Ma et al. BMC Genomics 2018, 19(Suppl 10):884
https://doi.org/10.1186/s12864-018-5277-6

BMC Genomics

RESEARCH Open Access

Constructing tissue-specific transcriptional regulatory networks via a Markov random field

Shining Ma[2], Tao Jiang[1,3*] and Rui Jiang[1*]

From 29th International Conference on Genome Informatics
Yunnan, China. 3-5 December 2018

Abstract

Background: Recent advances in sequencing technologies have enabled parallel assays of chromatin accessibility and gene expression for major human cell lines. Such innovation provides a great opportunity to decode phenotypic consequences of genetic variation via the construction of predictive gene regulatory network models. However, there still lacks a computational method to systematically integrate chromatin accessibility information with gene expression data to recover complicated regulatory relationships between genes in a tissue-specific manner.

Results: We propose a Markov random field (MRF) model for constructing tissue-specific transcriptional regulatory networks via integrative analysis of DNase-seq and RNA-seq data. Our method, named CSNets (cell-line specific regulatory networks), first infers regulatory networks for individual cell lines using chromatin accessibility information, and then fine-tunes these networks using the MRF based on pairwise similarity between cell lines derived from gene expression data. Using this method, we constructed regulatory networks specific to 110 human cell lines and 13 major tissues with the use of ENCODE data. We demonstrated the high quality of these networks via comprehensive statistical analysis based on ChIP-seq profiles, functional annotations, taxonomic analysis, and literature surveys. We further applied these networks to analyze GWAS data of Crohn's disease and prostate cancer. Results were either consistent with the literature or provided biological insights into regulatory mechanisms of these two complex diseases. The website of CSNets is freely available at http://bioinfo.au.tsinghua.edu.cn/jianglab/CSNETS/.

Conclusions: CSNets demonstrated the power of joint analysis on epigenomic and transcriptomic data towards the accurate construction of gene regulatory network. Our work provides not only a useful resource of regulatory networks to the community, but also valuable experiences in methodology development for multi-omics data integration.

Background

The complicated process of transcription in eukaryotes largely attributes to the collaboration among DNA regulatory elements, RNA polymerases, mediator and cohesion complexes, and sequence-specific transcription factors (TFs). Such collaboration is encoded in a comprehensive gene regulatory network that determines how the expression of a gene is regulated, what responses a cell would adopt to answer external stimuli, and which phenotypic consequences a genetic variation could result in [1]. In the recent years, gene regulatory networks has been widely applied to answer a variety of questions, including the explanation of gene expression [2], the identification of disease genes in genome-wide association studies (GWAS) [3, 4], functional annotations for biological pathways [5, 6], and many others [7–9].

A gene regulatory network is often inferred based on high-throughput assays about interactions among transcription factors and their target genes. RNA-seq, as a means of capturing a snapshot of the whole transcriptome, has provided the most abundant data in such

* Correspondence: jiang@cs.ucr.edu; ruijiang@tsinghua.edu.cn
[1]Ministry of Education Key Laboratory of Bioinformatics; Bioinformatics Division, Beijing National Research Center for Information Science and Technology; Department of Automation, Tsinghua University, Beijing 100084, China
Full list of author information is available at the end of the article

© The Author(s). 2018 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

studies. For example, Hu and Chen constructed a transcriptional regulatory network in memory CD8+ T cells with gene expression profiles and predicted TF information, and then identified the core TFs [10]. Li et al. constructed a human regulatory regulatory network in glioma with the expression data of TFs and observed the dynamic rewiring of regulators during the glioma progression [11]. Marbach et al. introduced a resource of 394 human gene regulatory networks by integrating TF binding motifs with Cap Analysis of Gene Expression (CAGE) data from the FANTOM5 project [12].

With the promise of detecting TF binding sites at high resolution, ChIP-seq has been used with RNA-seq data to infer gene regulatory networks. For example, Roy et al. constructed a mixed regulatory network that combines transcriptional regulation by TFs from ChIP experiments and posttranscriptional regulation by miRNAs [13]. Chen et al. developed an efficient Bayesian integration method for the inference of regulatory networks using ChIP-seq and RNA-seq profiles [14]. These studies have also suggested that TFs normally bind to their target sites and regulate downstream genes in a cell-type specific manner [15, 16]. Moreover, such specificity is closely related to biological functions and cellular properties [10, 11, 17, 18].

There are several difficulties that restrict large-scale applications of the ChIP-seq technology. Besides the restriction of suitable antibodies for TFs, the number and cost of experiments required by a large number of TFs also limit the feasibility to construct gene regulatory networks for a variety of phenotypes and species via ChIP-seq. To overcome these limitations, DNase-seq has been developed to enable the capture of chromatin accessibility in whole-genome scale [19, 20]. Taking advantage of such merits as free from the consideration of TF-specific antibodies, it has been shown that the regulatory network specific to a cell line can be constructed from a single DNase-seq experiment [21–23]. Moreover, the collection of abundant DNase-seq profiles for major human cell lines in such genomic studies as the ENCODE [24] and Roadmap [25] projects has made the large-scale construction of regulatory networks for a variety of cell lines and tissues possible.

Motivated by the above understanding, we propose in this paper a Markov random field (MRF) model, named CSNets (Cell-line Specific regulatory Networks), that integrates DNase-seq data with RNA-seq data towards large-scale inference of gene regulatory networks. In this method, we first roughly infer regulatory networks for individual cell lines using DNase-seq data alone. Then, we fine-tune these networks using an MRF model, based on pairwise similarity between cell lines derived from RNA-seq data. Focusing on data released by the ENCOODE project, we constructed regulatory networks specific to 110 cell lines and 13 major tissues for human. Using ChIP-seq experimental data as a gold standard, we showed the superior quality of our networks over that obtained by existing methods. Through functional enrichment analysis, we demonstrated that TFs and their predicted targets tend to share similar biological functions. Besides, integrative analysis of our networks with GWAS data of Crohn's disease and prostate cancer both suggested genes and genetic variants that were either consistent with the literature or provided biological insights into regulatory mechanisms of these two complex diseases.

