Published online 27 July 2021

Nucleic Acids Research, 2021, Vol. 49, No. 18 ***e108***
https://doi.org/10.1093/nar/gkab629

TSMiner: a novel framework for generating time-specific gene regulatory networks from time-series expression profiles

Mingfei Han [1,†], Xian Liu [1,†], Wen Zhang [1,2,†], Mengnan Wang [1], Wenjing Bu [1], Cheng Chang [1], Miao Yu [1], Yingxing Li [3], Chunyan Tian [1], Xiaoming Yang [1,*], Yunping Zhu [1,*] and Fuchu He [1,*]

[1]State Key Laboratory of Proteomics, Beijing Institute of Lifeomics, National Center for Protein Sciences (Beijing), Beijing 102206, P.R. China, [2]Tianjin Key Laboratory of Food Science and Biotechnology, School of Biotechnology and Food Science, Tianjin University of Commerce, Tianjin 300134, China and [3]Central Research Laboratory, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China

Received April 01, 2020; Revised June 30, 2021; Editorial Decision July 05, 2021; Accepted July 09, 2021

ABSTRACT

Time-series gene expression profiles are the primary source of information on complicated biological processes; however, capturing dynamic regulatory events from such data is challenging. Herein, we present a novel analytic tool, time-series miner (TSMiner), that can construct time-specific regulatory networks from time-series expression profiles using two groups of genes: (i) genes encoding transcription factors (TFs) that are activated or repressed at a specific time and (ii) genes associated with biological pathways showing significant mutual interactions with these TFs. Compared with existing methods, TSMiner demonstrated superior sensitivity and accuracy. Additionally, the application of TSMiner to a time-course RNA-seq dataset associated with mouse liver regeneration (LR) identified 389 transcriptional activators and 49 transcriptional repressors that were either activated or repressed across the LR process. TSMiner also predicted 109 and 47 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways significantly interacting with the transcriptional activators and repressors, respectively. These findings revealed the temporal dynamics of multiple critical LR-related biological processes, including cell proliferation, metabolism and the immune response. The series of evaluations and experiments demonstrated that TSMiner provides highly reliable predictions and increases the understanding of rapidly accumulating time-series omics data.

INTRODUCTION

Biological processes are often dynamic, and time-series expression profiles are the most abundant sources of information for such dynamic activity (1). Capturing underlying regulatory mechanisms from time-series gene expression data can help broaden our understanding of complex biological processes, including the cell cycle, stem cell differentiation and disease progression. Clustering methods (2–5) have been extensively applied to time-series data for module detection. Genes included in the same module likely share common regulators or participate in similar biological pathways; hence, clustering is often accompanied by enrichment analysis to identify the regulators or pathways significantly associated with each gene module. However, clustering methods cannot evaluate the temporal dynamics of regulators or pathways.

To overcome this limitation, Ernst *et al.* (6) developed Dynamic Regulatory Events Miner (DREM), which is capable of training a time-series expression matrix into a tree-structure model comprising multiple bifurcation points when two or more gene sets abruptly diverge from the same expression level. Gene sets divided from a bifurcation point are speculated to be derived from different regulatory events. Thus, TSMiner further predicts the transcription factors (TFs) involved in regulating each subset of genes. Multiple methods have been subsequently developed to expand the functions of DREM. For example, interactive DREM (iDREM) (7) provides support for additional data

*To whom correspondence should be addressed. Tel: +86 10 61777058; Fax: +86 10 61777058; Email: zhuyunping@gmail.com
Correspondence may also be addressed to Xiaoming Yang. Tel: +86 10 66931428; Email: xiaomingyang@sina.com
Correspondence may also be addressed to Fuchu He. Tel: +86 10 61777001; Email: hefc@nic.bmi.ac.cn
†The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

types (e.g. miRNA expression, proteomics, epigenomics and single-cell RNA-seq data). Additionally, Signalling and Dynamic Regulatory Events Miner (SDREM) (8) links the TFs predicted by DREM to a cascaded network. Although iDREM and SDREM provide additional capabilities, they follow the core algorithms of DREM, which remains the only method for capturing time-specific regulators from time-series data to date.

However, DREM and its extended methods have three limitations. First, DREM uses a version of the Baum–Welch algorithm to train the parameters of tree-structure models. The Baum–Welch training algorithm seeks local maxima that are highly dependent on initial parameters; however, DREM lacks an effective parameter initialization method, limiting the accuracy and robustness of the resulting model. Second, DREM uses a hypergeometric test to identify TFs whose target genes (TGs) are significantly enriched in subpaths derived from a specific bifurcation point; however, this method is limited in its ability to determine when these TFs exert their influence on the targets. Finally, SDREM can combine the predicted TFs into a network covering the entire time-series process but cannot distinguish subnetworks occurring in specific periods. Therefore, methods to construct time-specific gene regulatory networks from time-series expression data are still required.

To address these issues, we present a novel computational framework called the time-series miner (TSMiner) that integrates three types of input files [(i) a time-series expression matrix, (ii) TF–TG interaction information and (iii) gene sets of biological pathways] and constructs two TF–pathway interaction networks (one activated and one repressed) for each time point. TSMiner presents three key innovations to address the limitations of DREM. First, TSMiner uses a novel expression characteristic called the 'bifurcation pattern' to infer initial parameters for the Baum–Welch algorithm. A bifurcation pattern represents two groups of genes diverging from the same expression level at a given time point; for subsequent time points, they include activities where one group of genes is upregulated and the other is downregulated or vice versa. Compared with the bifurcation points used by DREM, those used by TSMiner consider expression information from more time points and provide better initial parameters for the Baum–Welch algorithm, allowing improved training from time-series data to generate a better tree-structure model. Second, in addition to using the hypergeometric test to identify significantly enriched TFs for each subpath, TSMiner applies a permutation test to determine TF activation/repression time points. Third, TSMiner can predict pathways showing significant mutual interactions with the TFs activated/repressed at a specific time to construct time-specific TF–pathway interaction networks. These networks provide not only predictions of dynamic TF pathway regulatory events but also insight into new TF–pathway interactions.

To evaluate the core functions of TSMiner, TF and pathway prediction, we downloaded an RNA-seq dataset associated with liver regeneration (LR) in mice following partial hepatectomy (PH) (9) and collected 113 TFs and 58 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways involved in LR by text mining and manual curation. By applying TSMiner, DREM (v. 2.0) and four widely used clustering tools for LR RNA-seq data, we demonstrated the highest sensitivity and accuracy of TSMiner in detecting known LR-associated TFs and pathways. Furthermore, TSMiner promoted a deeper understanding of LR by discovering a novel cascade of immune-regulatory pathways activated during the late stage of LR, followed by experimental validation. These results demonstrated the efficacy of TSMiner for supporting the deep mining of rapidly accumulating time-series omics datasets.

MATERIALS AND METHODS

Workflow of the analytical framework

TSMiner accepts the following three types of input data: (i) time-series expression data; (ii) TF–gene interaction data (from chromatin immunoprecipitation experiments or bioinformatics prediction) and (iii) pathway gene sets from the KEGG (https://www.genome.jp/kegg/pathway.html) or Reactome (https://reactome.org/) pathway database. TSMiner has modules with the following four key functions: (i) dividing a time-series expression matrix based on bifurcation patterns; (ii) predicting TFs undergoing time-specific activation/repression; (iii) expanding the time-specific TFs to upstream or downstream pathways and (iv) interactively visualizing the results (Figure 1).

