# Point-by-Point response letter for revision

**Reviewer 1**

-- Ref 1.1 –Math equations--

|  |  |
| --- | --- |
| Reviewer  Comment | Most of the math is very well explained, and the figures are extremely helpful. But I think section 2.4 and 2.5 are not as clear as the other sections. In particular, on page 8, I don’t understand how the equation in the 3rd line is related to the one in the first line. The eigen-decomposition of A^~ would give U\*S\*inv(U) (where U is the matrix of eigenvectors, one per column) and I cannot get how X^~^INT\_t is decomposed. In general, what the math is doing in this paper is clear, so I think it is just a matter of fixing (or explaining better) the equations on pages 8 and 9. |
| Author  Response | Thank the reviewer. In order to explain more clearly in Sections 2.4 and 2.5, we added the following paragraph which introduces the analytic solution to the first-order linear matrix difference equation in a general form at the beginning of Section 2.4. The analytic solution to the equation can be decomposed by a linear combination of eigenvalue time exponentials. Because the internally/externally driven components of meta-gene expression follows the first-order linear matrix difference equation, the components can be also decomposed by the linear combination of ’s eigenvalue time exponentials (i.e., defined PDPs in our study). |
| Excerpt From Revised Manuscript | “The analytic solution to a general first-order linear matrix difference equation [[20](#_ENREF_20)], *Yt*+1=*CYt* is  *Yt*=*CtY0*=(*PDP*-1) *t Y0*=*PDtP*-1*Y0=PDtS*, where the columns of the matrix *P* are eigenvectors of *C*, the diagonal elements of the diagonal matrix *D* are eigenvalues of *C* such that *CP*=*PD*, and the vector *S*= *P*-1*Y0*. Then, if we rewrite *Yt* by a linear combination of the time exponential of eigenvalues of *C*, we have that , where *mc* is the total number of eigenvalues of *C*, *αi* is the *i*th eigenvalue of *C*, *si* is the *i*th element of *S*, *Pi* is the *i*th eigenvector of *C* (i.e., the *i*th column of *P*), and *Ki*=*siPi* is the coefficient vector of *Yt* over the *t*th time exponential of *αi*.”  20. Cull P, Flahive ME, Robson RO (2005) Difference equations : from rabbits to chaos. New York: Springer. xiii, 392 p. p. |

-- Ref 1.2 –More applications--

|  |  |
| --- | --- |
| Reviewer  Comment | This paper is about a novel method for bioinformatics. Therefore, while the results presented on the embryogenesis datasets of worm and fly are good, to convince that the method is good I think it would be helpful to show at least another example. Maybe the authors could implement one of the examples they mention in the second paragraph of the discussion. |
| Author  Response | Thank the reviewer for this good suggestion. We applied DREISS to a  human cancer dataset: the time-series gene expression of 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 (E-TABM-631, ArrayExpress). 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. As the following supplemental Figure shows, the internally driven principal dynamic patterns of conserved genes include an oscillation trajectory whose period is roughtly equal to a full cell cycle (iPDP No. 4), but the externally driven oscillation trajectory (ePDP No. 4) oscillates more frequently than internal one, which implies that though the evolutionarily conserved TFs regulate the normal cell cycle, the human specific TFs potentially drive the abnormal cell cycle behaviors of conserved gene expression after hormonal stimulation.  Supplemental Figure – Internally and externally principal dynamic patterns of cross-species conserved gene expression during human breast cancer cell cycle after hormonal stimulation. The horizonal 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 normailzed PDP expression with vector norm being 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. |
| Excerpt From Revised Manuscript |  |

-- Ref 1.3 –Sensitivity--

|  |  |
| --- | --- |
| Reviewer  Comment | Could the authors comment on how sensitive the method is to small changes in the gene expression data? On page 11, in the last sentence of the first paragraph of section 3.2, they mention calculating eigenvalues while leaving one gene out. But this is not clearly explained and it would be good if the authors could expand on this. |
| Author  Response | Thank the reviewer. We elaborated our leave-one-out method at the end of Section 3.2 to test how iPDP/ePDPs are sensitive to the small changes as follows. |
| Excerpt From Revised Manuscript | “…In addition, we checked the sensitivity of iPDP/ePDPs 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/ePDP eigenvalues for the remaining genes. We repeated the leave-one-out method for all genes, and finally found the ranges in which iPDP/ePDP eigenvalues vary shown as error bars in Figure S1. We can see that the iPDP eigenvalues vary less than ePDP ones for both worm and fly, which implies that the principal dynamic patterns of worm-fly orthologous genes driven by their conserved regulatory network are more robust to small changes than ones driven by their species-specific regulatory networks.” |