Methods

Data collection

We extracted DNase-seq profiles for 110 human cell lines, representing 70 diverse cell types and 13 unique tissue lineages, from the ENCODE project [26]. We collected gene expression data of corresponding cell lines from the ENCODE project [24]. We derived binding motifs of 368 transcription factors from the JASPAR [27] and TRANSFAC [28] databases. We extracted 353 ChIP-seq experiments from the ENCODE project, corresponding to 108 transcription factors and 59 cell lines. We collected 1454 gene sets with gene ontology (GO) annotations from the MSigDB database [29], involving 233 GO terms of cellular component, 825 terms of biological process, and 396 terms of molecular function.

Principles of CSNETS

We proposed to construct a transcriptional regulatory network specific to a cell line by integrating DNase-seq data, transcription factor binding motif information, and gene expression data, as illustrated in Fig. 1.

We first followed a computational method called Protein Interaction Quantitation (PIQ) [23] to perform a whole-genome prediction of transcription factor binding sites from DNase-seq data. Briefly, PIQ relied on a machine-learning method called expectation propagation [30] to identify binding sites for transcription factors with known motif patterns. Using this method, we obtained the position and binding probability for each predicted binding site, by using DNase-seq data corresponding to the 110 cell lines, the reference sequence of *Homo sapiens* (GRCh37), and position weighted matrix of motif for the 368 transcription factors. Focusing on predicted binding sites in promoter regions (TSS ± 2 kb), we linked transcription factors to their target genes, and thus obtained preliminary regulatory networks specific to the 110 human cell lines. Second, we incorporated gene expression data and adopted a rigorous Markov random field model to fine-tune these preliminary networks. The basic assumption behind this model is that similar cell lines tend to share

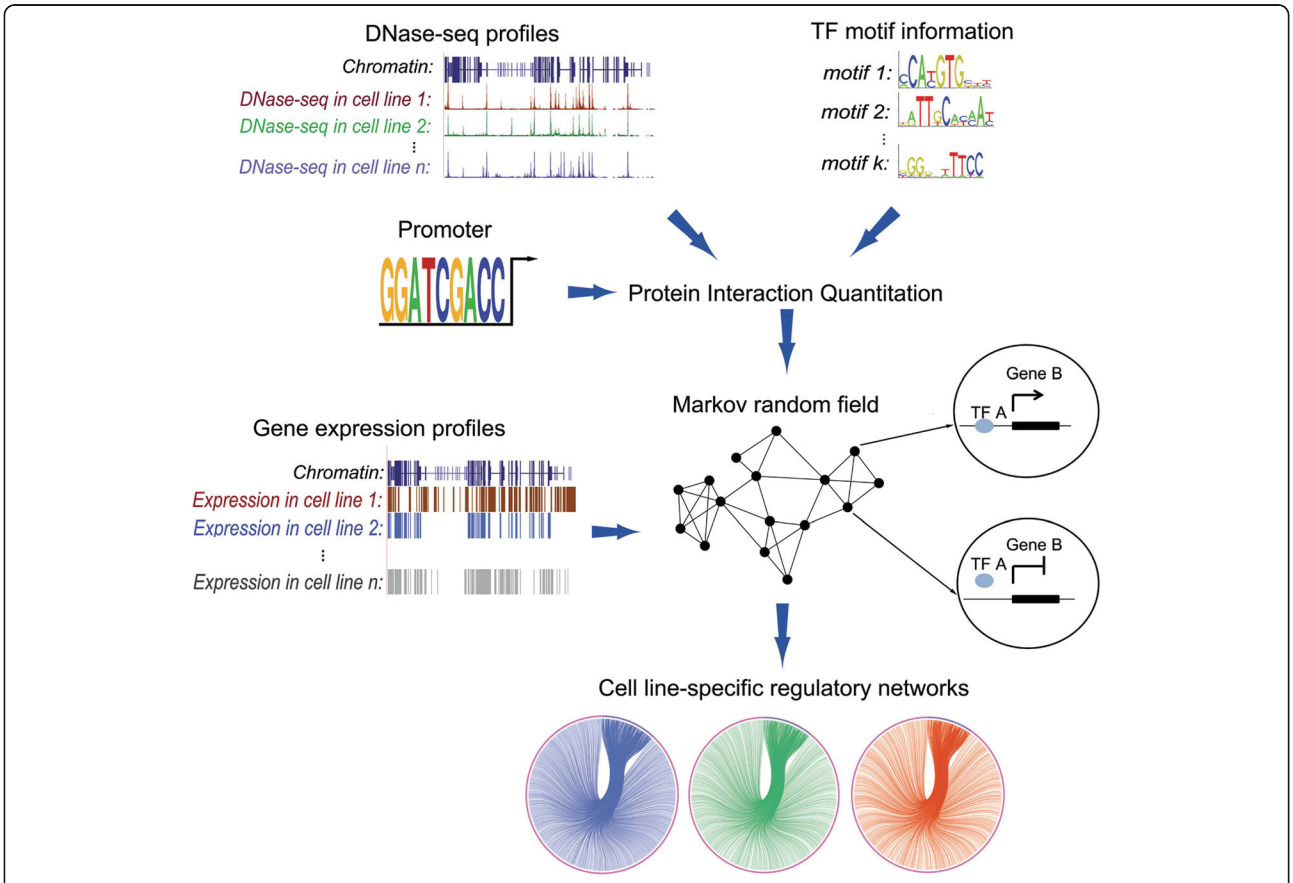

Fig. 1 Workflow of the construction of 110 cell line-specific regulatory networks. Firstly, perform PIQ on DNase-seq profiles and TF motifs to predict genome-wide transcription factor binding sites for 368 TFs in 110 cell lines, respectively. Secondly, map these transcription factor binding sites to the promoter regions of genes and thus link TFs to target genes. Thirdly, construct regulatory networks with a Markov random field (MRF) model based on the cell line similarity measured by the expression profiles of these cell lines. Finally, we get the cell line-specific regulatory network for each cell line

similar regulatory patterns. With this understanding, we used gene expression data to measure the similarity between cell lines, and then connect regulatory relationships of different cell lines by using a Markov random field (MRF) model based on the similarity. Detailed explanations of components in our method are given below.

