Genomics 94 (2009) 1–10

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Gene-set analysis identifies master transcription factors in developmental courses

Ying Liu, Bo Jiang [1], Xuegong Zhang *

MOE Key Laboratory of Bioinformatics and Bioinformatics Division, TNLIST, Department of Automation, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Article history:
Received 7 November 2008
Accepted 26 February 2009
Available online 9 March 2009

Keywords:
Gene set analysis
Transcriptional regulation
Gene expression time series
Development

ABSTRACT

Transcriptional regulation plays key roles in many biological processes. The regulation is dynamic in time and space. Identifying transcription factors that play major roles in a developmental time course is very important for understanding the regulation. This cannot be realized by studying the relation between the expression of individual genes. We developed a gene-set analysis approach to study master regulators and their actively regulated targets during a time course from gene expression data. We applied the method to a mouse liver development data and a mouse embryonic stem cell (mESC) development data, and identified 14 and 9 transcription factors that play major regulatory roles in the two development courses, respectively. Some transcription factors could not be identified as active in the process by studying their correlation with individual targets. The method was also extended for studying other regulation factors or pathways from time-course expression data.

Introduction

Biological processes rely on finely orchestrated regulation of gene expression. Different levels of regulation, including transcriptional regulation by transcription factors (TFs), post-transcriptional regulation by microRNAs (miRNAs) [1], etc, ensures biological processes on their rails.

Transcriptional regulation is a dynamic process. Sequence information has been used to discover targets of particular regulatory elements (TFs, miRNAs, etc) computationally [2]. ChIP-chip, RNAi and other similar experiments have also been employed for the discovery of TF targets [3–6]. These computational and experimental technologies have helped gain insights into the function of the regulatory elements at a certain spatial and temporal point. However, transcriptional regulation may be tissue specific and change from time to time according to the inner and outer environment of the cell [7,8], and these technologies provide a static view, or a snapshot of the complex scene under study.

Transcriptional regulation plays a key role in development processes and cell destiny determination, and gene expression profiles have been employed to study the regulation during developmental time course. For example, liver is a versatile organ and executes a wide variety of fundamental functions for life, and dynamic transcriptional regulation is essential to the development of such a complex organ. In a previous work, we discovered several TFs which function at different stages of the mouse liver development process respectively [8]. The differentiation and development of embryonic stem cells (ESCs) is a biological process of great interest. ESCs have the pluripotency property, which makes them able to differentiate into a wide range of cell types [9,10]. Both the maintenance of their undifferentiated state and triggering of their differentiation depend on precisely tuned gene expression which constantly varies over time [3,5,6,11]. There have been studies focusing on deciphering the regulatory networks for some functional TFs during mESC differentiation [12–14].

There have been several methods proposed to study dynamics of gene regulation in biological processes using expression time series. For example, Pearson Correlation Coefficient (PCC) [15], the local clustering (LC) coefficient [16], mutual information (MI) [17] and trend correlation (TC) scores [18] have been proposed to evaluate the correlation between time series of two genes. Assuming that associated genes have similar expression patterns, these methods could reveal possible functional or expressional associations and be used to construct biological networks [17], but TFs that play major roles in a time course are hard to be revealed from such results. It has also been shown that for many cases, the correlation between the expression time series of a single TF and its targets can be weak [18,19], and the absolute values of such measurements are often incomparable among different TFs, making it difficult to set thresholds for detecting active TFs from the correlation of individual gene pairs.

When studying development courses it is important to identify master TFs that play dominant roles in regulating the expression of their target genes. Considering the aggregation of targets of a regulator may help solve the aforementioned problems because the larger proportion of a regulator's targets show similar expression patterns to it, the more likely it is a master TF. Enlightened by gene-set enrichment analysis or GSEA for static expression data [20], we adopted a gene-set analysis strategy to discover master regulators and their actively

* Corresponding author. Fax: +86 10 6278 6911.
E-mail address: zhangxg@tsinghua.edu.cn (X. Zhang).
[1] Current address: Department of Statistics, Harvard University, Cambridge, MA 02138, USA.

0888-7543/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2009.02.005

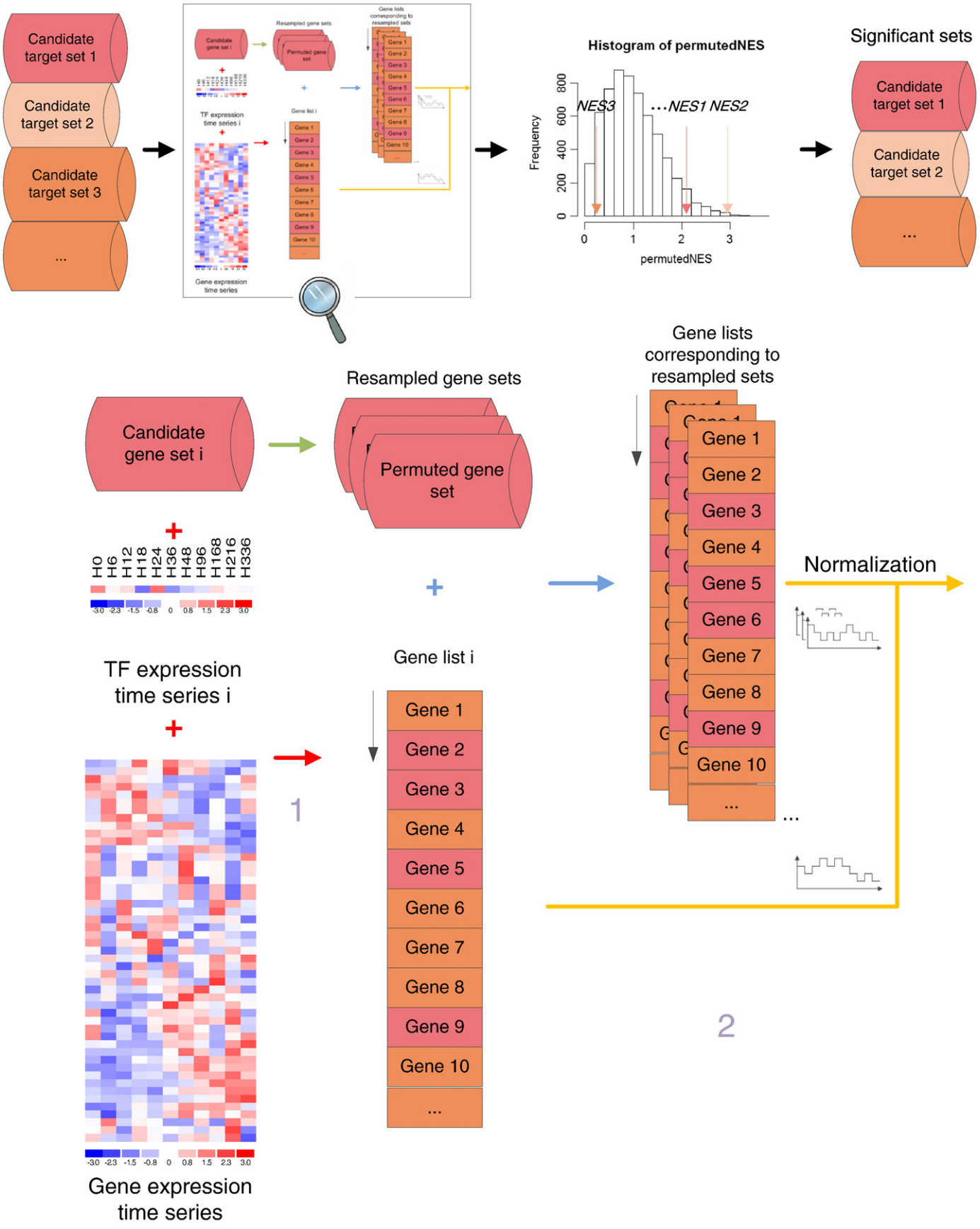

Fig. 1. The pipeline of the strategy. The strategy is executed in a two-step manner. In the first step, we use correlation scores between time series of a TF and genes to measure and rank the correlation between their expression profiles (red arrow). In the second step, the normalized enrichment score (*NES*) is obtained for a particular gene set. Gene-set resampling (green arrow) is used to evaluate the significance of enrichment (blue arrow) and normalize the original enrichment scores (yellow arrow). If there are more than one candidate target set, all the permuted *NES*s are used for the calculation of a *q*-value, which degenerates to *p*-value if there is only one candidate set (black arrows).

