BMC Bioinformatics

CrossMark

TIGER*i*: modeling and visualizing the responses to perturbation of a transcription factor network

Namshik Han[1,3*], Harry A. Noyes[2] and Andy Brass[3*]

Abstract

Background: Transcription factor (TF) networks play a key role in controlling the transfer of genetic information from gene to mRNA. Much progress has been made on understanding and reverse-engineering TF network topologies using a range of experimental and theoretical methodologies. Less work has focused on using these models to examine how TF networks respond to changes in the cellular environment.

Methods: In this paper, we have developed a simple, pragmatic methodology, TIGER*i* (Transcription-factor-activity Illustrator for Global Explanation of Regulatory interaction), to model the response of an inferred TF network to changes in cellular environment. The methodology was tested using publicly available data comparing gene expression profiles of a mouse p38α (Mapk14) knock-out line to the original wild-type.

Results: Using the model, we have examined changes in the TF network resulting from the presence or absence of p38α. A part of this network was confirmed by experimental work in the original paper. Additional relationships were identified by our analysis, for example between p38α and HNF3, and between p38α and SOX9, and these are strongly supported by published evidence. FXR and MYC were also discovered in our analysis as two novel links of p38α. To provide a computational methodology to the biomedical communities that has more user-friendly interface, we also developed a standalone GUI (graphical user interface) software for TIGERi and it is freely available at https://github.com/namshik/tigeri/.

Conclusions: We therefore believe that our computational approach can identify new members of networks and new interactions between members that are supported by published data but have not been integrated into the existing network models. Moreover, ones who want to analyze their own data with TIGER*i* could use the software without any command line experience. This work could therefore accelerate researches in transcriptional gene regulation in higher eukaryotes.

Keywords: Machine Learning, Transcriptional regulatory network, Transcription factor binding site, Gene expression

Background

Integrated functional genomics attempts to utilize the vast wealth of data produced by modern large scale genomic and post-genomic projects to understand the functions of cells and organisms [1]. The rapidly increasing amount of high throughput sequencing data makes it essential to develop new analytical tools that can systematically process and integrate those datasets. This presents both challenges and opportunities to the computer science community.

Transcription factor (TF) proteins bind to promoter elements on genomic DNA at TF binding sites (TFBS), to help control the transfer of genetic information from gene to mRNA [2]. Understanding the mechanisms underlying mRNA transcription is one of the "grand challenges" in modern biology. Experimental techniques allow direct measurement of individual gene transcription, but the contribution of multiple TFs is hard to determine [3–5]. Measuring the concentration of TF proteins and their affinity for the promoter region of genes is difficult because

* Correspondence: namshik.han@gurdon.cam.ac.uk;
andy.brass@manchester.ac.uk
[1]Gurdon Institute, University of Cambridge, Cambridge, UK
[3]School of Computer Science and School of Health Sciences, University of Manchester, Manchester, UK
Full list of author information is available at the end of the article

concentrations are low and protein-DNA interactions are subject to multiple controls, resulting in measurement artifacts [6–8]. Post transcriptional regulation compounds these difficulties because other molecules modify mRNA stability and hence the signals from the TFs [3, 9–11]. In such a complex environment, *in-silico* techniques can provide insights and hypotheses into the underlying TF regulatory activity, although they clearly have limitations.

Reverse-engineering of TF network and TFBS information

A number of techniques are available to uncover the topology of the TF network—the networks of complex reactions and interactions in the cell that control transcript levels [12]. One strategy is to use principals of reverse-engineering and use gene expression data to infer regulatory interactions [13, 14]. Various reverse-engineering methods can reduce the dimensionality of the classic combinatorial search problem and utilize genome sequence data to enhance the sensitivity and specificity of predictions. However, they have difficulties in describing regulatory control by mechanisms other than TFs. Reverse-engineering of TF networks in the lower eukaryotes has been well developed [15–17]. However, the problems in mapping the regulatory mechanisms in cells of higher eukaryotes have made such global studies either impossible or impractical. Some recent studies have begun to address this issue [18–20], but have tended to focus only understanding which TFs bind to which genes—not looking in detail at the nature of the TF/TFBS interaction. A recent study [21] identified key biological features in transcriptional changes, however this method has difficulties in inferring the dynamics of the interactions. Furthermore, TF concentrations were not considered during the identification of the features.

To date, various reverse-engineering methods can reduce the dimensionality of the reverse-engineering problem and utilize genome sequence data to enhance the sensitivity and specificity of predicted interactions. However, they have difficulties in describing regulatory control by mechanisms other than TFs. To address this issue TFBSs information is required to complement the gene expression data. We used a list of 132,654 TFBSs between 20,920 genes and 174 TFs that had been identified by searching an alignment of five mammal species for conserved 5' and 3' regions [22]. Connectivity data is notorious for high false positive rates; however, our connectivity data is robust against the problem because it extracts binding information from well conserved upstream regions. A more detailed explanation is addressed in the Methods and Results section and a schematic diagram of the connectivity data is presented in Fig. 1.

Identifying regulation type by combining TFA and TFC analysis

Transcription factor activities (TFAs) are the intensity of the interactions between a certain transcription factor (TF) and its targets at a certain experimental point [23]. Thus, the estimated strength of TFAs between each TF and its target gene are useful to know which TF is acting on which gene at a given time point or experiment condition. However, simply knowing the regulatory activities under a single experimental condition provides limited information about the transcriptional network. To understand the mechanism of regulatory interactions, we developed a method that identifies statistically significant differences in TFAs under two different conditions. The significant differences indicate the changing level of TFAs between two conditions, so varying trends of TFAs in whole experimental process are easily detected and can be used to identify TF-specific regulatory patterns (up and down-regulation).

A highly concentrated TF induces more gene expressions rather than a lower concentrated TF. High-affinity binding sites induce the gene expressions at any level of TF concentration (TFCs), but low-affinity binding sites require high level of TF concentration for induction [24]. Thus, we might assume that TF concentration level is an important factor for investigating TFAs, and TFA investigation with considering TFC provides more reliable and accurate results closer to the complex reality of biology. To address this problem we proposed a probabilistic variational inference method to infer the concentration of each TF protein (TFC) and the regulatory intensities (TFA) of each TF and gene pair [4].

Aside from the method, there have been some notable attempts to infer TFAs based on integrating gene expression data and TFBSs information. The approaches use various well-known statistical inference techniques such as network component analysis [25], support vector machine [26], multivariate regression plus backward variable selection [27] and partial least squares [28]. However, the TFAs, which are inferred by these methods, do not contain any information on the strength and the sign of the physical interaction between a TF and its target genes. Moreover, the regulatory interactions can change easily in response to changing experimental conditions and over time. Since the methods are not fully probabilistic, they are not ideal for investigating the stochastic interactions. A linear regression based probabilistic method to model the full probability distribution of each TFA on each gene was developed [23]. The limitation of this method, however, is that it does not infer the TFAs and TFCs separately. This is a serious problem in subsequent analysis and prediction.