Dividing a time-series expression matrix according to bifurcation patterns. Given a time-series expression matrix $M = (G, T, v)$, where G represents a set of genes, $T = \{t_1, t_2, t_3, t_4\}$ represents an ordered set of continuous time points, and $v : G \times T$ represents the expression-value matrix, we developed an expression pattern classifier capable of searching a complete set of bifurcation patterns and allowing the division of the expression matrix, M, into two submatrices. Using the division of M at time t_1 as an example (Supplementary Figure S1), we first calculated the fold changes of all genes between t_1 and t_2 and divided the genes into the following three classes accordingly: upregulated, downregulated and unchanged. We then compared the expression values of any two genes, g_1 and g_2, coming from two different classes at time points t_2, t_3 and t_4, generating a three-dimensional vector, nc (for $j \in \{t_2, t_3, t_4\}$, when $v_{1,j} > v_{2,j}$, $nc_j = 1$; otherwise, $nc_j = 0$). Notably, we ensured that the first element of nc was zero by exchanging the order of g_1 and g_2. Thus, nc had four possible values: (0, 0, 0), (0, 0, 1), (0, 1, 0) and (0, 1, 1). Each possible value of nc was defined as one bifurcation pattern of M starting at time t_1. Based on each bifurcation pattern, we divided the gene set G into two subsets based on the likelihood ratio test (see Section 1.1 in the Supplementary Methods).

To train the time-series expression matrix, M, into a tree-structure model comprising an optimal set of bifurcation patterns, TSMiner first models M into a single chain of Gaussian distributions based on an input–output hidden Markov model (6,10). This chain is considered the first path and is repeatedly divided into a tree structure in four steps: (i) assign all genes in M to the current paths using the Viterbi algorithm (11); (ii) divide each path in the current model

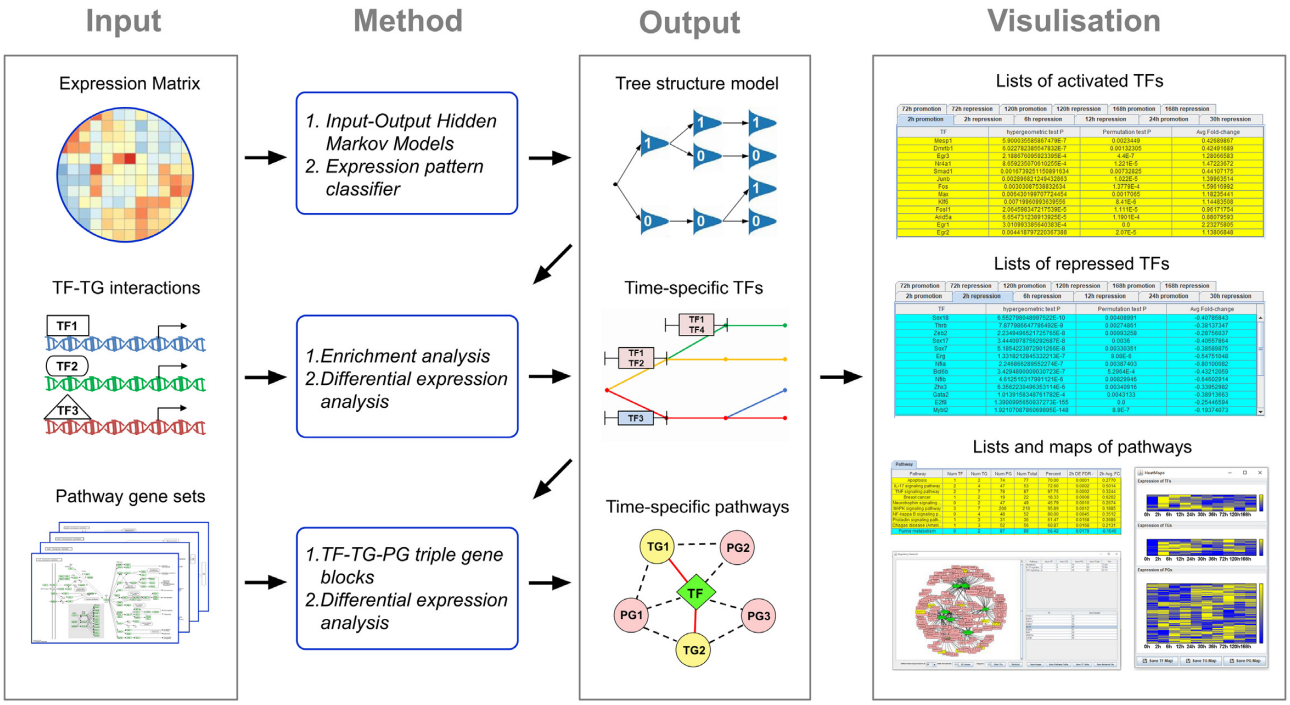

Figure 1. Overview of the TSMiner workflow. TSMiner accepts time-series expression profiles, TF–TG interaction data and pathway gene sets. The time-series gene expression matrix is first repeatedly divided to generate a tree-structure model comprising multiple bifurcation patterns, followed by identification of activated or repressed TFs with targets that are significantly enriched in an expression pattern and significantly upregulated/downregulated at a specific time. Time-specific TFs are then used to infer significantly associated biological pathways. TSMiner enables visualization of the results, including tables of the TFs/pathways activated or repressed at different time points, network maps showing TF–pathway interactions and heat maps showing TF and pathway gene expression levels.

based on a complete set of bifurcation patterns, calculate the initial parameters for each division and train the parameters by the Baum–Welch algorithm; (iii) update the current model to one that can achieve the best likelihood score and (iv) repeat steps 1–3 until the likelihood score converges (see Supplementary Figure S2 and Section 1.2 in the Supplementary Methods).

Predicting time-specific TFs. In the final tree-structure model, we first used the Viterbi algorithm to assign all genes in M to the most likely subpaths, followed by applying the hypergeometric test to identify TFs significantly regulating each of the subpaths (Supplementary Figure S3). The P-values were corrected for multiple testing using Benjamini–Hochberg (BH) correction. We considered that a TF, f, significantly regulating a subpath may exert its influence on the targets at any time point from the bifurcation point to the last time point. Therefore, we calculated the activation scores of f for each time point in the subpath according to the expression changes of its targets. The significance of the activation score was estimated using a permutation test, the P-values from which were corrected by BH correction. Finally, the TF f was assigned to the time points upon receiving significant permutation test P-values and subsequently classified as activated or repressed based on the average fold change of its targets. Notably, TSMiner can distinguish positively and negatively correlated interactions in TF–TG interaction data. A TF showing a positive correlation with its targets is identified as activated/repressed when its targets are significantly upregulated/downregulated, whereas the opposite is true for negative correlation.

Expanding time-specific TFs to upstream or downstream pathways. Assuming that TF tf_1 was identified as significantly activating four target genes, $TG = \{g_1, g_2, g_3, g_4\}$, at time point t_0, given a pathway comprising five genes, $P_1 = \{g_1, g_2, g_6, g_7, g_8\}$, we aimed to verify whether P_1 was significantly correlated with tf_1 and significantly upregulated at time point t_0 along with the activation of tf_1 (Supplementary Figure S4). We first divided P_1 into two parts: one containing overlapping genes with TG ($PG_1 = \{g_1, g_2\}$) and another containing the other genes in P_1 ($PG_2 = \{g_6, g_7, g_8\}$). We then built two TF–TG regulatory pairs using tf_1 and the genes in PG_1 (tf_1-g_1 and tf_1-g_2). Using these two TF–TG pairs and genes in PG_2, we built six TF–TG–PG triple-gene blocks (tf_1-g_1-g_6, tf_1-g_1-g_7, tf_1-g_1-g_8, tf_1-g_2-g_6, tf_1-g_2-g_7 and tf_1-g_2-g_8). Wei *et al.* (12) developed the mutual interaction measure (MIM), a valuable scoring method for evaluating triple-gene mutual interactions. Therefore, we calculated the MIM for each triple-gene block and then used a permutation procedure to determine the triple-gene blocks with significant MIMs (see Section 1.5 in the Supplementary Methods for details). A significant TF–TG–PG gene block indicates that the TG is directly regulated by the TF, whereas the PG is indirectly correlated with the TF through the TG.

