantly larger
average rank values of ePDP coefficients than iPDP ones in both worm (t-test $p<5.6\times10^{-11}$) and fly (t-test $p<8.3\times10^{-4}$), as shown in Fig 6. This result implies that the signaling gene expression is significantly more driven by the species-specific regulatory network than by the conserved regulatory network, which is consistent with the signaling genes being commonly associated with species-specific functions, such as body plan establishment and cell differentiation.
DNA replication and Proteasome machinery are enriched in orthologous genes with high coefficients for the dynamic patterns with fast growing canonical trajectories.

We next turn to the biological meaning of individual canonical temporal expression trajectory for iPDPs and ePDPs. For the fast-decaying pattern (2nd iPDP), we find that the DNA replication is significantly enriched in Top 300 (~10%) orthologous genes that have the most negative coefficients for this pattern, in both worm ($p<1.6e-8$) and fly ($p<4.5e-6$). The GO enrichment analysis was performed using DAVID [23]. The very negative coefficients for the fast decaying...
pattern mean high positive coefficients for a fast-growing pattern (vertically flipped 2nd iPDP) of worm and fly represent a fast-growing pattern), showing a drastic increase at the beginning of embryogenesis, then remain flat during the late embryogenesis (red curve). For the fast-decaying pattern (2nd iPDP), we found that the DNA replication is significantly enriched in Top 300 (~10%) orthologous genes that have the most negative coefficients for this pattern, in both worm (p<1.6e-8) and fly (p<4.5e-6). The very negative coefficients for the fast decaying pattern means high positive coefficients for a fast-growing pattern (red curve). The original expression patterns of those top orthologous genes actually do not have fast-growing patterns (black curve).

Fig 7. DNA replication is enriched in orthologous genes with high coefficients for the dynamic patterns with fast growing canonical trajectories. (A) The first principal component of Top 10% genes with most negative coefficients with 2nd worm iPDP (black curve). (B) The fast-growing iPDP (vertical flipped 2nd iPDP) showing a drastic increase at the beginning of embryogenesis, then remain flat during the late embryogenesis (red curve). For the fast-decaying pattern (2nd iPDP), we found that the DNA replication is significantly enriched in Top 300 (~10%) orthologous genes that have the most negative coefficients for this pattern, in both worm (p<1.6e-8) and fly (p<4.5e-6). The very negative coefficients for the fast decaying pattern means high positive coefficients for a fast-growing pattern (red curve). The original expression patterns of those top orthologous genes actually do not have fast-growing patterns (black curve).

doi:10.1371/journal.pcbi.1005146.g007

Besides a fast-growing pattern driven by conserved worm-fly orthologous TFs, we also identified a fast-growing pattern driven by non-conserved TFs for the two species. The Top 300 orthologous genes (~10%) with the fast-growing worm ePDP (ePDP No.2) (i.e., driven by species-specific regulatory networks) are enriched in 'proteasome' (p<9.8e-16). Protein degradation is not only a key process in apoptosis, but also throughout the entire course of development [26,27]. For example, eliminating proteins that are no longer needed is a vital process
during embryo development; e.g., the maternal proteins need to be cleaned as the embryogenesis proceeds). Previous reports also showed that different species usually have different maternal mRNA in the oocyte, which indicates that species-specific strategies might be utilized to regulate the protein degradation process [28]. In this study, after separating the effect of conserved and non-conserved regulatory networks, we observed that the protein degradation is significantly enriched in the genes majorly driven by species-specific TFs in worms. In contrast, the Top 300 orthologous genes with fast growing fly ePDP3 are enriched in 'mitotic cell cycle' \( (p<3.5e-29) \), 'translation' \( (p<1e-30) \) and 'mitochondrion' \( (p<7.7e-20) \). Those enriched function related to energy generation is probably indicative of the large energy requirement during fly embryogenesis [29], which did not provide the evolutionary conservation of this energy-related gene regulation. Our result reveals that the fly genes associated with respiration are more up-regulated by fly-specific TFs relative to conserved TFs, and that this up-regulation evolved after the separation of worm and fly.

Besides the fast-growing pattern driven by species-specific TFs, we also observed some other interesting patterns. For example, worm ePDP3 displays a dramatic peak about 5 hours after fertilization. Among the Top 300 worm orthologous genes of this pattern, genes involved in synaptic transmission \( (p<5.6e-9) \) and cell-cell signaling \( (p<1e-7) \) are over-represented, suggesting that they are transiently activated in this stage of embryogenesis by worm-specific TFs. This observation indicates the gene regulatory network for these genes have evolved after the speciation.