Methods

Transcription regulatory circuits and mathematical model

Transcription regulatory circuits can be thought of as having *trans-* and *cis*-inputs that are transformed into genetic information at mRNA level [29]. These circuits
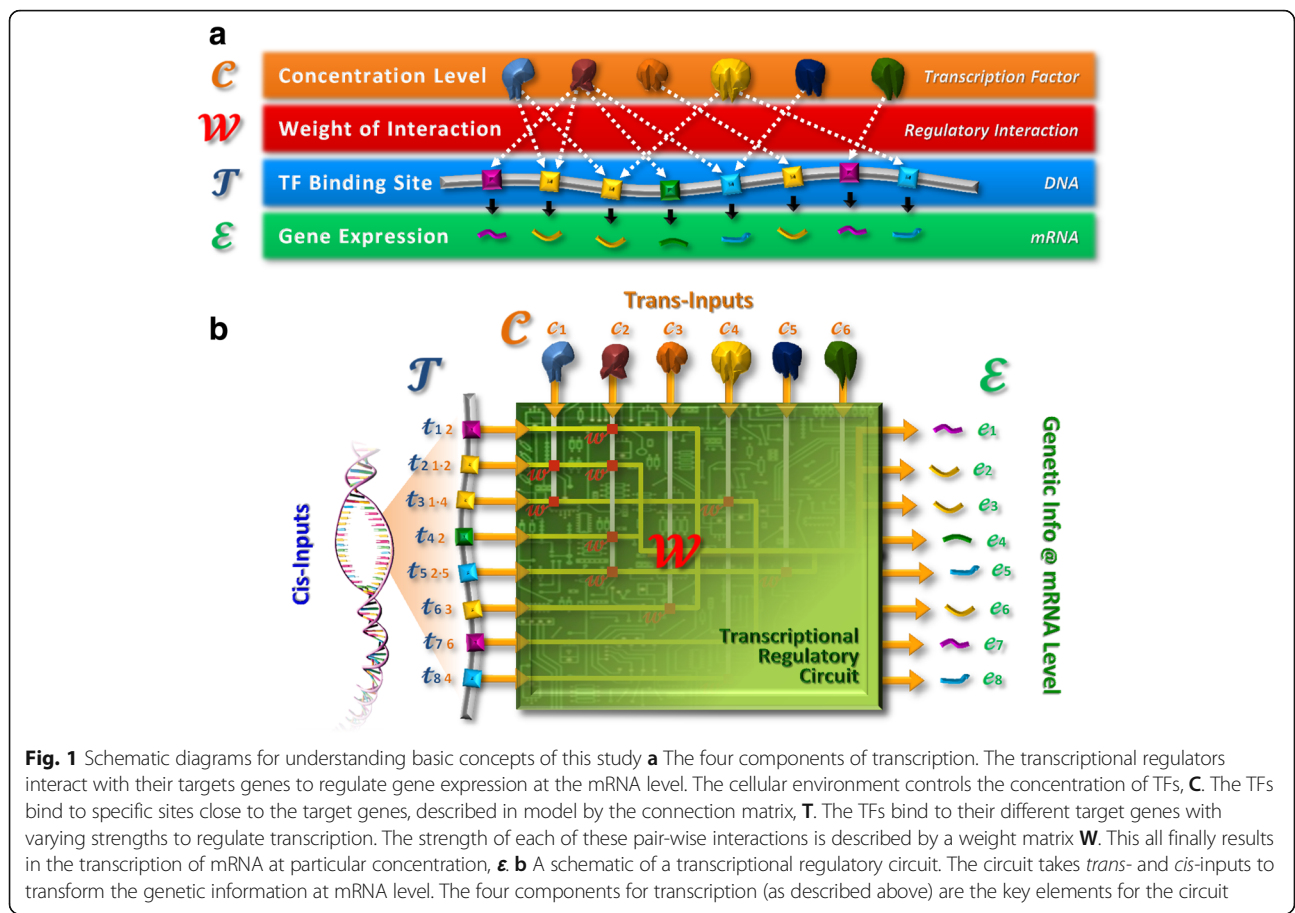
Fig. 1 Schematic diagrams for understanding basic concepts of this study **a** The four components of transcription. The transcriptional regulators interact with their targets genes to regulate gene expression at the mRNA level. The cellular environment controls the concentration of TFs, **C**. The TFs bind to specific sites close to the target genes, described in model by the connection matrix, **T**. The TFs bind to their different target genes with varying strengths to regulate transcription. The strength of each of these pair-wise interactions is described by a weight matrix **W**. This all finally results in the transcription of mRNA at particular concentration, **ε**. **b** A schematic of a transcriptional regulatory circuit. The circuit takes *trans*- and *cis*-inputs to transform the genetic information at mRNA level. The four components for transcription (as described above) are the key elements for the circuit

are a key component in the regulation of mRNA levels in the cell, and have a number of components (shown in Fig. 1): TFs, whose concentration can change, bind to TFBS upstream of genes with a strength that is a function of the particular TF-gene interaction, to control the concentration of mRNA produced. A number of mathematical models have been developed which attempt to describe these interactions [15–17, 19, 21, 30]. For example Sanguinetti et al. [4] model the log gene expression in the form:

$$\underline{e} = \underline{\underline{T}}\,\underline{\underline{W}}\,\underline{c} + \underline{v} \tag{1}$$

Where:

i) $\underline{e}$ is a set of logged gene expression measurements.
ii) $\underline{\underline{T}}$ is a binary matrix capturing the connection topology—the specific set of TFBS upstream of genes and the TFs that bind to them. If TF f binds upstream of gene g then $\underline{\underline{T}}_{gf} = 1$.
iii) $\underline{\underline{W}}$ is a weight matrix that captures the nature of the interaction strengths between TF-gene pairs in regulating expression of a specific gene.

iv) $\underline{c}$ is the vector of concentrations of each of the TFs.
v) $\underline{v}$ is a vector of independent and identically distributed variables modeling the noise in the system. The model assumes that a spherical Gaussian term could explain all noise on gene expression profiling data.

Typically, we have knowledge of $\underline{e}$ (from gene expression profiling experiments, such as microarray or RNA-seq) and would like to infer the set of TFC $\underline{c}$, and TFA $\underline{\underline{W}}$ giving rise to this signal. Given $\underline{\underline{T}}$ and $\underline{e}$ Sanguinetti et al. [4] then show how it is possible to solve for $\underline{c}$ and $\underline{\underline{W}}$ using a discrete time state space model (Eq. 1) with expectation-maximization (EM) algorithm. In the model, elements of the $\underline{c}$ matrix indicate the concentration level of a given TF protein (TFC) at a specific time. Elements of the $\underline{\underline{W}}$ matrix represent the regulatory intensity (TFA) between a given TF protein and its binding affinity to its target genes. The baseline expression level is the mean vector. The measurement noise $\underline{v}$ follows zero-mean i.i.d. Gaussian noise. To estimate the $\underline{c}$ and $\underline{\underline{W}}$ matrices, the model used posterior estimation of

Bayer's theorem. During this estimation, EM algorithm allowed the model to efficiently approximate the log likelihood. However, it is rare to have a complete knowledge of $\underline{\underline{T}}$ —we simply do not know the binding sites for all TFs in a typical higher eukaryotic cell. Recent experimental techniques, such as ChIP-chip and ChIP-seq can provide useful data to help construct the connection topology $\underline{\underline{T}}$ [31], however they have clear limitations if we are looking for a complete topology [32, 33]. A number of theoretical techniques are also available to uncover the connection topology [34–36]. The techniques generally use principals of reverse-engineering and use gene expression and genome sequence data to infer regulatory interactions.

Gene expression data

Gene expression datasets were downloaded from Gene Expression Omnibus (accession number GSE7342 for p38α and GSE36890 for STAT5) [37, 38]. The expression profiling data of GSE7342 dataset was normalized by the robust microarray average (RMA) method. The read counts of GSE36890 dataset was normalized to the reads per kilobase of exon per mega-base of library size (RPKM).

Generating $\underline{T}$, the connection topology

In this paper we have taken a conservative strategy for generating $\underline{\underline{T}}$ which looks at upstream region of genes that are well-conserved in multiple mammalian genomes. We used a published catalogue of common regulatory motifs that were overrepresented in gene upstream regions [22]. These motifs were identified by constructing genome-wide alignments for four mammalian species in promoter regions and 3' UTRs relating to well-annotated genes from the RefSeq database. The same TFs were assumed to bind the same TFBSs in mice since the TFBS had been discovered in an alignment of human, mouse, rat and dog promoter regions. TFBS upstream of human 13,330 RefSeq genes were predicted. Mouse genes corresponding to the published list of human genes [22] were identified using Ensembl mouse gene annotation.

Estimation of statistically significant changes

We were specifically interested in any TF activity that exhibits statistically significant changes between the two conditions. In particular, we are interested in changes that may be due to a change in activity of the TF, and not just in its concentration. We therefore scaled the TFA by the predicted TFC as a measure for changes in activity [24]. A joint analysis of TFA and TFC should provide more robust predictions of those TFs whose activity has changed for reasons beyond those of a simple change in concentration. To compare two different conditions, the normalized TFA by TFC in wild-type condition were subtracted by the normalized TFA in knock-out condition. We therefore determined those interactions for which:

$$\left[\left\|\frac{W_{gf}}{c_f}\right\|\right]_{WT} - \left[\left\|\frac{W_{gf}}{c_f}\right\|\right]_{KO} > \textbf{\textit{Cutoff}} \qquad (2)$$

