Supplementary materials and methods

We processed the datasets with a unified protocol to document the expression patterns and chromatin accessibility profiles from 144 cell subtypes of 42 tissue types across 47 developmental timepoints in human, 122 cell-types of 17 tissue types among 30 timepoints in mouse, 64 cell-types of 33 tissue types among 9 timepoints in zebrafish and 214 cell-types of 11 tissue types among 12 timepoints in nematode, respectively.

To identify genes with expression patterns that are similar to or distinct from those of a target gene in the anchor compared with those tissue-types/cell-types/timepoints that the target gene showed significantly differential expression, we further proceeded the datasets as (1) gene expression level: calculation of the median TPM values for each gene as the gene expression level in a selected tissue-type (with a recorded timepoint) or a particular cell-type from a tissue-type (as a cluster) was performed; (2) ratio of expression fold-change: for each gene, fold-change of expression level was calculated by using the value from the anchor [a tissue-type (setting cell-type as all) or a cell-type from a tissue-type and the corresponding timepoint] dividing by the average value from the other selected timepoints/cell-types; (3) upon a customized selection of fold-change [such as log2 transformation of absolute fold-change value (LogFC) as 1 indicating an absolute cutoff of 2-fold change] and the maximum number of genes being displayed (MaxNumber, prioritized by the absolute fold-change value), those genes with LogFC>1 were remarked as significantly higher expressed genes in the anchor, while those ones with LogFC<-1 were indicated as significantly lower expressed genes in the anchor; and (4) clustering of the expression levels for these significantly higher/lower expressed genes among the selected timepoints/cell-types.

To confirm the robustness of our analysis, we firstly compared the TEDD result with the one generated from the original study with a scRNA-seq dataset consisting of 87,065 cells from human cerebellum ((1), ***Supplementary Figure S1***). The UMAP

result showed a complete separation of different cell clusters (based on the cell types) consistently with the result from the original study, indicating a good reproducibility of our analysis. In addition, to further evaluate whether integration would introduce batch effects or artifacts, we compared the integrated results with the original studies using an overall 141,133 cells derived from liver tissue in human fetal and adult development ((1,2), ***Supplementary Figure S2***). Cells from different cell-types (such as hepatoblasts, myeloid cells, erythroid cells) were likely separated from each other although they were from different datasets and different developmental age groups. In addition, by labelling the cells from each study independently in the UMAP results (***Supplementary Figure S2B***), hepatoblasts, which are the bi-potential cells and dominant in fetal liver development, were absent in the adult datasets. In addition, hepatoblasts presented as two main subclusters which is consistently as shown in the original study (***Supplementary Figure S2***). Although there were a few small subclusters of hepatoblasts observed, by comparing the UMAP result labelled by developmental timepoints, it was contributed by the temporal expression differences (***Supplementary Figure S2***). The same scenario was also observed from the scATAC-seq datasets from human liver development (***Supplementary Figure S3***). Taken together, it indicated that the data processing, normalization, and integration conducted in our analysis unlikely introduced batch effects or artifacts across different studies.

Instruction on the user interface and functions

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 modules, and (iii) three diagrams of cell-

type/tissue/organ sampling sites in human (three phases: embryonic, fetal and adult) curated by TEDD (***Supplementary Figure S4***). If users are interested in a particular module such as ‘Dynamic Compositions’ under the ‘Single-cell Atlas’ domain, they could click on the hyperlink and go to the corresponding page. 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, which will direct to the page showing all datasets related to this cell-type/tissue-type/organ.

On the page of ‘Current Release’ under the ‘Datasets’ domain, a statistical summary of datasets curated in this database was provided. In addition, a list of these 257 scRNA-seq and scATAC-seq datasets was available from “Data Download”, and detailed metadata was also provided for each dataset; including data source (name of publication), technology (platform), species, tissue type (or cell type), timepoint, treatment (if available), the number of cells used for analysis and the PubMed ID of the publication. In addition, for each dataset, we also provided download function for the data matrix as well as the metadata. Lastly, in the page of ‘Release History’, the history of each version released with the updated content was provided.

Although the advancement of single-cell sequencing provides insights on developmental biology or diseases, limitations compared with bulk sample RNA-seq remain. Apart from the essential usage of live cells (from fresh samples), scRNA-seq adopts a dissociation process, which induces the expression of stress genes, causing "artifactual transcriptional stress responses" in cellular transcriptional patterns, likely resulting in transcriptional bias (3). For many solid tissues, such as brain (4,5) and kidney (6), proteases tend to dissociate cell types that are easy to dissociate, so cells that are not easy to dissociate will be lost (no expression can be seen in the sequencing data). In this regard, providing both bulk RNA-seq and scRNA-seq data for the users would be important to compare the consistency and difference of

expression patterns. The ‘drop-down’ menu of ‘Bulk Expression’ (***Supplementary Figure S5***) consists of three modules: ‘Gene Overview’, ‘Single-gene Enquiry’ and ‘Multi-gene Enquiry’. ‘Gene Overview’ includes pie charts of different categories of RNAs expressed in the pan-tissues of human and mouse, and a bar chart showing the percentages of human or mouse specific and shared RNAs in each category of expressed RNAs. Under the module of ‘Single-gene Enquiry’, researchers can search the expression patterns of gene of their interest in each tissue (presented as a violin plot) will be provided with functions of sorting the data by the median values or by the names of tissues (by the first character). In comparison, ‘Multi-gene Enquiry’ enables an input of gene list for identifying the gene expression patterns in any sample type(s) from the categories of cell lines, tissue, primary cells and in vitro differentiated cells for clustering of gene expressions. The clustering can be further sorted by genes or by sample/tissue type(s). In addition, TEDD also allows submission of one or multiple GO/KEGG IDs together with genes for expression clustering. To enable researchers understanding the expression patterns of the genes under the same GO or KEGG pathway of a given gene, TEDD allows identification of the GO/KEGG that the enquired gene belongs to, and profiles the expressions of all genes with clustering. Of note, since expression levels can be significantly different among different tissue types for certain genes, we provided log10 scale of the expressions (TPM format) in addition to the original values for visualization. As there are multiple genes involved in the cluster, TEDD also allows zoom in and out for visualization, and download function for figure and expression matrix.