**Human-specific transcription factors respond to hormonal stimulation during breast cancer cell cycle**

We applied DREISS to another example (also see supplement) about cancer. We are also interested to identify the gene expression dynamic patterns driven by conserved and human-specific regulatory networks during breast cancer cell cycle. Thus, we applied DREISS to a time-series gene expression data for human estrogen-responsive breast cancer cell line (ZR-75.1) before and after hormonal stimulation, which has 12 time points covering a complete mitotic cell cycle (0–32 hours) of hormonal stimulated cells [30]. The internal group, \( \Omega \) is defined as a set of cross-species conserved human genes (i.e., 1132 worm-fly-human orthologs including 150 orthologous TFs), and the external group, \( \Psi \) consists of 1870 human-specific TFs. As shown in S3 Fig, the internally driven principal dynamic patterns (iPDPs) of conserved human genes include an oscillation trajectory whose period is roughly equal to a full cell cycle (iPDP No. 4), but the externally driven patterns (ePDPs No. 2–4) oscillates more frequently than internal one, which suggests that though the evolutionarily conserved TFs regulate the normal cell cycle, the human specific TFs potentially drive the abnormal cycling behaviors of conserved gene expression responding to the hormonal stimulation.

**Discussion**

In this paper, we presented a novel computational method, DREISS, which decomposes time-series expression data of a group of genes into the components driven by the regulatory network inside the group (internal regulatory subsystem), and the components driven by the external regulatory network consisting of regulators outside the group (external regulatory subsystem). DREISS is a general-purpose tool that can be used to study the gene regulatory effects of any interested biological subsystems such as protein-coding transcription factors, microRNAs, epigenetic factors and so on. As an illustration, we applied DREISS to the time-series gene expression datasets for worm and fly embryonic developments from the modENCODE project [10], and compared the worm-fly orthologous gene expression dynamic patterns driven by the conserved regulatory network (i.e., regulation effects from orthologous TFs), with the
patterns driven by the species-specific regulatory networks (i.e., regulation effects from worm or fly specific TFs). We found that the conserved TFs drive similar genomic functions, but non-conserved TFs drive species-specific functions of orthologous genes between worm and fly, implying that, in addition to having ancient conserved functions, orthologous genes have been regulated by evolutionarily younger GRNs to execute species-specific functions during the evolution. This work can be easily extended to study the regulatory effects from orthologous TFs and species-specific TFs to species-specific genes. For example, one can find the expression dynamic patterns of worm/fly specific genes driven by specific TFs, and identify the genes with strong patterns associated with worm/fly specific functions, such as body formations. To the best of our knowledge, DREISS is the first method to reveal how the evolution of GRNs affects gene expression during embryogenesis.

We emphasize that DREISS is a general-purpose method (a free downloadable R tool available from github.com/gersteinlab/dreiss). Users can define the internal group ($\Omega$) and external group ($\Psi$) according to their interests. For example, if users want to identify the protein-coding expression patterns driven by miRNAs, they can define miRNAs as an external group and protein-coding genes as an internal group. Additionally, DREISS can be applied to more than two datasets, such as comparing worm, fly and human embryonic stem cell developmental data, and finding their conserved and specific developmental expression patterns. The expression patterns driven by human-specific regulatory factors will potentially help us understand human-specific developmental processes along with the associated human genes.

Due to the limited time samples in gene expression datasets, DREISS uses the simple linear state space model (i.e. the first order linear invariant difference equation) to model the temporal gene expression dynamics, and identify principal temporal dynamic patterns. This model assumes that the gene regulatory networks controlling temporal gene expression dynamics does not change across the entire biological process such as $(A, B)$ in Eq 1. Thus, based on the analytic analysis, the principal dynamic patterns (PDPs) must follow a small set of canonical temporal trajectories (Fig 3). With the rapidly increasing gene expression data, we can extend DREISS to more advanced models such as switched and hybrid system models, non-linear models [31], which will allow us to study the gene regulatory networks are time varying, and potentially find the more temporal gene expression patterns capturing the more complex gene regulatory activities.

