D1168–D1178 *Nucleic Acids Research, 2023, Vol. 51, Database issue* *Published online 9 November 2022*
https://doi.org/10.1093/nar/gkac978

TEDD: a database of temporal gene expression patterns during multiple developmental periods in human and model organisms

Ziheng Zhou[1,2,3,†], Cong Tan[4,5,†], Matthew Hoi Kin Chau[1,2,3,6], Xiaosen Jiang[5,7], Ziyuan Ke[5], Xiaoyan Chen[1,8], Ye Cao[1,2,3,9], Yvonne K. Kwok[1], Matthew Bellgard[10], Tak Yeung Leung[1,2,3,6,9], Kwong Wai Choy[1,2,3,6,9,*] and Zirui Dong[1,2,3,9,*]

[1]Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Hong Kong, China, [2]Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen 518057, China, [3]Hong Kong Hub of Paediatric Excellence, The Chinese University of Hong Kong, Hong Kong, China, [4]State Key Laboratory of Agricultural Genomics, BGI-Shenzhen, Shenzhen, 518083, China, [5]BGI-Shenzhen, Shenzhen 518083, China, [6]The Chinese University of Hong Kong-Baylor College of Medicine Joint Center for Medical Genetics, Hong Kong, China, [7]College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China, [8]Maternal-Fetal Medicine Institute, Shenzhen Baoan Women's and Children's Hospital, Shenzhen University, Shenzhen, China, [9]The Fertility Preservation Research Center, Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Hong Kong, China and [10]eResearch Office, Queensland University of Technology, Brisbane, QLD, Australia

Received August 09, 2022; Revised October 06, 2022; Editorial Decision October 14, 2022; Accepted October 15, 2022

ABSTRACT

Characterization of the specific expression and chromatin profiles of genes enables understanding how they contribute to tissue/organ development and the mechanisms leading to diseases. Whilst the number of single-cell sequencing studies is increasing dramatically; however, data mining and reanalysis remains challenging. Herein, we systematically curated the up-to-date and most comprehensive datasets of sequencing data originating from 2760 bulk samples and over 5.1 million single-cells from multiple developmental periods from humans and multiple model organisms. With unified and systematic analysis, we profiled the gene expression and chromatin accessibility among 481 cell-types, 79 tissue-types and 92 timepoints, and pinpointed cells with the co-expression of target genes. We also enabled the detection of gene(s) with a temporal and cell-type specific expression profile that is similar to or distinct from that of a target gene. Additionally, we illustrated the potential upstream and downstream gene–gene regulation interactions, particularly under the same biological process(es) or KEGG pathway(s). Thus, TEDD (Temporal Expression during Development Database), a value-added database with a user-friendly interface, not only enables researchers to identify cell-type/tissue-type specific and temporal gene expression and chromatin profiles but also facilitates the association of genes with undefined biological functions in development and diseases. The database URL is https://TEDD.obg.cuhk.edu.hk/.

INTRODUCTION

Cell fate decisions play a pivotal role in the development of multicellular organisms from a zygote to functionally differentiated cell types, tissues, and organs. Mammalian development consists of multiple stages, including embryonic, fetal, neonatal, childhood and adult stages. In humans, each of these stages involves complex cellular processes, including proliferation, differentiation and reprogramming. These complex programs are regulated by precise gene expression patterns and further translational regulation in cells (1). The dynamic expression of genes and regulatory networks in each cell define the cell fate and control its cellular processes (2). Developmental processes are precisely regulated, which is reflected in the dynamic and distinct gene expression patterns in each cell. The emergence of transformative technologies such as single-cell omics sequencing unlocked the capability to understand the underlying mechanisms of cell fate decisions and cellular processes and how they are controlled and determined during organ development. Thus,

[*]To whom correspondence should be addressed. Tel: +852 3505 3099; Fax: +852 2636 0008; Email: elvisdong@cuhk.edu.hk
Correspondence may also be addressed to Kwong Wai Choy. Tel: +852 3505 3099; Fax: +852 2636 0008; Email: richardchoy@cuhk.edu.hk
[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

we can gain insight into diseases in both human and model organisms (3–5).

Dramatic advances in single-cell RNA sequencing (scRNA-seq) technologies have occurred in the past decade. This technology provides an unprecedented opportunity to obtain the precise gene expression profiles of tens of thousands of cells at a single-cell resolution. The gene expression profiles of individual cells provide meaningful insights that could elucidate the mechanisms involved in organ development and disease pathogenesis (6). Owing to the practical challenges of sample collection, human development, particularly from fertilized ovum to term birth, has remained a poorly understood 'black box' (8–10). Model organisms (such as mouse and zebrafish) are widely utilized to study developmental biology, with the aim of recapitulating and understanding the mechanisms of human embryonic and organ development. A comprehensive atlas of the developing mouse neocortex has been generated by sampling the neocortex daily throughout embryonic corticogenesis and at early postnatal ages (7). Although the general driving mechanisms of implantation, gastrulation, organogenesis, and basic processes of forming diverse tissue types appear to be largely conserved among species (such as human and mouse), studies have suggested differences in cellular compositions and the time course of human tissue formation compared with other organisms. For example, superficial cortical layers are expanded in humans but not in mice (11), and clear timing differences in intestinal maturation stages are observed between mice and humans (12). In addition, comparative studies of humans and other mammals have revealed both conserved and divergent transcriptional programs, such as preimplantation and gastrulation processes (8,9), brain processes (10–12), immune system processes (13,14), and musculoskeletal processes (15).

Various types of single-cell omics databases have been developed to share, retrieve, analyze, and visualize raw data (including expression matrices) and value-added information (Table 1). First, the Gene Expression Omnibus (NCBI/GEO) (16), Single Cell Expression Atlas (EMBL-EBI/SCEA) (17), Mouse Cell Atlas (MCA) (18), and Single Cell Portal and Cell-omics Data Coordinate Platform (CDCP) (19) are devoted to archiving and sharing raw data or expression matrics produced in scRNA-seq and/or bulk RNA-seq studies. Second, several databases were developed for a certain project or species, such as the Human Cell Atlas Portal (HCA) (20) and Mouse Cell Atlas (MCA) (18). Third, a few value-added databases were developed for hosting curated gene expression profiles and integrating certain disease information in single-cell expression maps across various human cancers (CancerSCEM) (21), single-cell transcriptomes for human diseases (SC2disease) (22), single-cell atlases for exposing molecular characteristics of COVID-19 (SCovid) (23) and Deeply Integrated human Single-Cell Omics data (DISCO) (24). While these resources are essential, the establishment of a reference database that integrates the most up-to-date and comprehensive datasets involving the timepoints throughout the lifespan of human and model animals, such as mouse, zebrafish, and nematode (both fetal and adult), is urgently needed. With such an up-to-date reference repository, researchers can (i) characterize the emergence and dynamics of cell-type and tissue-type specific expression and chromatin accessibility patterns during development (such as tissue formation) in bulk-sample and single-cell levels, (ii) profile the convergence and divergence of gene−gene interactions (such as co-expression) in cell-type, timepoint and sex categories and (iii) identify gene(s) that have similar or distinct temporal and cell-type specific expression patterns from that of a target gene, particularly for those from the same Gene Ontology (GO) or KEGG pathway.

Herein, we present TEDD, the Temporal Expression during Development Database, which was constructed by collecting data from bulk-sample RNA-seq (2760 samples), scRNA-seq (3 814 231 cells from over 1000 samples) and scATAC-seq (1 329 392 cells from 174 samples) data from publicly available datasets in human and mulitple model organisms (mouse, zebrafish, and nematode). With an user-friendly interface of TEDD, users could easily explore the dynamics of gene expression and chromatin accessibility profiles across different developmental stages in each cell subtype to pinpoint the specific tissue-type(s), cell-type(s) and timepoint(s) expression patterns for a gene to understand its contribution to tissue/organ development and the mechanism(s) leading to diseases. In addition, the associations among certain genes (such as co-expression) at the cell-type, timepoint and sex categories by the expression correlation analysis provide a foundation to illustrate the potential upstream and downstream gene−gene regulation to uncover the biological functions related to development and diseases (28,29).