Quantification of cell line similarity

We adopted a measure, called TSI (Tissue Similarity Index) [31], to characterize relationships among cell lines and quantify their degree of similarity. First, we used SAM [32] to identify 592 genes that were differentially expressed (q-value = 0) in the 110 cell lines. Then, we applied the singular value decomposition (SVD) to expression data of these genes to perform a dimensionality reduction. In detail, expression values of a gene across the 110 cell lines was first normalized to zero mean and standard deviation one. The resulting expression matrix regarding the 592 genes and 110 cell lines were then decomposed into USV^T, where columns of U were called eigenarrays, diagonals in S singular values, and rows of V^T right singular vectors. Finally, we characterized the similarity between two cell lines as the Pearson's correlation coefficient of the first 16 dimensions of the eigenarrays in the SVD decomposition. Here, we calculated CSI values based on different numbers of dimensions and found that they were robust and slightly changed. We name such a similarity measure as the Cell-line Similarity Index (CSI).

Markov random field model

We proposed a Markov random field model to fine-tune the preliminary networks. Specifically, for a given regulatory relationship (e.g., TF A regulates target gene B), we constructed an MRF network $G = \{V_{AB}, E_{AB}\}$, where a node $v_{ABi} \in V_{AB}$, $(i = 1, 2, \ldots, 110)$ indicates the regulation of TF A on target gene B in cell line i, and an edge $(i, j) \in E_{AB}$ denotes the regulation coherence for TF A and gene B between cell line i and j. We introduced an

indicator variable $x_i \in X$ for the node v_{ABi}, indicating whether the regulation exists ($x_i = 1$) or not ($x_i = 0$) in cell line i. Suppose that the higher the degree of similarity between cell line i and j, the stronger the positive correlation of variable x_i and x_j, we define a criterion called the TF non-specific index (TNI) as the proportion of common targets for the TF involved in cell line i and j. The larger the TNI value, the more similar of the regulatory mechanism corresponding to the concerned TF. We define the edge weight, $w_{ij} \in W$, as the average of CSI(x_i, x_j) and TNI(x_i, x_j). We set a threshold c with default value 0.5, and regarded cell line i and j as connective if and only if $w_{ij} > c$.

We then followed the literature [33] to construct a pairwise MRF model that uses the similarity information between cell lines to assist the prediction on the existence of a transcriptional regulatory relationship in the cell lines. This model contains two types of potential functions. The first type is called the node function, defined as

$$\phi_i(x_i) = \begin{cases} P_{(1,i)}/P_{(0,i)} \text{ if } P_{(1,i)} > P_{(0,i)}, x_i = 1 \\ P_{(0,i)}/P_{(1,i)} \text{ if } P_{(0,i)} > P_{(1,i)}, x_i = 0 \\ 1, \text{otherwise} \end{cases}$$

where $P_{(1,i)}$ and $P_{(0,i)}$ denote the probability of existence ($x_i = 1$) or failure ($x_i = 0$) of the given regulatory relationship in cell line i, respectively. We used the probability of TF binding inferred from PIQ as the probability of $P_{(1,i)}$.

The second type of potential function is called the edge function, defined as

$$\psi_{(i,j)}(x_i, x_j) = \begin{cases} e^{(\mathrm{CSI}(x_i,x_j)+\mathrm{TNI}(x_i,x_j))/2}, \text{if } x_i = x_j, \\ 1, \text{otherwise.} \end{cases}$$

This function uses the CSI and TNI score mentioned above to measure the association between cell lines.

With the definition of two types of potential functions, the joint distribution of all indicator variables X can be denoted as

$$\Pr(X) = \frac{1}{Z} \prod_{(i,j)\in E_{AB}} \psi_{(i,j)}(x_i, x_j) \prod_{i=1}^{n} \phi_i(x_i)$$

where Z represents the partition function, making the sum of the probabilities equal to 1. Through a negative logarithmic transformation, the joint distribution of X can be written as

$$E(X) = -\gamma - \sum_{i=1}^{n} \ln \phi_i(x_i) - \sum_{(i,j)\in E_{AB}} \ln \psi_{(i,j)}(x_i, x_j)$$

where γ is a constant. $E(X)$ is named as pseudo-energy function. With this formulation, we transformed the problem of maximizing the joint distribution Pr(X) into that of minimizing the pseudo-energy function [33, 34]. We then applied iterated conditional modes [35] to further transform the problem of minimizing the pseudo-energy function into the maximum flow problem of networks.

In detail, firstly we define $\alpha_i(x_i) = \ln \phi_i(x_i)$and $\beta_{ij}(x_i, x_j) = \ln \psi_{(i,j)}(x_i, x_j)$. When the value of x_i is not consistent with its probability distribution (i.e. $P_{(0,i)} > P_{(1,i)}$ when$x_i = 1$ or $P_{(0,i)} < P_{(1,i)}$ when$x_i = 0$), the value of $\alpha_i(x_i)$ is 0, rather than $|\ln \phi_i(1) - \ln \phi_i(0)|$. When cell lines i and j are connective and x_i and x_j are of differential values, the value of $\beta_{ij}(x_i, x_j)$ is 0 rather than (CSI(x_i, x_j) + TNI(x_i, x_j))/2. It is verified that we can transform the problem of minimizing the pseudo-energy function into that of summing total losses of $\alpha_i(x_i)$and $\beta_{ij}(x_i, x_j)$ when the following inequality satisfies [36].

$$\begin{aligned} &\beta_{ij}(x_i = 1, x_j = 1) \\ &\quad + \beta_{ij}(x_i = 0, x_j = 0) \geq \beta_{ij}(x_i = 1, x_j = 0) \\ &\quad + \beta_{ij}(x_i = 0, x_j = 1) \end{aligned}$$

This equation suggests that the problem of minimizing the pseudo-energy function is transferred into the maximum flow problem of networks. We finally applied the loopy belief propagation algorithm [37] to calculate the probability distribution of X.