As shown in Supplementary Figure S4, $(tf_1 - g_1 - g_5)$ and $(tf_1 - g_2 - g_7)$ were identified as significant triple-gene

blocks, indicating that g_5 and g_7 were indirectly correlated with tf_1 through g_1 and g_2, respectively. Because gene sets $PG_1 = \{g_1, g_2\}$ and $PG_2 = \{g_5, g_7\}$ included genes directly and indirectly interacting with tf_1, respectively, we considered the number of genes present in the union of PG_1 and PG_2 to be the interaction score between tf_1 and pw_1 ($S = 4$). Additionally, we calculated the interaction probability between tf_1 and pw_1 as their interaction score divided by the total number of genes involved in pw_1 ($S/5 = 0.8$). We considered pw_1 to be significantly correlated with tf_1 when both the interaction score and probability were higher than the user-defined thresholds. The default thresholds for the interaction score and probability are 50 and 50%, respectively. For a pathway passing the two abovementioned cut-offs, we employed a permutation test to verify whether the genes involved in the pathway showed significant upregulation (downregulation) at the time point when their interacting TF was significantly activated (repressed).

Interactive visualization of the results. TSMiner provides two lists for each time point to display activated and repressed TFs. Users can select one or more items in a TF list and predict significantly associated pathways. The resulting pathways are then displayed in a list, allowing users to select one or more items for generating heat maps or network maps (see Supplementary Figure S5 and Section 1.6 in the Supplementary Methods).

Data collection and preprocessing

We downloaded an RNA-seq dataset associated with LR in mice following PH from the Gene Expression Omnibus (record: GSE95135) (9). Liver samples were collected at 12 time points post-PH along with a negative control group. Because several adjacent time points had intervals that were too small, we used only the samples profiled at 10 of the 12 time points (0, 1, 4, 10, 20, 36, 48, 72, 168 h and 4 weeks). Among the initial 37 991 genes profiled, we used only 22 707 protein-coding genes and filtered those with more than one missing value or those in which all the absolute fold changes between adjacent time points were < 1 ($\forall\, t > 0,\ |\log_2 v_t - \log_2 v_{t-1}| < 1$), resulting in 6253 genes.

We then constructed TF–TG interaction data for integration with the LR RNA-seq data. We collected 133 599 TF–TG interactions for 1222 TFs that might function in the mouse liver according to CellNet (13), including 119,461 positively correlated interactions and 14 138 negatively correlated interactions. We further excluded TFs that were not detected in the target RNA-seq data, consequently leaving 1027 TFs interacting with 13 578 and 5957 TGs through 87,386 positively and 13 395 negatively correlated interactions, respectively (Supplementary Table S1).

Searching known LR-associated TFs

Among the 1027 TFs that remained after preprocessing, we searched for those involved in LR by text mining. First, we used BANNER (14) to search 31 624 408 abstracts in MEDLINE with the keywords 'liver regeneration' and 'partial hepatectomy', which provided 186 274 sentences associated with LR. We found that 163 of the 1027 TFs appeared in these sentences and ultimately selected 113 TFs reportedly associated with LR through manual curation. The gene symbol, related text and PubMed ID of the 113 known LR-associated TFs are described in Supplementary Table S1.

Searching known LR-associated KEGG pathways

We first constructed a background pathway set containing 499 KEGG pathways downloaded from the NCBI BioSystems Database (https://www.ncbi.nlm.nih.gov/biosystems). In the background pathway set, we searched for the pathways associated with LR in three steps: (i) we obtained 186 274 sentences using BANNER (14) to search abstracts in MEDLINE using the keywords 'liver regeneration' and 'partial hepatectomy'; (ii) we identified 100 pathways that were mentioned in those sentences and included in the background pathway set and (iii) we selected 58 KEGG pathways reportedly associated with LR through manual curation. The name, related texts and PMIDs of the 58 LR-associated pathways are shown in Supplementary Table S2.

Mice experiments

We conducted a series of experiments to validate the immune cascade activated during the late stage of mouse LR predicted by TSMiner. First, we employed flow cytometry experiments to study the number of leukocytes, neutrophils, macrophages and T cells in mouse liver post-PH. Second, we detected the levels of several chemokines that recruit neutrophils, macrophages and T cells, respectively. Third, we performed reverse transcription-quantitative polymerase chain reaction (RT-qPCR) to evaluate the transcription levels of multiple immunity-related TFs and pathways during LR. The detailed experimental procedures are provided in the Supplementary Methods. All animal experiments were approved by the Institutional Animal Care and Use Committee of the Beijing Institute of Lifeomics (ID: IACUC-DWZX-2020–572).

RESULTS

Evaluating TF prediction ability

To evaluate TSMiner, we downloaded an RNA-seq dataset associated with mouse LR from the Gene Expression Omnibus (record: GSE95135) (9). We then constructed TF–TG interaction data containing 1027 TFs that might function in the mouse liver according to CellNet (13). Among the 1027 TFs, we further identified 113 TFs reportedly involved in LR by text mining and manual curation. The preprocessing of the RNA-seq data and TF–TG interaction data as well as the collection of known LR-associated TFs are described in the Materials and Methods. Because TSMiner and DREM are the only tools capable of detecting time-specific TFs from time-series expression profiles, we used them to integrate the LR RNA-seq data and TF–TG interaction data, respectively, and compared their performances in detecting known LR-associated TFs. Notably, DREM cannot distinguish between positively and negatively correlated TF–TG regulatory interactions; thus, we used only those that were positively correlated in the TF–TG interaction data, which accounted for 90% of the total interactions.

Moreover, to compare the two methods using multiple parameters, we set the 'Maximum number of subpaths out of a bifurcation point' to four values (3, 4, 5 and unlimited), which generated four tree-structure models from TSMiner and DREM (Supplementary Figures S6 and S7).

In each of the resulting models, we searched the minimum enrichment *P*-value (BH corrected) for each TF among all the subpaths (Supplementary Table S3) and found that TSMiner and DREM obtained lower *P*-values for 66 and 41 known LR-associated TFs, respectively, with the remaining six displaying equal *P*-values (Supplementary Figure S8). We then counted the number of known LR-associated TFs identified as being significant (Supplementary Figure S9). Using *P*-value <0.01 as the significance cut-off, DREM identified 60, 60, 64 and 62 known TFs under different parameter settings, with 48 overlapping TFs. In contrast, TSMiner identified 70, 64, 67 and 66 known TFs under different parameter settings, with 57 overlapping TFs. At *P*-value < 0.05, DREM identified 79, 75, 74 and 82 known TFs (66 overlaps), and TSMiner identified 80, 85, 81 and 82 known TFs (71 overlaps). These findings indicated that TSMiner identified more known LR-associated TFs than DREM (Figure 2A). At *P*-value cut-offs of 0.01 and 0.05, TSMiner uniquely identified 10 and 9 known TFs, respectively, whereas DREM provided only 1 and 4 unique results, respectively (Figure 2B). Additionally, TSMiner maintained superior sensitivity for detecting known LR-associated TFs at a *P*-value cut-off ranging from 0 to 1 (Supplementary Figure S10). Furthermore, TSMiner showed better robustness than DREM because the overlapping TFs accounted for a higher proportion of the TFs identified under different parameters (Figure 2C). These results suggested that the TSMiner-specific application of a bifurcation pattern promoted the training of time-series data into better tree-structure models than those of DREM.

We then evaluated the accuracy of TSMiner and DREM in determining the TF activation time. Taub (15) reported that in the post-PH mouse liver, the rate of DNA synthesis in hepatocytes increased after ∼12 h, peaked at ∼36–40 h and then continuously decreased until 168 h, when the liver mass was restored. This finding suggests the possibility of extracting two time periods, namely, cell proliferation (10–36 h post-PH) and growth suppression (72–168 h post-PH), from the LR RNA-seq data. Additionally, Kurinna *et al.* (16) reported 27 and 10 TFs activated in the cell proliferation and growth suppression periods of LR, respectively (Figure 2D). Therefore, we expected TSMiner and DREM to detect the 27 TFs associated with cell proliferation at 10, 24 or 36 h post-PH and to detect the 10 growth suppressors at 72 or 168 h post-PH. Figures 2E and 2F present the receiver operating characteristic (ROC) curves of TSMiner and DREM for selectively identifying or excluding TFs associated with cell proliferation or growth suppression at their designated periods. We found that TSMiner showed not only a larger area under the ROC curve (AUC) but also a smaller deviation between different parameters. These results showed that TSMiner not only outperformed DREM in detecting true-positive TFs but also in determining their activation time points.