The value of the ***Cutoff*** was chosen such that all differences at the 95% confidence interval were considered significant (±2 standard deviations). ±2SD limit is widely chosen as a normal limit because it fits well into two important categories: (1) confident interval and (2) testing hypothesis.
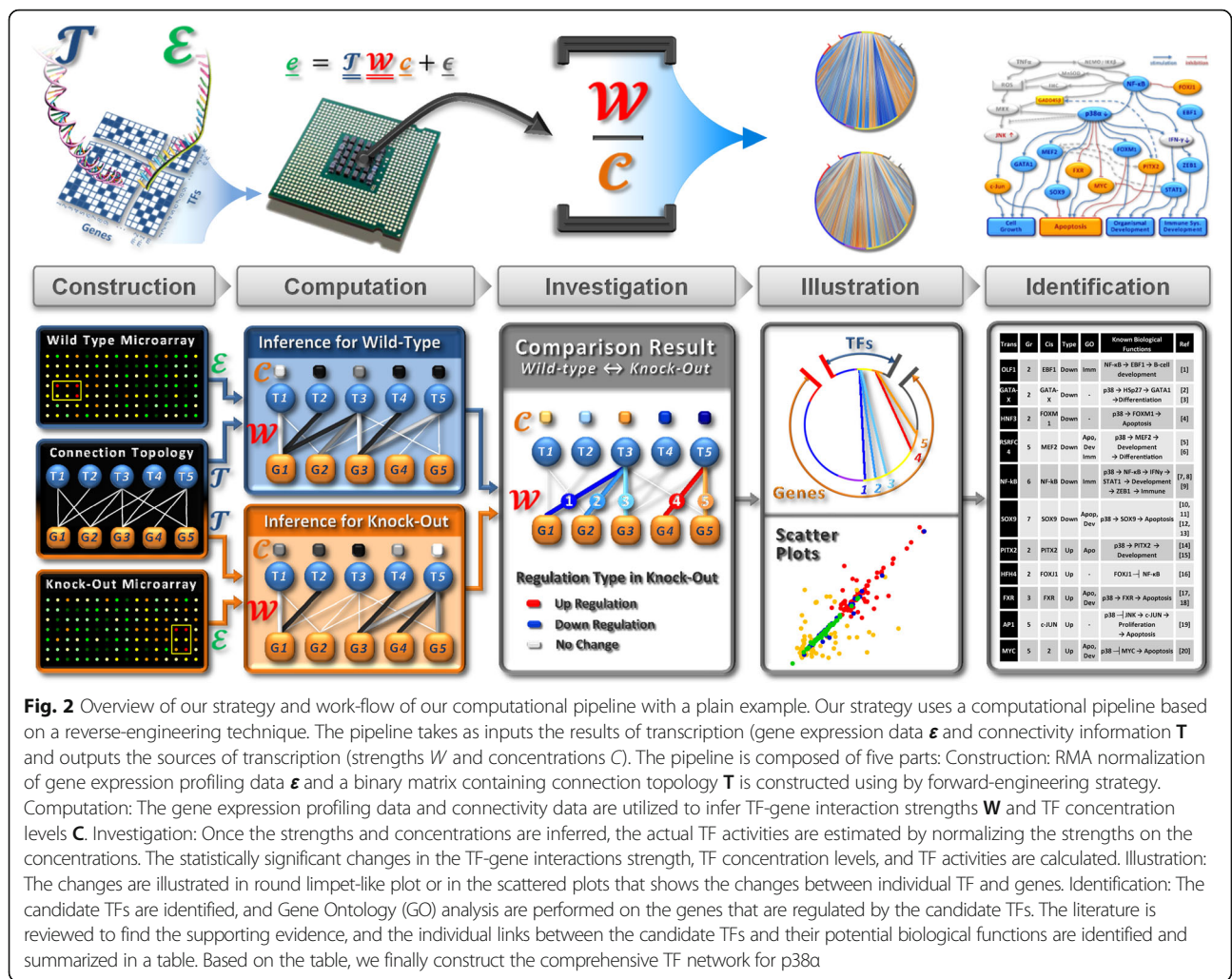
Gene Ontology (GO) analysis

GO analysis was performed by using DAVID [39]. The sets of genes showing significant changes identified in Eq. 2 were submitted to DAVID using the default parameters in order to obtain the GO term classifications of each gene. Our computational pipeline utilized the results to investigate the functionality of genes and their regulatory TFs. The detailed methods and the result figures of GO analysis are supplied in Additional file 1: Supplementary Text and Figure S1–S3.

Results

Estimating the responses to perturbation of transcription networks

We have developed a strategy which used forward-engineering to construct the connection topology (see Fig. 1, Methods, and Additional file 1: Supplementary Text and Figure S1–S3), based on a previous study of regions upstream of genes conserved in multiple mammalian genomes [22]. The structure of this network of transcriptional regulatory interactions between TFs and the genes whose transcription they control is described by a binary matrix $T \in \mathfrak{R}^{n \times m}$, where n is the number of TFs and m is the number of genes; An element (i, j) of the matrix is '1' if TF i binds to the upstream control region of gene j, '0' otherwise. We have then employed a mathematical model to integrate the connection topology data and a gene expression dataset from a higher eukaryote in which we are interested in modeling the changes that occur in the TF network in response to a change in the cellular environment (Fig. 2). Our approach could be seen as complementary to 'Integrative methods', as defined in [40], as it provides a strategy for creating an approximate connection topology if more detailed information is not available. The connection topology that is being used for this analysis contains many approximations and is certainly incomplete. However, it should be noted that we are looking at the differences between the models, for example

Trans	Gr	Cis	Type	GO	Known Biological Functions	Ref
OLF1	2	EBF1	Down	Imm	NF-κB → EBF1 → B-cell development	[1]
GATA-X	2	GATA-X	Down		p38 → Hsp27 → GATA1 → Differentiation	[2] [3]
HNF3	2	FOXM1	Down		p38 → FOXM1 → Apoptosis	[4]
RSRFC4	5	MEF2	Down	Apo, Dev, Imm	p38 → MEF2 → Development → Differentiation	[5] [6]
NF-kB	6	NF-kB	Down	Imm	p38 → NF-κB → IFN-γ → STAT1 → Development → ZEB1 → Immune	[7, 8] [9]
SOX9	7	SOX9	Down	Apop, Dev	p38 → SOX9 → Apoptosis	[10, 11] [12, 13]
PITX2	2	PITX2	Up	Apo	p38 → PITX2 → Development	[14] [15]
HFH4	2	FOXJ1	Up		FOXJ1 -| NF-κB	[16]
FXR	3	FXR	Up	Apo, Dev	p38 → FXR → Apoptosis	[17, 18]
AP1	5	c-JUN	Up		p38 -| JNK → c-JUN → Proliferation → Apoptosis	[19]
MYC	5	2	Up	Apo, Dev	p38 -| MYC → Apoptosis	[20]

Fig. 2 Overview of our strategy and work-flow of our computational pipeline with a plain example. Our strategy uses a computational pipeline based on a reverse-engineering technique. The pipeline takes as inputs the results of transcription (gene expression data ε and connectivity information **T** and outputs the sources of transcription (strengths W and concentrations C). The pipeline is composed of five parts: Construction: RMA normalization of gene expression profiling data ε and a binary matrix containing connection topology **T** is constructed using by forward-engineering strategy. Computation: The gene expression profiling data and connectivity data are utilized to infer TF-gene interaction strengths **W** and TF concentration levels **C**. Investigation: Once the strengths and concentrations are inferred, the actual TF activities are estimated by normalizing the strengths on the concentrations. The statistically significant changes in the TF-gene interactions strength, TF concentration levels, and TF activities are calculated. Illustration: The changes are illustrated in round limpet-like plot or in the scattered plots that shows the changes between individual TF and genes. Identification: The candidate TFs are identified, and Gene Ontology (GO) analysis are performed on the genes that are regulated by the candidate TFs. The literature is reviewed to find the supporting evidence, and the individual links between the candidate TFs and their potential biological functions are identified and summarized in a table. Based on the table, we finally construct the comprehensive TF network for p38α

between a wild-type and knock-out state, and those differences will be in parts of the model for which we do have data.

The results of our approach provide a set of TFs and their target genes which are related by significant up- or down-regulation in transcription. It provides a clear indication of the changes in TFA and TFC of TFs that are controlling transcriptional regulatory mechanisms in response to a specific stimulus. We therefore showed an "integrated" approach for network inference, based on a forward-engineered connection topology, can produce plausible and testable hypotheses about the responses to perturbation of transcription networks in higher eukaryotes.

Illustrating interpretable images of complex data
The visualization tools then make patterns apparent that would be difficult to detect in numerical data (Fig. 2). To distinguish regulation patterns between different experimental conditions, recognizing at a glance is

important. However, computing results are formed in large numerical matrix, thus it is not only difficult to navigate through the whole matrix but also impossible to present the results in one page.