The ‘Single-cell Atlas’ domain consists of five modules (***Supplementary Figure S6***), ‘Gene Overview’, ‘Single-gene Enquiry’, ‘Multi-gene Enquiry’, ‘Dynamic Composition’ and ‘Expression Principal Component Analysis’. The page of ‘Gene Overview’, it is similar to the one presented under ‘Bulk Expression’ domain but with the scRNA-seq data instead. For the modules of ‘Single-end Enquiry’ and ‘Multi-gene Enquiry’, they are similar with the ones shown under the ‘Bulk Expression’ domain but utilizing

scRNA-seq and scATAC-seq data instead. In addition, ‘Single-end Enquiry’ provides subclassification of each tissue type into further subgroups based on cell types and timepoints as well as the percentage of cells with target gene expression identified (***Supplementary Figure S6***). It serves as an intensive statistical summary of gene expression compared with the ones shown with bulk sample RNA-seq. In particular, under ‘Multi-gene Enquiry’, researchers can fill in a gene to identify putative GO or KEGG pathway that the gene participated in. A list of GO and KEGG pathway will be provided with lists of genes in each particular GO/KEGG available for download. Users can select genes (up to 50) as input to perform clustering. Under the module of ‘Dynamic Composition’, the compositions of each cell type (such as excitatory neuron and inhibitory neuron) presenting as cell number (Y axis) are shown based on the sampling timepoints provided (X axis). While with the module of ‘Principal Component Analysis’, when a researcher filling in a gene name, TEDD provides clustering of the dataset with four options of cell labelling (cell type, tissue type, timepoint and sex), respectively. In particular, researchers can opt out cells from cell-type(s), tissue-type(s), timepoint(s) or sex that are not interested, in order to remove those cells from the figure of clustering. In addition, TEDD also provides labels to indicate those cells with expression (scRNA-seq data) or chromatin accessibility (scATAC-seq data) of the target gene identified (adjustable cutoff) in the clusters. Furthermore, upon selection of cell type, tissue type, timepoint or sex (X axis), the numbers of cells and the ratio of expression of the target gene (calculated by the overall TPM dividing by the number of cells) are shown (Y axis), respectively, based on the selected category. In particular, in order 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. Of note, when queuing for a particular gene, for those categories (with the same species, sequencing type, tissue type) with only one dataset, raw count for each cell is shown. For those categories with more than one dataset recruited and with integration performed, raw normalized count

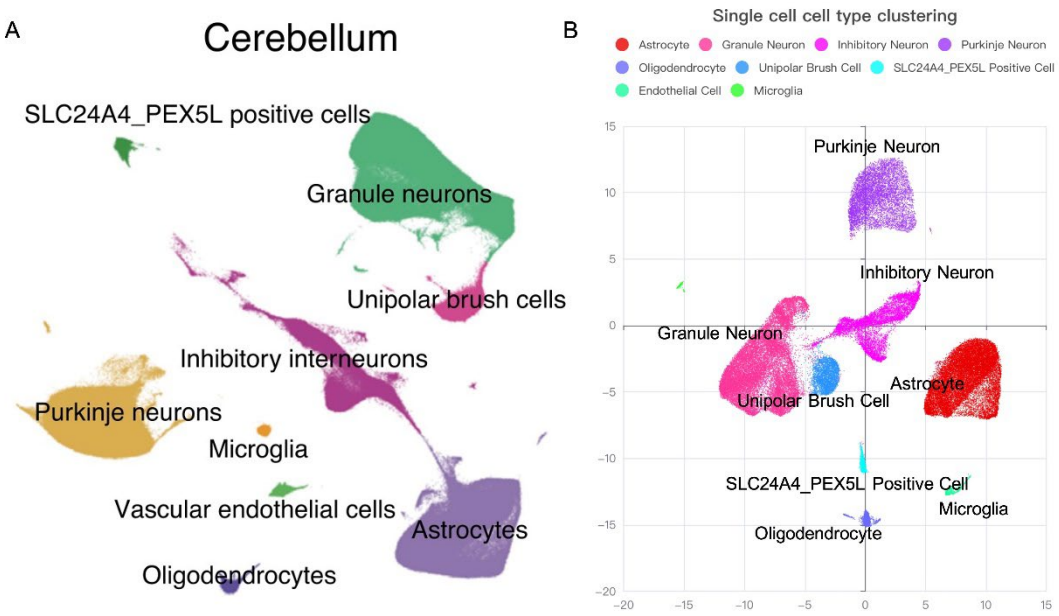
for each cell is provided.

In addition, within the ‘Co-expression’ domain, after defining the information of species, tissue type and the options of clustering (cell type, timepoint and sex), TEDD allows submission of multiple genes (up to 5) to investigate those cells with co-expressions of the submitted genes. Similar functions are provided in showing the clustering results of datasets used and the cells with co-expression identified in the clusters. Furthermore, upon selection of cell type, tissue type, timepoint or sex (X axis), the numbers of cells with co-expression identified are shown (Y axis), based on the selected category (***Supplementary Figure S7***).