DATA COLLECTION AND DATABASE CONTENT

To provide scientists with the most comprehensive expression profiles and signatures of key genes contributing to the regulation of embryogenesis, organogenesis and development in human and model species, three kinds of publicly accessible datasets produced in studies related to the transcriptomic regulation of development in humans and the three most important model organisms (mouse, zebrafish, and nematode) were collected and curated in this study (Figure 1A). The curated and integrated datasets include the scRNA-seq data of 3 814 231 cells from over 1000 samples, the scATAC-seq data of 1 329 392 cells from 174 samples and the bulk-sample RNA-seq data of 2760 samples.

ScRNA-seq datasets were retrieved from a literature search in PubMed with the keywords 'single-cell RNA seq' or 'single-cell ATAC seq', 'homo sapiens', 'mus musculus', 'danio rerio', 'nematode' or 'cross-tissue' and 'temporal'. Overall, 27 articles with scRNA-seq datasets and two with scATAC-seq datasets with data and information from Homo sapiens, Mus musculus, Danio rerio and Nematode from at least one timepoint were included in this study (Figure 1A, Supplementary Table S1). The datasets were downloaded from multiple scRNA-seq data repositories, including Gene Expression Omnibus (NCBI/GEO) (16), Human Cell Atlas Portal (HCA) (20), Single Cell Expression Atlas (EMBL-EBI/SCEA) (17) and Mouse Cell Atlas (MCA) (18).

All scRNA-seq datasets were processed with a unified analytical pipeline (Figure 1B) with functions in Seurat v4.0.6 (25). First, quality control for each dataset was performed

Table 1. Comparison of database features

Name of database	Species curated	Features of data resource	Features of database analytical and visualization functions
Temporal Expression during Development Database (TEDD)	human and model animal species, include mouse, zebrafish, and nematode	curation, integration and visualization of RNA-seq, scRNA-seq/scATAC-seq datasets from human and model animal species studies; focuses on providing temporal gene expression and chromatin accessibility profiles during embryonic, fetal, neonatal, childhood, and adult development.	1. temporal gene expressions in bulk-sample and single-cell levels, PCA analysis and UMAP visualization; 2. temporal chromatin accessibilities in single-cell level, PCA analysis, and UMAP visualization; 3. clustering of multiple genes (from the same GO or KEGG pathway) based on the expression levels; 4. co-expression analysis; 5. temporally regulated genes 6. stably expressed genes
Gene Expression Omnibus (NCBI/GEO)	all species include plants and animals	archiving of raw data or expression matrics for all transcriptome sequencing resource.	Not available
Single Cell Expression Atlas (EMBL-EBI/SCEA)	animals, protists, plants and Fungi	curation and visualization of scRNA-seq/scATAC-seq studies in model and non-model species.	1. gene expression in single-cell levels, and PCA analysis; 2. chromatin accessibilities in single-cell level, and PCA analysis; 3. clustering of marker genes
Mouse Cell Atlas (MCA)	Mouse	single-cell RNA sequencing of major mouse organs from early embryonic stage to the mature adult stage.	1. gene expression in single-cell levels, and PCA analysis; 2. clustering of marker genes; 3. identifying cell types in users' data
Cell-omics Data Coordinate Platform (CDCP)	23 organisms	curation and visualization of scRNA-seq/scATAC-seq studies in model animal species.	1. gene expression in single-cell levels, and PCA analysis; 2. clustering of marker genes; 3. online single-cell data analysis
Single Cell Portal	10 organisms	curation and visualization of scRNA-seq/scATAC-seq studies in model and non-model species.	1. collection of single-cell genomics data; 2. gene expression in single-cell levels, and PCA analysis
Human Gene Expression During Development (Descartes)	Human, mouse, worm and fly	scRNA-seq/scATAC-seq for human and model species development; for human, focuses on 15 organs in fetal samples	1. collection of scRNA and scATAC data; 2. gene expression in single-cell levels, and PCA analysis; 3. exploration on cell trajectory for organ development
Human Cell Atlas Portal (HCA)	Human	scRNA-seq/scATAC-seq/snRNA-seq for mapping the human body at the cellular level; focuses on data collection of single-cells derived from 9 human tissue-types and immune system	1. collection of scRNA and scATAC data from human tissues or diseases; 2. gene expression in single-cell levels, and PCA analysis;
Cancer Single-cell Expression Map (CancerSCEM)	Human	focuses on cancer samples	1. gene expression in single-cell levels, and PCA analysis; 2. gene correlation; 3. cell component comparison; 4. cell interaction network; 5. survival analysis

Table 1. Continued

Name of database	Species curated	Features of data resource	Features of database analytical and visualization functions
Single-Cell Transcriptome for human diseases (SC2disease)	Human	focuses on human multiple diseases and cancer studies	1. providing browse for the expression of interested genes; 2. searching for cell-type markers; 3. searching for the biomarkers of multiple diseases; 4. comparing the expression profiles of various types of cells in disease and non-disease states
Deeply Integrated human Single-Cell Omics data (DISCO)	Human	focuses on 14 tissue types in human and PBMC samples in COVID-19 atlas	1. data exploration at sample and integrated level; 2. CELLiD cell type identification; 3. online integration; 4. mapping user single-cell data to a specific atlas with Cell Mapper.
Single-cell atlases for exposing molecular characteristics of COVID-19 (SCovid)	COVID-19	focuses on COVID-19 across 10 human tissues	1. providing a browser of the molecular characteristics of COVID-19 on independent datasets of different tissues; 2. obtaining the molecular roles in different tissues by searching genes; 3. obtaining significantly differentially expressed genes in a specific cell type; 4. providing access of differentially expressed genes' expression profiles based on single-cell datasets of COVID-19 and controls
The Genotype-Tissue Expression (GTEx)	Human	genome sequencing and exome sequencing, RNA-Seq and snRNA-seq; focuses on differential expressions among different tissue-types for human adult cases	1. gene expression in bulk-sample and single-cell levels, and PCA analysis; 2. providing enquiry and visualization of the QTLs by gene/variant in the Locus Browser 3. providing a browser of H3K27ac ChIP-seq, m6A methylation, and WGBS DNA methylation data in IGV Browser 4. providing search and request for the GTEx archived biospecimens

by filtering out (i) cells with < 200 expressed genes and (ii) genes expressed in <10 cells. Then, 'NormalizedData' in Seurat v4.0 was applied to normalize the sparse single-cell gene expression matrix. We classified the datasets into different categories based on the species, sequencing types (scRNA-seq or scATAC-seq) and tissue types. Of note, integration was not performed for datasets generated from different species. For those categories with only one single dataset, no integration was performed, and the analyzed results were used directly for data query and visualization. For categories with more than one dataset, integration was performed by applying 'SCTransform' (26) for normalization and 'Harmony' to reduce batch correction (27) of gene expression data across different datasets. Notably, the data matrices consisting of gene-level chromatin accessibility scores from 69 scATAC-seq datasets (>1.3 million single-cells from 68 distinct cell types in humans) from 20 tissue types in adults and 14 tissue types in fetuses (28) were collected. Genes with highly variable expression patterns were identified by 'FindVariableGenes', and the top 2000 highly variably expressed genes were used for the dimensionality reduction using principal component analysis (PCA). The top 30 PCs were selected and used for clustering with FindNeighbors (SNN). The robustness of our analysis was further verified by comparing the results with those generated by the original studies (see Supplementary Materials, Supplementary Figures S1–S3).