Figure 3 shows a graphical representation of the significant changes in TFA matrices $\underline{\underline{W}}$ (n by m) and TFC vectors $\underline{c}$ (n) obtained from this analysis. The patterns of the responses to perturbation of TF networks are readily observed in this single-shot image that presents approximately 2000 significant changes of varying TF activities on the 132,654 TFBSs after deleting p38α. In upper part of the plots, the TFs place in the functional group order. The genes, which have at least one significant interaction with TFs, locate in bottom part of the plots. A line in the plots presents a regulatory interaction (normalized TFA by TFC) between a TF and its target gene, and line color indicates a significant difference between the strengths of the regulatory interaction of two conditions. For example, we can easily find in visualized format (Fig. 3) that TF group three has
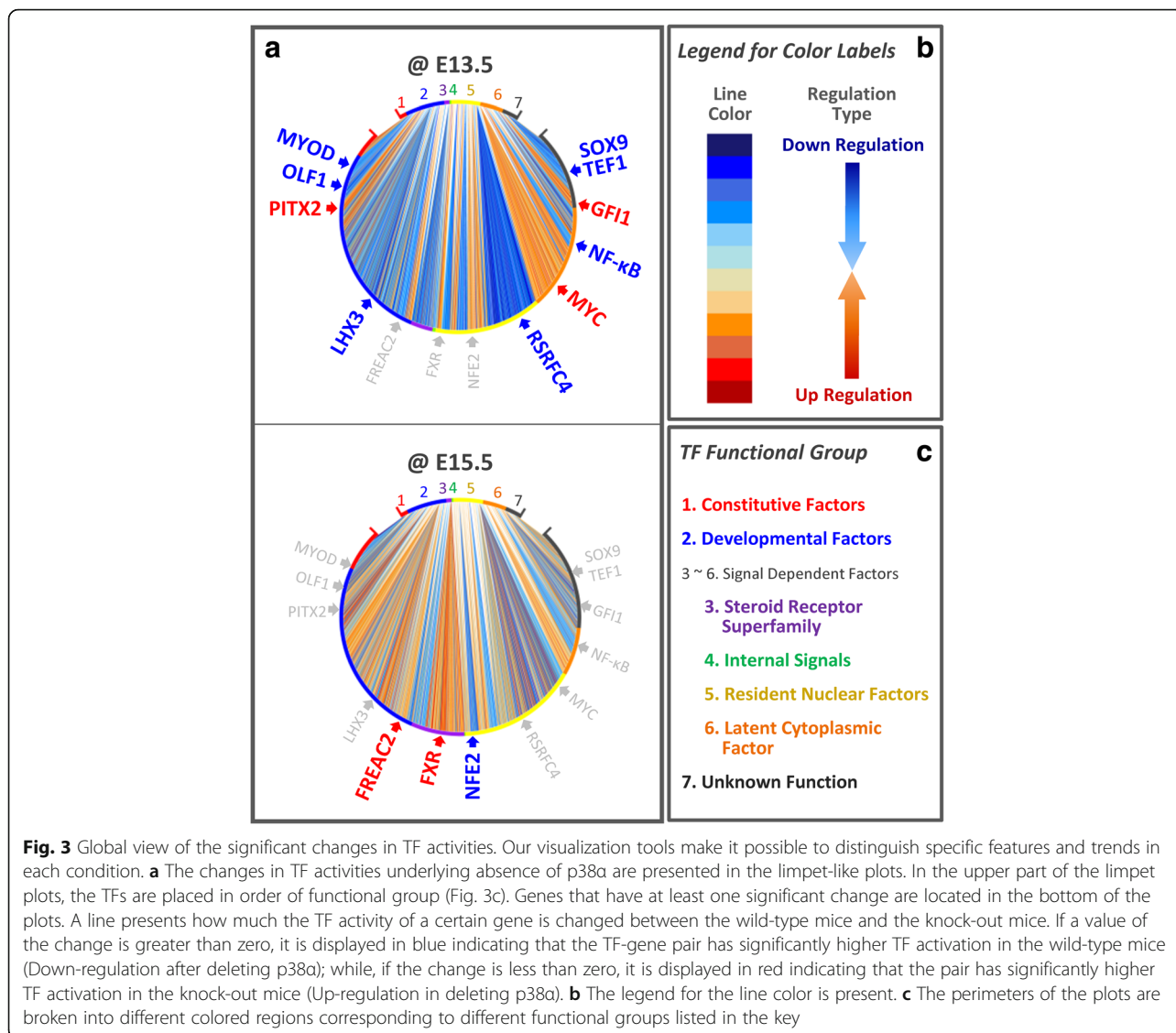
Fig. 3 Global view of the significant changes in TF activities. Our visualization tools make it possible to distinguish specific features and trends in each condition. **a** The changes in TF activities underlying absence of p38α are presented in the limpet-like plots. In the upper part of the limpet plots, the TFs are placed in order of functional group (Fig. 3c). Genes that have at least one significant change are located in the bottom of the plots. A line presents how much the TF activity of a certain gene is changed between the wild-type mice and the knock-out mice. If a value of the change is greater than zero, it is displayed in blue indicating that the TF-gene pair has significantly higher TF activation in the wild-type mice (Down-regulation after deleting p38α); while, if the change is less than zero, it is displayed in red indicating that the pair has significantly higher TF activation in the knock-out mice (Up-regulation in deleting p38α). **b** The legend for the line color is present. **c** The perimeters of the plots are broken into different colored regions corresponding to different functional groups listed in the key

distinct patterns (down-regulation at E13.5, up-regulation at E15.5) between two time points.

Modelling the changes of transcription factor network in p38α deficient mice

The computational pipeline as highlighted in Fig. 2 was applied to a published study of the effect p38α knock-out in mouse embryos [37]. This study developed four gene expression profiling datasets (Gene Expression Omnibus, accession number GSE7342) comprising of two time points at days 13.5 and 15.5 of embryonic development (E13.5 and E15.5) for p38α knock-outs and their wild-type controls. This data set was chosen for this study as it includes experimental measurements of gene expression in the wild-type and knock-out mice and showed that p38α deficient mice have significantly

different phenotype. Thus, the experimental datasets were used as positive controls for our theoretical study.

The TF-gene interaction strengths (TFAs) $\underline{\underline{W}}$ and TF concentration levels (TFCs) $\underline{c}$ in each of these four data sets were then inferred to produce four weight matrices of TFAs:

$$\left[\underline{\underline{W}}\right]_{WT@E13.5}, \left[\underline{\underline{W}}\right]_{WT@E15.5}, \left[\underline{\underline{W}}\right]_{KO@E13.5}, \left[\underline{\underline{W}}\right]_{KO@E15.5}$$

and four concentration vectors of TFCs:

$$\left[\underline{c}\right]_{WT@E13.5}, \left[\underline{c}\right]_{WT@E15.5}, \left[\underline{c}\right]_{KO@E13.5}, \left[\underline{c}\right]_{KO@E15.5}$$

From the TFA weight $\underline{\underline{W}}$ and connection topology $\underline{\underline{T}}$ matrices, the average strength of TFA, S_f for each TF in the datasets was calculated:

$$\mathcal{S}_f = \frac{\Sigma_g |W_{gf}|}{\Sigma_g |\mathcal{T}_{gf}|} \qquad (3)$$

By comparing the average strengths between wild-type and knock-out mice, it is possible to see which of the TFs have significantly changed as a consequence of the removal of p38α.

Figure 4a, b show the changes in TFA strengths $\mathcal{S}_f$ between wild-type and knock-out mice at E13.5 and E15.5. It can be seen that a number of TFs show a significant signal (>2 s.d.) in this data. These are shown with more detail in Table 1. Figure 4c, d show the inferred TFCs $\underline{c}$ obtained for the E13.5 and E15.5 time points. Again, from this graph it is possible to see that a number of TFs appear to be responding to the p38α status. These are shown in more detail in Table 2.

