Published online 9 May 2017 *Nucleic Acids Research, 2017, Vol. 45, Web Server issue* ***W103–W108***
doi: 10.1093/nar/gkx402

CSTEA: a webserver for the Cell State Transition Expression Atlas

Guanghui Zhu[1,†], Hui Yang[2,†], Xiao Chen[1], Jun Wu[1], Yong Zhang[2,*] and Xing-Ming Zhao[1,*]

[1]Department of Computer Science and Technology, Tongji University, Shanghai 201804, China and [2]Translational Medical Center for Stem Cell Therapy & Institute for Regenerative Medicine, Shanghai East Hospital, School of Life Science and Technology, Shanghai Key Laboratory of Signaling and Disease Research, Tongji University, Shanghai 200092, China

Received February 20, 2017; Revised April 14, 2017; Editorial Decision April 27, 2017; Accepted April 28, 2017

ABSTRACT

Cell state transition is one of the fundamental events in the development of multicellular organisms, and the transition trajectory path has recently attracted much attention. With the accumulation of large amounts of "-omics" data, it is becoming possible to get insights into the molecule mechanisms underlying the transitions between cell states. Here, we present CSTEA (Cell State Transition Expression Atlas), a webserver that organizes, analyzes and visualizes the time-course gene expression data during cell differentiation, cellular reprogramming and trans-differentiation in human and mouse. In particular, CSTEA defines gene signatures for uncharacterized stages during cell state transitions, thereby enabling both experimental and computational biologists to better understand the mechanisms of cell fate determination in mammals. To our best knowledge, CSTEA is the first webserver dedicated to the analysis of time-series gene expression data during cell state transitions. CSTEA is freely available at http://comp-sysbio.org/cstea/.

INTRODUCTION

Cell state transition is a dynamic process, in which the following three types of transitions are highly important: cell differentiation, cellular reprogramming and trans-differentiation. Cell differentiation is a process during which differentiation potential decreases progressively. In mammals, cell differentiation begins from a totipotent zygote and ends with hundreds of differentiated cell types that are essential for the normal functions of a complex organism (1). Through cellular reprogramming technologies, especially somatic cell nuclear transfer (SCNT) (2) and induced pluripotent stem cell (iPSC) (3) technologies, different types of somatic cells can be converted to highly pluripotent cell types (4). Cellular reprogramming technologies not only offer efficient and convenient tools for dissecting the principles of cell fate determination during normal development and disease dysfunctions (5) but also provide a valuable resource of patient-specific cells for the study and potential treatment of human diseases (6). Trans-differentiation is the process of lineage conversion between different somatic cells without an intermediate pluripotent state. For example, B cells can be reprogrammed to macrophages through induction with a transcription factor (7). Although the cells that are produced often have residual characteristics of the cell type of origin, trans-differentiation holds great promise for biomedical applications, such as regenerative medicine (8). Recently, the trajectory path, rather than the origin and destination of cell state transitions, has drawn much attention, especially regarding the features of uncharacterized intermediate states, which are usually unstable and reversible but are informative for revealing the mechanisms of cell fate determination (9).

Gene expression datasets are valuable resources for monitoring the process of cell state transition and further elucidating the pattern of cell fate determination. In databases such as the Gene Expression Omnibus (GEO) (10) and ArrayExpress (11), abundant datasets produced through the development of microarray and next-generation sequencing techniques have been deposited. The large amount of deposited data could be confusing for studies on cell state transition, as no transition-specific labels are provided in these data resources. Several websites specifically addressing gene expression data from different cell states have been established to make efficient use of public datasets. In the web-based platform Gene Expression Commons (http://gexc.riken.jp), gene expression data on cell types are deposited, including stem cells, progenitor cells, and differentiated cells in the haematopoietic system; this platform utilizes a large number of reference datasets to determine the gene expression level of a particular cell type (12). LifeMap Discov-

*To whom correspondence should be addressed. Tel: +86 21 65981196; Fax: +86 21 65981041; Email: yzhang@tongji.edu.cn
Correspondence may also be addressed to Xing-Ming Zhao. Tel: +86 21 69583959; Fax: +86 21 69583959; Email: xm_zhao@tongji.edu.cn
†These authors contributed to this work as first authors.