Publicly available bulk sample-based RNA-seq data produced in studies related to the transcriptomic regulation of development in humans and three model species and the corresponding sample information (such as sample type and collected timepoint) were collected from ENCODE for further analysis (29). The retrieval of RNA-seq datasets was performed with the following keywords: (i) 'total RNA-seq' and 'ployA plus RNA-seq', in combination with (ii) 'Homo sapiens' and 'Mus musculus'. In addition, all datasets were generated from tissues, primary cells, cell lines or *in vitro* differentiated cells of normal tissues (Figure 1A). As a result,

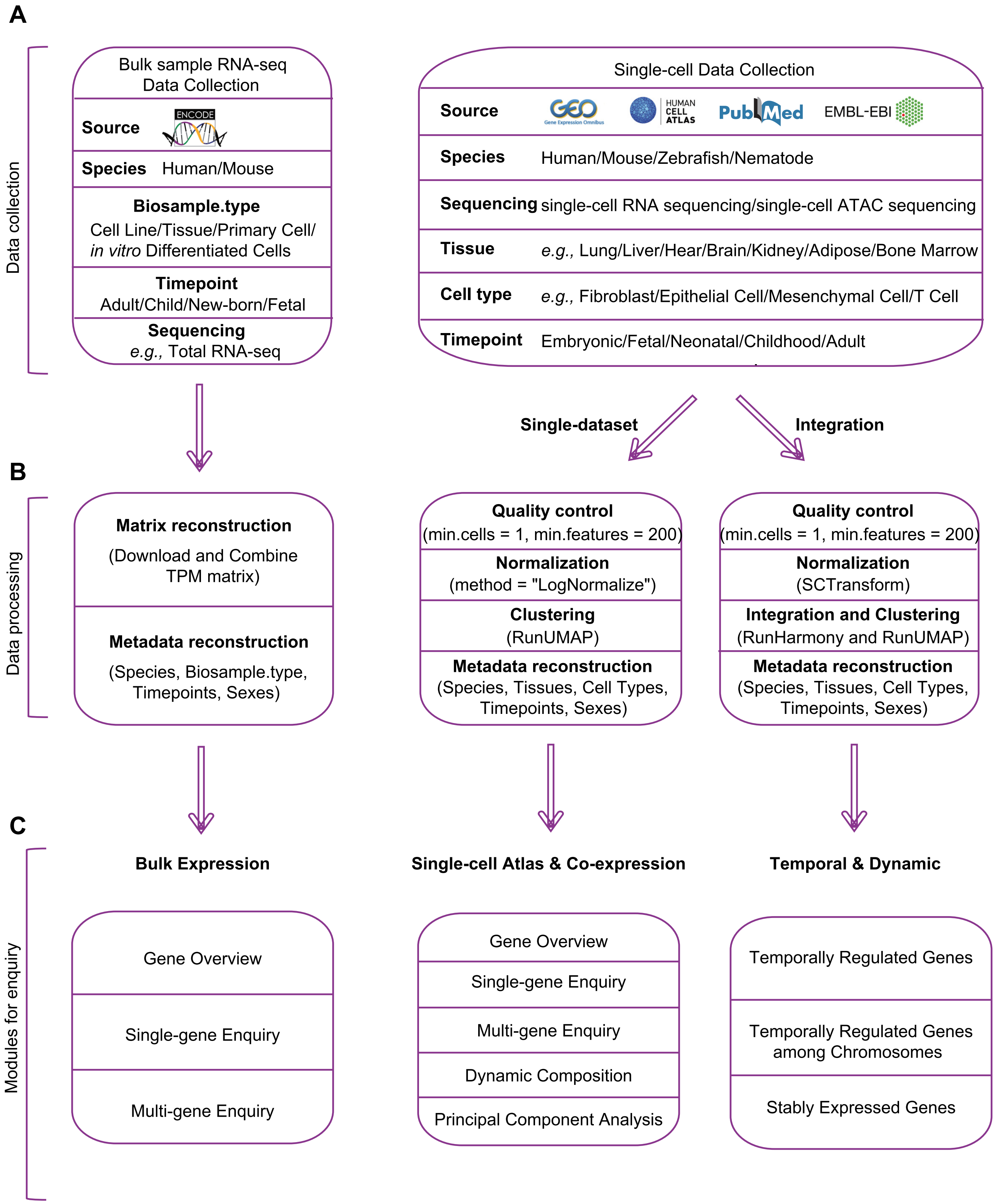
A
Data collection
Bulk sample RNA-seq Data Collection
Source
ENCODE
Species Human/Mouse
Biosample.type
Cell Line/Tissue/Primary Cell/ in vitro Differentiated Cells
Timepoint
Adult/Child/New-born/Fetal
Sequencing
e.g., Total RNA-seq
Single-cell Data Collection
Source
GEO Gene Expression Omnibus
HUMAN CELL ATLAS
PubMed
EMBL-EBI
Species Human/Mouse/Zebrafish/Nematode
Sequencing single-cell RNA sequencing/single-cell ATAC sequencing
Tissue e.g., Lung/Liver/Hear/Brain/Kidney/Adipose/Bone Marrow
Cell type e.g., Fibroblast/Epithelial Cell/Mesenchymal Cell/T Cell
Timepoint Embryonic/Fetal/Neonatal/Childhood/Adult
Single-dataset
Integration
B
Data processing
Matrix reconstruction
(Download and Combine TPM matrix)
Metadata reconstruction
(Species, Biosample.type, Timepoints, Sexes)
Quality control
(min.cells = 1, min.features = 200)
Normalization
(method = "LogNormalize")
Clustering
(RunUMAP)
Metadata reconstruction
(Species, Tissues, Cell Types, Timepoints, Sexes)
Quality control
(min.cells = 1, min.features = 200)
Normalization
(SCTransform)
Integration and Clustering
(RunHarmony and RunUMAP)
Metadata reconstruction
(Species, Tissues, Cell Types, Timepoints, Sexes)
C
Modules for enquiry
Bulk Expression
Gene Overview
Single-gene Enquiry
Multi-gene Enquiry
Single-cell Atlas & Co-expression
Gene Overview
Single-gene Enquiry
Multi-gene Enquiry
Dynamic Composition
Principal Component Analysis
Temporal & Dynamic
Temporally Regulated Genes
Temporally Regulated Genes among Chromosomes
Stably Expressed Genes

Figure 1. Overview of data collection, data processing and functional modules under each domain of TEDD database. (**A**) Diagram of bulk-sample RNA-seq and single-cell RNA-seq and ATAC-seq data collection. (**B**) Analytical steps of data processing. (**C**) Overview of functional modules under the domains of 'Bulk Expression', 'Single-cell Atlas', 'Co-expression' and 'Temporal & Dynamic'.

2760 RNA-seq datasets were included, as shown in Supplementary Table S2. The expression of each gene was calculated in TPM format (transcripts per million) by a unified analytical pipeline (the mapping of the reads was performed using the STAR program (30), and the quantification of genes and transcripts was conducted with the RSEM program (31)). The expression matrices were then integrated and indexed based on the sample information. Overall, the bulk RNA expression data of 52 tissue types, 77 primary cells, 43 cell lines and 15 *in vitro* differentiated cells from four developmental periods were curated.

CONSTRUCTION AND DESCRIPTION OF FUNCTIONS

TEDD is a publicly and freely accessible web-based browser for researchers, and the web application was deployed on a Linux server. The VUE framework (https://vuejs.org/) was used for rendering and the interactive operations of front-end pages, and CodeIgniter was used as the basic architecture of the backend system. In addition, MySQL served as a container for data storage, and Apache2 acted as a background server. Echarts and d3.js were adopted for construction of the interactive graphs, while a bootstrap table was used to build the data tables. The displaying modules also enable intuitive visualization of the gene expression signatures at the bulk-tissue level or at a single-cell resolution from multiple developmental periods. It provides gene expression patterns from across gestational weeks (for humans), days (for mice), hours (for zebrafish), or minutes (for nematodes) to demonstrate the regulatory expression patterns during embryogenesis and organogenesis. Figure 1 shows the schematic workflow (Figure 1B) and main functional modules of this database (Figure 1C).

Overall, the main focuses of this database (Figure 2) include the following: (i) allowing researchers to understand the differences in cell-type, tissue-type and timepoint specific expression patterns and chromatin accessibility profiles of an enquired gene at the bulk-sample and single-cell levels; (ii) providing an integrated network to identify cells with the co-expression of target genes in the cell-type, timepoint, and sex categories; (iii) profiling and clustering the expression patterns of gene(s), particularly for those under the same GO or KEGG pathway; and (iv) identifying gene(s) with an expression profile that is similar to or distinct from that of an enquired gene across the selected cell types, tissue types and timepoints.