Transcriptional regulatory network for p38α

Gene Ontology (GO) analysis on the target genes of the TFs with strongly changed activity showed enrichment for three GO terms and provided insight into the functional role of the TFs (see Methods, Tables 1 and 2, and Additional file 1: Supplementary Text and Figure S1–S3). The three GO terms are the regulations of the apoptosis (programmed cell death), the downward spiral of the developmental process, and the immune system development. The JNK-c-Jun pathway stimulates the apoptosis, and the I-kB kinase/NF-kB cascade acts as a suppressor of the JNK-c-Jun pathway [41]. Inhibition of p38α MAPK retards another JNK-c-Jun pathway inhibitor NF-kB cascade, but promotes JNK-c-Jun pathway which induces the apoptosis by expressing the Bcl2 protein family [20, 42]. On the other hand, developmental process related genes are down-regulated in the p38α knock-out mice. The study of p38α MAPK [37] reported that the p38α knock-out mice die within days after birth. We do not have enough gene expression profiling data (either other time points in the embryonic period or postnatal period) to investigate TFAs in whole developmental process of the p38α knock-out mice; we cannot confirm but suppose that it might be the reason of the death of the knock-out mice. Further, the genes interact with the TFs which are reported as crucial TFs in the developmental
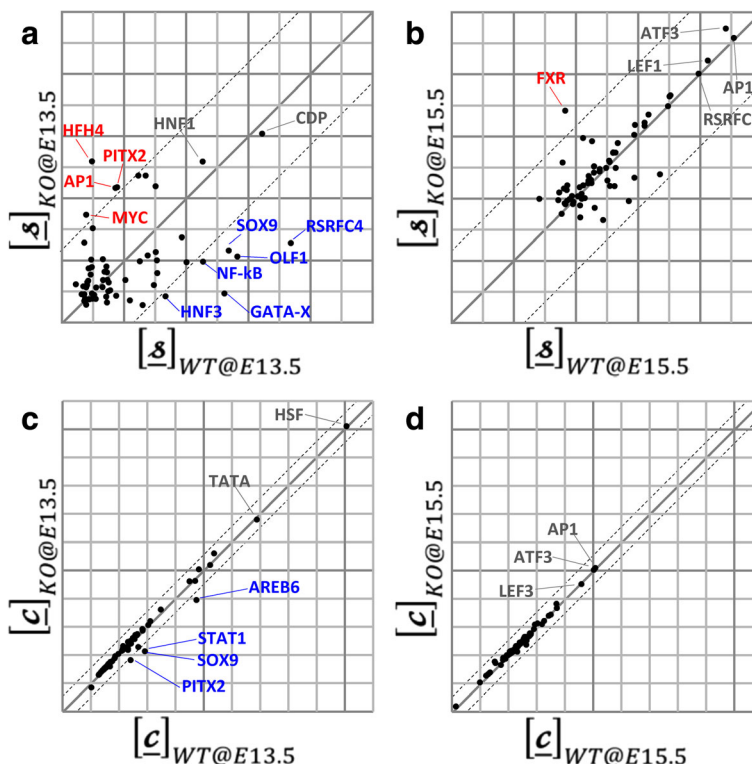

Fig. 4 a, b The average strength of changes in TF-gene interaction S and TF concentration levels c between wild-type and knock-out mice. The figures clearly show not only which TFs have strong interaction strengths or high concentrations (gray-colored TFs) but also which TFs have significant changes in their interaction pattern or concentration (blue- or red-colored TFs). The dotted lines indicate the standard deviation (=2) centered on the median value of the straight lines. (figure a for time point E13.5 and figure b for time point E15.5). Five TFs (shown in red) interact particularly strongly with their target genes in the p38α knout-mice. In contrast, six TFs (shown in blue) interact less-strongly in the p38α knock-out than wild-type. c, d The TF concentration levels 뷀 of wild-type and knock-out mice. TF concentration levels are plotted. The strongest signal was observed in E13.5, only. Deleting p38α induces a down regulation of AREB6, PITX2, STAT1 and SOX9

Table 1 TFs showing significant changes in interaction strength between wild-type and knock-out mice

Trans name	TF group	*Cis* name	Regul-ation type	GO analy-sis	Known biological functions	Ref.
OLF1	1.Dev	EBF1	Down	Imm	NF-κB → **EBF1** → B-cell development	[52]
GATA-X	1. Dev	GATA1 ~6	Down	-	p38 → HSp27 **GATA1** → Differentiation → **GATA1** → IL9 → Asthma	[53, 54]
HNF3	1. Dev	FOXM1	Down	-	p38 → **FOXM1** → Apoptosis	[45]
RSRFC4	5. Res	MEF2	Down	Apo, Dev Imm	p38 → **MEF2** → Development → Differentiation	[46, 55]
NF-κB	6. Lat	NF-κB	Down	Imm	p38 → **NF-κB** → IFNγ → STAT1 → Development → ZEB1 → Immune	[56–58]
SOX9	7. Unk	SOX9	Down	Apop, Dev	p38 → **SOX9** → Apoptosis	[59–62]
PITX2	1. Dev	PITX2	Up	Apo	p38 → **PITX2** → Development → Apoptosis	[63, 64]
HFH4	1. Dev	FOXJ1	Up	-	**FOXJ1** ⊣ NF-κB	[65]
FXR	3. Ste	FXR	Up	Apo, Dev	p38 → **FXR** → Apoptosis	[66, 67]
AP1	5. Res	c-JUN	Up	-	p38 ⊣ JNK → **c-JUN** → Proliferation → Apoptosis	[37]
MYC	5. Res	MYC	Up	Apo, Dev	p38 ⊣ **MYC** → Apoptosis	[68]

The TFs predicted to have significantly different behaviors between the wild-type and knock-out mice. "Trans Name"—the official gene symbol of the TF. "Cis Name"—the name given to the binding site of the TF. TFs were characterized into different functional groupings (see Fig. 3c for details: Dev—cell-type specific developmental TFs, Res—signal dependent resident nuclear factors, Lat—signal dependent latent cytoplasmic factors, Ste—signal dependent steroid receptor group, Unk—unknown). "Regulation Type", the way in which the TF regulates its target genes in the absence of p38α. "GO Analysis", provides more functional classification for the identified TFs (see Methods and Additional file 1: Figure S1 to S3 for details). The abbreviations of the GO terms are: Apo, Apoptosis; Dev, Developmental Process; Imm, Immune System Development. "Known Biological Functions" summarizes the findings from the recent biological literature as shown in the "References"
The boldface ones are the main node in Fig. 5

process and the immune system development. Our results are therefore in broad accordance with the experimentally validated results, so it confirmed that our pipeline produces reliable results.

Combining the data in Tables 1 and 2 with those obtained from the literature it is possible to build a putative model for the effects of p38α knock-out (Fig. 5). This figure shows TFs with a strong response in our analysis as nodes, with links that demonstrate regulatory interactions between them. The TF network therefore comprehensively shows the biological consequences of p38α knock-out at transcriptional level.

Discussion
We have developed a novel strategy for discovering changes in transcriptional regulatory networks of higher eukaryotes. It integrates methods for inferring TF-gene interaction strengths (TFAs) and TF concentration levels (TFCs); identifying statistically significant changes in TFAs and TFCs; analyzing the changes; classifying TFs into functional

groups; and visualizing the changes. To our knowledge, this is the first ensemble approach for characterizing the transcriptional function of TF proteins and their target genes in higher eukaryotes. Reverse-engineering of TF networks has been well developed in the lower eukaryotes [15, 17]. However, the problems in mapping the regulatory mechanisms in cells of higher eukaryotes have made such global studies either impossible or impractical. Some recent studies have begun to address this issue [16, 19, 30], but have tended to focus only on understanding which TFs bind to which genes—not looking in detail at the nature of the TF-gene interaction. Other studies [5, 21] identified key biological features in transcriptional changes, however the methods have difficulties in inferring the dynamics of the interactions. A recent review [40] has categorized techniques for network inference and listed their limitations.

We validated our computational pipeline using the p38α gene expression profiling data and our connectivity data. The study of p38α MAPK [37] used various experimental methods including a gene expression profiling analysis to

Table 2 TFs showing significant changes in concentration between wild-type and knock-out mice

Trans name	TF group	*Cis* name	Level	GO analy-sis	Known biological functions	Ref.
AREB6	1.Dev	ZEB1	Down	-	p38 → IFNγ → **ZEB1** → Immune	[58]
PITX2	1.Dev	PITX2	Down	Apo	p38 → **PITX2** → Development → Apoptosis	[63, 64]
STAT1	6. Lat	STAT1	Down	-	**STAT1** → Development → Immune	[57, 69]
SOX9	7. Unk	SOX9	Down	Apo, Dev	p38 → **SOX9** → Apoptosis	[59–62]

TFs changing their concentration levels significantly between wild-type and knock-out mice. "Level", the changes of TF concentration level in the absence of p38α. Other column headings and abbreviations are the same as those in Table 1
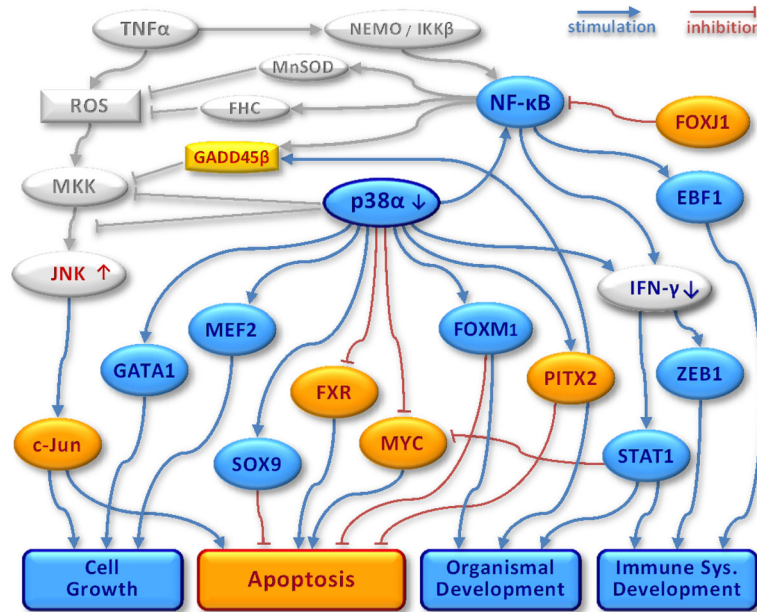The boldface ones are the main node in Fig. 5