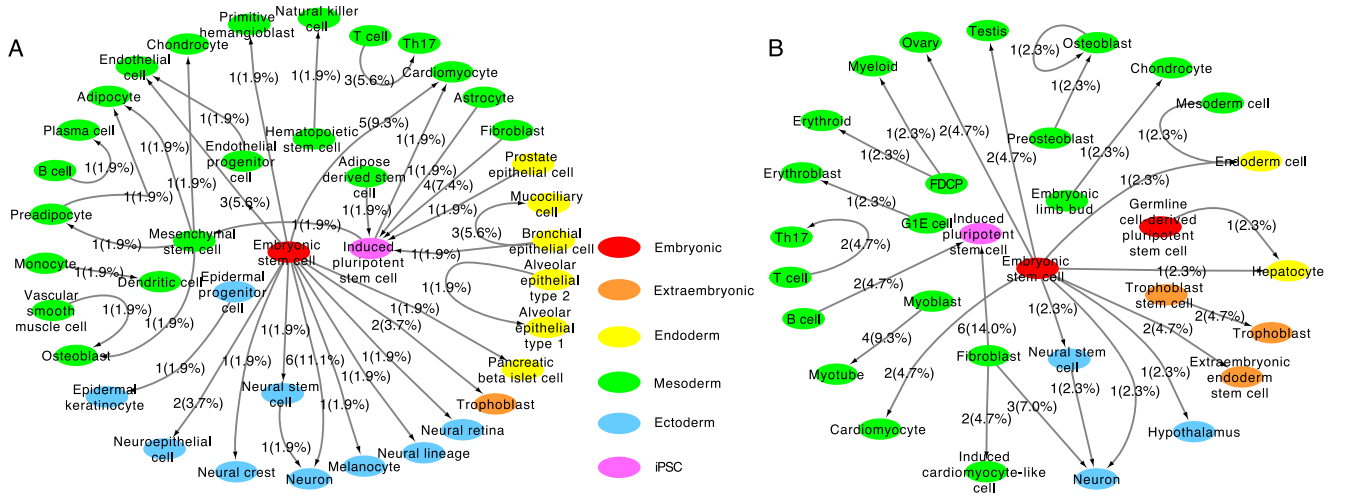

Figure 1. The landscape of cell state transitions in (**A**) human and (**B**) mouse, where the nodes denote cell types, and the edges are the transition paths between cell types. The number accompanying each edge is the number of datasets describing the transition, and the number in the bracket denotes the percentage of datasets for the transition.

ery (http://discovery.lifemapsc.com) is a database providing expression datasets related to embryonic development and manually curated information on the induction of differentiation (13). GenomicScape (www.genomicscape.com) provides visualization, clustering and differential expression analysis of the gene expression profiles of different cell populations (14). However, cell state transition is a complicated dynamic process involving time-dependent regulatory events that cannot be simply reflected by gene expression data for several static cell types. As specific analyses of time-series data are of great value in illuminating the still unclear mechanism of cell fate determination, we have developed the CSTEA, a webserver for the **C**ell **S**tate **T**ransition **E**xpression **A**tlas, which focuses on providing comprehensive analysis and visualization of time-series gene expression data not only for the original and destination cell states but also for un-characterized intermediate cell states during cell transitions.