Evaluating pathway prediction

Although some clustering methods can segregate genes into subsets and identify their significantly enriched pathways, only TSMiner assesses the temporal dynamics of pathways from time-series data. Here, we collected 58 KEGG pathways involved in LR (see Materials and Methods) and compared the sensitivity and accuracy of TSMiner with those of clustering methods in detecting known LR-associated pathways without considering their temporal dynamics.

Using each resulting model generated by TSMiner (Supplementary Figure S6), we predicted the pathways activated or repressed at different time points from a background pathway set containing 499 KEGG pathways. Supplementary Table S4 shows the minimum *P*-value for each pathway among all the time points. We then applied four widely used clustering-analysis packages [hclust (R package for hierarchical clustering), *k*-means (R package for *k*-means clustering), STEM (JAR file for clustering short time-series expression data) (17) and WGCNA (R package for identifying co-expressed gene modules) (18)] to the RNA-seq data. We performed enrichment analysis to identify significant pathways (see Section 3 in the Supplementary Methods). The minimum enrichment *P*-values for each of the 499 background pathways are shown in Supplementary Table S4. Compared with the clustering methods, we found that TSMiner displayed consistently and significantly higher sensitivity in detecting known LR-associated pathways at *P*-value thresholds ranging from 0 to 0.05 (Figure 2G). We also examined pathways that were uniquely identified by one method and not detected by the others, finding that TSMiner uniquely identified 20 and 5 known LR-associated pathways at *P*-value cut-offs of 0.01 and 0.05, respectively, whereas the other methods provided almost no unique results (Figure 2H). These results indicated that TSMiner exhibited significantly improved sensitivity in pathway discovery relative to the tested clustering methods.

In many cases, high sensitivity is accompanied by a high false-positive rate. Therefore, we generated a false-positive pathway set that was unlikely to participate in LR and evaluated the ability of TSMiner to exclude false-positive results. Pathways are often interconnected and act as regulatory cascades, with overlapping genes being a feature of interconnected pathways. Therefore, we searched all the pathways either directly or indirectly associated with the 58 known LR-associated pathways in four steps: (i) we built a positive-pathway set containing the 58 LR-associated pathways; (ii) we searched pathways in the parent set with at least one gene overlapping with the current positive-pathway set; (iii) we inserted the pathways identified in step 2 into the positive-pathway set and (iv) we repeated steps 2 and 3 until the positive-pathway set converged. After processing, the positive-pathway set contained 421 pathways possibly involved in LR. After excluding these pathways from the parent set, the remaining 78 pathways in the parent set were highly unlikely to participate in LR. We then generated ROC curves for TSMiner and the clustering methods depending on their ability to detect the 58 LR-associated pathways and exclude the 78 negative pathways, finding that TSMiner showed a significantly higher AUC than the clustering methods (Figure 2I). These results indicated that

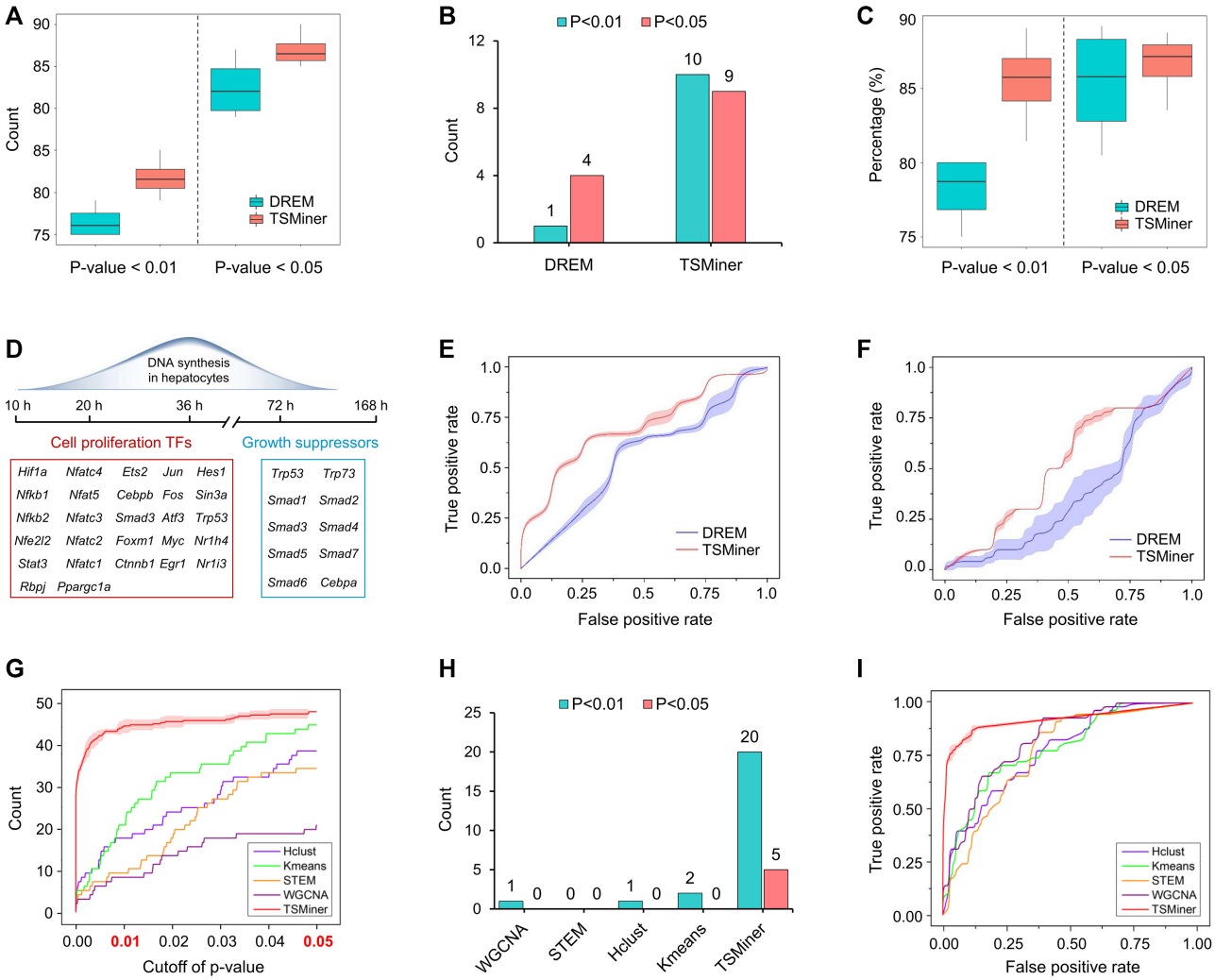

Figure 2. Comparison of TSMiner and DREM for detecting known LR-associated TFs and pathways. (**A**) Box plots for the known LR-associated TFs identified by DREM and TSMiner under different parameters at *P*-value cut-offs of 0.01 and 0.05, respectively. (**B**) Number of known LR-associated TFs uniquely identified by TSMiner or DREM and undetected by the other at *P*-value cut-offs of 0.01 (turquoise cubes) and 0.05 (pink cubes). (**C**) Proportion of overlapping TFs among the TFs identified from different parameters using *P*-value cut-offs of 0.01 and 0.05. (**D**) The 27 TFs known to regulate cell proliferation 10 to 36 h post-PH and the 10 growth suppressors known to function from 72 to 168 h post-PH. (**E**) ROC curves of TSMiner and DREM assessing their ability to identify TFs associated with cell proliferation and excluding those associated with growth suppression from 10 to 36 h post-PH. The curve width represents the standard deviation of multiple parameter settings. (**F**) ROC curves of TSMiner and DREM assessing their ability to identify TFs associated with growth suppression and excluding those associated with cell proliferation from 48 to 168 h post-PH. (**G**) Number of known LR-associated pathways identified by TSMiner and four clustering methods at a *P*-value cut-off ranging from 0 to 0.05. The curve width (for TSMiner) represents the standard deviation of multiple parameter settings. (**H**) Number of known LR-associated pathways uniquely identified by one method and undetected by the others at *P*-value cut-offs of 0.01 (turquoise cubes) and 0.05 (pink cubes). (**I**) ROC curves of TSMiner and clustering methods assessing their ability to identify known LR-associated pathways and excluding those highly unlikely to participate in LR.