Evaluation using ChIP-seq data

We collected 353 ChIP-seq experiments, regarding 108 TFs in 59 cell lines, from the ENCODE project. We then evaluated the contribution of the MRF model as follows.

We first generated a gold standard of target genes for a TF in a cell line from the corresponding ChIP-seq experiment. To achieve this objective, we mapped binding sites identified in the experiment to promoter regions (TSS ± 2Kbps) of protein coding genes and assigned experimental scores of the binding sites to the mapped genes, which were used as candidate target genes. To further reduce false positives in these genes, we identified the median size (M) of three gene sets, which include candidate target genes according to ChIP-seq data, target genes of the TF according to the network constructed by the MRF model for the given cell line, and target genes of the TF according to the preliminary network for the given cell line. Finally, we ranked candidate target genes according to their scores and used those ranked among top M as the gold standard of target genes for the TF in the given cell line.

We then performed a ROC analysis to evaluate the quality of the networks constructed by our method. Given a TF and a cell line, we used target genes identified by the corresponding ChIP-seq experiment as the positive set, and the reset genes as the negative set. Focusing on the list of target genes for a TF given by our

Table 1 Property for 13 tissue-specific networks

Tissue	Node	Edge	In-degree	Out-degree
Epithelial	19,535	388,433	19.98	1063.34
Fibroblast	19,345	385,268	20.02	1051.29
Muscle	19,349	386,523	20.08	1056.50
Brain	19,602	443,826	22.55	1268.87
Hematopoietic	19,316	452,915	23.44	1277.08
Primitive	19,369	377,249	19.57	1083.85
Skin	19,430	380,265	19.65	1088.57
Stem	20,803	628,780	30.02	1735.75
Endothelial	19,319	454,054	23.62	1234.11
Cervix	19,830	495,002	25.09	1404.84
Liver	19,354	486,529	25.17	1437.27
Prostate	19,146	431,096	22.52	1171.46
Mammary	20,019	496,650	24.91	1380.05

method, at a cut-off value of the regulatory probability, we calculated the sensitivity as the proportion of positives whose regulatory probability is higher than the cut-off, and the specificity as the proportion of negatives whose regulatory probability is lower than the cut-off. Varying the cut-off value, we drew a receiver operating characteristic (ROC) curve (sensitivity versus 1-specificity) and calculated the AUC score as the area under this curve. In a similar way, we obtain the AUC score of for the preliminary network. The relative change of these two AUC scores is then used to compare the performance of these two networks, for the given TF in the give cell line.

We further evaluated quality of the constructed networks by checking the overlap between target genes in the networks and those identified by ChIP-seq experiments. This was done by filtering out low confidence target genes that were ranked below the threshold M and then counting the number of genes shared in the remaining target genes. Using a similar strategy, we obtained an overlapping score for the preliminary network. The relative change of these two scores can then be used to compare the performance of these two networks, for the given TF in the give cell line.

Results

Regulatory networks specific to 110 cell lines and 13 tissues

We constructed regulatory networks specific to 110 human cell lines, and we further merged networks specific to cell lines belonging to the same tissues to obtain 13 tissue-specific regulatory networks, as summarized in Table 1. From the table, we observe that the network specific to stem cell has the largest in-degree and out-degree. This phenomenon can probably be explained by the pluripotency nature of stem cells. We also notice that the liver-specific network also has high degrees.

We then extracted sub-networks regarding transcription factors only from the 13 networks and illustrated 6 of such networks in Fig. 2(a). In comparison with other tissues, tissue-specific regulatory relationships in the hematopoietic tissue and stem cells tend to present more frequently, indicating their high degree of the tissue-specificity. In stem cells, we collected TFs and genes closely correlated with the pluripotency from

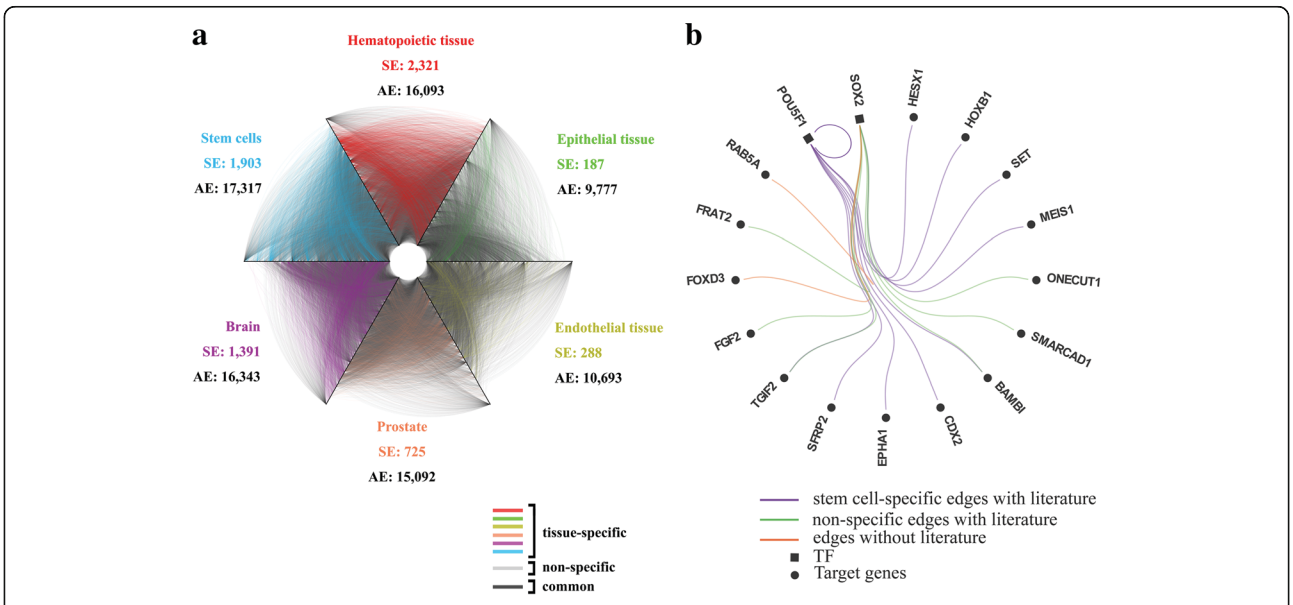

Fig. 2 a TF-TF tissue-specific regulatory networks and (**b**) core transcriptional regulatory network in human embryonic stem cells. SE means specific edges and AE means all edges

literature [38, 39] and further analyzed the regulatory interactions among them in Fig. 2(b). We notice that most of the regulatory interactions derived from the network specific to stem cell are testified by literature [38, 39] (shown in purple and green) and present a high degree of the tissue specificity (shown in purple), indicating that regulatory relationships specific to stem cells are highly correlated with the property of pluripotency.