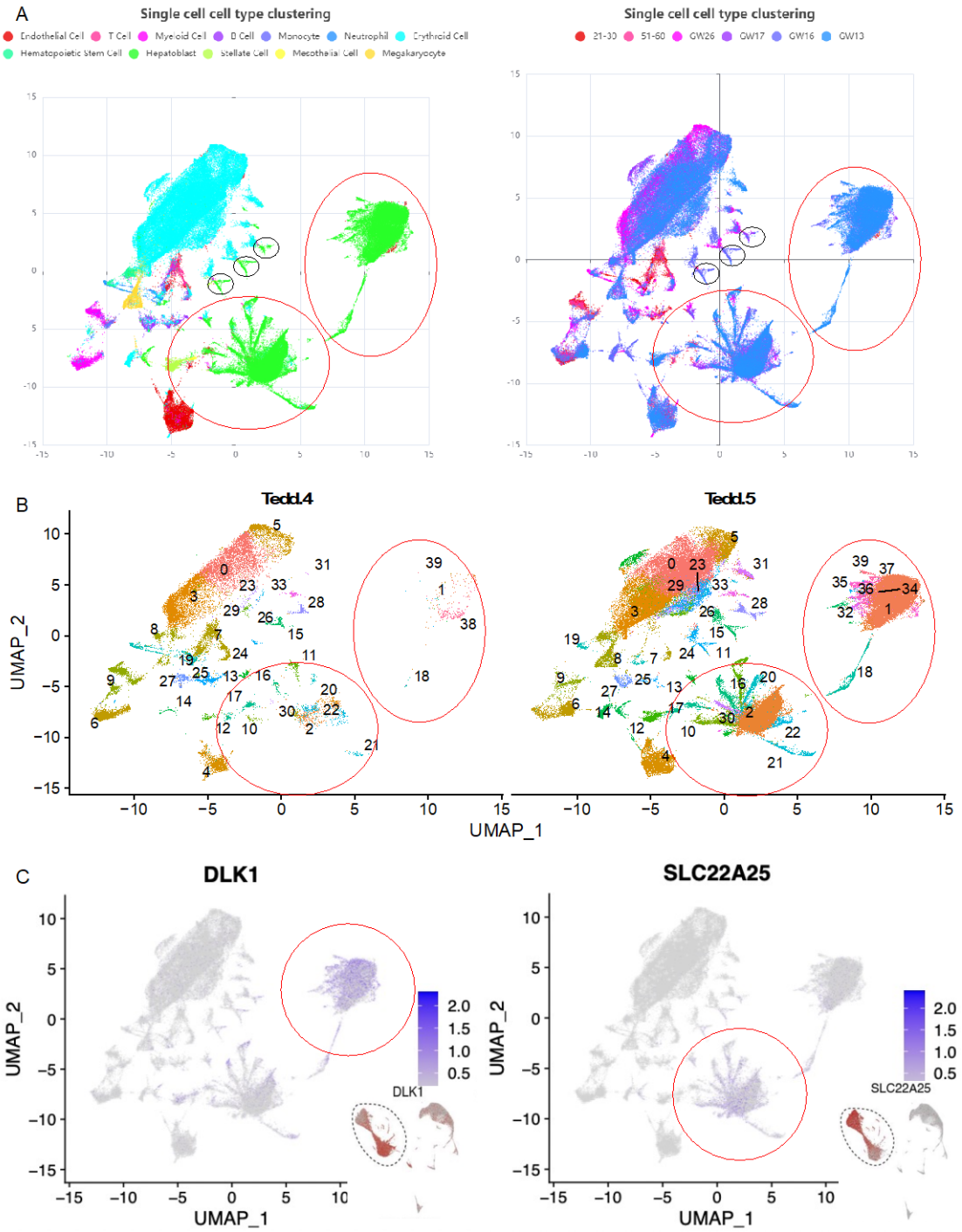
The ‘Temporal & Dynamic’ domain (***Supplementary Figure S8***) includes three modules, ‘Temporally Regulated Genes’, ‘Temporally Regulated Genes among Chromosomes’ and ‘Stably Expressed Genes’. These three modules are only confined to the scRNA-seq data. Firstly, ‘Temporally Regulated Genes’ provides an identification of genes with differential expressions (with a customized cutoff: logFC representing an absolute log2 scale of the fold changed) among the selected timepoints (with tissue type and cell type selected). It provides option to set a gene as an anchor to further investigate gene(s) with a similar or distinct expression pattern with the target gene. In addition, a maximum number of temporally regulated genes to be shown (N) can be customized and only those N genes with the most significantly differential expressions will be shown if there are more than N genes remained with a set logFC for filtering of fold-change. Secondly, ‘Temporally Regulated Genes among Chromosomes’ module provides a genome-wide distribution of temporally regulated genes with a customized cutoff (logFC) and a maximum number of genes being displayed. Lastly, under the module of ‘Stably Expressed Genes’, researchers can identify gene(s) with expressions identified among the tissue-types or timepoints selected.

Lastly, in the ‘Help’ page, a graphical operation guide is prepared for new users to get

started with each function, which will help them fully utilize the resources in TEDD.