CSTEA—the Cell State Transition Expression Atlas

The CSTEA is a web server that aims to provide systematic and comprehensive analysis of time-series gene expression data during cell state transitions. The gene expression datasets used by the CSTEA were collected from the Gene Expression Omnibus (GEO) database (10), where all of the datasets have been manually curated, and only datasets with at least five time points were retained. Datasets produced using the same platform were pre-processed via the same procedure, while normalization was conducted for each dataset individually. For microarray datasets, after the raw data was retrieved from GEO, the probe IDs were converted to Refseq IDs with the Brainarray Chip Description Files (CDFs) (http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF). Subsequently, RMA with R package affy (15), normexp with R package limma (16) and variance stabilizing transform with R package lumi (17) were respectively employed for preprocessing the data generated with Affymetrix, Agilent and Illumina microarrays. For RNA-seq datasets, after downloaded from NCBI Sequence Read Archive (SRA), all the reads were firstly aligned to mouse (mm10) and human (hg19) reference genome assembly using STAR with default parameters (18), and the GFOLD algorithm was then used to count the number of reads mapped to each transcript and to calculate the RPKM value (19). In particular, descriptions of the conditions under which the data were generated were also added to CSTEA to facilitate the understanding of the datasets. For each dataset, genes that were differentially expressed between two consecutive time points were identified first. The differentially expressed genes (DEGs) were required to have at least two-fold change and p-value less than 0.05 according to the t-test if more than three replicates were available for each time point for microarray data, and the same fold-change and P-value cutoffs obtained through the GFOLD algorithm were established for RNA-seq data. If a gene was identified as a DEG for at least two consecutive time points, the gene was regarded as a transition-specific DEG, and the set of transition-specific DEGs was used as the gene signature for the transition procedure. In total, 97 datasets were collected, 54 for human and 43 for mouse (Table 1). Figure 1 shows the landscapes of cell state transitions in human and mouse based on the gene expression data deposited in the CSTEA, from which it can be observed that differentiation from embryonic stem cells is the most widely studied transition.

As shown in Figure 2, the CSTEA provides a powerful and user-friendly interface for the analysis and visualization of gene expression data during cell state transitions. In addition to simply browsing the cell state transitions in the CSTEA, users can query the server based on the gene or transition path of interest. In particular, the CSTEA can be queried with a single gene, transition path or gene signature. Users can query a single gene against the CSTEA to determine which cell state transitions might involve the gene. All of the datasets in which the gene is expressed will be listed. If a specific dataset is chosen, the expression pro-

Table 1. Statistics of time-series gene expression datasets deposited in the CSTEA

Species	Technology	Datasets	Cell types	Average intermediate time points
Human	Microarray	46	38	6.0
	RNA-seq	8	10	6.9
Mouse	Microarray	30	28	7.8
	RNA-seq	13	10	5
Total		97	59	6.5

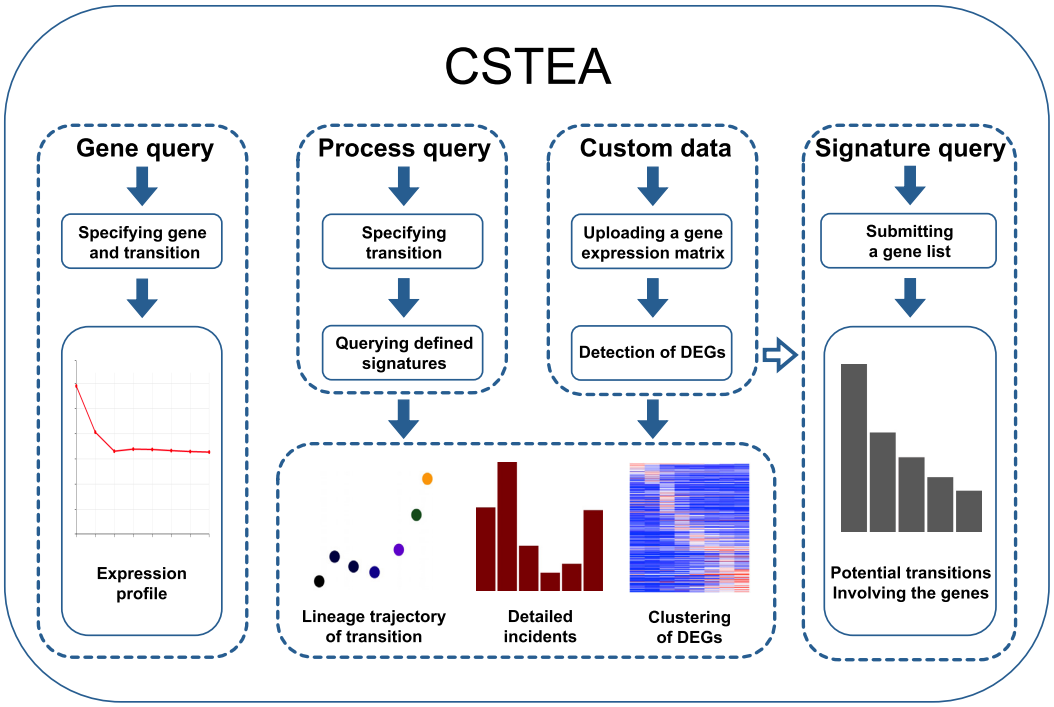

Figure 2. The schematic demonstration of the CSTEA server. The server can be queried with single genes, transition process and signature. The users can upload their custom data for analysis and visualization.