Constructed networks are consistent with ChIP-seq data

We evaluated contributions of the MRF model using 353 ChIP-seq experiments collected from the ENCODE project. Briefly, we first identified a gold standard of target genes for a TF in a cell line from the corresponding ChIP-seq experiment. Then, we evaluated the improvement of a network constructed by using the MRF model over the corresponding preliminary one in terms of relative changes in the AUC score and the overlapping score, as detailed in Methods.

As shown in Fig. 3(a), the MRF model improves the accuracy in recovering true target genes for a TF in a specific cell line, according to criterion of the relative change in AUC scores. In 225 out of the 353 (64%) experiments, networks constructed by using the MRF model show higher consistency with ChIP-seq data than the preliminary network. Such positive contribution of the MRF model is further supported by the one-sided binomial exact test (p-value = 8.4e-08). In terms of the relative change in the overlapping score, as shown in Fig. 3(b), networks constructed by using the MRF model show higher consistency with ChIP-seq data in 287 out of the 353 (81.3%) experiments. The positive contribution of the MRF model is again supported by the one-sided binomial exact test (p-value< 2.2e-16).

We further performed a TF level evaluation by aggregating ChIP-seq experiments according to TFs and averaging a criterion over corresponding cell lines. As shown in Fig. 3(c), the MRF model shows higher consistency with ChIP-seq data for 78 out of the 108 (72%) TFs, and the positive contribution of the MRF model is supported by the one one-sided binomial exact test (p-value = 1.1e-05). In terms of the relative change in the overlapping score, networks constructed by using the MRF

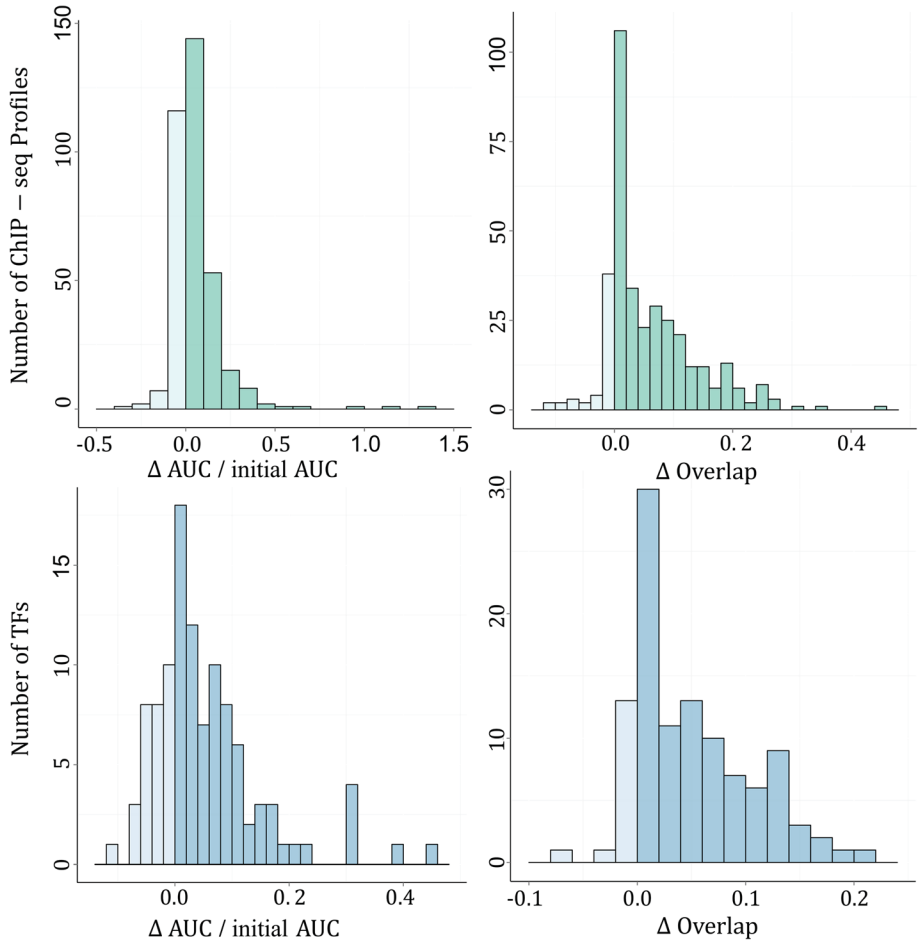

Fig. 3 Comparison of the consistency between preliminary/MRF-based networks and ChIP-seq profiles

model show higher consistency with ChIP-seq data for 88 out of the 108 (81.5%) TFs. Again, he positive contribution of the MRF model is supported by the one-sided binomial exact test (p-value< 6.4e-15).

Constructed networks are consistent with taxonomy

We testified the rationality of the tissue similarity measured by gene expression data and observed the consistency between regulatory networks and the human cell hierarchical taxonomy graph. After extracting the directed acyclic subgraph of the human cell hierarchical taxonomy graph from the Foundational Model of Anatomy Database [40] in the Unified Medical Language System [41] (shown in Additional file 1: Figure S1), we performed the hierarchical clustering on tissues and cell lines based on gene expression and regulatory relationships respectively, and compare them with the human taxonomy graph.