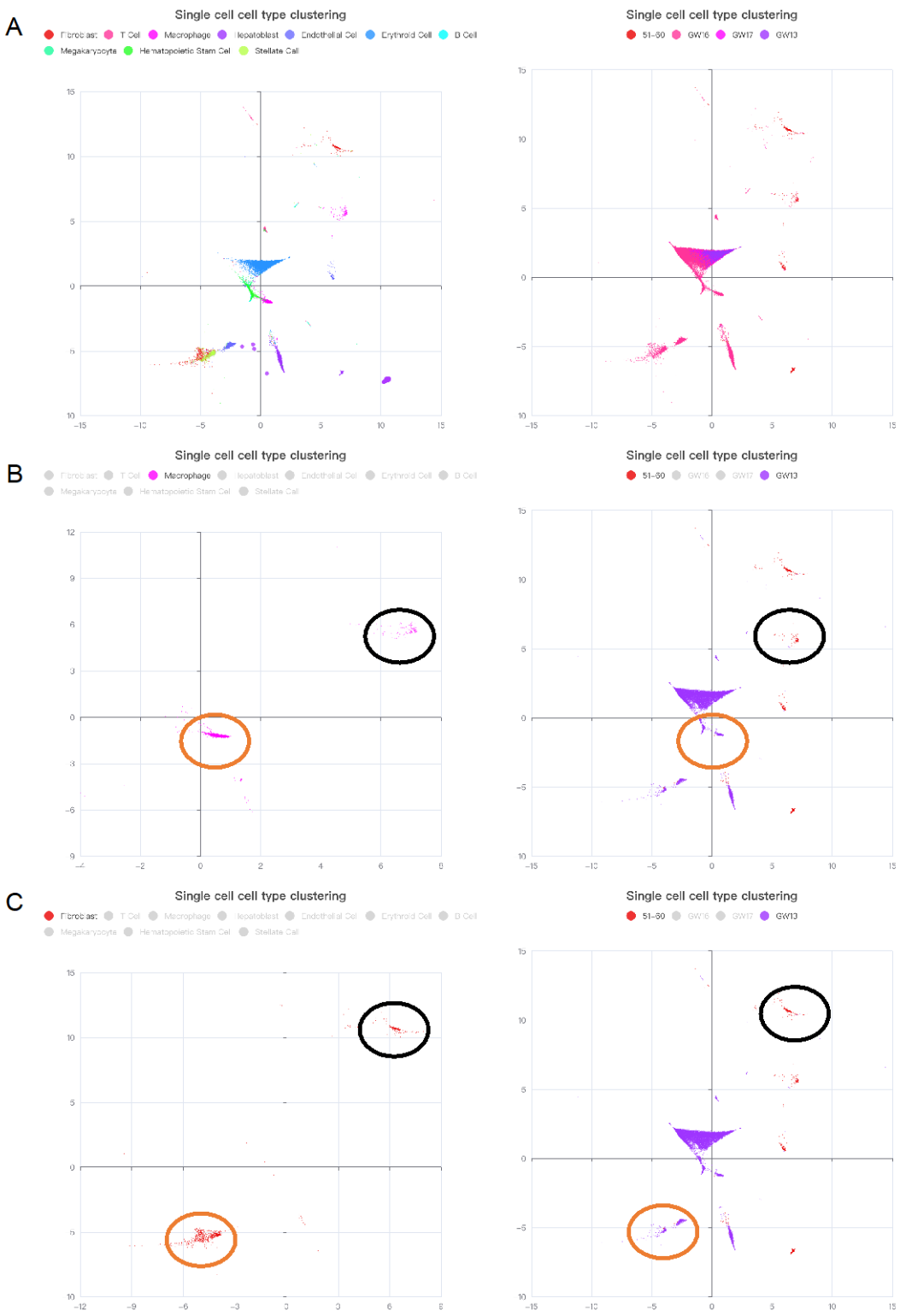


Supplementary Figure S1. **Comparison of UMAP result from a dataset between the published results and TEDD.** (A) Original published UMAP result, and (B) analyzed result from TEDD. Each cell type is labelled in a color and indicated by the name near the cell cluster.



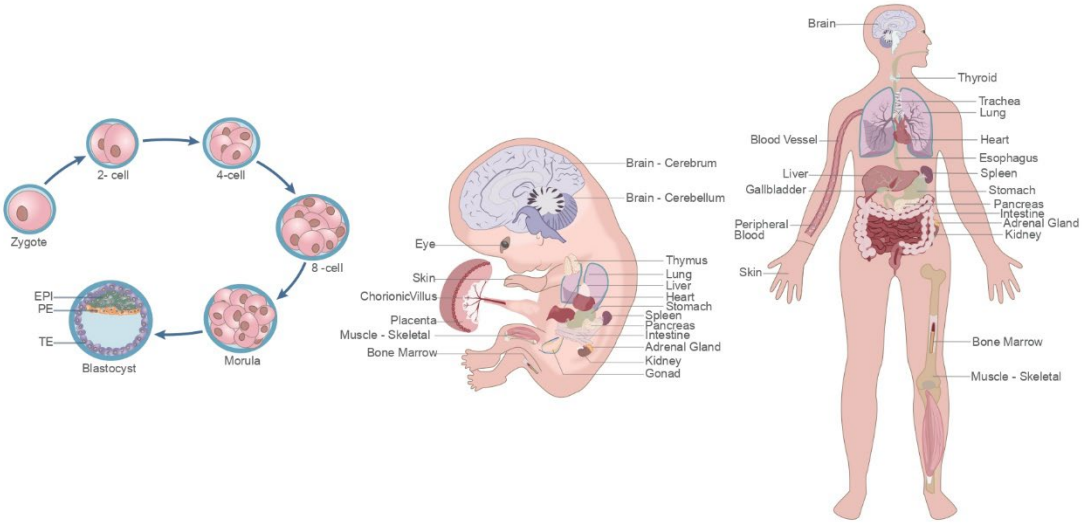
Supplementary Figure S2. Demonstration on the robustness of integration performed by our analysis. (A) UMAP result after integration is labelled by cell types (left) and by the developmental stages (right). The figure legends are shown in the upper panel to indicate the cell color and corresponding cell type or developmental stage. Five circles are shown to indicate different subclusters of hepatoblast: black circles indicate the three small subclusters and red circles reveal the two main

subclusters. (B) Cell labelling for each study: tedd 4 (left) and tedd 5 (right). (C) Cell labelling with expression of *DLK1* (left) and *SLC22A25* genes (right). The UMAP result from the original study labelled with expression of *DLK1* (left) and *SLC22A25* genes (right) is shown in the right bottom of each figure for comparison. For figure (B) and (C), two red circles reveal the region of two main subclusters of hepatoblast shown in figure A.