file of the gene across all time points will be shown so that the users can easily investigate how the gene regulates the transition process. If the gene queried is a DEG in a specific transition, all of the DEGs for the transition will be shown so that the users can investigate how the gene interacts with other DEGs during the cell state transition.

If the users are interested in a specific transition between a pair of cell types and want to determine which genes might regulate the transition, they can use the process query function to view the genes that play potential important roles in the transition process. When a specific dataset is chosen for the transition of interest, the trajectory of gene expression changes during the transition will be shown based on principle component analysis of the gene expression profiles. At the same time, a dendrogram will be shown so that it can be clearly visualized which stages are more similar to each other. Furthermore, the DEGs for each time point will be obtained by comparing each time point to the previous one, and the number of DEGs will be visualized as a bar chart for all time points. Using this chart, the critical transition stage may be identified. The detailed list of DEGs for every time point can be downloaded, and functional processes (from Gene Ontology (20)) or pathways (from KEGG (21)) that may involve the DEGs are shown by performing functional enrichment analysis. In addition, a heatmap is available for the transition-specific DEGs so that how those genes are expressed at each time point can be clearly visualized.

Under the signature query function, the users can also submit a gene list obtained from their own studies to check which transitions might involve these genes. In the CSTEA, a signature has been defined for each cell state transition, with all transition-specific DEGs identified for all datasets describing the same origin and destination cell types. With the cell state transition signatures defined, it will be straightforward to determine which transitions involve the queried list of genes by performing enrichment analysis using Fisher's exact test against all transition signatures. A transition process whose gene signature is enriched for the user-defined gene set will be regarded as a potential cell state transition regulated by the gene set. For each potential transition, a score defined as $-\log_{10}(P\text{-value})$ will be employed to quantify the relevance of the gene list to the transition, where the P-value is obtained via Fisher's exact test. The top five candidate cell state transitions and their corresponding scores are shown as a bar chart, and the genes from the gene list that are involved in each transition will also be shown.

Table 2. The functions enriched in genes that are up-regulated during cardiac muscle cell development. Only the second day (D2), third day (D3) and eighth day (D8) are listed

Genes	Enriched function and pathways	*P*-value	DEGs annotated
Up-regulated genes of D2	GO:0060038 cardiac muscle cell proliferation	0.002	FOXC1, TENM4
	GO:0009880 embryonic pattern specification	0.004	LHX1, RIPPLY2
	GO:0060912 cardiac cell fate specification	0.004	TENM4
	GO:2000691 negative regulation of cardiac muscle cell myoblast differentiation	0.004	PRICKLE1
	GO:0003241 growth involved in heart morphogenesis	0.009	MESP1
Up-regulated genes of D3	GO:0055010 ventricular cardiac muscle tissue morphogenesis	0.0002	HAND1, PKP2, TPM1
	GO:0055014 atrial cardiac muscle cell development	0.001	FHL2
	GO:0048739 cardiac muscle fiber development	0.003	MYH11
	GO:0007507 heart development	0.008	GATA5, HAND1, ITGA3, PKP2
	hsa04550 signalling pathways regulating pluripotency of stem cells	0.040	FZD4, ID2, MEIS1
Up-regulated genes of D8	GO:0006942 regulation of striated muscle contraction	0.00005	MYBPC3, MYL3
	GO:0002026 regulation of the force of heart contraction	0.0003	ADM, MYL3
	GO:0006936 muscle contraction	0.0003	CKMT2, CRYAB, MYOM1
	GO:0055010 ventricular cardiac muscle tissue morphogenesis	0.0005	MYBPC3, MYL3
	GO:2000291 regulation of myoblast proliferation	0.003	KLHL41

Especially, the functions and pathways enriched in the common genes between query gene list and the signatures for the top five candidate cell state transitions will be shown. If a public dataset was chosen for a certain transition, the expression patterns of the common genes across different time points will also be shown so that the users can check whether their custom data is similar to the public dataset.