regulated targets during a specific time course. The strategy assumes no prior knowledge of the function or expression of regulators, allows both positive and negative correlation and potential time lag in the TF-target regulation, and enables the discovery of functional TFs whose expressions are weakly correlated with targets even in a short time course. Applying the strategy on a mouse liver development dataset [8], we identified 14 master TFs in the liver development course. Most of these TFs have been reported to participate in the liver development process except two (IRF8 and IRF4). They have not been reported as related to liver development, but this study shows that these two TFs also play key roles in liver development. We also explored a dataset of mouse ESC (mESC) differentiation process [21] with the method and highlighted nine master TFs, eight of which have been reported to play important roles in stem cell differentiation except GC, which is shown functional in this process by this study. Some of the identified master TFs cannot be discovered as active if their effects on targets were not studied in the gene-set analysis manner. The method was also extended to analyze other types of regulators and active pathways from time-course expression data, and was developed as a software *TAGS* (Time-series Analysis for Gene Set) for the convenience of biologist users. The software is available at http://bioinfo.au.tsinghua.edu.cn/member/yliu/TAGS.

Results and discussion

Identifying master TFs and their actively regulated targets in a time course

We use a two-step strategy to discover master TFs and their targets actively regulated by them during a specific time course from a set of candidate regulators and the sets of their validated or predicted target genes. In the first step, for each TF, we use absolute cross-correlation function (CCF) to measure the correlation between its expression time series and those of all the genes on a microarray. We choose absolute CCF to take into account both positive and negative correlation and potential time lag between the expression profiles of a TF and its target. The genes are ranked according to the maximums of their absolute CCFs defined as correlation scores (see Materials and methods for the definition). The relative strength of the correlation is used for ranking the genes rather than these values *per se*. This is suitable even if the correlation measurements (CCFs) are not very large. In the second step, we study the target set of each TF with a running-sum statistic as in the GSEA method [20]. Based on the assumption that master TFs dominate in regulating their targets' expression, we study the TF's target set to see whether a significant proportion of its targets behave more similarly to the TF than other genes. An enrichment score is calculated to reflect the enrichment of a TF's targets toward the top of the corresponding gene ranking generated in the first step. This circumvents the potential need to determine a threshold of the correlation with regard to whether a target gene is regulated by the TF, and enables the discovery of TFs whose correlation with targets are weak but significant. A permutation strategy is used for significance evaluation. The strategy can produce two types of discoveries: master TFs and their actively regulated targets during a time course. Those actively regulated targets constitute the so-called 'leading-edge subset' of the target genes. Fig. 1 shows the pipeline of the method. More details are described in the Materials and methods section. As examples, we applied the strategy to a mouse liver development dataset and a mESC differentiation dataset to study transcriptional regulation during these two processes.

Master TFs and their regulated targets in mouse liver development

We applied the strategy on a mouse liver development dataset [8]. The data contain 14 time points across the developmental course, which consist of E11.5 (embryonic day 11.5), E12.5, E13.5, E14.5, E15.5, E16.5, E17.5, E18.5, Day0 (the day of birth), Day3, Day7, Day14, Day21, and normal adult. 303 TFs and their experimentally validated target sets got from TRANSFAC were analyzed with the strategy, and 100 runs of permutation were done for significance evaluation. We identified 14 significant TFs with FDRs less than 10.0%. Table 1 gives these TFs and the description of their known functions associated with liver development.

It can be seen from Table 1 that most of these 14 TFs can be divided into three functional groups: (1) TFs involved in the developmental process; (2) TFs with hematopoietic function which is an important function of fetal liver; (3) TFs participating in liver metabolic processes such as fatty acid, glucose and lipid metabolic processes. Some of these TFs are well-known liver-specific TFs, such as HNF4a and CEBPA, and have been reported to be critical to liver development.

HNF4a plays critical roles in liver development, including controlling hepatic epithelial structure and liver sinusoidal organization [22]. It binds to almost half of the transcribed genes tested in the adult mouse liver studied in a recent ChIP-array study [4]. Although HNF4a's validated target set only contains 14 genes according to the current version of TRANSFAC, it is enough for the strategy to identify this important TF with FDR 0.000. Fig. 2a shows the expression profiles for HNF4a and its targets. The targets in the leading-edge subset are indicated in Fig. 2, and these genes show expression patterns more similar to HNF4a than the other genes, so they are more likely to be actively tuned by HNF4a and are participators in the development process. HNF4a can both activate and repress the expression of its targets according to our data (Fig. 2a). Table 2 gives the function

Table 1
Master TFs obtained on the mouse live development data and validated target sets.

TF	Function description[a]	Normalized enrichment score	*q*-value
HNF4a	Lipid metabolic process, liver development [22]	3.460	0.000
NFE2L2	Regulation of embryonic development	3.501	0.000
SFPI1	Granulocyte differentiation, lymphocyte differentiation, macrophage differentiation	2.715	0.044
CEBPA	Liver development, regulation of cell proliferation, macrophage differentiation, negative regulation of cell proliferation	2.673	0.049
POU5F1	(Endodermal and mesodermal) cell fate commitment	2.720	0.050
PPARA	Fatty acid, glucose and lipid metabolic processes, regulation of fatty acid metabolic process, epidermis development	2.731	0.060
IRF8	Myeloid cell differentiation	2.753	0.068
GATA1	The dominant action of GATA1s leads to hyperproliferation of a unique, previously unrecognized yolk sac and fetal liver progenitor [37], critical hematopoietic transcription factor [38]	2.579	0.070
ZBTB7B	Cell differentiation, multicellular organismal development	2.523	0.070
KLF3	Adipogenesis with CtBP [39], highly expressed in erythroid cells [40]	2.780	0.073
SOX17	Angiogenesis, vasculogenesis	2.529	0.075
NR5A2	Bile acid metabolic process, cholesterol homeostasis, primitive streak morphogenesis	2.443	0.090
IRF4	Myeloid dendritic cell differentiation	2.425	0.092
SP1	Definitive hemopoiesis, liver development	2.321	0.098

[a] Function annotations are retrieved from GO (http://www.geneontology.org) and the referenced literatures.

description of each of the 14 HNF4a targets. Comparing this information with Fig. 2a could help filter out some irrelevant genes which do not reside in the leading-edge subset and probably do not function in the liver development process.

SP1 is another TF important to liver development discovered with FDR 0.098. SP1 has been reported to be essential to embryonic development, and SP1 knockout mice die around E10.5. A very recent study shows that SP1/SP3 compound heterozygous mice are not viable, and suffer from a spectrum of developmental abnormalities of different organs [23]. This implies that unlike HNF4a, which is specific to liver, SP1 may play a more extensive role in the development procedure, and the development of different organs may share some common molecular mechanisms. SP1 is also involved in definitive hemopoiesis (according to GO), which is a function of the fetal liver. Fig. 2b shows the time series of SP1 and its targets. It has been reported that liver becomes a major site of fetal hemopoiesis around E10.5 to E12.5, and this is consistent with SP1's highest expression value at E12.5 according to our data (Fig. 2b). Like HNF4a, SP1 is able to both enhance and repress its targets' expression, while the more targets compared to those of HNF4a may indicate the more comprehensive function of SP1.