Fig. 5 A comprehensive transcriptional regulatory network for the p38α. The TFs depicted in gray is already known to p38α [41, 51]. The TFs identified in our analysis were then added to the figure and colored blue if their activity was down regulated in the absence of p38α, and colored orange if their activity was up regulated in the absence of p38α. Lines in the figure represent interactions that are known in the literature (listed in more detail in Tables 1 and 2)

show that p38α negatively regulates cell proliferation by antagonizing the JNK-c-Jun pathway. We utilized the published gene expression profiling dataset from their study, to demonstrate that our computational pipeline is able to infer from the gene expression profiling data the same *in-silico* conclusions that the authors obtained from their in-vitro experiments. Therefore, our analysis focused on the JNK-c-Jun pathway to validate the accuracy, robustness and reliability of our strategy. The results are consistent with the experimentally validated inhibitory effect of p38α on transcriptional networks [37]. Their published data confirmed that the most important TF involved in the response to the knock-out was c-Jun, with a clear change observed in both its activation and concentration. In our theoretical work, we also showed a significant change in TFA of c-Jun, but we did not see any corresponding change in the predicted TFC, which is disappointing.

The p38α MAPK is one of many signal transduction pathways and works in both cell-type specific and cell-context specific manner. It plays a pivotal role in converting extra-cellular signal into a wide range of cellular response [43]. We classified a set of TFs that responded to the deletion of p38α into functional groups (Tables 1 and 2, Fig. 3c), that are either developmental factors (group 2) or extra-cellular signal dependent factors (group 3). Developmental factors are also dependent on extra-cellular signals because cells may require such signals to generate developmental

factors [44]. In Fig. 3a, it can be seen that the main factors that responded to the knock-out are the extra-cellular signal dependent factors. None of the TFs that significantly respond in the knock-out are constitutive factors. Our results are consistent with recent publications on the JNK-c-Jun pathway (see citations in Tables 1 and 2).

Our analyses generated a comprehensive transcriptional regulatory network for p38α. The network and a detailed description are shown in Fig. 5. The nodes in the graph were generated from our analysis of responding TFs. The edges in this network were derived from the literature or GO analysis (citations in Tables 1 and 2). The edges or links in the network of p38α regulated TFs have mostly been previously reported, but none of the reports had integrated all these p38α related TFs into a single comprehensive network diagram. Together these results predict a set of TFs that are in some way regulated by p38α, a set somewhat larger than that identified in the original paper. For example, we predict that Foxm1 (HNF3) responds to the p38α status. Recent papers, published since the original study, provide some support for this hypothesis [45, 46]. Most parts of the network are reported in numerous biological studies. However, our network reveals novel links such as p38α—FXR and p38α—MYC. The inferred links are supported by direct experimental evidence so validating the approach, but that in addition novel links have been proposed that are now testable.

The data shown in Tables 1 and 2 and visualized in Fig. 5, provide evidence that the methodology described in this paper is capable of generating plausible hypotheses about linkage between p38α and a range of different TFs. The hypotheses presented in this table have been generated solely from our input data (the connection topology data and gene expression profiling data), but are well-supported by the literature. The methodology has therefore demonstrated that it can produce plausible and testable hypotheses, even if the specific details of those interactions may not be completely accurate. This is not surprising given the fact that we only have an incomplete model of the transcription process. Any *in-silico* techniques which uses predicted TF/TFBSs interactions can provide only a limited view of the complete complexity of transcription control due to the nature of the binding between the TF and the TFBS and the complex effect of gene expression on the TFBS—for example dependent on the epigenetic factors, such as the pattern of histones or DNA methylation at the binding site—as well as the state and concentration of the TF itself. Analysis is complicated by the fact that there are other processes in the cell that act to control mRNA concentration. Such as the rate of RNAi regulated mRNA degradation [9, 10] or susceptibility to attack by RNAses [3, 11]. TFBS can hidden by histones [7, 8], or made more accessible by genomic uncoiling [6]. Furthermore, most TF binding may be cell or species specific not all sites are functional even if occupied, and many functional sites have low levels of conservation [47]. This rather undermines the commonly accepted assumption that TFBSs can be discovered by conservation [22]. However although the exact binding sites may not be conserved the set of TFs that bind a gene somewhere probably is.

p38α deficient mice showed significantly different phenotype which indicates its role is critical. The p38α study also provided the gene expression profiling dataset of wild-type mice as well as p38α deficient mice, so that we could apply our pipeline on the dataset to investigate TFAs and TFCs. It allowed us to directly compare our *in-silico* results to the experimental in-vitro results, and it validated our findings. However, the experiment was done on two time-points that could limit our validation. Thus, we tested our pipeline on a larger dataset from a recent STAT5 transcription factor study [38] which is consisted of 18 samples in five time-points. This study showed the critical role of STAT5-tetramer in immune system. To do this, the authors made STAT5-tetramer deficient mice by generating STAT5A-STAT5B double-knockin mice. Interleukin 2 (IL2) and IL15 are two of well-known upstream regulators of STAT5A-STATB, so they measured IL2- and IL15-induced gene expression profiling in both wild-type mice and STAT5-tetramer deficient mice. We downloaded the RNA-seq gene

expression dataset from this study and analyzed with our pipeline. TF activities were decreased in STAT5-tetramer deficient mice (both IL2- and IL15-induced), particularly at 4, 24, 48 h (Fig. 6). This general trend is well-corresponded to the experimental findings as the author reported IL2- and IL15-induced gene expression were both down-regulated in STAT5-tetramer deficient
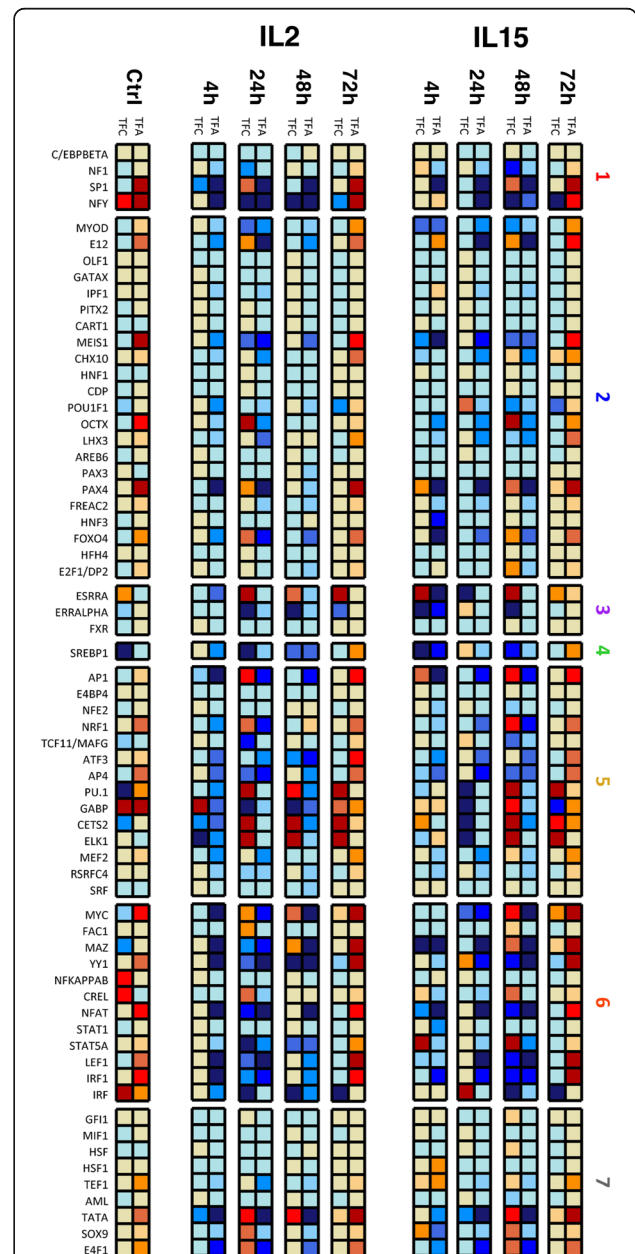

Fig. 6 TFA and TFC changes in STAT5-tetramer deficient mice. TFA and TFC of 65 TFs were estimated from IL2- and IL5-induced RNA-seq datasets and compared between wild-type and STAT5-tetramer deficient mice. Thus, TFA or TFC of a given TF is shown in *red* color if it is higher in STAT5-tetramer deficient mice than wild-type mice. If the level of TFA or TFC is higher, the color is darker. The numbers in right-side of heat-map indicates TF functional group (please see legend in Fig. 3c)

mice. However, most TFs were re-activated at the last time-points which is also exactly same observation in the experimental result. In particular, STAT5A is in our TF list so we closely investigated its activation patterns. STAT5A showed weak activities and concentration level at control sample, but it dramatically de-activated at 4, 24, 48 h and then re-activated at 72 h. More interestingly, all TFs in cytoplasmic factor group (marked as number 6 in Fig. 6) including STAT5A have same activation patterns with STAT5. Moreover, there are a few interesting TFs which are not classified as cytoplasmic factor but also followed same up- and down-regulated patterns with STAT5 (e.g. SP1, NFY, E12, MEIS1, PAX4, AP1, NRF1, TCF11, AP4, GABP, TATA, E4F1). We considered these are new findings and could lead new insight and testable hypotheses.