For the TEDD interface, the navigation menu contains seven drop-down menus, including 'Home', 'Datasets', 'Bulk Expression', 'Single-cell Atlas', 'Co-expression', 'Temporal & Dynamic' and 'Help', which could lead users to the corresponding interfaces. On the 'Home' page, there are three main elements in addition to the header and the main navigation menu, including (i) a direct search engine for single gene expression profiles in bulk-sample and single-cell levels, (ii) a comprehensive navigation panel showing each hyperlinked module, and (iii) three diagrams of cell-type/tissue/organ sampling sites in humans (three phases: embryonic, fetal and adult) curated by TEDD (Supplementary Figure S4). If a researcher would like to focus on a particular cell type/tissue type/organ at specific stages (e.g. embryonic, fetal, and adult), they could click on the cell type/tissue-type/organ of interest in the diagrams. This will direct the researcher to the page showing all datasets related to this cell type/tissue-type/organ. A detailed description of the functions and usage guidance under each module is provided in the Supplementary Materials (Supplementary Figures S5–S8).

The 'Single-cell Atlas' domain provides comprehensive and interactive functions to illustrate gene expression and chromatin accessibility profiles at the single-cell level. It consists of five modules (Supplementary Figure S6). The 'Single-gene Enquiry' module provides the subclassification of each tissue type into further subgroups based on cell types and timepoints as well as the percentage of cells with the targeted gene expression identified. In addition, under the 'Multi-gene Enquiry', researchers can fill in a gene to identify putative GO or KEGG pathways in which the gene participated, and they can further input a list of genes (up to 50) for clustering of expression data after selection of the tissue type, cell type and timepoint. Furthermore, the 'Dynamic Composition' module provides the cell-type compositions based on the sampling timepoints from each dataset. Most importantly, under the 'Principal Component Analysis' module, when a researcher fills in a gene name, TEDD provides UMAP results of the dataset with four options of cell labeling (cell type, tissue type, timepoint and sex), and the cells with expression of the targeted gene are labeled by the expression abundance. To conveniently compare the patterns of expression or chromatin accessibility of the same gene in different datasets (with different parameters of species, cell types/tissue types, timepoints and data types) or with another gene, a concurrent interface is set if users click the 'Add Compare' button.

Within the 'Co-expression' domain, after defining the species, tissue-type and the options of clustering (cell type, timepoint and sex), TEDD allows the submission of multiple genes (up to 5) to investigate those cells with co-expressions of the submitted genes from scRNA-seq and scATAC-seq datasets. The overall number of each cell-type and the percentage of cells with the identified co-expressions are provided for the researchers to investigate whether the targeted genes are expressed in the same cell-type at a particular timepoint and tissue-type.

The 'Temporal & Dynamic' domain includes three modules. 'Temporally Regulated Genes' provides the identification of genes with differential expression among the selected timepoints (with tissue-type and cell-type selected). It provides an option to set a gene at a specific tissue-type (and cell-type) and timepoint as an anchor to identify any other tissue-types/cell-types/timepoints that the targeted gene in the anchor shows significantly differential expression, and to further investigate gene(s) with an expression pattern that is similar to or distinct from that of the targeted gene (such as with significantly higher or lower expression) in the anchor compared with these tissue-types/cell-types/timepoints. The 'Temporally Regulated Genes among Chromosomes' module provides a genome-wide distribution of these temporally regulated genes with a customized cutoff (logFC) and a maximum number (MaxNumber) of genes being displayed. Under the 'Stably Expressed Genes' module, researchers can iden-

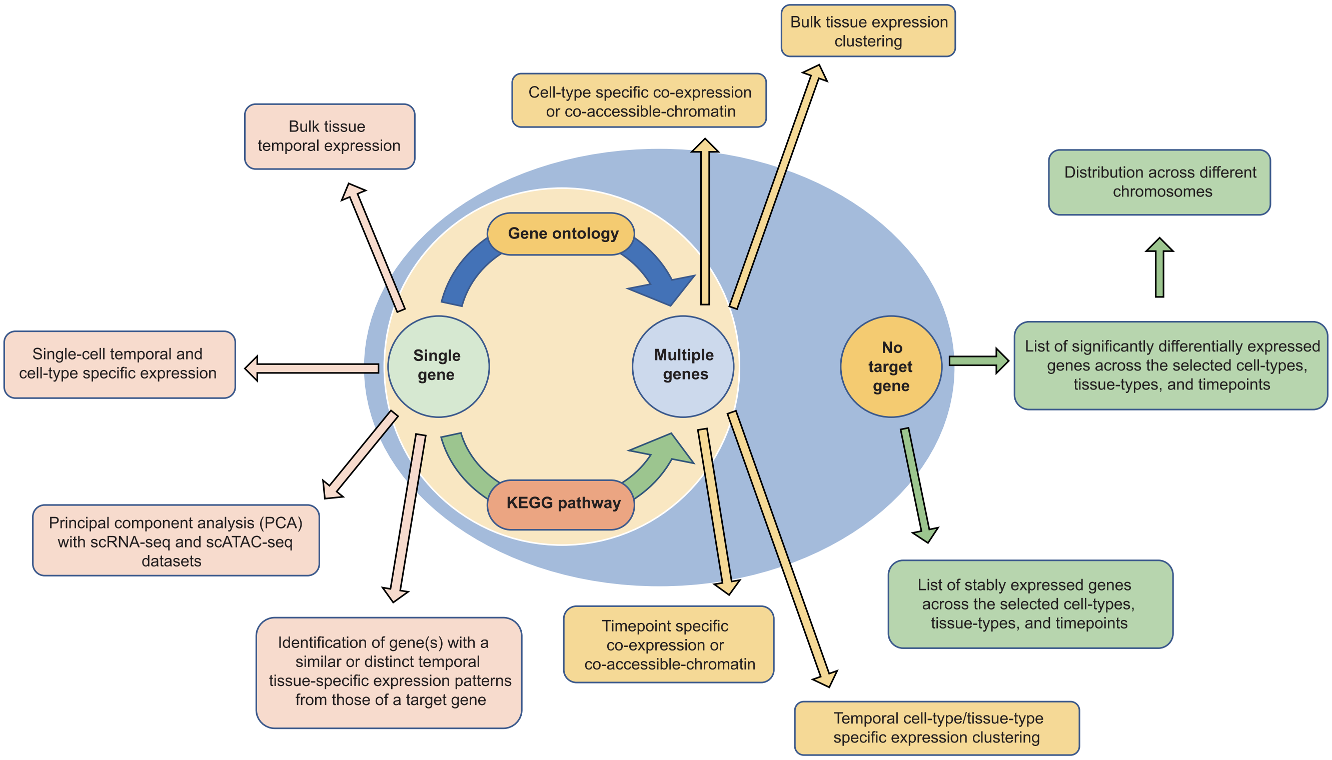
Bulk tissue expression clustering
Cell-type specific co-expression or co-accessible-chromatin
Bulk tissue temporal expression
Distribution across different chromosomes
Gene ontology
Single gene
Multiple genes
No target gene
Single-cell temporal and cell-type specific expression
List of significantly differentially expressed genes across the selected cell-types, tissue-types, and timepoints
KEGG pathway
Principal component analysis (PCA) with scRNA-seq and scATAC-seq datasets
List of stably expressed genes across the selected cell-types, tissue-types, and timepoints
Identification of gene(s) with a similar or distinct temporal tissue-specific expression patterns from those of a target gene
Timepoint specific co-expression or co-accessible-chromatin
Temporal cell-type/tissue-type specific expression clustering

Figure 2. Summary of analytical utility of TEDD. Enquiries are shown in circles and actions are shown in rectangles. TEDD provides analytical utility for users when there is target of single-gene, multiple genes or even no targeted gene.

tify gene(s) with expression patterns identified among the tissue-types or timepoints selected.

Last, on the ‘Help’ page, a graphical operation guide is provided for new users to start each function, and this will help them fully utilize the resources in TEDD.