First, we performed hierarchical clustering of tissues according to gene expression profiles. Figure 4(a) shows that the hematopoietic tissue is the most distal to the other tissues. The fibroblast and muscle are clustered together, and the prostate tissue is in short distance with the liver. The endothelial tissue and cervix are very close. The epithelial tissue is the parent node of the skin and brain in the human cell hierarchical taxonomy graph, and they are clustered together as well. Therefore, we conclude that it is reasonable to measure the tissue similarity based on gene expression profiles.

Further, hierarchical clustering was performed for the 110 cell lines based on their similarity index, with results shown in Fig. 4(c). We find that cell lines from the same tissues tend to be clustered together, indicating that such cell lines in general have higher similarity than those from different tissues. Moreover, in the 110 cell lines, there are 25 cancer cell lines, 65 cell lines with normal karyotype and 20 unidentified cell lines. Cell lines with the same karyotype are more likely to be in the same cluster.

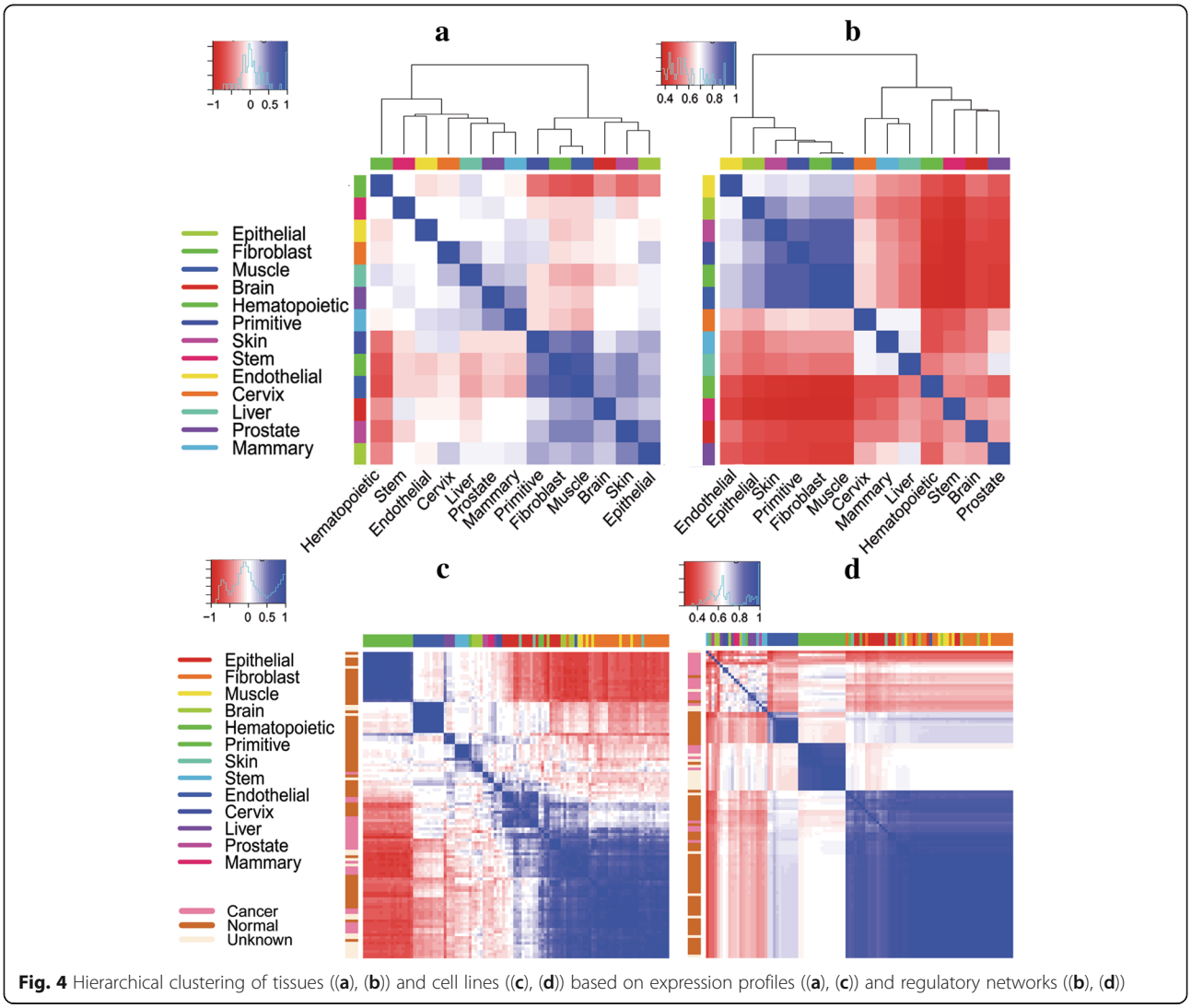

Fig. 4 Hierarchical clustering of tissues ((**a**), (**b**)) and cell lines ((**c**), (**d**)) based on expression profiles ((**a**), (**c**)) and regulatory networks ((**b**), (**d**))

Next, we inspected the consistency between regulatory networks and the human cell hierarchical taxonomy graph to evaluate the rationality of the tissue similarity measured by regulatory networks. For this objective, we merged regulatory networks specific to cell lines belonging to the same tissue, used the Jaccard coefficient of tissue-specific regulatory networks to measure the similarity between tissues, and then perform hierarchical clustering on the 13 tissues, as shown in Fig. 4(b). The skin, fibroblast, muscle, endothelial and epithelial tissues are clustered together, of which fibroblast and muscle are the closest. The epithelial tissue is close with skin, and these four tissues are relatively distant from the endothelial tissue. The hematopoietic tissue is the most distal one from the others, followed by stem cells. These results are consistent with the human cell hierarchical taxonomy graph. We assert that our regulatory networks describe explainable tissue similarity relationships.