Except for querying, the users can upload their custom time-series gene expression data, which can be analysed and visualized in CSTEA, including detection of DEGs, visualization of transition trajectory path, signature based comparison with deposited datasets, etc. Specifically, a gene signature will be generated for the uploaded custom dataset, and this gene signature will be compared with those defined for different cell state transitions. By comparing the gene signatures, the users can easily check what potential cell state transitions have been involved in their custom data, which could give the users clues for further exploration.

Case study 1—transition path from human embryonic stem cells to cardiac muscle cells

In this section, the state transition from human embryonic stem cells (ESCs) to cardiomyocytes (CMs) is taken as an example. The BMP and WNT pathways play important roles in the directed cardiac differentiation process from ESCs (22). When the transition of human ESCs→CMs is queried against the CSTEA, the GSE67152 dataset is shown to describe this transition, which involves two transition processes induced at different time points (23). For the process induced by IWP-2 on the second day (D2) or third day (D3) during differentiation, the transition trajectory path is shown in Figure 3A, where the trajectory path is visualized with the first two principle components of the gene expression profiles. Based on the trajectory, it appears that the fourth day (D4) is important for the state transition, since gene activities change drastically at that time. This phenomenon is consistent with a report in the literature indicating that putative cardiac precursor cells and the induction of an early cardiomyocyte-like fate were identified on D4 and D5, respectively, when induction was performed on the second or third day (23).

The CSTEA also provides the DEGs across the transition process and a bar chart of the number of DEGs at different time points, as shown in Figure 3B. By examining the functions that are enriched among these DEGs, we can better understand the transition process. For example, the up-regulated DEGs identified on D2 and D3 are observed to be enriched in cardiac cell fate specification, ventricular cardiac muscle tissue morphogenesis and heart development, while the DEGs up-regulated on D8 are enriched in muscle contraction and ventricular cardiac muscle tissue morphogenesis (Table 2). The decreasing number of DEGs on D6 indicates that the transition processes may be completed on the sixth day. By further examining the DEGs enriched with cardiac-associated terms, we found that some genes indeed play important roles in differentiation. For example, FZD4 has been reported to be involved in the induction of cardiac differentiation (24), and ID2 and MEIS1 have been identified as cardiomyocyte-specific transcriptomic gene signatures (25).

Case study 2—gene signatures for the transition from mouse MEFs to iPSCs

In this section, we show how to identify potential cell state transitions when given a gene list of interest. As described above, for each state transition, the CSTEA defines a gene signature. By comparing the queried gene list to these defined gene signatures, the cell state transitions that involve