**Supporting Information**

**S1 Fig. Principal dynamic patterns and their eigenvalues.** Internal principal dynamic patterns (iPDPs) of orthologs during worm and fly embryonic development. Barplots show the eigenvalues of iPDPs. The error bar for each eigenvalue tells the its variation range. We left one gene out, and calculated eigenvalues for the remaining genes thus obtaining the eigenvalue variations. The curves show the canonical temporal expression trajectories of iPDPs. (TIF)

**S2 Fig. The expression dynamic patterns driven by the interactions between worm-fly orthologs and species-specific TFs.** The first five singular vectors (>95% covariance in total) of $[X_{\text{INTER}}^t, t = 1, 2, 3, \ldots, T]$ defined at the end of Section "Identification of internally and externally driven principal dynamic expression patterns of meta-genes (ca-nonical temporal expression trajectories)." (TIF)

**S3 Fig. Internally and externally principal dynamic patterns of cross-species conserved gene expression during human breast cancer cell cycle after hormonal stimulation.**
horizontal axis represents 12 time points from 0 to 32 hours during a complete mitotic breast cancer cell cycle (E-TABM-631, ArrayExpress). The vertical axis represents the normalized PDP expression with the vector norm equal to one. The internal group is defined as a set of cross-species conserved human genes (i.e., 1132 worm-fly-human orthologs; including 150 orthologous TFs), and the external group consists of 1870 human-specific TFs.

S1 Table. Examples of internal and external regulatory networks.

Author Contributions

Conceived and designed the experiments: DW MG.

Performed the experiments: DW.

Analyzed the data: DW FH.

Contributed reagents/materials/analysis tools: DW FH SM.

Wrote the paper: DW FH MG.

References

1. Kim PM, Tidor B (2003) Subsystem identification through dimensionality reduction of large-scale gene expression data. Genome Res 13: 1706–1718. doi: 10.1101/gr.903503 PMID: 12840046
2. Vilar JM (2006) Modularizing gene regulation. Mol Syst Biol 2: 2006 0016. doi: 10.1038/msb4100058 PMID: 16738561
3. Peter IS, Davidson EH (2011) Evolution of gene regulatory networks controlling body plan development. Cell 144: 970–985. doi: 10.1016/j.cell.2011.02.017 PMID: 21414487
4. Chen K, Rajewsky N (2007) The evolution of gene regulation by transcription factors and microRNAs. Nat Rev Genet 8: 93–103. doi: 10.1038/nrg1990 PMID: 17230196
5. Levin M, Hashimshony T, Wagner F, Yanai I (2012) Developmental milestones punctuate gene expression in the Caenorhabditis embryo. Dev Cell 22: 1101–1108. doi: 10.1016/j.devcel.2012.04.004 PMID: 22560298
6. Kalinka AT, Varga KM, Gerrard DT, Preibisch S, Corcoran DL, et al. (2010) Gene expression divergence recapitulates the developmental hourglass model. Nature 468: 811–814. doi: 10.1038/nature09634 PMID: 21150996
7. Yanai I, Peshkin L, Jorgensen P, Kirschner MW (2011) Mapping gene expression in two Xenopus species: evolutionary constraints and developmental flexibility. Dev Cell 20: 483–496. doi: 10.1016/j.devcel.2011.03.015 PMID: 21497761
8. Irie N, Kuratani S (2011) Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. Nat Commun 2: 248. doi: 10.1038/ncomms1248 PMID: 21497719
9. Casci T (2011) Development: Hourglass theory gets molecular approval. Nat Rev Genet 12: 76. doi: 10.1038/nrg2940 PMID: 21173773
10. Gerstein MB, Rozowsky J, Yan KK, Wang D, Cheng C, et al. (2014) Comparative analysis of the transcriptome across distant species. Nature 512: 445–448. doi: 10.1038/nature13424 PMID: 25164755
11. Brogan WL (1991) Modern control theory. Englewood Cliffs, N.J.: Prentice Hall. xviii, 653 p. p.
12. Rangel C, Angus J, Ghahramani Z, Lioumi M, Sotheran E, et al. (2004) Modeling T-cell activation using gene expression profiling and state-space models. Bioinformatics 20: 1361–1372. doi: 10.1093/bioinformatics/bth093 PMID: 14962938
13. Bansal M, Deila Gatta G, di Bernardo D (2006) Inference of gene regulatory networks and compound mode of action from time course gene expression profiles. Bioinformatics 22: 815–822. doi: 10.1093/bioinformatics/bti003 PMID: 16418235
14. Huang S, Ingber DE (2006) A non-genetic basis for cancer progression and metastasis: self-organizing attractors in cell regulatory networks. Breast Dis 26: 27–54. doi: 10.3233/bd-2007-26104 PMID: 17473364
15. Saey Y, Inza I, Larrañaga P (2007) A review of feature selection techniques in bioinformatics. Bioinformatics 23: 2507–2517. doi: 10.1093/bioinformatics/btm344 PMID: 17720704