Finally, we performed hierarchical clustering on the 110 cell lines based on their specific regulatory networks. The results, in Fig. 4(d) shows that cell lines from the hematopoietic and endothelial tissues are clustered, respectively, and those from the skin, fibroblast, muscle, endothelial and epithelial tissues are clustered together, which is consistent with the human cell hierarchical taxonomy graph. Similar to the observation from the results of expression profiles, cell lines of the same karyotype are more likely to be closely clustered.

Correlation between expression of TFs and target genes

We evaluated whether expression of TFs exhibited positive correlation with their predicted target genes in a cell line-specific manner. For this objective, we collected gene expression data for the GM12878, K562, MCF-7, and SK-N-SH cell lines. For each of these cell lines, we calculated Pearson's correlation coefficient between a TF and each of its target genes, and we plot the statistical significance of the correlation coefficient in Fig. 5. From the figure, we clearly see that TFs show stronger correlation with their target genes in networks constructed by using the MRF model, and one-sided Wilcoxon tests support this observation (p-values < 2.2e-16 for all these cell lines). These results suggest that the networks constructed by using our method can well characterize the regulatory relationship between transcription factors and their target genes.

GO enrichment analysis

With the assumption that a TF and its target genes tend to share common biological functions, we performed a functional enrichment analysis on the target genes of a TF. To achieve this objective, we collected 1454 gene sets from the MSigDB database, covering

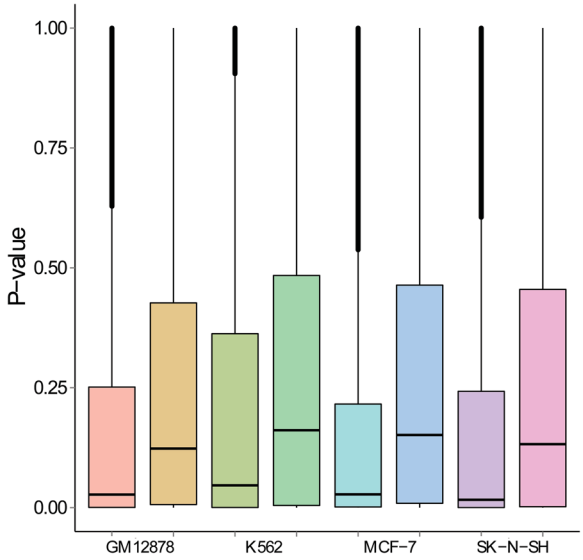

Fig. 5 The correlation of expression for TF-gene pairs that are included and not included in the four cell line-specific regulatory networks

825 biological process terms, 396 molecular function terms, and 233 cellular component terms in GO. We then derived two criteria to characterize the degree that a TF and its target genes share a common function, as follows.

Given a TF, a cell line and a GO term, we identified target genes of the TF in the network specific to the cell line, and we performed a Fisher's exact test to see whether the target genes were enriched in the function corresponding to the GO term. With results of such an analysis for every TF, in every cell line, and for every GO term collected, we were able to derive a statistic to indicate consistency between functions of TFs and their target genes. Particularly, we defined such a statistic as the proportion of significant tests (FDR ≤ 0.2) over all tests performed, and we calculated a statistic for each of the three GO categories, biological process, molecular function, and cellular component, separately. Because all the 1454 GO terms were used in the above analysis, we referred to such a statistic as the total enrichment score. By contrast, we repeated the above enrichment analysis with GO terms not relevant to a TF discarded, and we referred to this formulation as positive enrichment analysis.

As shown in Fig. 6, it is clear that for each the three GO categories, the positive enrichment analysis exhibits much higher score than the total enrichment analysis, suggesting that TFs indeed tend to have similar functions as their target genes. One-sided Wilcoxon rank sum test further support this conclusion, in that p-values are as small as 5.60×10^{-7}, 7.95×10^{-3} and $7.80 \times 10^{-4}$for the cellular component, biological process and molecular function, respectively.

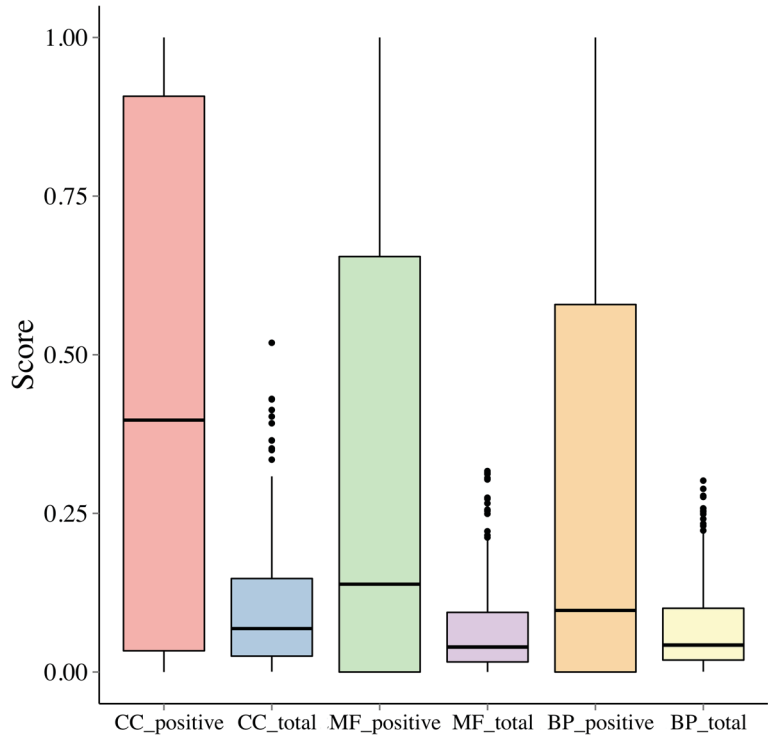

Fig. 6 Enrichment results for positive/entire GO terms of three categories

Identification of TFs with differential regulation

We further analyzed whether target genes of a TF exhibit different functions in different cell lines, especially in normal and cancer cell lines. To achieve this objective, we identified 65 normal cell lines and 25 cancer cell lines. For each TF, we collected its target genes, and we performed functional enrichment analysis to see whether functions of the TF were enriched in its target genes, for the normal and cancer cell lines, separately. We showed TFs and GO terms with most significant enrichment *p*-values in Additional file 1: Table S1, from which we can see that many of these TFs have been verified to associate closely with various types of cancer.