Besides TFs relevant to development, there are also significant TFs which take part in some metabolic processes that take place in the liver. These TFs include PPARA, KLF3 and NR5A2 (Table 1). HNF4a also takes part in the lipid metabolic process according to its GO annotation. This is consistent with its up-regulation during the time course (Fig. 2a), because in the late phase of embryonic development the liver begins to function as a metabolic organ.

There are two other TFs, IRF8 and IRF4, which play a part in myeloid cell differentiation according to existing literature (Table 1) and seem not directly relevant to liver development. Besides the common hematopoietic function of liver and myeloid, further literature mining shows that normal mouse liver can form myeloid cell clusters which contain dendritic cell progenitors in vitro [24]. Since there may be common mechanisms for a variety of development processes according to the above analysis, and IRF8 and IRF4 both have smaller FDRs than SP1, the crucial TF for liver development (Table 1), we infer that these two TFs may play key roles in mouse liver development, which deserves further study.

We compared the significant TFs discovered here with the TFs which were considered over-represented at different stages of liver development in our previous study [8]. There are two common TFs, HNF4a and SP1. These two TFs were considered to be over-represented in more than one stage in [8] (HNF4a in Stage I and IV, and SP1 in Stage II and IV). This indicates that the TFs discovered in this study play dominant roles constantly during the whole liver development process. The TFs only functioning in a particular stage in [8] are not of interest in the current work, because liver development is considered as a whole process here. The TFs in Table 1 are critical to liver development for they always dominate in regulating their targets' expression during this process for its normal running. Note that the above results do not mean that other TFs (such as the ones discovered in [8]) do not contribute to liver development. The 14 significant TFs are 'master' TFs. This means that their regulatory effect can be reflected by a comparatively large part of its TFs' expression, which implies they may have master functions during the development process.

We then studied TFs with their predicted targets under the same framework, and investigated TFs with motifs from TRANSFAC and their predicted targets with p-values lower than 10^{-5} by the Staden's method [2]. We obtained eight significant TFs with FDRs less than 10.0%. Table S1 gives these TFs and the description of their known functions associated with liver development. The two most relevant TFs, HNF4a and SP1, are both in the list of master TFs resulting from the predicted targets as expectation. Fig. S1 shows the time series of HNF4a and its predicted targets. Independent liver-specific ChIP-array data has reported that HNF4a bounds to 1575 genes in the liver [4], and we used it as a validation of the predicted targets in the leading-edge subset. 103 targets in the leading-edge subset are validated by the experiment in [4]. We then applied the chi-square test to check whether the leading-edge subset is enriched with targets validated by experiments. The obtained chi-square statistic is 90.0122 with p-value of almost 0. We can see apparently from the expression profiles that expression patterns of genes in the leading-edge subset either positively or negatively correlate with that of HNF4a (Fig. S1 (A)). We infer these genes are regulated by HNF4a, either enhanced or repressed, during mouse liver development. Note that this does not mean that other genes are definitely not targets of HNF4a, but only that those genes are not actively regulated by it during the specific development process (Fig. S1 (B)). The consistency between our results and the experiments together with the gene function annotations in Table 2 and expression profiles in Fig. S1 (A) confirms that the targets in the predicted leading-edge subset are actively regulated by HNF4a and play important roles during liver development.

Master TFs and their regulated targets in mESC differentiation

We then used a mESC differentiation dataset generated by Hailesellasse Sene et al. [21]. The data contain gene expression time series for three biologically equivalent but genetically distinct mESC lines (R1, J1, and V6.5) at 11 time points (0 h which represents undifferentiated mESCs, 6 h, 12 h, 18 h, 24 h, 36 h, 48 h, 4 d, 7 d, 9 d, and 14 d). We explored the potential regulation relationships between 62 TFs and their predicted targets ranked top 100 by the Staden's method [2] under 100 runs of permutation. We identified 15, 3 and 23 significant TFs with FDRs less than 15.0% for R1, J1 and V6.5 cell lines, respectively (We also analyzed similarly TFs and their validated target sets, and Table S2 gives the results). Table 3 shows nine TFs considered significant according to the results of at least two of the cell lines with

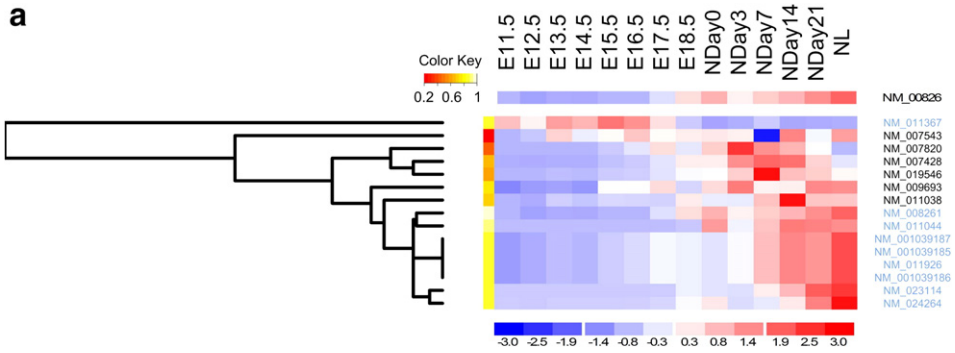

Fig. 2. Expression for HNF4a and SP1 and their validated targets in the liver development data. The targets with blue names are in the corresponding leading-edge subsets. The correlation scores are indicated by color on the left side of the heatmaps. (a) HNF4a and its targets. (b) SP1 and its targets.