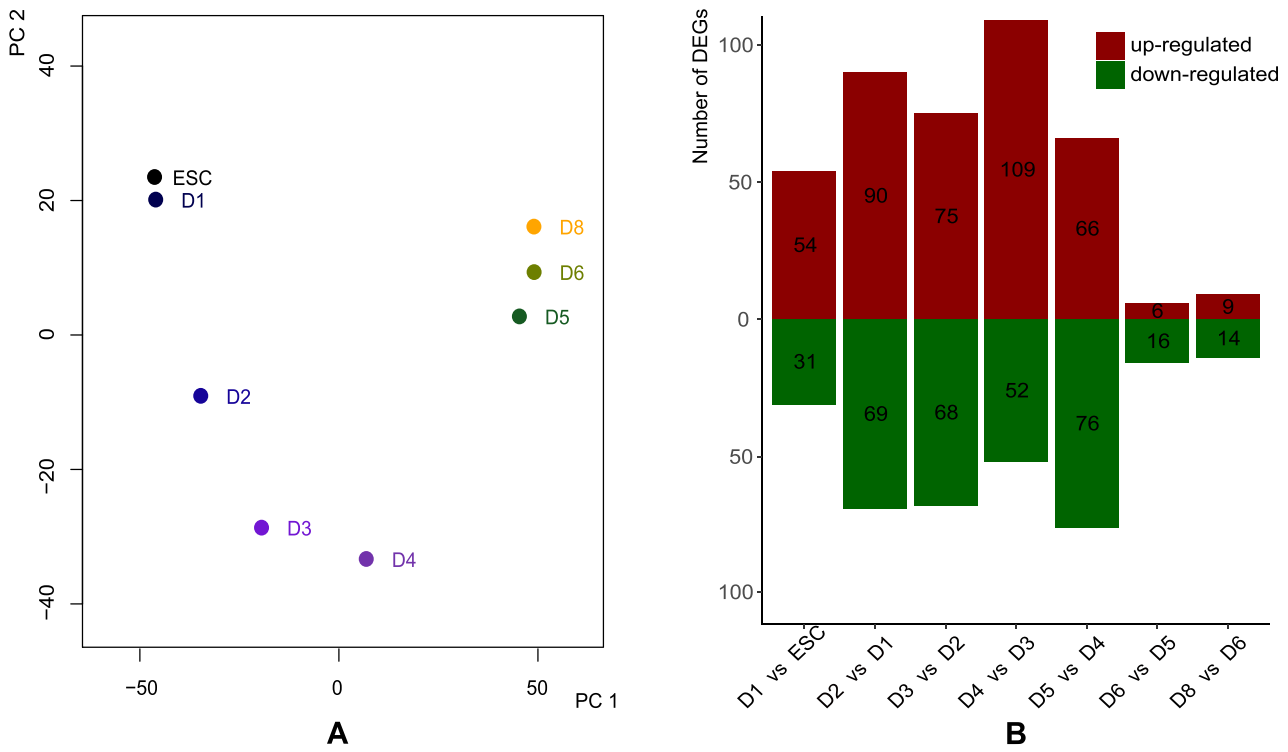

Figure 3. Visualization of trajectory path and detailed incidents of the GSE67152 dataset on cell differentiation induced by IWP-2 on the second or third day during the transition. (**A**) Trajectory of the gene expression profiles are visualized as the first two principal components (PCs) in principal component analysis. (**B**) Up- and down-regulated DEGs are identified for different time points during differentiation.

the list of genes can be identified. For example, Polo *et al.* identified 323 genes that are transiently either up- or down-regulated (clusters IV and VIII) during the reprogramming of mouse embryonic fibroblasts (MEFs) to iPSCs (26). Taking this list of 323 genes as an example, we sought to determine whether the CSTEA could recover this process by querying the gene list against the CSTEA. We found that the 323 genes were enriched in the gene signatures defined based on datasets GSE50206 (27) and GSE46532 (28), which were generated during the reprogramming of MEFs to iPSCs, indicating that the gene signatures defined by the CSTEA for cell state transitions are indeed useful.

By further examining the genes that overlap between the gene signatures and the queried gene list, additional details about the cell state transition can be uncovered. Out of the gene signature associated with the transition of MEFs to iPSCs, the expression of nine genes (Scg5, Insm1, Nnat, Elavl4, Mapk10, Spsb4, Bcl11a, 6330403K07Rik and Col2a1) reaches a peak on D26 based on the GSE46532 dataset, and these genes are enriched in the functions of cell fate commitment and pattern specification process. As reported in the literature, pluripotency-related genes are highly expressed on D26 (28), indicating the important roles of these genes. In particular, among these genes, Col2a1, Bcl11a and Insm are related to cell differentiation and organ development. These findings indicate that the 26th day may be the critical time point at which cells are converted to the desired pluripotent state.

DISCUSSION

We developed the CSTEA to provide both visualization and analysis of valuable time-series gene expression data during cell state transitions. The CSTEA focuses on expounding important genes and their corresponding functions at intermediate time points during the dynamic transition process, which is distinct from websites that provide analyses of gene expression data on static cell states. We utilized text mining to collect public datasets to offer a comprehensive roadmap describing the diverse cell state transitions. The datasets were then manually curated to not only filter out the false positive datasets produced during text mining but also to provide concise experimental descriptions and annotations of time points for every sample in the datasets, facilitating the efficient utilization of data. A particular transition from one cell type to another can be achieved through different inductions, and the detailed trajectory paths can differ greatly. Collecting datasets that are as complete as possible is crucial for describing the uncharacterized intermediate stages of each trajectory path, which may benefit the quantitative study of Waddington's epigenetic landscape (29,30).