APPLICATION CASE

To exemplify how TEDD can help researchers understand the temporal patterns of expression and chromatin accessibility of the gene(s) of interest, we used the *AFP* (Alpha Fetoprotein) gene as an example. Alpha fetoprotein is a well-established marker for the diagnosis and prognosis of hepatocellular carcinoma (3), and its role primarily is to transport heavy metal ions and various insoluble molecules in fetal blood circulation (32). As a paralog of *AFP*, the *ALB* gene is found to have a similar expression pattern in embryonic stem cell-derived hepatocytes (6).The *AFP* gene is involved in the Hippo signaling pathway (hsa04390), which regulates organ size, cell fate, and carcinogenesis in the liver (5) through the activation of YAP (gene *YAP1*) and TAZ (gene *WWTR1*) (33). By activating YAP, which is likely the upstream regulator of *AFP* (Supplementary Figure S9A), significant upregulation of the fetal hepatoblast marker *AFP* was observed compared with the control in primary hepatocytes and ±Dox (doxycycline) YAP Tg organoids ((34), Supplementary Figure S9B).

First, by enquiring the *AFP* gene using the ‘Single-gene Enquiry’ module of the ‘Bulk Expression’ domain, the expression levels of the *AFP* gene were significantly high in human fetal hepatocytes and HepG2 cancer cell lines derived from hepatocellular carcinoma but was absent in adult liver tissue (Figure 3A). In comparison, users can carry out ‘Principal Component Analysis’ from the ‘Single-cell Atlas’ domain with scRNA-seq or scATAC-seq datasets. After selection of ‘Homo sapiens’ (species), ‘Liver’ (tissue), ‘21–30 (year old), 51–60 (year old), GW13 (gestational weeks), GW16, GW17, GW26’ (timepoint), ‘Tedd.10_Liver_scRNA’ (sequencing type), and filling in ‘*AFP*’ as the enquiry, the UMAP result showed a high expression level of *AFP* in hepatoblasts of liver tissue from humans in the early gestational weeks (such as GW13). This high expression level continued to decline during development (from GW13 to GW17), and no expression was identified in adulthood. This is consistent with the results shown from the bulk sample RNA-seq data. In addition, the result of the scATAC-seq data with selection of ‘Homo sapiens’ (species), ‘Liver’ (tissue), ‘51–60 (year old), GW13, GW16, GW17’ (timepoint), ‘Tedd.10_Liver_scatac (sequencing type)’ and with ‘*AFP*’ as the enquiry, also shows a significantly higher level of chromatin accessibility scores in hepatoblasts from fetal livers (Figure 3B), supporting the evidence that *AFP* is involved in fetal liver development. Furthermore, by enquiring ‘*AFP*’ in the ‘Single-gene Enquiry’ from the ‘Single-cell Atlas’ domain, the distributions of cell type-specific and timepoint-specific expression patterns were provided (Supplementary Figure S10A and B).

TEDD also provides a function to directly identify genes with expression patterns that are similar to or distinct from those of a targeted gene (i.e. *AFP*) in the anchor compared with those tissue-types/cell-types/timepoints that the

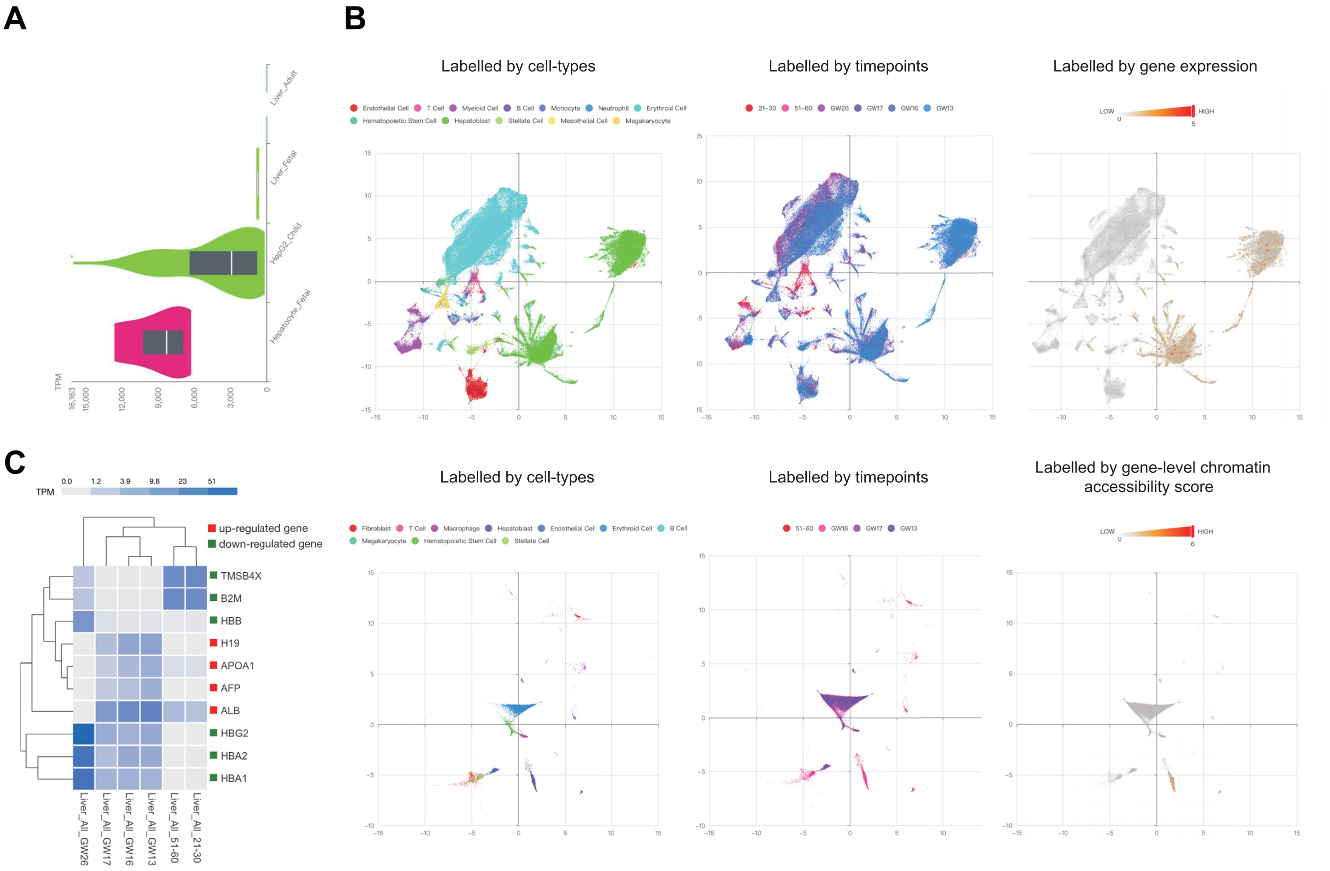

Figure 3. Investigation of temporally and tissue/cell-type specifically expressed gene(s). (**A**) Bulk-sample RNA-seq data indicate the absence of *AFP* gene expression in adult liver tissues but a significantly higher expression identified in fetal hepatocytes. Y axis indicates the sample type with specific timepoint, and X axis reveals the abundance of gene expression (in TPM format). (**B**) Principal Component Analysis with scRNA-seq and scATAC-seq datasets shows the cells labelled by cell-type (left) and by timepoint (middle) as well as the cells with *AFP* gene expression (or gene-level chromatin accessibility scores, right). The upper panel shows the UMAP results from scRNA-seq datasets, while the lower panel reveals the UMAP results from scATAC-seq datasets. (**C**) Identification of significantly differentially expressed gene(s) and clustering based on the expression levels. A box next to each gene is colored by red or green, respectively, if the gene is significantly higher or lower expressed in the selected tissue-type and timepoint (as an anchor) compared with other tissue-types/timepoints. Each box in the clustering figure indicates the abundance of gene expression in $\log_{10}$ transformation.

targeted gene showed significantly differential expression. With the 'Temporally Regulated Genes' function under the 'Temporal & Dynamic' domain, after selecting 'Homo sapiens' (species) and '*AFP*' (as gene), a list of clusters (tissue-type and timepoint) in which *AFP* showed significantly temporally expressed among the same tissue-type was provided. Alternatively, users can directly select a cluster (with defined tissue-type, cell-type and timepoint) as an anchor. For instance, after selection of a cluster [such as 'Liver_ALL_GW13 (fetal liver tissue collected in 13 gestational weeks with all cell-types selected)'] as an anchor for comparison (see Supplementary Methods), we further selected the tissue-type ('Liver'), cell-type ('All'), and the timepoints of interest ['21–30 (year old)', '51–60 (year old)', 'GW13', 'GW16', 'GW17' and 'GW26'], a clustering of genes with significantly differential expressions was shown (Figure 3C). Both *AFP* and *ALB* genes showed a significantly higher expressed in fetal liver tissue at GW13 compared with the other selected tissue-types/timepoints, echoing the current knowledge that the *AFP* and *ALB* genes are likely co-expressed in fetal liver tissues (35).