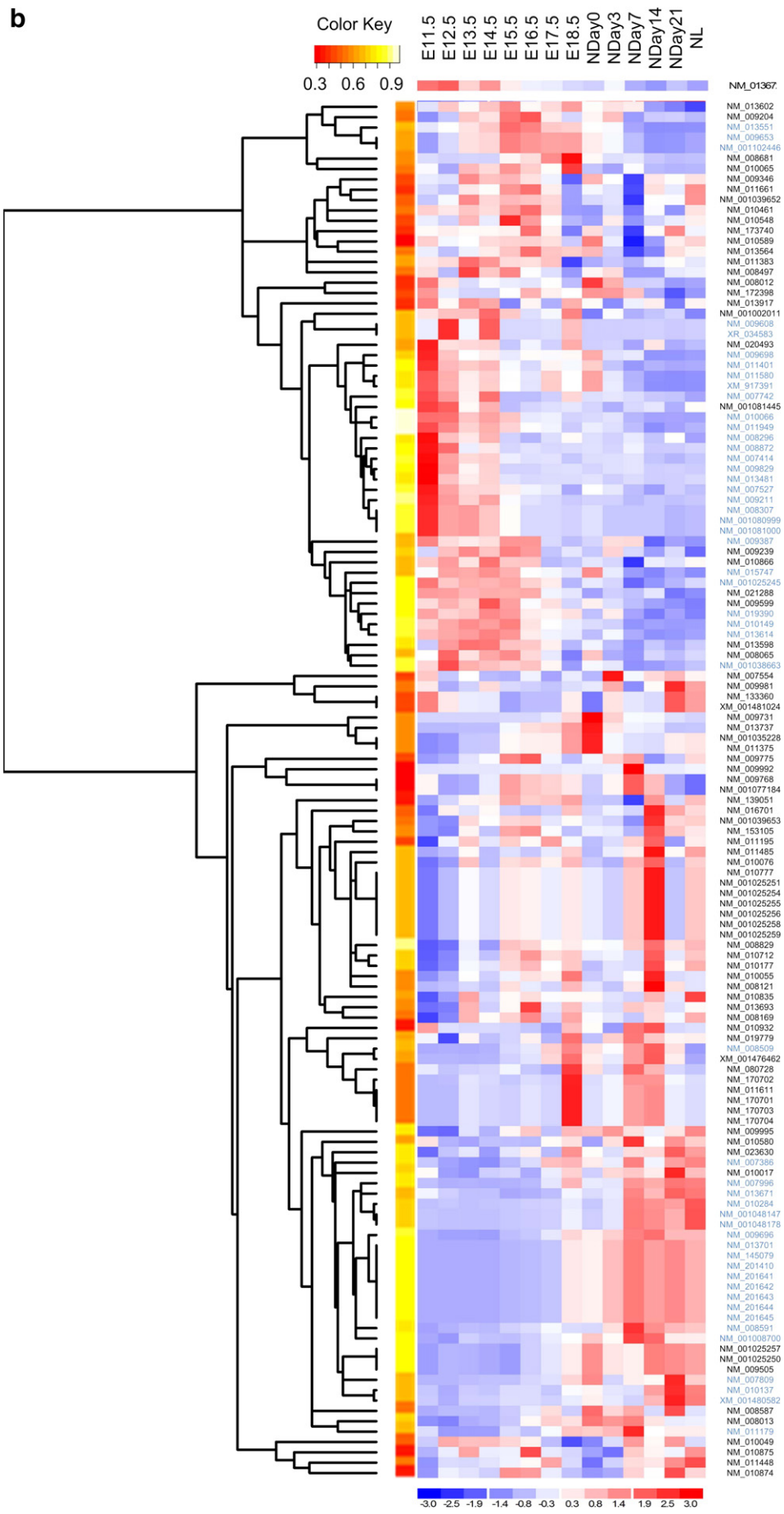
b
Color Key
0.3 0.6 0.9
E11.5
E12.5
E13.5
E14.5
E15.5
E16.5
E17.5
E18.5
NDay0
NDay3
NDay7
NDay14
NDay21
NL
NM_01367
NM_013602
NM_009204
NM_013551
NM_009653
NM_001102446
NM_008681
NM_010065
NM_009346
NM_011661
NM_001039652
NM_010461
NM_010548
NM_173740
NM_010589
NM_013564
NM_011383
NM_008497
NM_008012
NM_172398
NM_013917
NM_001002011
NM_009608
XR_034583
NM_020493
NM_009698
NM_011401
NM_011580
XM_917391
NM_007742
NM_001081445
NM_010066
NM_011949
NM_008296
NM_008872
NM_007414
NM_009829
NM_013481
NM_007527
NM_009211
NM_008307
NM_001080999
NM_001081000
NM_009387
NM_009239
NM_010866
NM_015747
NM_001025245
NM_021288
NM_009599
NM_019390
NM_010149
NM_013614
NM_013598
NM_008065
NM_001038663
NM_007554
NM_009981
NM_133360
XM_001481024
NM_009731
NM_013737
NM_001035228
NM_011375
NM_009775
NM_009992
NM_009768
NM_001077184
NM_139051
NM_016701
NM_001039653
NM_153105
NM_011195
NM_011485
NM_010076
NM_010777
NM_001025251
NM_001025254
NM_001025255
NM_001025256
NM_001025258
NM_001025259
NM_008829
NM_010712
NM_010177
NM_010055
NM_008121
NM_010835
NM_013693
NM_008169
NM_010932
NM_019779
NM_008509
XM_001476462
NM_080728
NM_170702
NM_011611
NM_170701
NM_170703
NM_170704
NM_009995
NM_010580
NM_023630
NM_007386
NM_010017
NM_007996
NM_013671
NM_010284
NM_001048147
NM_001048178
NM_009696
NM_013701
NM_145079
NM_201410
NM_201641
NM_201642
NM_201643
NM_201644
NM_201645
NM_008591
NM_001008700
NM_001025257
NM_001025250
NM_009505
NM_007809
NM_010137
XM_001480582
NM_008587
NM_008013
NM_011179
NM_010049
NM_010875
NM_011448
NM_010874
-3.0 -2.5 -1.9 -1.4 -0.8 -0.3 0.3 0.8 1.4 1.9 2.5 3.0

Table 2
Function annotations for the validated targets of HNF4a got from TRANSFAC.

Transcripts	Official symbol	Function description[a]	In the leading-edge subset of HNF4a[b]
NM_011044	Pck1	Gluconeogenesis, lipid metabolic process, glycerol biosynthetic process from pyruvate	Y
NM_001039185 NM_001039186 NM_001039187 NM_011926	Ceacam1	The sole receptor for mouse hepatitis virus A59 in both liver and brain, and its deletion from the mouse renders the mouse completely resistant to infection by this virus [41]; regulates insulin clearance in liver [42]	Y
NM_011367	Shbg	Monosaccharide-induced lipogenesis reduced hepatic HNF4a levels, which in turn attenuated Shbg expression [43]	Y
NM_023114	Apoc3	Cholesterol metabolic process, triacylglycerol catabolic process, triacylglycerol metabolic process, triacylglycerol mobilization	Y
NM_024264	Cyp27a1	Oxidation reduction, both its genotype and gender affected the regulation of hepatic bile acid, cholesterol, and fatty acid metabolism [44]	Y
NM_007543	Ceacam2	Component of cell surface	N
NM_007820	Cyp3a16	Oxidation reduction	N
NM_007428	Agt	Positive regulation of fatty acid biosynthetic process, kidney development and function [45]	N
NM_019546	Prodh2	Glutamate biosynthetic process, oxidation reduction, proline catabolic process, proline metabolic process	N
NM_009693	Apob	Cholesterol homeostasis, lipid metabolic process, lipoprotein metabolic process, triacylglycerol catabolic process, triacylglycerol mobilization, in utero embryonic development	N
NM_011038	Pax4	Organ morphogenesis, positive regulation of cell differentiation	N

[a] Function annotations are retrieved from GO and the referenced literatures.
[b] Whether this gene is in the leading-edge subset of the validated targets.

the description of their functions. Table S3 shows detailed information and time series shown by heatmaps of the significant TFs and their targets in the leading-edge subsets according to each of the three cell lines respectively.

The TFs in Table 3 are considered to be master TFs for the general mESC differentiation process since they are significant in at least two of the three cell lines, while cell-line-specific significant TFs may reflect inherent genetic discrepancies and unique regulation in each cell line (Table S3). Most of the master TFs in Table 3 are critical to the balance of self-renewal and differentiation into different cell types, which are the two most important characteristics of ESCs [9,10].

NANOG is reported to be significant by both R1 and V6.5 mESC lines. It is a leading TF for ESC to maintain self-renewal and pluripotency according to its GO annotations and literature. Fig. 3 shows the time series of NANOG and its predicted targets in R1 mESC line. It can be seen that NANOG can both activate and suppress its targets' transcription, which is consistent to the independent study of Loh et al. [6]. Most of its predicted targets are in the intersection of the leading-edge subsets obtained from R1 and V6.5 lines and show similar expression patterns to NANOG, and this may be an indication of NANOG's major role in mESC. These targets are more likely to be true targets and actively regulated by NANOG following the similar reasoning for liver development. Independent ChIP-chip [3], ChIP-PET [6] and RNAi [5] data for NANOG are available and we used them to validate our predication. There are 13 targets that are in the leading-edge subset and are reported to be bound by NANOG with high confidence in those data. All genes outside the leading-edge are not direct targets of NANOG according to those experiments. Chi-square test was used to check whether the leading-edge subset is enriched with targets validated by experiments. The obtained chi-square statistic is 6.8555 with p-value 0.008837. We infer the other targets not yet validated in the leading-edge subset are also controlled by NANOG during mESC differentiation.

One of the nine TFs in Table 3, GC, takes part in vitamin D (VD) metabolism, which seems not directly relevant to ESC differentiation, whereas it is considered significant by R1 and V6.5 mESC lines. Interestingly, a recent study showed that exposure to VD_3 can enhance the differentiation of mESCs into osteoblasts, and osteoinduction is an intricate dynamic process [25]. This indicates that GC may also take part in this process, which deserves further study.