Beyond the comprehensive analysis of time-series gene expression data during cell state transitions, the CSTEA also defines the gene signatures for any given cell state transition process. For every signature defined in this study, genes with clear expression changes during a cell state transition process are included, which could be very different from the list of DEGs between the original and destination

cell types. In other words, for any defined signature, at least some genes reflect the features of uncharacterized intermediate states, which are usually unstable and reversible but informative for revealing the mechanisms of cell fate determination. With the signatures defined here, users can carry out enrichment analysis of any user-defined gene list (for example, a list of DEGs identified during a specific type of tumourigenesis) and obtain the cell state transition processes in which those genes may play a role. Such analysis could contribute to the linkage of two different biological processes or even to the discovery of novel mechanisms of disease progression. Besides, as comprehensive expression datasets during cell state transitions has been collected in CSTEA, the signatures from these datasets provide valuable references for users to compare with their own dataset. Different to the analysis solely based on their own datasets, such comparisons performed in CSTEA enable the users to identify transition processes with similar gene expression dynamic patterns, which may give useful clues for further exploration. In general, with the functions of the CSTEA designed specifically for time-series data during cell state transitions, we believe that this webserver will be a valuable resource for both experimental and computational biologists who are interested in revealing the mechanisms underlying cell state transition.

FUNDING

National Natural Science Foundation of China [61572363, 91530321, 61602347, 31371288, 31571365, 31322031]; National Key Research and Development Program of China [2016YFA0100400]; Specialized Research Fund for the Doctoral Program of Higher Education [20130072110032]; Program of Shanghai Academic Research Leader [17XD1403600]; Natural Science Foundation of Shanghai [17ZR1445600]. Funding for open access charge: National Natural Science Foundation of China [91530321].
Conflict of interest statement. None declared.