TSMiner maintained high sensitivity while also effectively excluding false positives.

Constructing TF–pathway interaction networks across LR

After comparing TSMiner with various state-of-the-art methods, we evaluated whether TSMiner could detect important regulators and pathways in LR and accurately determine their activation/repression time points. Among the resulting models obtained from different parameters, we selected the model with the parameter 'Maximum number of subpaths out of a bifurcation point' set as 3 (Supplementary Figure S6A), which performed the best in detecting both known TFs and pathways in LR at a *P*-value cut-off of 0.01 (Supplementary Figures S9B and S11A). Because TSMiner can distinguish between positively and negatively correlated TF–TG regulatory interactions, we evaluated the results obtained from the positive and negative interactions separately.

Using positively and negatively correlated TF–TG interactions, we obtained 389 transcriptional activators and 49 transcriptional repressors that were either activated or repressed at different time points (Supplementary Table S5). The activators were mostly repressed in the early and late periods (1–4 h and 72 h to 4 weeks post-PH) and activated in the middle period (10, 20, 36 and 48 h post-PH), whereas

the repressors showed exactly opposite temporal characteristics (Supplementary Figures S12A and B). These findings indicated that the activators and repressors had synergistic effects on their TGs because gene upregulation can result from both activation of their positive regulators and repression of their negative regulators and vice versa. We also predicted 109 and 47 KEGG pathways that showed significant mutual interactions with the activators and repressors, respectively, and were significantly activated or repressed during LR (Supplementary Tables S6 and S7). The pathways interacting with the activators and repressors showed a similar repression–activation–repression trend during LR (Supplementary Figure S12C and D) that resembled the temporal characteristics of the activators and differed from those of the repressors. This finding was consistent with the liver possibly shutting down multiple biological processes to preserve cell survival following PH, subsequently reactivating them following liver tissue regeneration and preventing their activation after liver mass restoration (19,20).

The resulting pathways were classified into multiple classes according to the KEGG Orthology, including metabolism, genetic information processing, cellular processes and organismal systems. In the present study, we focused on the temporal dynamics of the pathways associated with metabolism, cell proliferation and the immune system. First, we found that all three types of pathways showed a similar repression–activation–repression trend during LR (Figures 3A and B, Supplementary Figures S13A and B). Second, all the pathways interacting with the repressors were included in those interacting with the activators, which may be due to the limited number of negative TF–TG interactions. Third, the genes involved in these pathways tended to show common upregulation/downregulation with the activators interacting with them at the predicted activation/repression time points (heat maps in Figure 3C) but displayed opposite expression changes with the repressors interacting with them (heat maps in Supplementary Figure S13C). These findings were consistent with the positive and negative regulatory effects of the two types of TFs in their interacting pathways, indicating the reliability of the predictions that we obtained from both the positive and negative TF–TG interactions.

We then explored the most representative pathways associated with metabolism, cell proliferation and the immune system (Figure 3C and Supplementary Figure S13C). First, numerous metabolic pathways, including energy metabolism (e.g., oxidative phosphorylation), amino acid metabolism (e.g. arginine and proline metabolism), carbohydrate metabolism (e.g. glycolysis/gluconeogenesis), lipid metabolism (e.g. steroid hormone biosynthesis) and nucleotide metabolism (e.g. purine metabolism and pyrimidine metabolism), were suppressed from 1 to 4 h post-PH and reactivated after 10 h. These findings were consistent with disrupted metabolism preceding hepatocellular proliferation and ultimately being resolved with liver restoration (20). Second, pathways associated with hepatocellular proliferation are divided into three classes: (i) genetic information processing (e.g. DNA replication and mismatch repair), (ii) cell growth (e.g. cell cycle) and (iii) cell death (e.g. p53 signalling pathway and apoptosis). We found that the genetic information processing pathways were suppressed from 1 to 4 h post-PH, which may be due to the disruption of nucleotide metabolism. Then, genetic information processing and cell growth were both reactivated during 10 to 36 h post-PH, consistent with the rate of DNA synthesis in mouse hepatocytes peaking at ∼36 h post-PH (15). The subsequent level of genetic information processing began to decrease at 72 h post-PH, but that of cell growth decreased later, at 168 h. Third, immunity-related pathways were activated across the entire LR process (Figure 3A), which included known mechanisms, such as activation of the IL-17 signalling pathway, Toll-like receptor signalling pathway, NOD-like receptor signalling pathway and the downstream MAPK signalling pathway during the priming phase of LR (21–23). However, TSMiner also provided novel findings, including a cascade of immune pathways activated from 36 to 168 h post-PH. Finally, the expression of all three types of pathways was restored to normal levels following the restoration of liver mass at 4 weeks post-PH.

Immune cascade during late-stage LR

Apoptosis and apoptotic cell clearance, which may regulate tissue homeostasis, are involved in LR (24,25); however, the detailed immune responses and their activation time in LR remain elusive. In this study, we identified significant activation of nine immunity-related pathways in the late stage of LR: two pathways (i.e. p53 signalling pathway and apoptosis) were activated at 36 h post-PH; five pathways (i.e. chemokine signalling pathway, leukocyte transendothelial migration, Fc gamma R-mediated phagocytosis, endocytosis and phagosome) were activated at 48 h post-PH; and two pathways (i.e. cell adhesion molecule and Th1 and Th2 cell differentiation) were activated at 168 h post-PH.

These pathways were closely interrelated and composed a cascade process involving phagocytes and T cells (Figure 4), which might provide important insights into the immune responses in LR. At 36 h post-PH, the p53 signalling pathway and apoptosis were involved in a cascade process (Supplementary Figure S14) that mediated cell cycle arrest and the death of hepatocytes or other cells (26,27). At 48 h post-PH, the chemokine signalling pathway and leukocyte transendothelial migration pathway (Supplementary Figure S15) were activated to recruit leukocytes and induce leukocyte migration from the blood to the liver tissue (28,29). Three phagocytosis-related pathways (i.e. Fc gamma R-mediated phagocytosis, endocytosis and phagosome) were then activated in the leukocytes that migrated to the liver (Supplementary Figure S16) and mediated engulfment of the apoptotic bodies generated at 36 h (30–32). At 168 h post-PH, the $CD4^+$ T cells received cues from the phagocytes and differentiated into T helper type 1 (Th1) and Th2 cells with the help of the cell adhesion molecule pathway and Th1 and Th2 cell differentiation pathway (Supplementary Figure S17) (33,34). In summary, we reconstructed an immunity-related process that likely played important roles in terminating LR and promoting liver homeostasis restoration.