Investigating the correlation between expression time series of a TF and each of its targets individually and comparing it with the results from gene-set approaches in this paper could be helpful to highlight the power of the latter. The correlation score between each target and NANOG is indicated in Fig. 3, and it shows that only a small portion of the targets have their expression strongly correlated with that of NANOG. In fact, the correlation scores for 38 out of the 69 targets are below 0.4, with the lowest only 0.136 despite of NANOG's definitely important roles during ESC differentiation. If we make inferences based only on such individual correlation, it is unlikely to discover TFs such as NANOG. However, this critical TF is detected significant when we apply gene-set analysis. Considering that the values of correlation measurements can be small between the expression time series of a single TF and its targets [18,19], gene-set-based strategies such as the one in this paper can help to reveal some of the important TFs and their regulated targets.

Table 3
Master TFs obtained on the mESC data and predicted target sets.

TF	Function description[a]	Normalized enrichment score[b]	q-value[b]
HOXA7	Multicellular organismal development, embryonic skeletal morphogenesis	4.399	0.000
EVI1	In utero embryonic development, post-embryonic development, multicellular organismal development, forebrain development, embryonic forelimb morphogenesis	4.101	0.000
ESR1	Cell growth, negative regulation of mitosis	4.083	0.000
EN1	Multicellular organismal development, neuron differentiation, embryonic forelimb morphogenesis	3.908	0.000
E4F1	Embryonic development	3.587	0.002
NANOG	Embryonic development, stem cell differentiation and maintenance, multicellular organismal development	3.428	0.007
TBX5	Embryonic forelimb morphogenesis, multicellular organismal development, morphogenesis of an epithelium	2.843	0.025
MEIS1	Multicellular organismal development, definitive hemopoiesis	3.119	0.050
GC	Vitamin D metabolic process	2.055	0.144

[a] Function annotations are retrieved from GO.
[b] Results got from the cell line whose result reports the corresponding TF with the lowest FDR.

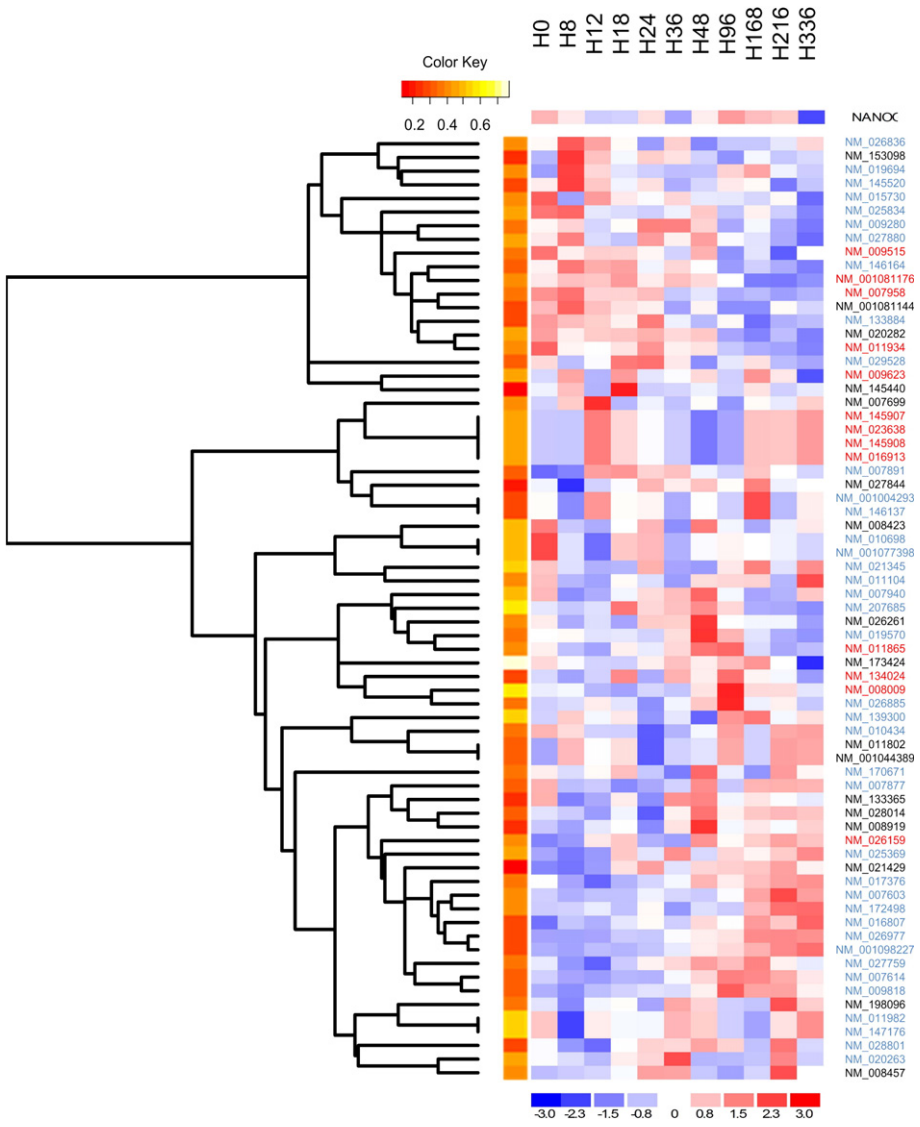

Fig. 3. Expression for NANOG and its predicted targets in R1 mESC line. The targets with red and blue names are in the intersection of the leading-edge subsets by R1 and V6.5 mESC lines, the results from which report NANOG significant, where targets with red names are also validated by independent experiments and actively regulated by NANOG during mESC differentiation. The correlation scores are indicated by color on the left side of the heatmap.

Identifying other regulators and their regulated targets in a time course

Besides TFs, other regulators can be interrogated in the similar way presented above, for example, miRNAs, which are approximately 22 nt RNAs that exert their function after transcription by targeting mRNAs for cleavage or translational repression [1]. But it is more difficult to detect a relatively strong correlation between the expression of miRNA and its target, especially for the repression mechanism. This makes the ranking criterion based on correlation less reasonable for the analysis of miRNAs. Besides, the expression time series of regulators under study such as miRNAs may be unavailable due to experimental limitation. Considering the above situations, we generalized the strategy for analyzing regulation using only targets' expression time series.

The difference between the current extension for miRNA analysis and the basic procedure lies in the ranking criteria. Specifically, we rank the genes according to their differential expression rather than expression correlation with the regulator so that genes that are differentially expressed (up- or down-regulated) over time will be ranked at the top of the list, and there is only one ranking for all gene sets if multiple sets are analyzed. The underlying rationale is that if expression of most targets of a regulator varies significantly across time during a time course, this regulator is likely to be an active player in the corresponding biological process. Different criteria can be used for the ranking task. A simplest one is the variance of each gene's expression across the time course. Some existing methods, such as EDGE [26], maSigPro [27] and MESA [28] could also provide gene ranks according to the genes' differential expression.

We analyzed miRNAs and their target sets with the mouse liver development dataset. Table S4 gives significant miRNAs and their functions. Interestingly, there is one significant miRNA, miR-451, which is regulated by GATA1, and GATA1 is also a master TF discovered by our transcriptional regulation analysis (Table 1). This gives a clue that different levels of regulation mechanisms may cross-talk with each other to form a comprehensive regulatory network in one process, and analysis of the dynamics of different regulatory mechanisms with the framework in this paper could endow a promise for unveiling at least part of this whole picture.