Conclusions

Our objective was to develop an effective computational pipeline which produces reliable and explicit models of transcriptional regulatory networks. Even though the TFBS information is incomplete due to the difficulties in identifying them, our pipeline predicts new biological hypotheses on a genome-wide scale by combining TFBS and gene expression information. TIGERi is publicly available as a stand-alone GUI software, so ones who have their own gene expression profiling data could easily use the TIGERi software to analyze their data on their fingertips. It would facilitate transcriptional gene regulation researches in the biomedical community.

Our approach can be applied to other gene expression datasets to provide a display of the transcriptional regulatory networks and identify novel candidate genes and TFs underlying specific phenotypes. For example, our methodology has been successfully applied to three recent studies [48–50]. The pipeline would be particularly valuable if it were run on large-scale multi-time point genomic data. It is also the case that we might expect the method to become increasingly predictive with improved connection topologies created from large scale experimentally validated TF/TFBS datasets as opposed to those generated from simple conservation data.

Additional file

Additional file 1: Supplementary Text and Figures. (DOC 7227 kb)

Availability of data and materials
TIGERi is a standalone GUI software for all platforms. The TIGERi software and manual can be freely downloaded at https://github.com/namshik/tigeri/.

Authors' contributions
Conceptualization, NH, HAN and AB; Methodology, NH; Software, NH; Analysis, NH and AB; Writing, NH, HAN and AB; Supervision, AB; Funding Acquisition, NH and AB All authors have read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
NA.

Ethics approval and consent to participate
NA.

About this supplement
This article has been published as part of BMC Bioinformatics Volume 18 Supplement 7, 2017: Proceedings of the Tenth International Workshop on Data and Text Mining in Biomedical Informatics. The full contents of the supplement are available online at https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-18-supplement-7.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Gurdon Institute, University of Cambridge, Cambridge, UK. [2]School of Biological Sciences, University of Liverpool, Liverpool, UK. [3]School of Computer Science and School of Health Sciences, University of Manchester, Manchester, UK.

Published: 31 May 2017

References
1. Olivera DJ, Nikolaub B, Wurtelea ES. Functional genomics: high-throughput mRNA, protein, and metabolite analyses. Metab Eng. 2002;4(1):98–106.
2. Scott MP. Development: the natural history of genes. Cell. 2000;100(1):27–40.
3. Amrani N, Sachs MS, Jacobson A. Early nonsense: mRNA decay solves a translational problem. Nat Rev Mol Cell Biol. 2006;7:415–25.
4. Sanguinetti G, Lawrence ND, Rattray M. Probabilistic inference of transcription factor concentrations and gene-specific regulatory activities. Bioinformatics. 2006;22(22):2755–81.
5. Cheng C, Yan X, Sun F, Li LM. Inferring activity changes of transcription factors by binding association with sorted expression profiles. BMC Bioinf. 2007;8:452.
6. Strick TR, Croquette V, Bensimon D. Single-molecule analysis of DNA uncoiling by a type II topoisomerase. Nature. 2000;404:901–4.
7. Jagannathan I, Cole HA, Hayes JJ. Base excision repair in nucleosome substrates. Chromosome Res. 2006;14:27–37.
8. Sonehara H, Nagata M, Aoki F. Roles of the first and second round of DNA replication in the regulation of zygotic gene activation in mice. J Reprod Dev. 2008;54(5):381–4.
9. Keene JD. RNA regulons: coordination of post-transcriptional events. Nat Rev Genet. 2007;8:533–43.
10. Anderson P, Kedersha N. RNA granules: post-transcriptional and epigenetic modulators of gene expression. Nat Rev Mol Cell Biol. 2009;10:430–6.
11. Brogna S, Wen J. Nonsense-mediated mRNA decay (NMD) mechanisms. Nat Struct Mol Biol. 2009;16:107–13.
12. Marko-Varga G. Pathway proteomics: global and focused approaches. Am J Pharmacogenomics. 2005;5(2):113–22.
13. Gardner TS, Faith JJ. Reverse-engineering transcription control networks. Phys Life Rev. 2005;2(1):65–88.
14. Bansal M, Belcastro V, Ambesi-Impiombato A, Bernardo D. How to infer gene networks from expression profiles. Mol Syst Biol. 2007;3(78):78–87.

15. Herrgård MJ, Lee B-S, Portnoy V, Palsson BØ. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in Saccharomyces cerevisiae. Genome Res. 2006;16(5):627–35.

16. Ramsey SA, Klemm SL, Zak DE, Kennedy KA, Thorsson V, Li B, Gilchrist M, Gold ES, Johnson CD, Litvak V, et al. Uncovering a macrophage transcriptional program by integrating evidence from motif scanning and expression dynamics. PLoS Comput Biol. 2008;4(3):e1000021.

17. Ye C, Galbraith SJ, Liao JC, Eskin E. Using network component analysis to dissect regulatory networks mediated by transcription factors in yeast. PLoS Comput Biol. 2009;5(3):e1000311.

18. Pham H, Ferrari R, Cokus SJ, Kurdistani SK, Pellegrini M. Modeling the regulatory network of histone acetylation in saccharomyces cerevisiae. Mol Syst Biol. 2007;3:153.

19. Gatta GD, Bansal M, Ambesi-Impiombato A, Antonini D, Missero C, Bernardo D. Direct targets of the TRP63 transcription factor revealed by a combination of gene expression profiling and reverse engineering. Genome Res. 2008;18(6):939–48.

20. Cai B, Chang SH, Becker EBE, Bonni A, Xia Z. p38 MAP kinase mediates apoptosis through phosphorylation of BimEL at ser-65. J Biol Chem. 2006;281(35):25215–22.

21. Oliveira AP, Patil KR, Nielsen J. Architecture of transcriptional regulatory circuits is knitted over the topology of bio-molecular interaction networks. BMC Syst Biol. 2008;2:17.

22. Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, KLindblad-Toh V, Lander ES, Kellis M. Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. Nature. 2005;434:338–45.

23. Sanguinetti G, Lawrence ND, Rattray M. A probabilistic dynamical model for quantitative inference of the regulatory mechanism of transcription. Bioinformatics. 2006;22(14):1753–9.

24. Driever W, Thoma G, Nüsslein-Volhard C. Determination of spatial domains of zygotic gene expression in the drosophila embryo by the affinity of binding sites for the bicoid morphogen. Nature. 1989;340:363–7.

25. Liao JC, Boscolo R, Yang Y-L, Tran LM, Sabatti C, Roychowdhury VP. Network component analysis: reconstruction of regulatory signals in biological systems. Proc Natl Acad Sci. 2003;100(26):15522–7.

26. Alter O, Golub GH. Integrative analysis of genome-scale data by using pseudoinverse projection predicts novel correlation between DNA replication and RNA transcription. Proc Natl Acad Sci. 2004;101:16577–82.

27. Gao F, Foat BC, Bussemaker HJ. Defining transcriptional networks through integrative modeling of mRNA expression and transcription factor binding data. BMC Bioinforma. 2004;5(31):31.

28. Boulesteix A-L, Strimmer K. Predicting transcription factor activities from combined analysis of microarray and ChIP data: a partial least squares approach. Theor Biol Med Model. 2005;2(23):23.

29. Kim HD, Shay T, O'Shea EK, Regev A. Transcriptional regulatory circuits - predicting numbers from alphabets. Science. 2009;325:429–32.

30. Honkela A, Girardot C, Gustafson EH, Liu Y-H, Furlong EEM, Lawrence ND, Rattray M. Model-based method for transcription factor target identification with limited data. Proc Natl Acad Sci U S A. 2010;107(17):7793–8.

31. Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan K-K, Cheng C, Mu XJ, Khurana E, Rozowsky J, Alexander R, et al. Architecture of the human regulatory network derived from ENCODE data. Nature. 2012;489(7414):91–100.