In addition, there were another two genes (*APOA1* and *H19*) showing significantly higher expressions (similar to *AFP* gene), whereas another six genes showed significantly lower expressions (distinct from *AFP* gene). The temporal expression indicated by our analysis was supported by the reported studies. For instance, *HBA1*, *HBA2*, *HBB* and *HBG2* are known to play an important role in human haematopoiesis from fetal liver (36). Previous study demonstrated that all of them showed an increasing expression pattern during fetal development (36). In addition, as erythropoiesis mainly taking place in the bone marrow in late gestation and after birth (37), it explains the reason why there was absence of expression of these genes in adult liver in our result. In comparison, Tβ4 encoded by gene *TMSB4X* did not reveal any significant reactivity in the vast majority of liver cells during human gestation and at birth (38), while a following study demonstrated a strong diffuse accumulation of in the hepatocytes in some adults (39). It supported the significantly lower expression of *TMSB4X* in fetal liver at GW13 compared with the other timepoints. Lastly, significantly higher expression of *H19* in liver tissues in fetal stages compared to adults (nearly absent) has been shown (40),

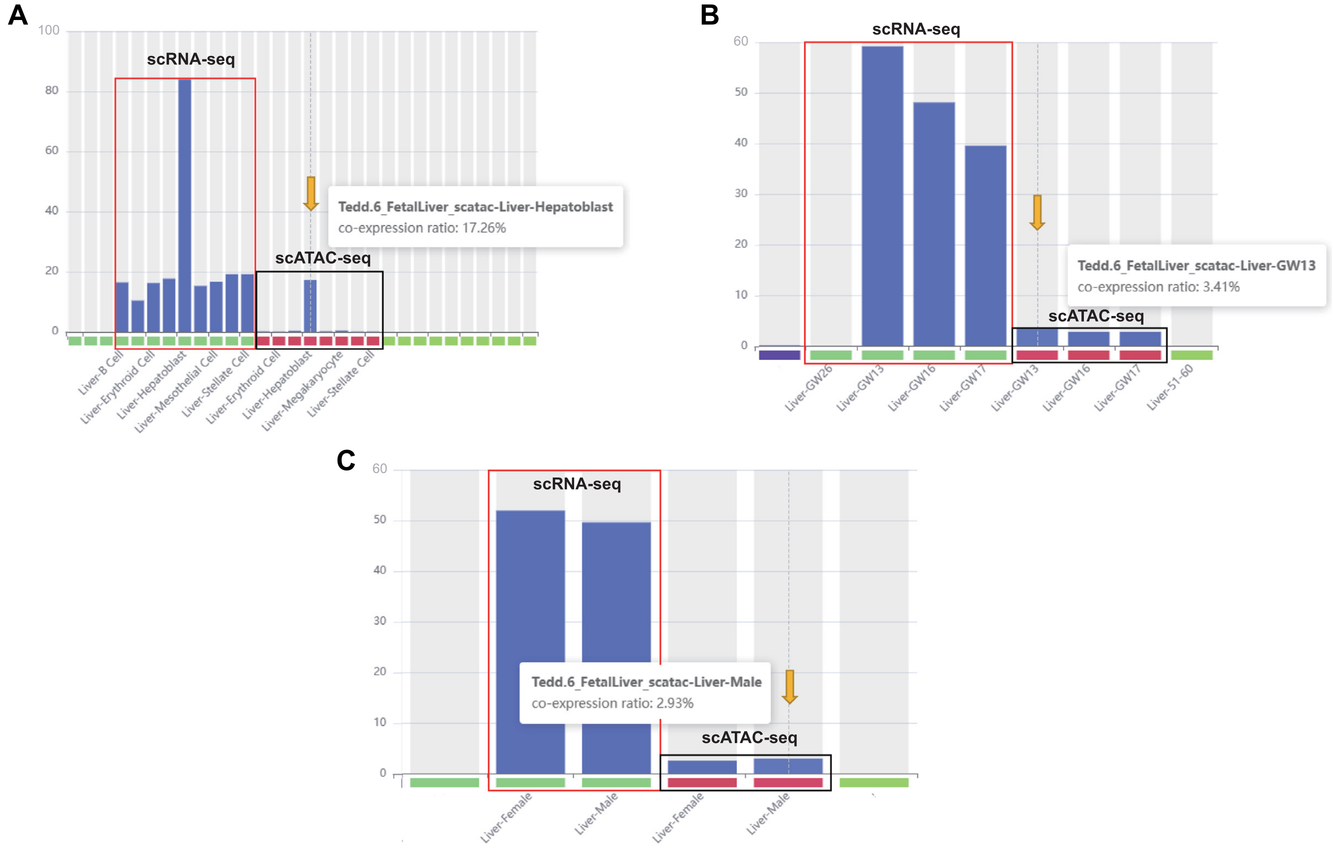

Figure 4. Co-expression network. Temporal and cell-type specific co-expression with multiple genes exemplified by AFP and ALB genes. Identification of cells with co-expression of the targeted genes in cell-type (**A**), timepoint (**B**) and sexual categories (**C**). Each bar indicates the percentage of cells in each category with co-expression identified. In each figure, the column with the percentage of cells with co-expression of AFP and ALB gene identified shown is indicated by an orange arrow.

which also supported our finding. Taking together, these studies supported the robustness of our analysis.

To detect the temporal expression patterns of *AFP* and its relation to other genes from the GO biological functions or KEGG pathways that it is involved, we first identified those GO or KEGG pathways in humans that *AFP* was anticipated to participate in by using the 'Multigene Enquiry'. There was one KEGG pathway (Hippo signaling pathway: hsa04390) and 10 GO processes in which *AFP* was involved. Genes from the same GO or KEGG pathway, or even no relationship reported before, were used as input for the clustering of the temporal, tissue-type, and cell-type specific expressions (gene number up to 50). We investigated the expression patterns of the genes *AFP*, *YAP1* and *WWTR1* (hsa04390) in human liver development across fetal and adult stages. After selection of species ('Homo sapiens'), tissue-type ('Liver'), cell-type (all types) and timepoints ['21–30 (year old)', '51–60 (year old)', 'GW13', 'GS16', 'GW17' and 'GW26'], the results indicated that the expression levels of the genes *YAP1* and *WWTR1* were both detected from GW13 to 17, but both were absent from GW26 toward adult stages (Supplementary Figure S9C and D). A similar expression pattern was observed for the *AFP* gene. Therefore, TEDD analysis showed a result that is consistent with the current knowledge that the activation of YAP, the potential upstream regulator of AFP, results in the significant upregulation of the *AFP* gene (34).

Last, researchers can also use the module of 'Co-expression' to investigate the cell type(s) and timepoint(s) with co-expression of the targeted genes (up to 5, Figure 4). Under the domain of 'Co-expression', by selecting species ('Homo sapiens'), tissue-type ('Liver'), and cluster-type ('cell-type') and entering '*AFP'* and '*ALB*' as genes stepwise, the results showed that 83.89% of fetal hepatoblasts with the expression of both '*AFP'* and '*ALB*' were identified from the scRNA-seq data (Figure 4A). In addition, the scATAC-seq data showed that 17.26% of cells had accessible chromatins identified in both the '*AFP'* and '*ALB*' genes (Figure 4A). Furthermore, when changing 'timepoint' as the cluster-type, the result showed the highest percentage of cells with the expression of both genes occurred in GW13 but declined from GW13 to GW17 (Figure 4B), while there was no significantly difference of the percentage of cells with co-expressions identified between female and male in both scRNA-seq and scATAC-seq datasets (Figure 4C). This result indicates that similar expression patterns were observed between the two genes and suggests that the expression levels of both (in hepatoblasts) are temporally regulated (during fetal liver development). As the data were archived by combining all datasets (scRNA-seq and scATAC-seq) related to liver, TEDD also provides UMAP results of independent datasets to show the cells with the co-expression of the targeted genes.