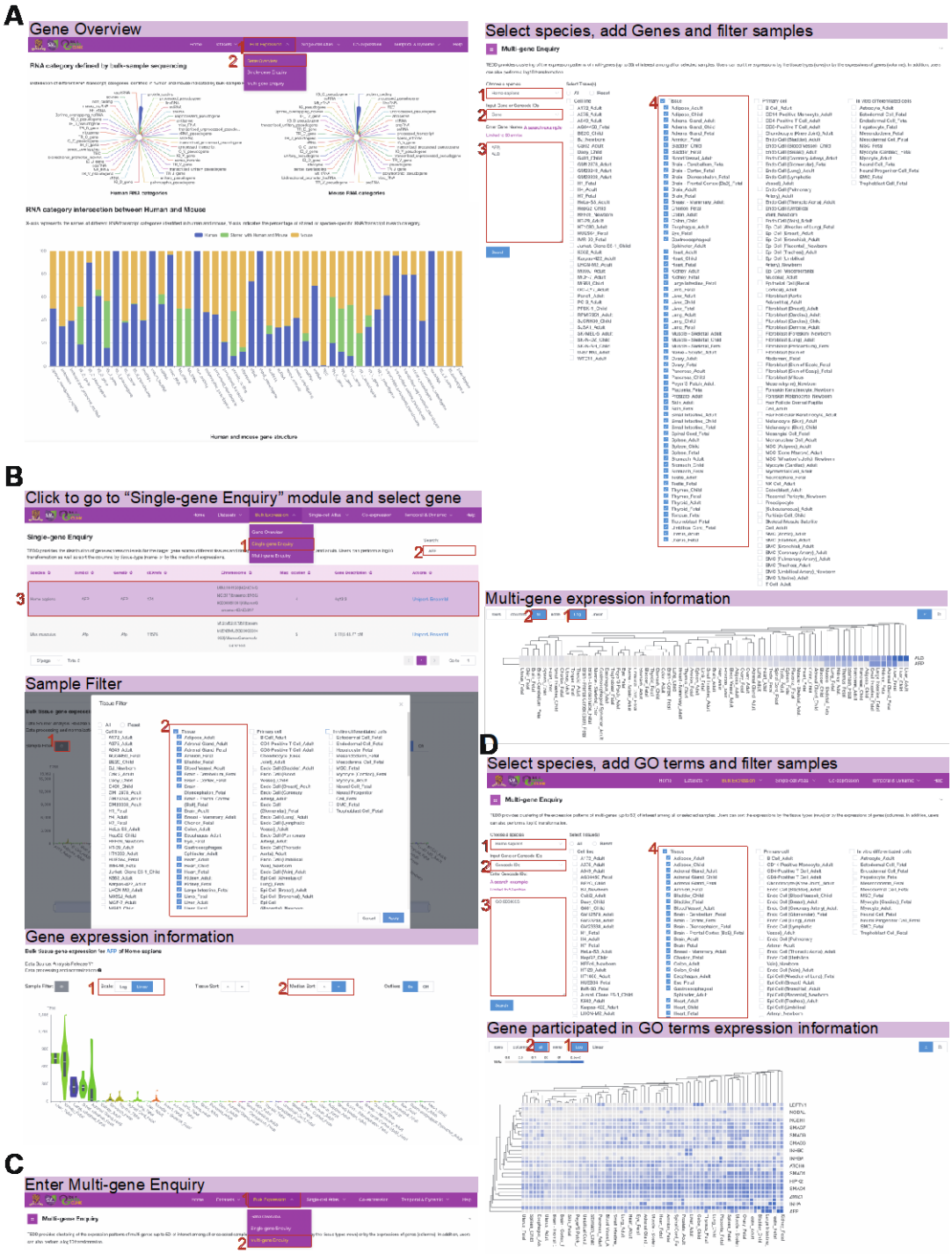


Supplementary Figure S3. Different subclusters observed in each cell type in scATAC-seq. (A) UMAP result after integration is labelled by cell types (left) and by the developmental stages (right). The figure legends are shown in the upper panel to indicate the cell color and corresponding cell type or developmental stage. (B) Cell labelling for macrophage (left) and different developmental stages (GW13 and 51-60

yo). (C) Cell labelling for fibroblast (left) and different developmental stages (GW13 and 51-60 yo). For figure (B) and (C), two circles in black and orange indicate the contribution of more than one subclusters were due to the different patterns in different developmental stages.