We developed a software package called Time-series Analysis for Gene Set or *TAGS* implementing the whole framework for gene-set analysis of expression time series, and used it to analyze a variety of expression time series for different purposes. For example, pathway analysis incorporates either pathway or functional annotations in search for consistent changes in gene expression. Methods for pathway analysis include GSEA [20], SAFE or Significance Analysis of

Function and Expression [29], SAM-GS, an extension of the SAM (Significance Analysis of Microarray) method to gene-set analyses [30], GSA [31], Sub-GSE [32], etc. These methods have been applied to analyze a wide range of biological data, including data from static microarray experiments [20], eQTL [33] and data for genomewide association study [34]. However, no applications to time-course microarray experiments have yet been published to the authors' best knowledge. We used *TAGS* to study pathways functioning in human brain ageing by the same strategy for the miRNA analysis, and obtained several significant pathways. We also analyzed the cytogenetic sets (C_1) downloaded from the Molecular Signature Database (MSigDB) [20] to study structurally associated genes with *TAGS*. The software is available for free academic use at http://bioinfo.au.tsinghua.edu.cn/member/yliu/TAGS.

Other correlation-based methods have been developed for the construction of biological networks based on the pairwise correlation between genes. Examples include local clustering (LC) [16], relevance networks [17] and trend correlation (TC) [18]. The observed correlation relationships with such methods are effects of all possible functional or expressional associations and they are not designed to reveal the TFs that play a major role in a time course. The *TAGS* strategy discovers master regulators during a particular time course by investigating the enrichment of their actively regulated targets towards the top of ranks of the correlation score. We expect that the current work will constitute an advance in exploring the dynamic characteristics of biological processes.

Materials and methods

Ranking genes with expression time series

Measuring the correlation between a TF and a gene

The basic assumption concerning the dynamic regulation is that if a TF plays an important role during a time course, its targets should exhibit similar expression patterns, although there will be some time lag because it may take some time for the TF to function. We use cross-correlation function (CCF) for the evaluation of the correlation between a TF and a gene with consideration of the possible time lag. Specifically, let $X_i(t)$ and $X_j(t)$ be time series of a TF and a gene, respectively, $t = 1, \ldots, n$. The CCF between the two time series is

$$\mathrm{CCF}(\tau) = \frac{\sum_s \left\{ \left[X_i(s+\tau) - \overline{X}_i\right]\left[X_j(s) - \overline{X}_j\right] \right\}}{\sqrt{\sum_s \left[X_i(s) - \overline{X}_i\right]^2} \sqrt{\sum_s \left[X_j(s) - \overline{X}_j\right]^2}}.$$

Then the correlation score for the considered gene is defined as

$$C = \max_{\tau \le 0} \mathrm{abs}(\mathrm{CCF}(\tau)).$$

The range of correlation scores is [0,1].The absolute value takes both positive and negative correlation into account, and only non-positive τs are considered since only TFs rather than targets can have phase lead. For a particular TF, genes are ranked according to their correlation scores.

Finding differentially expressed genes across the time course

Several methods for ranking the genes according to their differential expression in the dataset can be used for the miRNA analysis in this study. Adopting the terms used in classification studies, genes whose expressions change significantly across the time course can be called differentially expressed genes or DE genes.

As a simple and straightforward method, the variance of each gene's expression across the time series can be calculated as a measure of the gene's differential expression. We can rank the genes according to their variances. The variance can reflect the overall fluctuation of the expression values across the time course. This strategy is suitable for scenarios in which one does not care about the order of the expression values, e.g., when studying the gene expression profiles in a series of experimental situations. In some cases, even if the investigation is some time-relevant, variance is a suitable measure when only very few time points were observed.

Gene ranks generated from existing methods can also be used. Examples of such methods include EDGE [26] for significance analysis of time-course microarray experiments, maSigPro [27] that can handle multi-group time series data, and MESA [28] for cell-cycle types of time series data. Actually, any method can be used to provide gene ranking as long as it can produce an ordered list of genes that are meaningful for the underlying study. *TAGS* integrates some of the above ranking criteria.

Evaluating the enrichment of a target set

An enrichment score *ES* for each target set (corresponding to a regulator) is calculated with a running-sum statistic, following the strategy similar to the GSEA method [20]. Let S be the studied target set of N_H targets of a TF, and let $G = \{g_j; j = 1, 2, \ldots, N\}$ be the ordered list of genes corresponding to the TF, where j is the gene order index and N is the number of genes in the list. We can calculate two sums $P_{hit}(\mathrm{S,i})$ and $P_{miss}(\mathrm{S,i})$ as

$$P_{\mathrm{hit}}(S,i) = \sum_{\substack{g_j \in S \\ j \le i}} \frac{1}{N_{\mathrm{H}}},$$

$$P_{\mathrm{miss}}(S,i) = \sum_{\substack{g_j \notin S \\ j \le i}} \frac{1}{N - N_{\mathrm{H}}}.$$

A counter is calculated and its value at position i is

$$counter(S,i) = P_{\mathrm{hit}}(S,i) - P_{\mathrm{miss}}(S,i),$$

and the enrichment score of target set S is

$$ES(S) = \max_i \{counter(S,i)\}.$$

The range of *ES* values is [0,1]. We denote the position where the maximum is reached as i_0. The genes in set S ranked above or at position i_0 are called the leading-edge subset of the target set. Note that there is no hard boundary between targets in the leading-edge subset and those not. In this paper, we just take advantage of leading-edge subset to retrieve targets under the control of their regulator in a particular time course for further study.

We adopt a permutation strategy to assess the significance of the enrichment score. Specifically, a number of randomly selected gene sets with the same size are generated, which is defined as gene-set permutation, and the enrichment scores of these gene sets on the corresponding rank are calculated. The null distribution of the enrichment score and the permutation p-value are estimated according to these scores.

It is also possible to do other types of permutation. For example, by randomly shuffling the time points of the time series, any patterns in the gene expression across the time course will be randomized. But this kind of time-point permutation does not take effect for the calculation of CCF in our basic scenario. However, the extension for the miRNA analysis could apply this permutation strategy so it is also implemented in *TAGS*.

Multiple comparisons

It has been reported that there might be bias toward assigning higher enrichment scores to gene sets of large size [35]. We adjust for

such set-size variation following the similar strategy of GSEA by normalizing the enrichment scores with the mean of permuted scores [20]. Let the enrichment score for a target set S generated from a permutation p_0 be $ES(S, p_0)$. The original and permuted normalized enrichment score $NES(S)$ and $NES(S, p_0)$ are given by

$$NES(S) = \frac{ES(S)}{\overline{ES(S,p)}},$$

$$NES(S,p_0) = \frac{ES(S,p_0)}{\overline{ES(S,p)}},$$

where $\overline{ES(S,p)}$ is the mean of all the permuted enrichment scores for target set S.

As in other genome-scale investigations, Bonferroni correction can be applied to control the family-wise error rate when performing multiple testing, which is stringent but may result in too few or no discoveries when the number of candidate target sets (regulators) m is large. Another widely accepted correction is to control the False Discovery Rate or FDR [36]. We adopt the FDR correction in our experiments. More specifically, the FDR for a set S_0 is calculated as

$$\mathrm{FDR}(S_0) = \frac{\sharp_{S,p}\{NES(S,p) \geq NES(S_0)\} / (m \times t)}{\sharp_S\{NES(S) \geq NES(S_0)\} / m}.$$

where t is the permutation time for each target set.

Adjustment and variations of the strategy

Weighted enrichment scores can be calculated to emphasize genes of more importance. This can be useful when values of a particular ranking criterion vary abruptly along the gene lists. We also provide an option in the strategy to define tie genes when the difference between two criterion values is small. Both the above options are implemented in *TAGS*. See Supplementary material for the formulation and technical details of these options.

Mouse liver development data

We first applied the strategy to a time series expression dataset of mouse liver development [8]. It contains the expression profiles of 21561 mouse Refseq transcripts at 14 time points during the liver development process, which include E11.5 (embryonic day 11.5), E12.5, E13.5, E14.5, E15.5, E16.5, E17.5, E18.5, Day0 (the day of birth), Day3, Day7, Day14, Day21, and normal adult. The expression is measured with Mouse 430 microarrays (Affymetrix).