SUMMARY AND FUTURE PERSPECTIVES

To date, a rapidly increasing number of studies have investigated gene expression patterns and transposase-accessible chromatin across different species, cell-types, and tissue-types during different developmental timepoints by the state-of-the art sequencing technologies (bulk-sample RNA-seq, scRNA-seq and scATAC-seq) (41,42). We have devised and established a reference database (such as multiorgan developmental atlases) by systematically summarizing and reanalyzing datasets that enable the deciphering the underlying mechanism(s) across different developmental periods. We believe this resource will not only enable cross tissues/organs and timepoint analyses to pinpoint the critical cell-type(s), tissue-type(s) and timepoint(s) of each gene but also facilitate the identification of co-expression patterns for deciphering the potential contributions of genes with unknown biological functions related to development/diseases. Here, we presented a freely accessible online database, TEDD, to address such a demand with the integration of scRNA-seq and scATAC-seq data. In this study, we curated data from 481 cell types, 79 tissue types and 92 timepoints in humans and multiple model organisms. Of note, we applied the recommended modules (43), including SCtransform (26) and Harmony (27), for normalization and integration of the datasets to recapitulate the expression and chromatin accessibility patterns of the whole tissues/organs or whole embryos/fetuses. Batch effects among studies have been significantly reduced but cannot be excluded completely. Therefore, we retained some modules with selections of the datasets from a single study to minimize such potential effects. In addition, although TEDD has incorporated scRNA-seq, scATAC-seq data and all related bulk RNA-seq data from multiple tissues/organs and even from whole embryos in humans and most important model organisms, it still lacks datasets from spatial transcriptomic analysis. Further improvement for incorporating such data is warranted in the future. Nonetheless, researchers are advised to keep these limitations in mind.

To further improve the database, we seek feedback and suggestions from the community, which will be addressed in a timely manner to improve the performance and scientific value of the database. Furthermore, datasets from the state-of-the-art technologies, such as temporal-spatial multiomics from multiple tissues, during various pivotal timepoints will also be curated and integrated. In addition, continuing curation of datasets generated in future studies will be conducted and incorporated on a regular basis. This database is the foundation for continuing genetic research on human diseases and developmental biology for the scientific community. Thus, we expect to update the datasets semiannually with the most up-to-date published datasets and to develop and incorporate new analytical functions biannually. Overall, we developed a freely accessible online database, namely, TEDD, with the integration of RNA-seq and ATAC-seq data to facilitate the understanding of gene expression and chromatin accessibility profiles as well as co-expression networks during development (such as embryogenesis and organogenesis) and to provide a reference for understanding the etiologies of genetic defects.

DATA AVAILABILITY

TEDD is a database providing the dynamics of gene expression and chromatin accessibility profiles across different developmental stages in human and multiple model organisms (https://TEDD.obg.cuhk.edu.hk/). The program codes used in this study have been made available at https://github.com/ZihengCUOG/TEDD.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Shenzhen Science and Technology Innovation Committee [2021Szvup146]; Health and Medical Research Fund [08190226, 09200236]; Collaborative Research Fund [C4062-21GF]; Guangdong Genomics Data Center [2021B1212100001]; Drs Cao and Dong thanks the support from Direct Grant, The Chinese University of Hong Kong. Funding for open access charge: Health and Medical Research Funds [08190226, 09200236].

Conflict of interest statement. None declared.

REFERENCES

1. Wang,Z.Y., Leushkin,E., Liechti,A., Ovchinnikova,S., Mößinger,K., Brüning,T., Rummel,C., Grützner,F., Cardoso-Moreira,M., Janich,P. *et al.* (2020) Transcriptome and translatome co-evolution in mammals. *Nature*, **588**, 642–647.
2. Larsen,W.J., Schoenwolf,G.C., Bleyl,S.B., Brauer,P.R. and Francis-West,P.H. (2009) In: *Larsen's Human Embryology*. Churchill Livingstone.
3. Fei,L., Chen,H., Ma,L., E,W., Wang,R., Fang,X., Zhou,Z., Sun,H., Wang,J., Jiang,M. *et al.* (2022) Systematic identification of cell-fate regulatory programs using a single-cell atlas of mouse development. *Nat. Genet.*, **54**, 1051–1061.
4. Chen,A., Liao,S., Cheng,M., Ma,K., Wu,L., Lai,Y., Qiu,X., Yang,J., Xu,J., Hao,S. *et al.* (2022) Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell*, **185**, 1777–1792.
5. Han,X., Zhou,Z., Fei,L., Sun,H., Wang,R., Chen,Y., Chen,H., Wang,J., Tang,H., Ge,W. *et al.* (2020) Construction of a human cell landscape at single-cell level. *Nature*, **581**, 303–309.
6. Garcia-Alonso,L., Lorenzi,V., Mazzeo,C.I., Alves-Lopes,J.P., Roberts,K., Sancho-Serra,C., Engelbert,J., Marečková,M., Gruhn,W.H., Botting,R.A. *et al.* (2022) Single-cell roadmap of human gonadal development. *Nature*, **607**, 540–547.
7. Di Bella,D.J., Habibi,E., Stickels,R.R., Scalia,G., Brown,J., Yadollahpour,P., Yang,S.M., Abbate,C., Biancalani,T., Macosko,E.Z. *et al.* (2021) Molecular logic of cellular diversification in the mouse cerebral cortex. *Nature*, **595**, 554–559.
8. Boroviak,T., Stirparo,G.G., Dietmann,S., Hernando-Herraez,I., Mohammed,H., Reik,W., Smith,A., Sasaki,E., Nichols,J. and Bertone,P. (2018) Single cell transcriptome analysis of human, marmoset and mouse embryos reveals common and divergent features of preimplantation development. *Development*, **145**, dev167833.
9. Lu,X., Zhang,Y., Wang,L., Wang,H., Xu,Q., Xiang,Y., Chen,C., Kong,F., Xia,W., Lin,Z. *et al.* (2021) Evolutionary epigenomic analyses in mammalian early embryos reveal species-specific innovations and conserved principles of imprinting. *Sci. Adv.*, **7**, eabi6178.