We further studied the TFs that significantly interacted with the abovementioned pathways. First, the two apoptosis-related pathways significantly interacted with

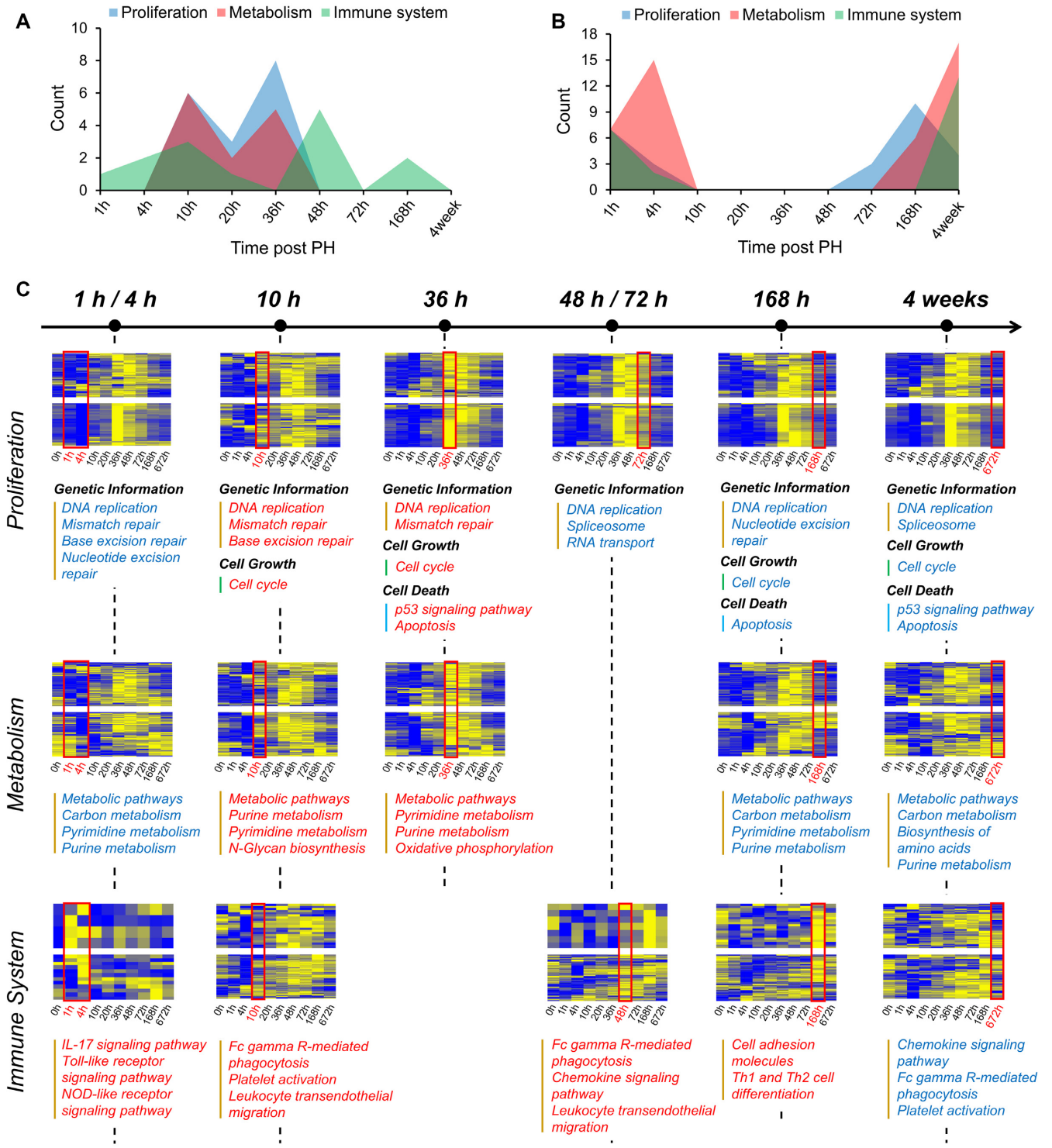

Figure 3. Overview of the KEGG pathways interacting with transcriptional activators. Number of pathways related to cell proliferation, metabolism and the immune system and (**A**) activated or (**B**) repressed at different time points post-PH. (**C**) Representative pathways activated (red) or repressed (blue) at different time points. The upper half of each heat map shows the expression levels of TFs associated with the representative pathways. The lower half of each heat map shows the expression levels of genes involved in the representative pathways. Activation or repression time points are marked in red boxes.

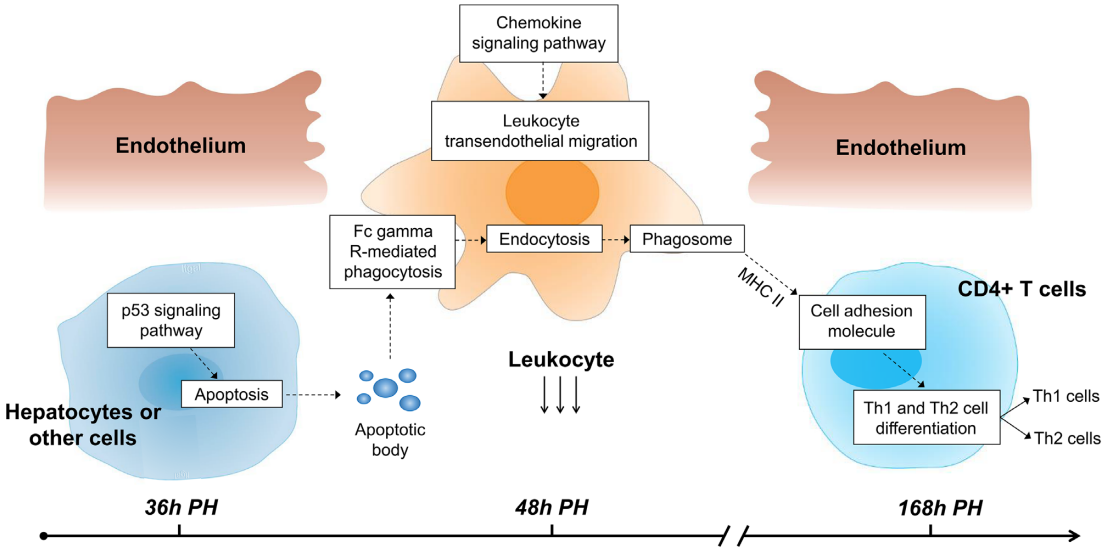

Figure 4. Schematic of the immunity-related cascade activated in LR from 36 to 168 h post-PH. Two pathways related to cell apoptosis were activated at 36 h post-PH, and five pathways related to leukocyte migration into the liver and phagocytosis were activated at 48 h post-PH. In addition, two pathways related to T-cell differentiation were activated at 168 h post-PH.

61 and 38 TFs, with 33 overlapping TFs (Figure 5A). Second, 10, 9, 8, 9 and 11 TFs significantly interacted with the 5 pathways activated in the leukocytes, with 7 overlapping TFs (Figure 5B). Finally, 21 and 18 TFs significantly interacted with the cell adhesion molecule pathway and Th1 and Th2 cell differentiation pathway, respectively, with 17 overlapping TFs (Figure 5C). These results indicated that the pathways involved in the same regulatory cascade were likely to share common regulators, consistent with our experience. Moreover, the gene expression heat maps (Figures 5D–F) revealed that most of the pathways that we predicted and the TFs significantly interacting with them showed common upregulation at the predicted activation time points.

We performed a series of experiments to validate the immune pathways and their interacting TFs (see Supplementary Methods). First, we examined the number of leukocytes per gram of mouse liver post-PH by flow cytometry and found that the number of leukocytes in the regenerating liver peaked at 48 h post-PH (Supplementary Figure S18A). Analysis of the neutrophil and macrophage counts among the leukocytes revealed significant increases in both at 48 h post-PH (Supplementary Figures S18B–E). Additionally, detection of the levels of chemokine (C–X–C motif) ligand 1 (CXCL1; a major chemotactic factor for neutrophils) (35), monocyte chemoattractant protein-1 (MCP-1; a chemotactic factor that regulates monocyte/macrophage migration into the liver) (36) and macrophage inflammatory protein-1α (MIP-1α; a chemotactic factor involved in neutrophil and macrophage recruitment) (37) in liver homogenates indicated that they all peaked at 48 h post-PH (Supplementary Figures S18F–H). These results strongly supported TSMiner predictions of activated leukocyte recruitment and phagocytosis at 48 h post-PH. Moreover, we identified significant increases in the number of T cells, including $CD4^+$ and $CD8^+$ T cells, as well as the levels of their chemotactic factor, RANTES (CCL5) (38), at 168 h post-PH (Supplementary Figure S19), further demonstrating that the immunity-related cascade involving phagocytes and T cells might play key roles in the late stage of LR.