-- Ref 1.4 –Step A--

|  |  |
| --- | --- |
| Reviewer  Comment | The methods section subdivides DREISS into 5 steps. However, Step A is not really a step of the method per se; it is just the definition/assumption of the model. Maybe it would be clearer if it would be described in that way. |
| Author  Response | Thank the reviewer. Besides modeling, the Step A also requires the users to define internal and external gene groups of interest. Thus we would like to keep this step as an initiative step. We added two sentences in Step A as follows for clarification. |
| Excerpt From Revised Manuscript | “…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….” |

-- Ref 1.5 –Math notations--

|  |  |
| --- | --- |
| Reviewer  Comment | In line 10 of section 2.2, I believe it should be: ‘; i.e. the values of the projection of X\_t onto the first M\_1 singular vectors’ (and similarly on line 12). |
| Author  Response | Thank the reviewer. We rephrased the descriptions about meta-gene states and controls via the singular value decomposition (SVD) as follows. |
| Excerpt From Revised Manuscript | “…the “meta-gene state” at time t, includes M1 (<< N1 and <T) meta-gene expression levels; i.e., the first M1 elements of the tth row of the matrix whose columns are right-singular vectors of the matrix  in group X by the singular value decomposition (SVD) [19]; … the “meta-gene input or control” at time t, includes M2 (<< N2 and <T) meta-gene expres-sion levels; i.e., the first M2 elements of the tth row of the matrix whose columns are right-singular vec-tors from SVD of the matrix  in group U;”  [19] Golub GH, Van Loan CF (1996) Matrix computations. Baltimore: Johns Hopkins University Press. xxvii, 694 p. p. |

-- Ref 1.6 –English--

|  |  |
| --- | --- |
| Reviewer  Comment | The first sentence on page 8, and the sentence starting with ‘The meta-gene dynamic patterns. . .’ on page 11 do not seem to be well formed in English. |
| Author  Response | Thank the reviewer. We fixed the English expressions as follows. |
| Excerpt From Revised Manuscript | On Page 8, “From Equation (4), the internally driven components of meta-gene states at two adjacent time points have X ̃\_(t+1)^INT=A ̃X ̃\_t^INT∈R^(M\_1×1). According to the analytic solution, the components of meta-gene expressions in X driven by effective internal regulations are linear combinations of M1 dynamic patterns determined by the eigenvalues of the effective system matrix as follows:...”  On Page 11, “…The meta-gene canonical temporal expression trajectories driven by the worm-specific regulatory network; i.e., worm ePDPs, include…” |

-- Ref 1.7 –Typo--

|  |  |
| --- | --- |
| Reviewer  Comment | In the third paragraph of page 14, ‘interssting’ should be ‘interesting’. |
| Author  Response | Thank the reviewer for pointing this typo. We fixed this typo. |
| Excerpt From Revised Manuscript | “…we also noted a couple of interesting observations.” |

**Reviewer 2**

-- Ref 2.1 –Gene Set Enrichment Analysis --

|  |  |
| --- | --- |
| Reviewer  Comment | The authors used DAVID to analyze the enrichment of biological processes for individuals PDP patterns by taking the top 10% of genes with the largest coefficients. Since each gene has an associated coefficient for each PDP, they could use the Gene Set Enrichment Analysis (GSEA) to identify significantly enriched pathways without the need to apply an arbitrary cutoff to decide the number of genes to include in the analysis. |
| Author  Response | Thank reviewer for this good suggestion. We reran GSEA analysis and found the similar results. For example, the DNA replication is also enriched with the fast decaying patterns (2nd iPDPs) with *p*<0.0001 from GSEA analysis for both worm and fly. In our study, because we just wanted to demonstrate the bioloigcal meanings of PDPs, we looked at the enriched DAVID functions of top10% genes, which might be arbitrary though, but is very straighforward for for a broader audience. |

Supplemental Table I – Examples of internal and external regulatory networks

|  |  |  |
| --- | --- | --- |
| Interested system | Internal regulatory network | External regulatory network |
| Cross-species conserved genes | Orthologous transcriptional factors (TFs) | Species-specific TFs |
| Protein-coding genes | TFs | micro-RNAs |
| Individual’s protein coding genes | Wild-type TFs | Somatic mutated TFs |
| Protein-coding genes in brain | Commonly expressed TFs | Brain-specific expressed TFs |
| Protein-coding genes in development | House-keeping TFs | Developmental TFs |