10. Shi,Y., Wang,M., Mi,D., Lu,T., Wang,B., Dong,H., Zhong,S., Chen,Y., Sun,L., Zhou,X. *et al.* (2021) Mouse and human share conserved transcriptional programs for interneuron development. *Science*, **374**, eabj6641.
11. Eze,U.C., Bhaduri,A., Haeussler,M., Nowakowski,T.J. and Kriegstein,A.R. (2021) Single-cell atlas of early human brain development highlights heterogeneity of human neuroepithelial cells and early radial glia. *Nat. Neurosci.*, **24**, 584–594.
12. La Manno,G., Gyllborg,D., Codeluppi,S., Nishimura,K., Salto,C., Zeisel,A., Borm,L.E., Stott,S.R.W., Toledo,E.M., Villaescusa,J.C. *et al.* (2016) Molecular diversity of midbrain development in mouse, human, and stem cells. *Cell*, **167**, 566–580.
13. Shay,T., Jojic,V., Zuk,O., Rothamel,K., Puyraimond-Zemmour,D., Feng,T., Wakamatsu,E., Benoist,C., Koller,D., Regev,A. *et al.* (2013) Conservation and divergence in the transcriptional programs of the human and mouse immune systems. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 2946–2951.
14. Suo,C., Dann,E., Goh,I., Jardine,L., Kleshchevnikov,V., Park,J.E., Botting,R.A., Stephenson,E., Engelbert,J., Tuong,Z.K. *et al.* (2022) Mapping the developing human immune system across organs. *Science*, **376**, eabo0510.
15. Ferguson,G.B., Van Handel,B., Bay,M., Fiziev,P., Org,T., Lee,S., Shkhyan,R., Banks,N.W., Scheinberg,M., Wu,L. *et al.* (2018) Mapping molecular landmarks of human skeletal ontogeny and pluripotent stem cell-derived articular chondrocytes. *Nat. Commun.*, **9**, 3634.
16. Barrett,T., Wilhite,S.E., Ledoux,P., Evangelista,C., Kim,I.F., Tomashevsky,M., Marshall,K.A., Phillippy,K.H., Sherman,P.M., Holko,M. *et al.* (2013) NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res.*, **41**, D991–D995.
17. Papatheodorou,I., Moreno,P., Manning,J., Fuentes,A.M., George,N., Fexova,S., Fonseca,N.A., Füllgrabe,A., Green,M., Huang,N. *et al.* (2020) Expression atlas update: from tissues to single cells. *Nucleic Acids Res.*, **48**, D77–D83.
18. Han,X., Wang,R., Zhou,Y., Fei,L., Sun,H., Lai,S., Saadatpour,A., Zhou,Z., Chen,H., Ye,F. *et al.* (2018) Mapping the mouse cell atlas by microwell-seq. *Cell*, **173**, 1307.
19. Li,Y., Yang,T., Lai,T., You,L., Yang,F., Qiu,J., Wang,L., Du,W., Hua,C., Xu,Z. *et al.* (2022) CDCP: a visualization and analyzing platform for single-cell datasets. *J Genet Genomics*, **49**, 689–692.
20. Regev,A., Teichmann,S.A., Lander,E.S., Amit,I., Benoist,C., Birney,E., Bodenmiller,B., Campbell,P., Carninci,P., Clatworthy,M. *et al.* (2017) The human cell atlas. *Elife*, **6**, e27041.
21. Zeng,J., Zhang,Y., Shang,Y., Mai,J., Shi,S., Lu,M., Bu,C., Zhang,Z., Li,Y., Du,Z. *et al.* (2022) CancerSCEM: a database of single-cell expression map across various human cancers. *Nucleic Acids Res.*, **50**, D1147–D1155.
22. Zhao,T., Lyu,S., Lu,G., Juan,L., Zeng,X., Wei,Z., Hao,J. and Peng,J. (2021) SC2disease: a manually curated database of single-cell transcriptome for human diseases. *Nucleic Acids Res.*, **49**, D1413–D1419.
23. Qi,C., Wang,C., Zhao,L., Zhu,Z., Wang,P., Zhang,S., Cheng,L. and Zhang,X. (2022) SCovid: single-cell atlases for exposing molecular characteristics of COVID-19 across 10 human tissues. *Nucleic Acids Res.*, **50**, D867–D874.
24. Li,M., Zhang,X., Ang,K.S., Ling,J., Sethi,R., Lee,N.Y.S., Ginhoux,F. and Chen,J. (2022) DISCO: a database of deeply integrated human single-cell omics data. *Nucleic Acids Res.*, **50**, D596–D602.
25. Hao,Y., Hao,S., Andersen-Nissen,E., Mauck,W.M., Zheng,S., Butler,A., Lee,M.J., Wilk,A.J., Darby,C., Zager,M. *et al.* (2021) Integrated analysis of multimodal single-cell data. *Cell*, **184**, 3573–3587.
26. Hafemeister,C. and Satija,R. (2019) Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.*, **20**, 296.
27. Korsunsky,I., Millard,N., Fan,J., Slowikowski,K., Zhang,F., Wei,K., Baglaenko,Y., Brenner,M., Loh,P.R. and Raychaudhuri,S. (2019) Fast, sensitive and accurate integration of single-cell data with harmony. *Nat. Methods*, **16**, 1289–1296.
28. Zhang,K., Hocker,J.D., Miller,M., Hou,X., Chiou,J., Poirion,O.B., Qiu,Y., Li,Y.E., Gaulton,K.J., Wang,A. *et al.* (2021) A single-cell atlas of chromatin accessibility in the human genome. *Cell*, **184**, 5985–6001.
29. Sloan,C.A., Chan,E.T., Davidson,J.M., Malladi,V.S., Strattan,J.S., Hitz,B.C., Gabdank,I., Narayanan,A.K., Ho,M., Lee,B.T. *et al.* (2016) ENCODE data at the ENCODE portal. *Nucleic Acids Res.*, **44**, D726–D732.
30. 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.
31. Li,B. and Dewey,C.N. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf.*, **12**, 323.
32. Sauzay,C., Petit,A., Bourgeois,A.M., Barbare,J.C., Chauffert,B., Galmiche,A. and Houessinon,A. (2016) Alpha-foetoprotein (AFP): a multi-purpose marker in hepatocellular carcinoma. *Clin. Chim. Acta*, **463**, 39–44.
33. Patel,S.H., Camargo,F.D. and Yimlamai,D. (2017) Hippo signaling in the liver regulates organ size, cell fate, and carcinogenesis. *Gastroenterology*, **152**, 533–545.
34. Yimlamai,D., Christodoulou,C., Galli,G.G., Yanger,K., Pepe-Mooney,B., Gurung,B., Shrestha,K., Cahan,P., Stanger,B.Z. and Camargo,F.D. (2014) Hippo pathway activity influences liver cell fate. *Cell*, **157**, 1324–1338.
35. Jochheim,A., Hillemann,T., Kania,G., Scharf,J., Attaran,M., Manns,M.P., Wobus,A.M. and Ott,M. (2004) Quantitative gene expression profiling reveals a fetal hepatic phenotype of murine ES-derived hepatocytes. *Int. J. Dev. Biol.*, **48**, 23–29.
36. Popescu,D.M., Botting,R.A., Stephenson,E., Green,K., Webb,S., Jardine,L., Calderbank,E.F., Polanski,K., Goh,I., Efremova,M. *et al.* (2019) Decoding human fetal liver haematopoiesis. *Nature*, **574**, 365–371.
37. Wood,W.G. (1976) Haemoglobin synthesis during human fetal development. *Br. Med. Bull.*, **32**, 282–287.
38. Nemolato,S., Cabras,T., Cau,F., Fanari,M.U., Fanni,D., Manconi,B., Messana,I., Castagnola,M. and Faa,G. (2010) Different thymosin beta 4 immunoreactivity in foetal and adult gastrointestinal tract. *PLoS One*, **5**, e9111.
39. Nemolato,S., Van Eyken,P., Cabras,T., Cau,F., Fanari,M.U., Locci,A., Fanni,D., Gerosa,C., Messana,I., Castagnola,M. *et al.* (2011) Expression pattern of thymosin beta 4 in the adult human liver. *Eur. J. Histochem.*, **55**, e25.
40. Wang,S., Wu,X., Liu,Y., Yuan,J., Yang,F., Huang,J., Meng,Q., Zhou,C., Liu,F., Ma,J. *et al.* (2016) Long noncoding RNA H19 inhibits the proliferation of fetal liver cells and the wnt signaling pathway. *FEBS Lett.*, **590**, 559–570.
41. Cao,J., O'Day,D.R., Pliner,H.A., Kingsley,P.D., Deng,M., Daza,R.M., Zager,M.A., Aldinger,K.A., Blecher-Gonen,R., Zhang,F. *et al.* (2020) A human cell atlas of fetal gene expression. *Science*, **370**, eaba7721.
42. Domcke,S., Hill,A.J., Daza,R.M., Cao,J., O'Day,D.R., Pliner,H.A., Aldinger,K.A., Pokholok,D., Zhang,F., Milbank,J.H. *et al.* (2020) A human cell atlas of fetal chromatin accessibility. *Science*, **370**, eaba7612.
43. Chen,W., Zhao,Y., Chen,X., Yang,Z., Xu,X., Bi,Y., Chen,V., Li,J., Choi,H., Ernest,B. *et al.* (2021) A multicenter study benchmarking single-cell RNA sequencing technologies using reference samples. *Nat. Biotechnol.*, **39**, 1103–1114.