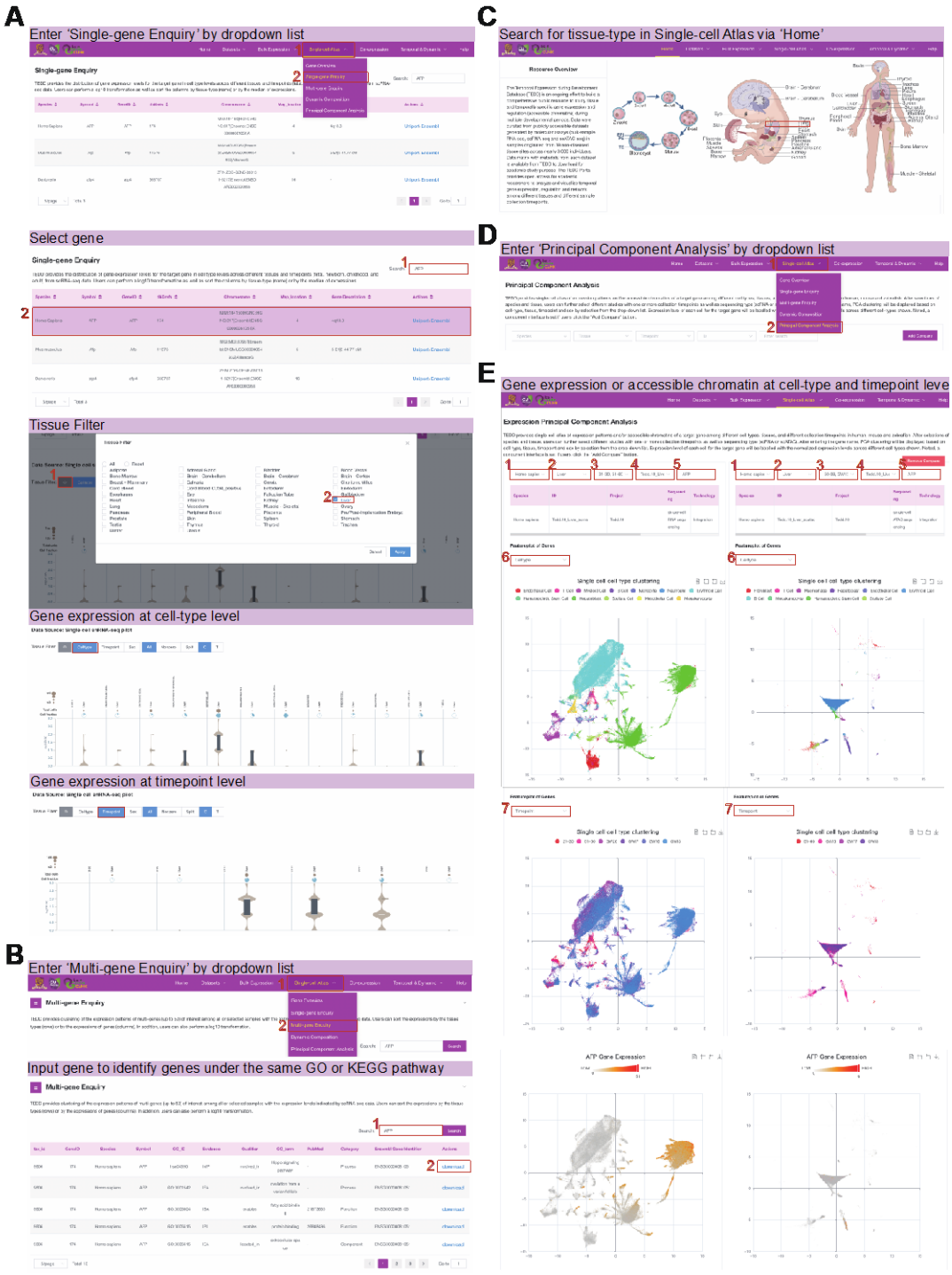


Supplementary Figure S4. Distribution of datasets originated cell-types and tissue-types collected in human in this study. The figures from left to right show diagrams of different stages (cleavage, morula, and blastocyst stages) in embryonic development, fetal development and adult stage with legends showing the cell-type or tissue-type collected (datasets).



Supplementary Figure S5. Demonstration of the 'Bulk Expression' domain. (A) Screenshot of Gene Overview page. (B) Operation to obtain gene expression profiles of *AFP* in bulk-tissue RNA-seq data. (C) Operation to view expressions of genes *AFP* and *ALB* in 'Multi-gene Enquiry' page. (D) Operation to view the expression patterns

with input of GO term(s) in ‘Multi-gene Enquiry’ page.