We used BioMart (http://www.biomart.org/biomart/martview/) and MGI (http://www.informatics.jax.org/) to convert the TF identifiers in TRANSFAC into RefSeq IDs, and then searched the expression matrix for the expression time series of the TFs and got 303 time series of TF transcripts and 63 time series corresponding to motifs. We then selected from the expression matrix transcripts with their expression values higher than 1.5-fold or lower than 1/1.5-fold compared to the mean at any time point (8640 in total) for the correlation analysis.

mESC differentiation data

We then used a mESC differentiation dataset generated by Hailesellasse Sene et al. [21]. The dataset has time series of 22690 genes in three biologically equivalent but genetically distinct mESC lines (R1, J1, and V6.5), each line with 11 time points (0 h which represents undifferentiated mESCs, 6 h, 12 h, 18 h, 24 h, 36 h, 48 h, 4 d, 7 d, 9 d, and 14 d) and 3 replicates at each time point. Two GeneChips (MOE430A and MOE430B, Affymetrix) are used for each cell line. MOE430B measures genes that are not very well characterized than those on MOE430A according to the data analysis by Hailesellasse Sene et al. [21], so we used data generated by MOE430A for our analysis. The data generated from R1, J1 and V6.5 cell lines were downloaded from the GEO database (http://www.ncbi.nlm.nih.gov/geo/) with accession numbers GDS2666, GDS2668 and GDS2671, respectively. The 3 replicates were averaged to form one time series for each cell line.

The expression matrix contains 62 time series corresponding to the motifs in TRANSFAC. We used the proposed strategy to analyze the TFs for each of the three cell lines.

Target sets for the regulatory analysis

For each TF, we used its validated or predicted target genes as the corresponding gene set. TFs and their true targets were retrieved from the mouse/rat/human-subset of TRANSFAC FACTOR TABLE (Release 2008.2). Rat/human targets in the table were converted to their mouse homologues using BioMart (http://www.biomart.org/biomart/martview/). 344 gene sets corresponding to 344 TF transcripts were constructed, each of which contains validated targets of a TF. Note that typically these targets are confirmed to be bound by their TFs at a specific space and time point, so whether they (and which subset of them) are regulated in a particular time course needs to be studied.

We got the promoter DNA sequences from the UCSC database (http://genome.ucsc.edu/). Each promoter sequence was taken 1000 bp upstream to 200 bp downstream from the transcription start site (TSS). All the motifs are represented by the Position Weight Matrices (PWMs). Known motifs were got from the vertebrate subset of TRANSFAC (Version 9.3). The targets were predicted by the Staden's method [2]. Given a motif M, base composition f and a match score S, the Staden's method can calculate the p-value of the match score S, that is, the probability that a randomly selected site has a score at least as high as S, and genes are ranked according to their p-values. The base composition f is calculated from the promoter sequences. The match score S is obtained by scanning the promoter sequence of each gene with the PWM of M. We constructed two editions of predicted target sets with genes whose p-values are lower than 10^{-5} or which rank top 100, respectively, and each edition contains 240 gene sets corresponding to 240 TFs.

Predicted sets are likely to contain some false targets which may bring noise to the analysis and encumber effective discoveries. Making the threshold of the corresponding prediction methods more rigid (like 10^{-5} used in our study) could solve this problem. Validated sets are more confident but may contain too few targets (sometimes under 10). Such sets can be considered as subsets sampled from all the potential targets of a TF. The difference between the numbers of available and true targets may cause unstable results. The limited validated targets may also make detection of master TFs difficult especially if the expression time series of available targets all have small correlations with that of their TF, but as long as a TF is considered significant with a relatively small target set (one sample from the population), it is more likely to be a true discovery (as seen for HNF4a). In a word, the usefulness of gene set approaches may rely on the perfectness of the available gene sets.

Acknowledgments

We thank Dr. Sheng Zhong for his help on the mESC data analysis, and thank Tingting Li and Yunfei Pei for their help on the mouse liver development data analysis. This study is partially supported by NSFC grants 30625012, 60575014 and 60721003, the National Basic Research Program (2004CB518605) and Hi-tech Research and Development Program (2006AA02Z325) of China.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2009.02.005.