We also performed RT-qPCR to validate the pathways activated at 48 h post-PH (i.e. the chemokine signalling pathway, Fc gamma R-mediated phagocytosis, endocytosis and phagosome) as well as the TFs showing shared associations with them (Figures 5B and E). We randomly selected eight genes from each of the abovementioned pathways and used RT-qPCR to evaluate their transcription levels during LR. The expression of eight, seven, seven and eight of the selected genes showed significant upregulation at 48 h compared to 36 h post-PH, and all the selected genes were significantly upregulated at 48 h post-PH relative to the control group (Supplementary Figures S20–S23). Additionally, we selected three TFs (Ikzf1, Sp100 and Trim30a) that shared associations with these pathways and were previously unreported as being related to LR and evaluated their transcription levels, as well as those of eight randomly selected TGs for each TF, by RT-qPCR. The results showed that the three TFs and nearly all their TGs were significantly upregulated at 48 h post-PH (Supplementary Figures S24–S26). In summary, the pathways related to phagocytosis and their correlated TFs predicted by TSMiner likely play critical regulatory roles in the regenerating liver at approximately 48 h post-PH. All the experimental results are shown in Supplementary Tables S8–S12.

DISCUSSION

Our evaluation of TSMiner using mouse LR data revealed its potential efficacy; however, some issues require further investigation. First, fewer results were obtained from the negative TF–TG interactions than from the positive interactions because the known repressors of many genes were limited. Future collection of negatively correlated TF–TG interactions may help address this issue. In addition to inte-

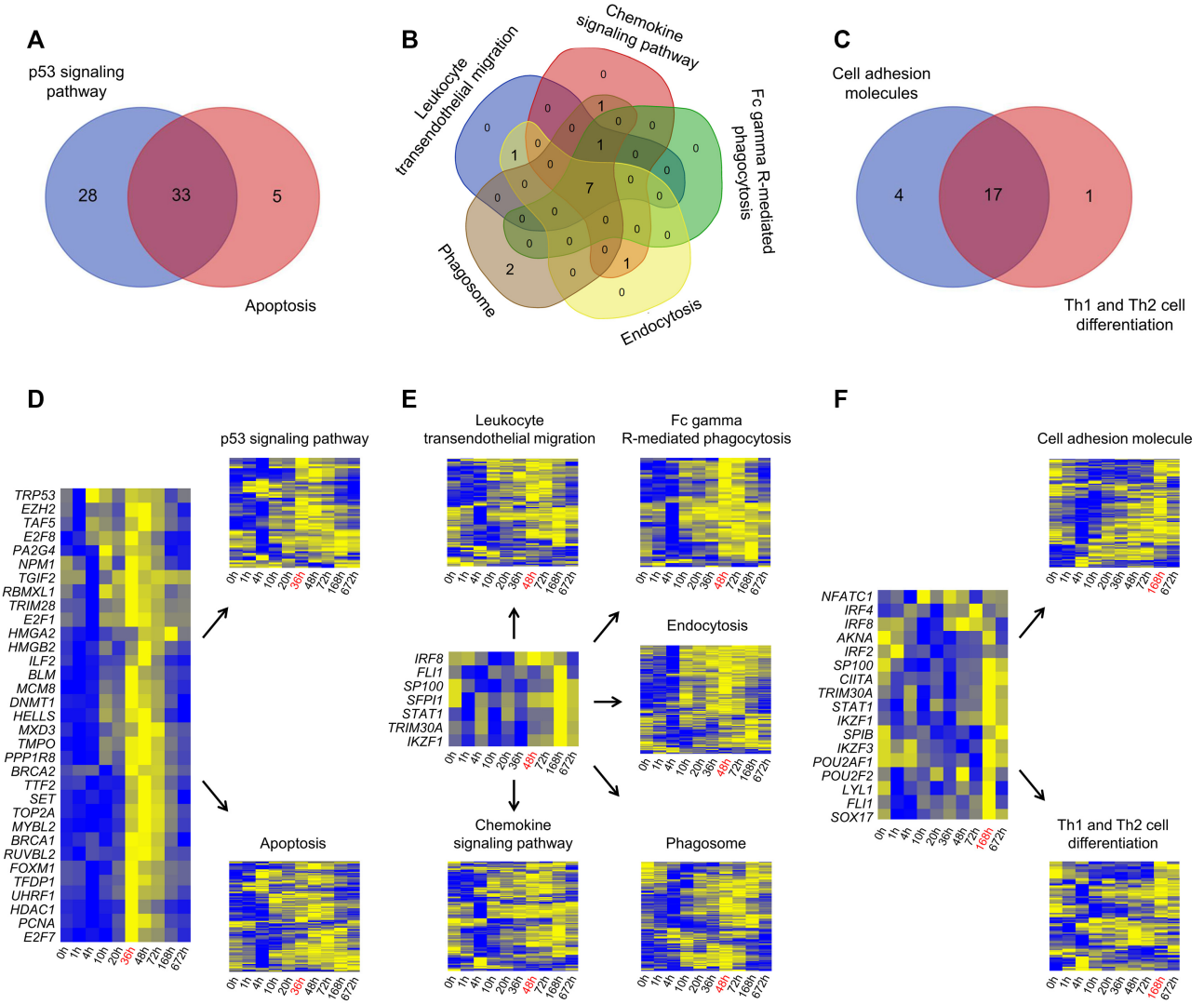

Figure 5. Venn diagrams of the TFs interacting with (**A**) the two pathways activated at 36 h post-PH, (**B**) the five pathways activated at 48 h post-PH and (**C**) the two pathways activated at 168 h post-PH. (**D**) Heat maps of the two pathways activated at 36 h post-PH and the 33 TFs showing shared associations with the two pathways. (**E**) Heat maps of the five pathways activated at 48 h post-PH and the seven TFs showing shared associations with the five pathways. (**F**) Heat maps of the two pathways activated at 168 h post-PH and the 17 TFs showing shared associations with the two pathways.

grating time-series gene expression profiles and TF–TG interaction information, TSMiner may be applicable to other types of time-series omics data, such as in integrating proteomics time-series data and protein–protein interaction (PPI) information to construct PPI networks activated or repressed at different time points. This will be an interesting area for future research. Finally, although all the experimental results strongly supported our predictions, the precise regulatory mechanisms require more rigorous evaluation.

With the development of sequencing technologies, time-series expression profiles have been rapidly accumulating; however, the development of methods capable of appropriately analysing these datasets has lagged. Here, we present TSMiner, a novel analytic tool that captures dynamic TF pathway regulatory events from time-series expression data. The application of TSMiner to the LR RNA-seq data revealed both known mechanisms of LR and mechanisms that have not been previously reported. Specifically, we discovered an immune response cascade that included cell apoptosis, apoptotic cell clearance and T-cell differentiation. A series of evaluations determined these results to be highly reliable, expanding our knowledge of the LR process. We note that TSMiner is the first tool to provide detailed insight into the role of critical regulatory events in time-series processes and their dynamic activation/repression.

DATA AVAILABILITY

The RNA-seq dataset analysed during the current study is available in the NCBI GEO database with the accession number GSE95135. The flow cytometry data from this publication has been deposited to the FlowRepository database (flowrepository.org) and assigned the identifiers FR-FCM-Z3RJ and FR-FCM-Z3RK. The other experimental data are included in the Supplementary Data. The source codes and the complied tool of TSMiner can be downloaded at https://github.com/free1234hm/tsminer-tool.git.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Author contributions: Y.P.Z., F.C.H. and X.M.Y. conceived and supervised the project. M.F.H. and X.L. designed and implemented the TSMiner software. W.J.B., W.Z. and M.Y. implemented partial hepatectomy and flow cytometry analysis. W.Z. performed cytokine detection and RT-qPCR analysis. M.N.W. and C.Y.T. participated in the evaluation of TSMiner. C.C. and Y.X.L. participated in writing the manuscript. All authors read and approved the final manuscript.

FUNDING

National Key Research Program of China [2016YFB0201702, 2017YFA0505002]; State Key Laboratory of Proteomics [SKLP-K201404]; National Basic Research Program of China [2013CB910801]; Innovation Program [18-163-15-ZT-001-006-01, Z181100004118004]. Funding for open access charge: Key Program for HPC Environment [2016YFB0201702].
Conflict of interest statement. None declared.