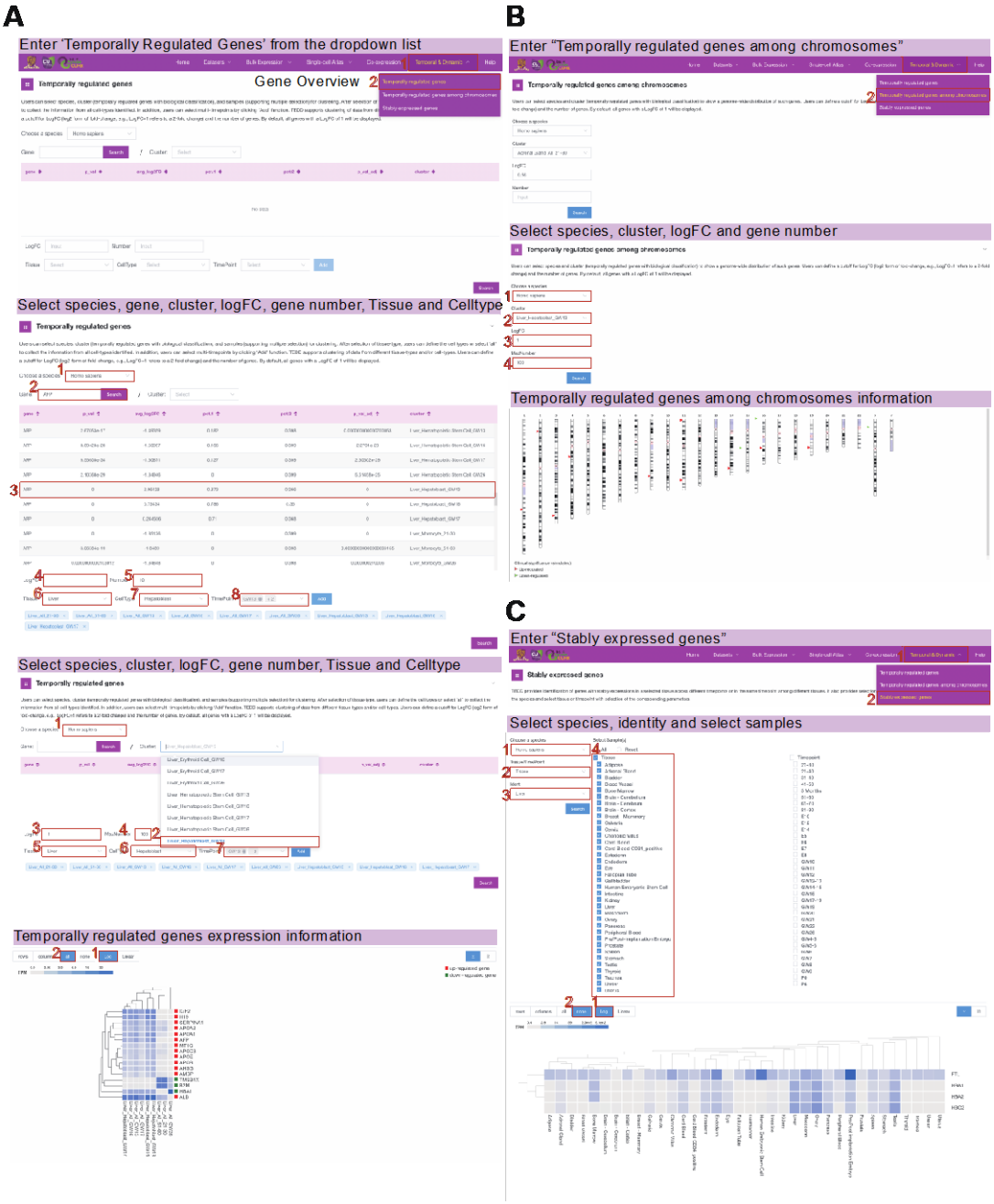


Supplementary Figure S6. Demonstration of the 'Single-cell Atlas' domain. (A) Operation to obtain gene expression profiles of *AFP* in scRNA-seq data. (B) Operation to identify genes under the same GO or KEGG pathway with the target gene in order to perform expression cluttering of multi-genes in 'Multi-gene Enquiry' page. (C)

Direction of performing ‘Principal Component Analysis’ in a particular cell/tissue/organ type by clicking that cell/tissue/organ the diagram in the front page or (D) via the dropdown list in ‘Single-cell Atlas’ domain. (E) Operation and concurrent windows to show the *AFP* gene expression and accessible chromatin profiles from scRNA-seq and scATAC-seq data at cell-type level and at timepoint level.

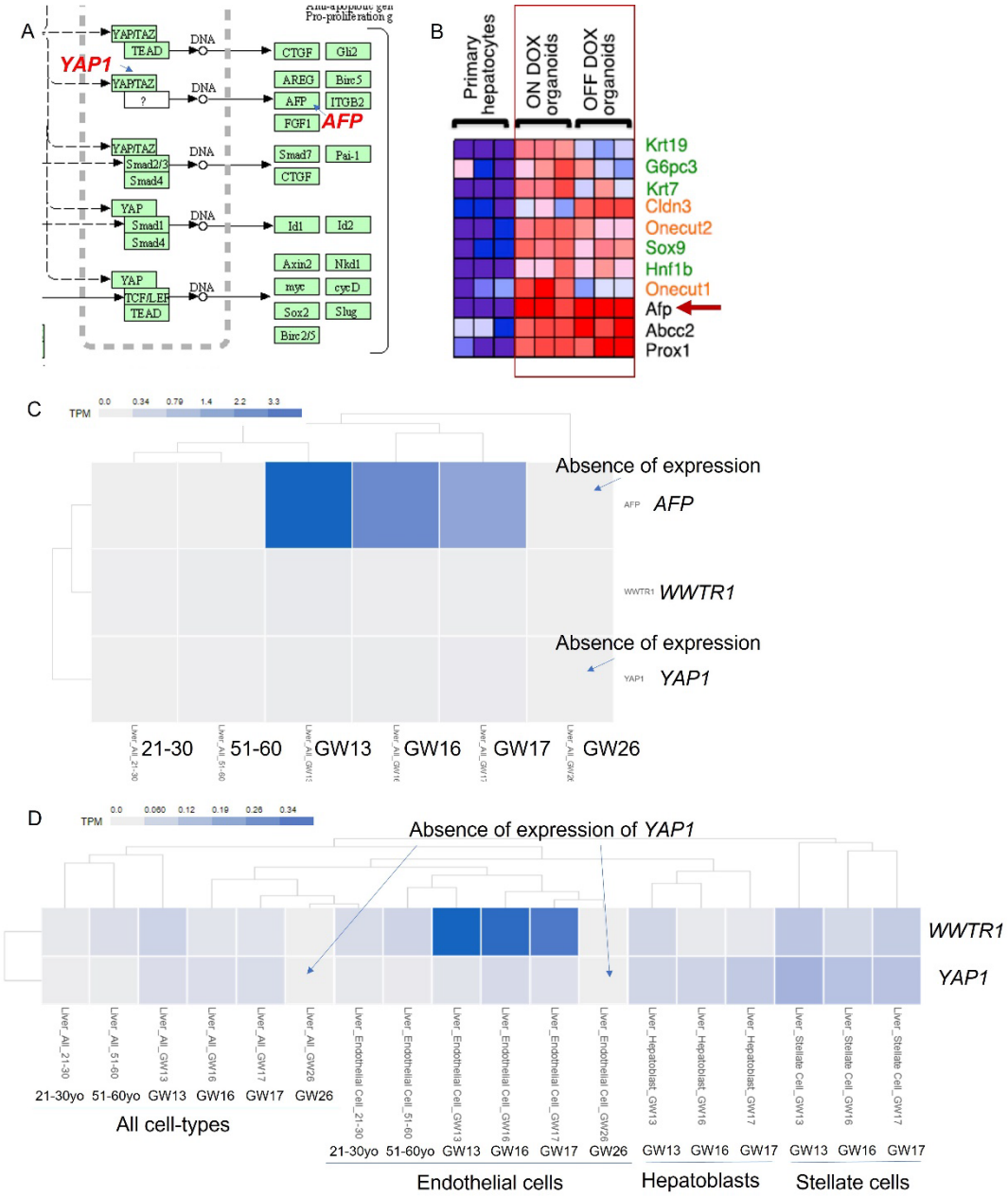
Supplementary Figure S7. Demonstration of the 'Co-expression' domain. (A) Entering the 'Co-expression' page. (B) Selection of species, tissue-type, cluster-type and entering the first gene (*AFP*) and (C) Entering another gene (*ALB*) and performing analysis with clicking on 'search'. (D) Co-expression profiles of *AFP* and *ALB* genes at

cell-type level and (E) at timepoint level.