16. Chu S, DeRisi J, Eisen M, Mulholland J, Botstein D, et al. (1998) The transcriptional program of sporulation in budding yeast. Science 282: 699–705. doi: 10.1126/science.282.5389.699 PMID: 9784122

17. Wang D, Arapostathis A, Wilke CO, Markey MK (2012) Principal-oscillation-pattern analysis of gene expression. PLoS One 7: e28805. doi: 10.1371/journal.pone.0028805 PMID: 22253697

18. Wang D, Markey MK, Wilke CO, Arapostathis A (2012) Eigen-genomic system dynamic-pattern analysis (ESDA): modeling mRNA degradation and self-regulation. IEEE/ACM Trans Comput Biol Bioinformatics 9: 430–437. doi: 10.1109/TCBB.2011.150 PMID: 22084146

19. Golub GH, Van Loan CF (1996) Matrix computations. Baltimore: Johns Hopkins University Press. xxvii, 694 p. p.

20. Cull P, Flahive ME, Robson RO (2005) Difference equations: from rabbits to chaos. New York: Springer. xiii, 392 p. p.

21. Reece-Hoyes JS, Deplancke B, Shingles J, Grove CA, Hope IA, et al. (2005) A compendium of Caenorhabditis elegans regulatory transcription factors: a resource for mapping transcription regulatory networks. Genome Biol 6: R110. doi: 10.1186/gb-2005-6-13-r110 PMID: 16420670

22. Shazman S, Lee H, Socol Y, Mann RS, Honig B (2014) OnTheFly: a database of Drosophila melanogaster transcription factors and their binding sites. Nucleic Acids Res 42: D167–171. doi: 10.1093/nar/gkt1165 PMID: 24271386

23. Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4: 44–57. doi: 10.1038/nprot.2008.211 PMID: 19131956

24. Bate M, Martinez Arias A (1993) The Development of Drosophila melanogaster. Plainview, N.Y.: Cold Spring Harbor Laboratory Press.

25. Baugh LR, Hill AA, Sionim DK, Brown EL, Hunter CP (2003) Composition and dynamics of the Caenorhabditis elegans early embryonic transcriptome. Development 130: 889–900. doi: 10.1242/dev.00302 PMID: 12938516

26. DeRenzo C, Seydoux G (2004) A clean start: degradation of maternal proteins at the oocyte-to-embryo transition. Trends Cell Biol 14: 420–426. doi: 10.1016/j.tcb.2004.07.005 PMID: 15308208

27. Du Z, He F, Yu Z, Bowerman B, Bao Z (2015) E3 ubiquitin ligases promote progression of differentiation during C. elegans embryogenesis. Dev Biol 398: 267–279. doi: 10.1016/j.ydbio.2014.12.009 PMID: 25523993

28. Shen-Or SS, Pilpel Y, Hunter CP (2010) Composition and regulation of maternal and zygotic transcriptomes reflects species-specific reproductive mode. Genome Biol 11: R58. doi: 10.1186/gb-2010-11-6-r58 PMID: 20515465

29. Tennessen JM, Bertagnotti NM, Evans J, Sieber MH, Cox J, et al. (2014) Coordinated metabolic transitions during Drosophila embryogenesis and the onset of aerobic glycolysis. G3 (Bethesda) 4: 839–850. doi: 10.1534/g3.114.010652 PMID: 24622932

30. Mutarelli M, Cicatiello L, Ferraro L, Grober OM, Ravo M, et al. (2008) Time-course analysis of genome-wide gene expression data from hormone-responsive human breast cancer cells. BMC Bioinformatics 9 Suppl 2: S12. doi: 10.1186/1471-2105-9-S2-S12 PMID: 18387200

31. Schaft AJvD, Schumacher JM (2000) An introduction to hybrid dynamical systems. London; New York: Springer. xi, 174 p. p.