REFERENCES

1. Bar-Joseph,Z., Gitter,A. and Simon,I. (2012) Studying and modelling dynamic biological processes using time-series gene expression data. *Nat. Rev. Genet.*, **13**, 552–564.
2. Ramoni,M.F., Sebastiani,P. and Kohane,I.S. (2002) Cluster analysis of gene expression dynamics. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 9121–9126.
3. Schliep,A., Schonhuth,A. and Steinhoff,C. (2003) Using hidden Markov models to analyze gene expression time course data. *Bioinformatics*, **19**, i255–263.
4. Ernst,J., Nau,G.J. and Bar-Joseph,Z. (2005) Clustering short time series gene expression data. *Bioinformatics*, **21**, i159–168.
5. Ma,P., Castillo-Davis,C.I., Zhong,W. and Liu,J.S. (2006) A data-driven clustering method for time course gene expression data. *Nucleic Acids Res.*, **34**, 1261–1269.
6. Ernst,J., Vainas,O., Harbison,C.T., Simon,I. and Bar-Joseph,Z. (2007) Reconstructing dynamic regulatory maps. *Mol. Syst. Biol.*, **3**, 74.
7. Ding,J., Hagood,J.S., Ambalavanan,N., Kaminski,N. and Bar-Joseph,Z. (2018) iDREM: Interactive visualization of dynamic regulatory networks. *PLoS Comput. Biol.*, **14**, e1006019.
8. Gitter,A., Carmi,M., Barkai,N. and Bar-Joseph,Z. (2013) Linking the signaling cascades and dynamic regulatory networks controlling stress responses. *Genome Res.*, **23**, 365–376.
9. Rib,L., Villeneuve,D., Minocha,S., Praz,V., Hernandez,N., Guex,N., Herr,W. and Cycli,X.C. (2018) Cycles of gene expression and genome response during mammalian tissue regeneration. *Epigenetics Chromatin*, **11**, 52.
10. Bengio,Y. and Frasconi,P. (1996) Input-output HMMs for sequence processing. *IEEE Trans. Neural Netw.*, **7**, 1231–1249.
11. Durbin,R. (1998) In: *Biological Sequence Analysis: Probabalistic Models of Proteins and Nucleic Acids*. Cambridge University Press, NY.
12. Gunasekara,C., Zhang,K., Deng,W., Brown,L. and Wei,H. (2018) TGMI: an efficient algorithm for identifying pathway regulators through evaluation of triple-gene mutual interaction. *Nucleic Acids Res.*, **46**, e67.
13. Cahan,P., Li,H., Morris,S.A., Lummertz da Rocha,E., Daley,G.Q. and Collins,J.J. (2014) CellNet: network biology applied to stem cell engineering. *Cell*, **158**, 903–915.
14. Leaman,R. and Gonzalez,G. (2008) BANNER: an executable survey of advances in biomedical named entity recognition. *Pac. Symp. Biocomput.*, **13**, 652–663.
15. Taub,R. (2004) Liver regeneration: from myth to mechanism. *Nat. Rev. Mol. Cell Biol.*, **5**, 836–847.
16. Kurinna,S. and Barton,M.C. (2011) Cascades of transcription regulation during liver regeneration. *Int. J. Biochem. Cell Biol.*, **43**, 189–197.
17. Ernst,J. and Bar-Joseph,Z. (2006) STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics*, **7**, 191.
18. Langfelder,P. and Horvath,S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 559.
19. Nagaoka,M. and Duncan,S.A. (2010) Transcriptional control of hepatocyte differentiation. *Prog. Mol. Biol. Transl. Sci.*, **97**, 79–101.
20. Huang,J. and Rudnick,D.A. (2014) Elucidating the metabolic regulation of liver regeneration. *Am. J. Pathol.*, **184**, 309–321.
21. Piobbico,D., Bartoli,D., Pieroni,S., De Luca,A., Castelli,M., Romani,L., Servillo,G. and Della-Fazia,M.A. (2018) Role of IL-17RA in the proliferative priming of hepatocytes in liver regeneration. *Cell Cycle*, **17**, 2423–2435.
22. Furuya,S., Kono,H., Hara,M., Hirayama,K., Tsuchiya,M. and Fujii,H. (2013) Interleukin-17A plays a pivotal role after partial hepatectomy in mice. *J. Surg. Res.*, **184**, 838–846.
23. Murata,S., Ohkohchi,N., Matsuo,R., Ikeda,O., Myronovych,A. and Hoshi,R. (2007) Platelets promote liver regeneration in early period after hepatectomy in mice. *World J. Surg.*, **31**, 808–816.
24. Sakamoto,T., Liu,Z., Murase,N., Ezure,T., Yokomuro,S., Poli,V. and Demetris,A.J. (1999) Mitosis and apoptosis in the liver of interleukin-6-deficient mice after partial hepatectomy. *Hepatology*, **29**, 403–411.
25. Horst,A.K., Tiegs,G. and Diehl,L. (2019) Contribution of macrophage efferocytosis to liver homeostasis and disease. *Front. Immunol.*, **10**, 2670.
26. Harris,S.L. and Levine,A.J. (2005) The p53 pathway: positive and negative feedback loops. *Oncogene*, **24**, 2899–2908.
27. Schleich,K. and Lavrik,I.N. (2013) Mathematical modeling of apoptosis. *Cell Commun Signal*, **11**, 44.
28. Wong,M.M. and Fish,E.N. (2003) Chemokines: attractive mediators of the immune response. *Semin. Immunol.*, **15**, 5–14.
29. Muller,W.A. (2003) Leukocyte-endothelial-cell interactions in leukocyte transmigration and the inflammatory response. *Trends Immunol.*, **24**, 327–334.
30. Swanson,J.A. and Hoppe,A.D. (2004) The coordination of signaling during Fc receptor-mediated phagocytosis. *J. Leukoc. Biol.*, **76**, 1093–1103.
31. Grant,B.D. and Donaldson,J.G. (2009) Pathways and mechanisms of endocytic recycling. *Nat. Rev. Mol. Cell Biol.*, **10**, 597–608.
32. Stuart,L.M. and Ezekowitz,R.A. (2005) Phagocytosis: elegant complexity. *Immunity*, **22**, 539–550.
33. Montoya,M.C., Sancho,D., Vicente-Manzanares,M. and Sanchez-Madrid,F. (2002) Cell adhesion and polarity during immune interactions. *Immunol. Rev.*, **186**, 68–82.
34. Knosp,C.A. and Johnston,J.A. (2012) Regulation of CD4+ T-cell polarization by suppressor of cytokine signalling proteins. *Immunology*, **135**, 101–111.
35. Chintakuntlawar,A.V. and Chodosh,J. (2009) Chemokine CXCL1/KC and its receptor CXCR2 are responsible for neutrophil chemotaxis in adenoviral keratitis. *J. Interferon Cytokine Res.*, **29**, 657–666.
36. Baeck,C., Wehr,A., Karlmark,K.R., Heymann,F., Vucur,M., Gassler,N., Huss,S., Klussmann,S., Eulberg,D., Luedde,T. *et al.* (2012) Pharmacological inhibition of the chemokine CCL2 (MCP-1) diminishes liver macrophage infiltration and steatohepatitis in chronic hepatic injury. *Gut*, **61**, 416–426.
37. Petray,P., Corral,R., Meckert,P. and Laguens,R. (2002) Role of macrophage inflammatory protein-1alpha (MIP-1alpha) in macrophage homing in the spleen and heart pathology during experimental infection with *Trypanosoma cruzi*. *Acta Trop.*, **83**, 205–211.
38. Hidi,R., Riches,V., Al-Ali,M., Cruikshank,W.W., Center,D.M., Holgate,S.T. and Djukanovic,R. (2000) Role of B7-CD28/CTLA-4 costimulation and NF-kappa B in allergen-induced T cell chemotaxis by IL-16 and RANTES. *J. Immunol.*, **164**, 412–418.