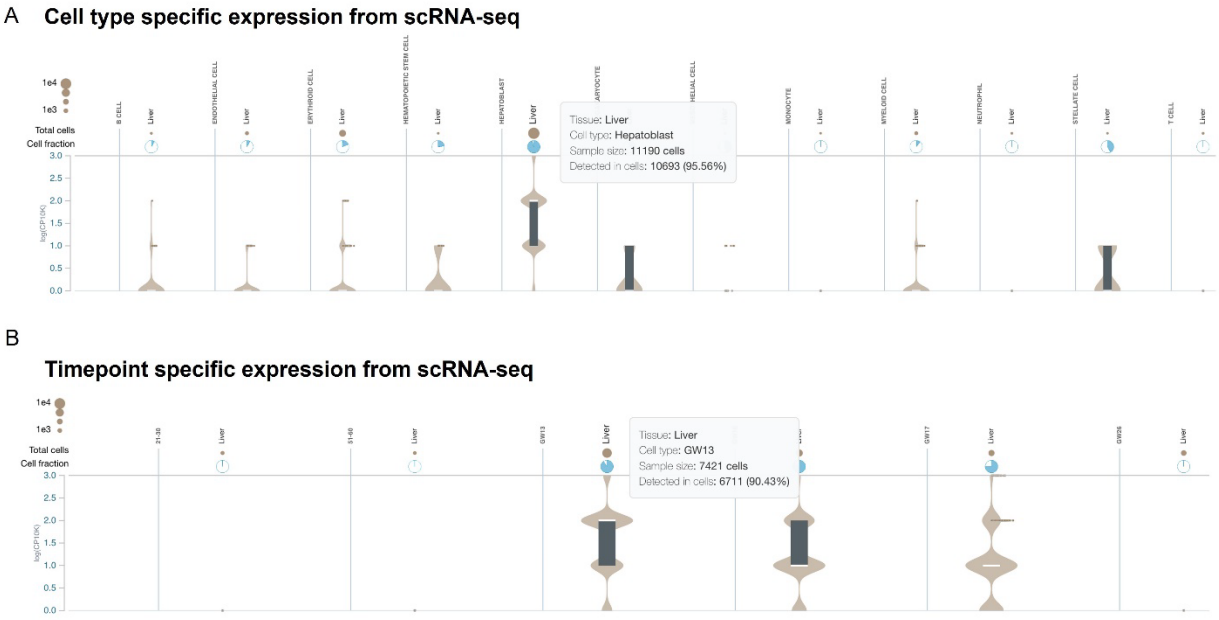


Supplementary Figure S8. Demonstration of the 'Temporal & Dynamic' domain.

(A) Operations and Screenshot of performing enquiry of 'Temporally Regulated Genes'. (B) Distribution of 'Temporally Regulated Genes among Chromosomes'. (C) Performing enquiry for 'Stably Expressed Genes'.



Supplementary Figure S9. Regulation of AFP by YAP. (A) Partial landscape of the Hippo Signaling pathway (hsa04390). It indicates YAP/TAZ are likely the upstream regulators for AFP. (B) RNA-expression results (7). Microarray gene expression comparison analysis of biliary and hepatoblast markers in primary hepatocytes and ± Dox YAP Tg organoids. Orange, green, and black letterheads represent HPC-, biliary- and hepatoblast correlated genes, respectively. (C) Expression clustering of genes *AFP1*, *WWTR1* and *YAP1*, and (D) expression clustering of gene *WWTR1* and *YAP1* only. Absence of expression is indicated by each arrow. A color scale is provided in figures (C) and (D).



Supplementary Figure S10. Cell-type and timepoint specific expression patterns. (A) Cell type specific and (B) timepoint specific expressions of AFP gene. Each blown circle indicates the number of cells in each cell-type/timepoint, while the pipechat shows the fraction of cells with AFP expression identified.

References:

1. 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**.
2. 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.
3. van den Brink, S.C., Sage, F., Vértesy, Á., Spanjaard, B., Peterson-Maduro, J., Baron, C.S., Robin, C. and van Oudenaarden, A. (2017) Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat Methods*, **14**, 935-936.
4. Grindberg, R.V., Yee-Greenbaum, J.L., McConnell, M.J., Novotny, M., O'Shaughnessy, A.L., Lambert, G.M., Araúzo-Bravo, M.J., Lee, J., Fishman, M., Robbins, G.E. *et al.* (2013) RNA-sequencing from single nuclei. *Proc Natl Acad Sci U S A*, **110**, 19802-19807.
5. Habib, N., Li, Y., Heidenreich, M., Swiech, L., Avraham-Davidi, I., Trombetta, J.J., Hession, C., Zhang, F. and Regev, A. (2016) Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. *Science*, **353**, 925-928.
6. Wu, H., Kirita, Y., Donnelly, E.L. and Humphreys, B.D. (2019) Advantages of Single-Nucleus over Single-Cell RNA Sequencing of Adult Kidney: Rare Cell Types and Novel Cell States Revealed in Fibrosis. *J Am Soc Nephrol*, **30**, 23-32.
7. 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.