REFERENCES

1. Hochedlinger,K. and Plath,K. (2009) Epigenetic reprogramming and induced pluripotency. *Development*, **136**, 509–523.
2. Gurdon,J.B. (1963) The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. *J. Embryol. Exp. Morphol.*, **10**, 622–640.
3. Takahashi,K. and Yamanaka,S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, **126**, 663–676.
4. Pujadas,E. and Feinberg,A.P. (2012) Regulated noise in the epigenetic landscape of development and disease. *Cell*, **148**, 1123–1131.
5. Stadtfeld,M. and Hochedlinger,K. (2010) Induced pluripotency: history, mechanisms, and applications. *Genes Dev.*, **24**, 2239–2263.
6. Wu,S.M. and Hochedlinger,K. (2011) Harnessing the potential of induced pluripotent stem cells for regenerative medicine. *Nat. Cell Biol.*, **13**, 497–505.
7. Xie,H., Ye,M., Feng,R. and Graf,T. (2004) Stepwise reprogramming of B cells into macrophages. *Cell*, **117**, 663–676.
8. Tabar,V. and Studer,L. (2014) Pluripotent stem cells in regenerative medicine: challenges and recent progress. *Nat. Rev. Genet.*, **15**, 82.
9. Moris,N., Pina,C. and Arias,A.M. (2016) Transition states and cell fate decisions in epigenetic landscapes. *Nat. Rev. Genet.*, **17**, 693.
10. Edgar,R. and Lash,A. (2002) The Gene Expression Omnibus (GEO): a gene expression and hybridization repository. *National Center for Biotechnology Information*.
11. Brazma,A., Parkinson,H., Sarkans,U., Shojatalab,M., Vilo,J., Abeygunawardena,N., Holloway,E., Kapushesky,M., Kemmeren,P. and Lara,G.G. (2003) ArrayExpress—a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res.*, **31**, D553–D555.
12. Seita,J., Sahoo,D., Rossi,D.J., Bhattacharya,D., Serwold,T., Inlay,M.A., Ehrlich,L.I., Fathman,J.W., Dill,D.L. and Weissman,I.L. (2012) Gene Expression Commons: an open platform for absolute gene expression profiling. *PLoS One*, **7**, 398–398.
13. Edgar,R., Mazor,Y., Rinon,A., Blumenthal,J., Golan,Y., Buzhor,E., Livnat,I., Ben-Ari,S., Lieder,I. and Shitrit,A. (2013) LifeMap Discovery™: the embryonic development, stem cells, and regenerative medicine research portal. *PLoS One*, **8**, e66629.
14. Kassambara,A., Jourdan,M., Fest,T., Hose,D., Tarte,K. and Klein,B. (2015) GenomicScape: an easy-to-use web tool for gene expression data analysis. Application to investigate the molecular events in the differentiation of B cells into plasma cells. *PLoS Comput. Biol.*, **11**, e1004077.
15. Gautier,L., Cope,L., Bolstad,B.M. and Irizarry,R.A. (2004) affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*, **20**, 307–315.
16. Smyth,G. (2005) Limma: linear models for microarray data. *Bioinform. Comput. Biol. Sol. R Bioconductor*, 397–420.
17. Du,P., Kibbe,W.A. and Lin,S.M. (2008) lumi: a pipeline for processing Illumina microarray. *Bioinformatics*, **24**, 1547–1548.
18. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
19. Feng,J., Meyer,C.A., Wang,Q., Liu,J.S., Shirley,L.X. and Zhang,Y. (2012) GFOLD: a generalized fold change for ranking differentially expressed genes from RNA-seq data. *Bioinformatics*, **28**, 2782.
20. Camon,E., Barrell,D., Lee,V., Dimmer,E. and Apweiler,R. (2003) The Gene Ontology Annotation (GOA) Database-an integrated resource of GO annotations to the UniProt Knowledgebase. *In Silico Biol.*, **4**, 5–6.
21. Kanehisa,M. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.*, **27**, 29–34.
22. Burridge,P.W., Keller,G., Gold,J.D. and Wu,J.C. (2012) Production of de novo cardiomyocytes: human pluripotent stem cell differentiation and direct reprogramming. *Cell Stem Cell*, **10**, 16.
23. Rao,J., Pfeiffer,M.J., Frank,S., Adachi,K., Piccini,I., Quaranta,R., Arauzo-Bravo,M., Schwarz,J., Schade,D., Leidel,S. *et al.* (2016) Stepwise clearance of repressive roadblocks drives cardiac induction in human ESCs. *Cell Stem Cell*, **18**, 341–353.
24. Abdulghani,M., Dufort,D., Stiles,R., Repentigny,Y.D., Kothary,R. and Megeney,L.A. (2010) Wnt11 promotes cardiomyocyte development by caspase-mediated suppression of canonical Wnt signals. *Mol. Cell. Biol.*, **31**, 163–178.
25. Doss,M.X., Gaspar,J.A., Winkler,J., Hescheler,J., Schulz,H. and Sachinidis,A. (2012) Specific gene signatures and pathways in mesodermal cells and their derivatives derived from embryonic stem cells. *Stem Cell Rev. Rep.*, **8**, 43–54.
26. Polo,J., Anderssen,E., Walsh,R., Schwarz,B., Nefzger,C., Lim,S.M., Borkent,M., Apostolou,E., Alaei,S. and Cloutier,J. (2012) A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell*, **151**, 1617–1632.
27. Takahashi,K., Tanabe,K., Ohnuki,M., Narita,M., Sasaki,A., Yamamoto,M., Nakamura,M., Sutou,K., Osafune,K. and Yamanaka,S. (2014) Induction of pluripotency in human somatic cells via a transient state resembling primitive streak-like mesendoderm. *Nature*, **5**, 3678.
28. Ho,R., Papp,B., Hoffman,J.A., Merrill,B.J. and Plath,K. (2013) Stage-specific regulation of reprogramming to induced pluripotent stem cells by Wnt signaling and T cell factor proteins. *Cell Rep.*, **3**, 2113–2126.
29. Waddington,C.H. (1942) Canalization of development and the inheritance of acquired characters. *Nature*, **150**, 563–565.
30. Waddington,C.H. (1957) The strategy of the genes. A discussion of some aspects of theoretical biology. *CAB Direct*. George Allen & Unwin Ltd., London, 262.