32. Buck MJ, Lie JD. ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. Genomics. 2004;83(3):349–60.

33. Park PJ. ChIP-seq: advantages and challenges of a maturing technology. Nat Rev Genet. 2009;10(10):669–80.

34. Zhu Z, Shendure J, Church GM. Discovering functional transcription-factor combinations in the human cell cycle. Genome Res. 2005;15:848–55.

35. Chang L-W, Nagarajan R, Magee JA, Milbrandt J, Stormo GD. A systematic model to predict transcriptional regulatory. Genome Res. 2006;16:405–13.

36. Davies SR, Chang L-W, Patra D, Xing X, Posey K, Hecht J, Stormo GD, Sandell LJ. Computational identification and functional validation of regulatory motifs in cartilage-expressed genes. Genome Res. 2007;17:1438–47.

37. Hui L, Bakiri L, Mairhorfer A, Schweifer N, Haslinger C, Kenner L, Komnenovic V, Scheuch H, Beug H, Wagne EF. p38alpha suppresses normal and cancer cell proliferation by antagonizing the JNK–c-Jun pathway. Nat Genet. 2007;39:741–9.

38. Lin JX, Li P, Liu D, Jin HT, He J, Ata Ur Rasheed M, Rochman Y, Wang L, Cui K, Liu C, et al. Critical role of STAT5 transcription factor tetramerization for cytokine responses and normal immune function. Immunity. 2012;36:586–99.

39. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44–57.

40. Smet RD, Marchal K. Advantages and limitations of current network inference methods. Nat Rev Microbiol. 2010;8(10):717–29.

41. Perkins ND. Integrating cell-signalling pathways with NF-kappaB and IKK function. Nat Rev Mol Cell Bio. 2007;8:49–62.

42. Brichese L, Cazettes G, Valette A. JNK is associated with Bcl-2 and PP1 in mitochondria: paclitaxel induces its activation and its association with the phosphorylated form of Bcl-2. Cell Cycle. 2004;3(10):1312–9.

43. Wagner EF, Nebreda ÁR. Signal integration by JNK and p38 MAPK pathways in cancer development. Nat Rev Cancer. 2009;9(8):537–49.

44. Brivanlou AH, Jr JED. Signal transduction and the control of gene expression. Science. 2002;295(5556):813–8.

45. Behren A, Muhlen S, Sanhueza GAA, Schwager C, Plinkert PK, Huber PE, Abdollahi A, Simon C. Phenotype-assisted transcriptome analysis identifies FOXM1 downstream from Ras-MKK3-p38 to regulate in vitro cellular invasion. Oncogene. 2010;29(10):1519–30.

46. Sanchez-Calderon H, Rosa LR-d, Milo M, Pichel JG, Holley M, Varela-Nieto I. RNA microarray analysis in prenatal mouse cochlea reveals novel IGF-I target genes: implication of MEF2 and FOXM1 transcription factors. PLoS One. 2010;5(1):e8699.

47. Schmidt D, Wilson MD, Ballester B, Schwalie PC, Brown GD, Marshall A, Kutter C, Watt S, Martinez-Jimenez CP, Mackay S, et al. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. Science. 2010;328(5981):1036–40.

48. McMaster A, Jangani M, Sommer P, Han N, Brass A, Beesley S, Lu W, Berry A, Loudon A, Donn R, et al. Ultradian cortisol pulsatility encodes a distinct, biologically important signal. PLoS One. 2011;6(1):e15766.

49. Han N, Dol Z, Vasieva O, Hyde R, Liloglou T, Raji O, Brambilla E, Brambilla C, Martinet Y, Sozzi G, et al. Progressive lung cancer determined by expression profiling and transcriptional regulation. Int J Oncol. 2012;41(1):242–52.

50. Darieva Z, Han N, Warwood S, Doris KS, Morgan BA, Sharrocks AD. Protein kinase C regulates late cell cycle-dependent gene expression. Mol Cell Biol. 2012;32(22):4651–61.

51. Papa S, Bubici C, Zazzeroni F, Pham CG, Kuntzen C, Knabb JR, Dean K, Franzoso G. The NF-kappaB-mediated control of the JNK cascade in the antagonism of programmed cell death in health and disease. Cell Death Differ. 2006;13:712–29.

52. Guo F, Tanzer S, Busslinger M, Weih F. Lack of nuclear factor-kappa B2/p100 causes a RelB-dependent block in early B lymphopoiesis. Blood. 2008;112(3):551–9.

53. Thonel A, Vandekerckhove J, Lanneau D, Selvakumar S, Courtois G, Hazoume A, Brunet M, Maurel S, Hammann A, Ribeil JA, et al. HSP27 controls GATA-1 protein level during erythroid cell differentiation. Blood. 2010;116(1):85–96.

54. Stassen M, Klein M, Becker M, Bopp T, Neudorfl C, Richter C, Heib V, Klein-Hessling S, Serfling E, Schild H, et al. p38 MAP kinase drives the expression of mast cell-derived IL-9 via activation of the transcription factor GATA-1. Mol Immunol. 2007;44(5):926–33.

55. Angelis L, Zhao J, Andreucci JJ, Olson EN, McDermott GCJC. Regulation of vertebrate myotome development by the p38 MAP kinase–MEF2 signaling pathway. Dev Biol. 2005;283(1):171–9.

56. Ramsauer K, Sadzak I, Porras A, Pilz A, Nebreda AR, Decker T, Kovarik P. p38 MAPK enhances STAT1-dependent transcription independently of Ser-727 phosphorylation. Proc Natl Acad Sci. 2002;99(20):12859–64.

57. Bradham C, McClay DR. p38 MAPK in development and cancer. Cell Cycle. 2006;5(8):824–8.

58. Li S, Gallup M, Chen Y-T, McNamara NA. Molecular mechanism of proinflammatory cytokine-mediated squamous metaplasia in human corneal epithelial cells. Invest Ophthalmol Vis Sci. 2010;51(5):2466–75.

59. Tew SR, Hardingham TE. Regulation of SOX9 mRNA in human articular chondrocytes involving p38 MAPK activation and mRNA stabilization. J Biol Chem. 2006;281(51):39471–9.

60. Zhang R, Murakami S, Coustry F, Wang Y, Crombrugghe B. Constitutive activation of MKK6 in chondrocytes of transgenic mice inhibits proliferation and delays endochondral bone formation. Proc Natl Acad Sci. 2006;103:352–70.

61. Akiyama H, Chaboissier M-C, Martin JF, Schedl A, Crombrugghe B. The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6. Genes Dev. 2002;16:2813–28.

62. Seymour PA, Freude KK, Tran MN, Mayes EE, Jensen J, Kist R, Scherer G, Sander M. SOX9 is required for maintenance of the pancreatic progenitor cell pool. Proc Natl Acad Sci. 2007;104(6):1865–70.

63. Acharya M, Lingenfelter DJ, Huang L, Gage PJ, Walter MA. Human PRKC apoptosis WT1 regulator is a novel PITX2-interacting protein that regulates PITX2 transcriptional activity in ocular cells. J Biol Chem. 2009;284(50):34829–38.

64. Toro R, Saadi I, Kuburas A, Nemer M, Russo AF. Cell-specific activation of the atrial natriuretic factor promoter by PITX2 and MEF2A. J Biol Chem. 2004;279(50):52087–94.

65. Li C-S, Chae S-C, Lee J-H, Zhang Q, Chung H-T. Identification of single nucleotide polymorphisms in FOXJ1 and their association with allergic rhinitis. J Hum Genet. 2006;51(4):292–7.

66. Gottardi AD, Dumonceau J-M, Bruttin F, Vonlaufen A, Morard I, Spahr L, Rubbia-Brandt L, Frossard J-L, Dinjens WN, Rabinovitch PS, et al. Expression of the bile acid receptor FXR in Barrett's esophagus and enhancement of apoptosis by guggulsterone in vitro. Mol Cancer. 2006;5(48):48.

67. Fujino T, Murakami K, Ozawa I, Minegishi Y, Kashimura R, Akita T, Saitou S, Atsumi T, Sato T, Ando K, et al. Hypoxia downregulates farnesoid X receptor via a hypoxia-inducible factor-independent but p38 mitogen-activated protein kinase-dependent pathway. FEBS J. 2009;276(5):1319–32.

68. Desbiens KM, Deschesnes RG, Labrie MM, Desfosses Y, Lambert H, Landry J, Bellmann AK. c-Myc potentiates the mitochondrial pathway of apoptosis by acting upstream of apoptosis signal-regulating kinase 1 (Ask1) in the p38 signalling cascade. Biochem J. 2003;372:631–41.

69. Shuai K, Liu B. Regulation of JAK-STAT signalling in the immune system. Nat Rev Immunol. 2003;3(11):900–11.