References

[1] D.P. Bartel, MicroRNAs: genomics, biogenesis, mechanism, and function, Cell 116 (2004) 281–297.
[2] R. Staden, Methods for calculating the probabilities of finding patterns in sequences, Comput. Appl. Biosci. 5 (1989) 89–96.
[3] L.A. Boyer, T.I. Lee, M.F. Cole, S.E. Johnstone, S.S. Levine, J.P. Zucker, M.G. Guenther, R.M. Kumar, H.L. Murray, R.G. Jenner, D.K. Gifford, D.A. Melton, R. Jaenisch, R.A. Young, Core transcriptional regulatory circuitry in human embryonic stem cells, Cell 122 (2005) 947–956.
[4] D.T. Odom, N. Zizlsperger, D.B. Gordon, G.W. Bell, N.J. Rinaldi, H.L. Murray, T.L. Volkert, J. Schreiber, P.A. Rolfe, D.K. Gifford, E. Fraenkel, G.I. Bell, R.A. Young, Control of pancreas and liver gene expression by HNF transcription factors, Science 303 (2004) 1378–1381.
[5] N. Ivanova, R. Dobrin, R. Lu, I. Kotenko, J. Levorse, C. DeCoste, X. Schafer, Y. Lun, I.R. Lemischka, Dissecting self-renewal in stem cells with RNA interference, Nature 442 (2006) 533–538.
[6] Y.H. Loh, Q. Wu, J.L. Chew, V.B. Vega, W. Zhang, X. Chen, G. Bourque, J. George, B. Leong, J. Liu, K.Y. Wong, K.W. Sung, C.W. Lee, X.D. Zhao, K.P. Chiu, L. Lipovich, V.A. Kuznetsov, P. Robson, L.W. Stanton, C.L. Wei, Y. Ruan, B. Lim, H.H. Ng, The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells, Nat. Genet. 38 (2006) 431–440.
[7] E. Segal, T. Raveh-Sadka, M. Schroeder, U. Unnerstall, U. Gaul, Predicting expression patterns from regulatory sequence in *Drosophila* segmentation, Nature 451 (2008) 535–540.
[8] T. Li, J. Huang, Y. Jiang, Y. Zeng, F. He, M.Q. Zhang, Z. Han, X. Zhang, Multi-stage analysis of gene expression and transcription regulation in C57/B6 mouse liver development, Genomics 93 (2009) 235–242.
[9] S. Pease, P. Braghetta, D. Gearing, D. Grail, R.L. Williams, Isolation of embryonic stem (ES) cells in media supplemented with recombinant leukemia inhibitory factor (LIF), Dev. Biol. 141 (1990) 344–352.
[10] G.R. Martin, Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells, Proc. Natl. Acad. Sci. U. S. A. 78 (1981) 7634–7638.
[11] M. Golan-Mashiach, J.E. Dazard, S. Gerecht-Nir, N. Amariglio, T. Fisher, J. Jacob-Hirsch, B. Bielorai, S. Osenberg, O. Barad, G. Getz, A. Toren, G. Rechavi, J. Itskovitz-Eldor, E. Domany, D. Givol, Design principle of gene expression used by human stem cells: implication for pluripotency, FASEB J. 19 (2005) 147–149.
[12] C.C. Chen, S. Zhong, Inferring gene regulatory networks by thermodynamic modeling, BMC Genomics 9 (Suppl. 2) (2008) S19.
[13] C.C. Chen, X.G. Zhu, S. Zhong, Selection of thermodynamic models for combinatorial control of multiple transcription factors in early differentiation of embryonic stem cells, BMC Genomics 9 (Suppl. 1) (2008) S18.
[14] Q. Zhou, H. Chipperfield, D.A. Melton, W.H. Wong, A gene regulatory network in mouse embryonic stem cells, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 16438–16443.
[15] Y.H. Taguchi, Y. Oono, Relational patterns of gene expression via non-metric multidimensional scaling analysis, Bioinformatics 21 (2005) 730–740.
[16] J. Qian, M. Dolled-Filhart, J. Lin, H. Yu, M. Gerstein, Beyond synexpression relationships: local clustering of time-shifted and inverted gene expression profiles identifies new, biologically relevant interactions, J. Mol. Biol. 314 (2001) 1053–1066.
[17] A. Lindlof, Z. Lubovac, Simulations of simple artificial genetic networks reveal features in the use of Relevance Networks, In Silico Biol. 5 (2005) 239–249.
[18] F. He, A.P. Zeng, In search of functional association from time-series microarray data based on the change trend and level of gene expression, BMC Bioinformatics 7 (2006) 69.
[19] F. He, J. Buer, A.P. Zeng, R. Balling, Dynamic cumulative activity of transcription factors as a mechanism of quantitative gene regulation, Genome Biol. 8 (2007) R181.
[20] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, J.P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 15545–15550.
[21] K. Hailesellasse Sene, C.J. Porter, G. Palidwor, C. Perez-Iratxeta, E.M. Muro, P.A. Campbell, M.A. Rudnicki, M.A. Andrade-Navarro, Gene function in early mouse embryonic stem cell differentiation, BMC Genomics 8 (2007) 85.
[22] F. Lemaigre, K.S. Zaret, Liver development update: new embryo models, cell lineage control, and morphogenesis, Curr. Opin. Genet. Dev. 14 (2004) 582–590.
[23] I. Kruger, M. Vollmer, D.G. Simmons, H.P. Elsasser, S. Philipsen, G. Suske, Sp1/Sp3 compound heterozygous mice are not viable: impaired erythropoiesis and severe placental defects, Dev. Dyn. 236 (2007) 2235–2244.
[24] L. Lu, J. Woo, A.S. Rao, Y. Li, S.C. Watkins, S. Qian, T.E. Starzl, A.J. Demetris, A.W. Thomson, Propagation of dendritic cell progenitors from normal mouse liver using granulocyte/macrophage colony-stimulating factor and their maturational development in the presence of type-1 collagen, J. Exp. Med. 179 (1994) 1823–1834.
[25] N.I. zur Nieden, F.D. Price, L.A. Davis, R.E. Everitt, D.E. Rancourt, Gene profiling on mixed embryonic stem cell populations reveals a biphasic role for beta-catenin in osteogenic differentiation, Mol. Endocrinol. 21 (2007) 674–685.
[26] J.D. Storey, W. Xiao, J.T. Leek, R.G. Tompkins, R.W. Davis, Significance analysis of time course microarray experiments, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 12837–12842.
[27] A. Conesa, M.J. Nueda, A. Ferrer, M. Talon, maSigPro: a method to identify significantly differential expression profiles in time-course microarray experiments, Bioinformatics 22 (2006) 1096–1102.
[28] C.J. Langmead, C.R. McClung, B.R. Donald, A maximum entropy algorithm for rhythmic analysis of genome-wide expression patterns, Proc. IEEE. Comput. Soc. Bioinform. Conf. 1 (2002) 237–245.
[29] W.T. Barry, A.B. Nobel, F.A. Wright, Significance analysis of functional categories in gene expression studies: a structured permutation approach, Bioinformatics 21 (2005) 1943–1949.
[30] I. Dinu, J.D. Potter, T. Mueller, Q. Liu, A.J. Adewale, G.S. Jhangri, G. Einecke, K.S. Famulski, P. Halloran, Y. Yasui, Improving gene set analysis of microarray data by SAM-GS, BMC Bioinformatics 8 (2007) 242.
[31] B. Efron, R. Tibshirani, On testing the significance of sets of genes, Ann. Appl. Stat. 1 (2007) 107–129.
[32] X. Yan, F. Sun, Testing gene set enrichment for subset of genes: sub-GSE, BMC. Bioinformatics 9 (2008) 362.
[33] C. Wu, D.L. Delano, N. Mitro, S.V. Su, J. Janes, P. McClurg, S. Batalov, G.L. Welch, J. Zhang, A.P. Orth, J.R. Walker, R.J. Glynne, M.P. Cooke, J.S. Takahashi, K. Shimomura, A. Kohsaka, J. Bass, E. Saez, T. Wiltshire, A.I. Su, Gene set enrichment in eQTL data identifies novel annotations and pathway regulators, PLoS Genet. 4 (2008) e1000070.
[34] K. Wang, M. Li, M. Bucan, Pathway-based approaches for analysis of genomewide association studies, Am. J. Hum. Genet. 81 (2007) 1278–1283.
[35] D. Damian, M. Gorfine, Statistical concerns about the GSEA procedure, Nat. Genet. 36 (2004) 663 author reply 663.
[36] J.D. Storey, A direct approach to false discovery rates, J. R. Stat. Soc. B. 64 (2002) 479–498.
[37] Z. Li, F.J. Godinho, J.H. Klusmann, M. Garriga-Canut, C. Yu, S.H. Orkin, Developmental stage-selective effect of somatically mutated leukemogenic transcription factor GATA1, Nat. Genet. 37 (2005) 613–619.
[38] A.B. Cantor, S.H. Orkin, Transcriptional regulation of erythropoiesis: an affair involving multiple partners, Oncogene 21 (2002) 3368–3376.
[39] N. Sue, B.H. Jack, S.A. Eaton, R.C. Pearson, A.P. Funnell, J. Turner, R. Czolij, G. Denyer, S. Bao, J.C. Molero-Navajas, A. Perkins, Y. Fujiwara, S.H. Orkin, K. Bell-Anderson, M. Crossley, Targeted disruption of the basic Kruppel-like factor gene (Klf3) reveals a role in adipogenesis, Mol. Cell. Biol. 28 (2008) 3967–3978.
[40] J. Turner, M. Crossley, Mammalian Kruppel-like transcription factors: more than just a pretty finger, Trends Biochem. Sci. 24 (1999) 236–240.
[41] B.B. Singer, I. Scheffrahn, R. Heymann, K. Sigmundsson, R. Kammerer, B. Obrink, Carcinoembryonic antigen-related cell adhesion molecule 1 expression and signaling in human, mouse, and rat leukocytes: evidence for replacement of the short cytoplasmic domain isoform by glycosylphosphatidylinositol-linked proteins in human leukocytes, J. Immunol. 168 (2002) 5139–5146.
[42] M.N. Poy, Y. Yang, K. Rezaei, M.A. Fernstrom, A.D. Lee, Y. Kido, S.K. Erickson, S.M. Najjar, CEACAM1 regulates insulin clearance in liver, Nat. Genet. 30 (2002) 270–276.
[43] D.M. Selva, K.N. Hogeveen, S.M. Innis, G.L. Hammond, Monosaccharide-induced lipogenesis regulates the human hepatic sex hormone-binding globulin gene, J. Clin. Invest. 117 (2007) 3979–3987.
[44] S. Dubrac, S.R. Lear, M. Ananthanarayanan, N. Balasubramaniyan, J. Bollineni, S. Shefer, H. Hyogo, D.E. Cohen, P.J. Blanche, R.M. Krauss, A.K. Batta, G. Salen, F.J. Suchy, N. Maeda, S.K. Erickson, Role of CYP27A in cholesterol and bile acid metabolism, J. Lipid Res. 46 (2005) 76–85.
[45] F. Massiera, M. Bloch-Faure, D. Ceiler, K. Murakami, A. Fukamizu, J.M. Gasc, A. Quignard-Boulange, R. Negrel, G. Ailhaud, J. Seydoux, P. Meneton, M. Teboul, Adipose angiotensinogen is involved in adipose tissue growth and blood pressure regulation, FASEB J. 15 (2001) 2727–2729.