For example, EP300 plays an important role in regulating cell growth and blocking the promotion of cancerous tumors. The targets of EP300 are enriched in normal cells rather than cancer cells in two GO terms, corresponding to apoptotic process and programmed cell death respectively, indicating the function of EP300 is altered in cancer cells (Fig. 7). As for FOSL1, the enrichment degree of its targets in the cell proliferation term is significantly different between normal and cancer cell lines. Therefore, it is reasonable to presume that the regulatory mechanisms of these TFs are perturbed in various cancer cells and further affect the growth and promotion of cancers through the matched GO annotated functions.

Locating TFs for Crohn's disease

We applied the constructed regulatory networks to analyze a GWAS data set of Crohn's disease, demonstrating the potential of these networks in identifying disease-related TFs and their regulatory mechanisms. We first select a regulatory network that is specific to Crohn's disease from the 110 networks. On one hand, it is generally thought that the inflammatory reaction in Crohn's disease is driven by the activated type 1 helper T cells (Th1) [42]. On the other hand, from the 1000 Genomes Project [43], we observed that the similarity between Th2 cells and Crohn's disease is the highest among all ENCODE cell lines (Additional file 1: Figure S2). A comparative study on the regulatory networks of these two cell lines shows that these networks share 98.6% edges. We therefore selected the regulatory network of Th1 cell line according to the literature.

We then collected GWAS data for this disease from the literature [44] and calculated for each gene a *p*-value that indicates the strength of association between the gene and this disease, using the tool Pascal [45]. By ranking genes based on their *p*-values, we obtained a gene list, in which top ranked genes can be treated as candidate disease genes. To avoid determining true disease genes based on a hard cut-off of the *p*-value, we resorted to GSEA [46] to assign an enrichment score to a TF, measuring whether its target genes are enriched in highly ranked candidate genes. We further ranked TFs

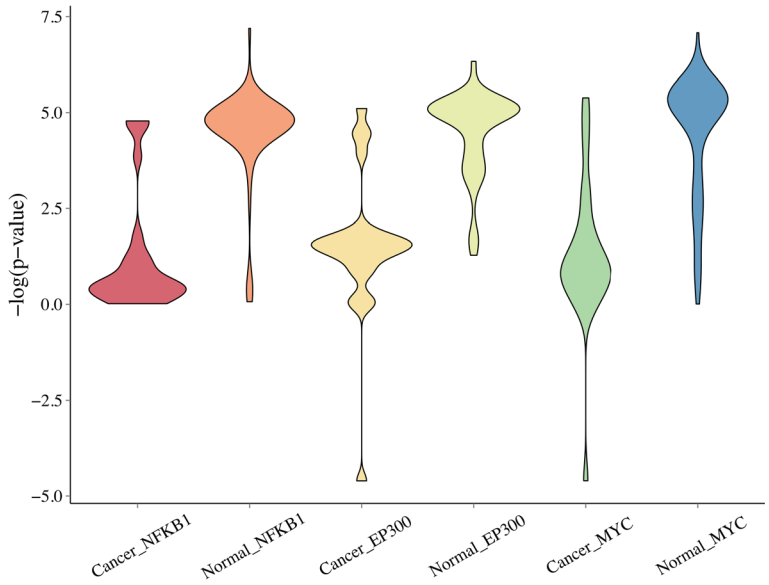

Fig. 7 The enrichment results of NFKB1, EP300 and MYC in their corresponding GO terms

according to their enrichment scores and obtained a list of 114 TFs, in which top-ranked TFs are considered as relevant to Crohn's disease. From the ranking list, we find that TFs of the NFκB family are relatively ranked high, with NFKB1 ranked first (Table 2). Previous studies [47–50] have shown that the NFκB family is activated and plays a key role in the inflammatory bowel diseases, especially Crohn's disease.

We further explored the importance of the cell line specificity of a regulatory network in identifying the NFκB family in the above analysis. Briefly, we extracted target genes of the NFκB family for each of the 110 networks and used Fisher's exact test to measure the degree of enrichment of these target sets in the candidate genes identified above for Crohn's disease. Results show that the target gene set of the NFκB family in Th1 cell line is ranked 2nd, and specifically, the target gene set of NFKB1 in Th1 cell line is ranked 1st, among the 110 cell lines (Additional file 1: Figure S3). We therefore conclude that the cell line-specificity of regulatory networks plays an important role in detecting disease-associated TFs.

To gain understanding about the mechanism of Crohn's disease, we further investigated target genes of the NFκB family in the Th1 cell line and surveyed their correlation with Crohn's disease according to literature.

Table 2 The rank results of the NFκB family

TF	*p*-value from GSEA	Rank
NFKB1	2.13E-05	1
REL	3.50E-03	6
NFKB	2.53E-02	17
RELA	1.10E-01	31

We found that Franke et al. confirmed 71 distinct loci for Crohn's disease and listed functionally interesting candidate genes [44]. We then listed the target genes supported by this literature in Table 3. We suspected that the mutations near these functional candidate genes may affect the binding affinity of the NFκB family and hence implicate in Crohn's disease pathogenesis.

Differential regulation analysis on the LNCaP cell line

We studied whether an environmental stimulus would affect the regulatory network related to a phenotype by using androgen-treated and untreated LNCaP cells as an example. To achieve this objective, we first compared regulatory networks specific to these two cell lines and found the Jaccard index between edges of these two

Table 3 Targets of the NFκB family correlated with Crohn's a supported by literature

TF	Target gene	dbSNP ID	Chr.	Risk allele
RELA	PTGER4	rs11742570	5p13	C
NFKB1/RELA	IRF1	rs12521868	5q31	T
NFKB1	SNAPC4	rs4077515	9q34	T
RELA	JAK2	rs10758669	9p24	C
REL/RELA	SMAD3	rs17293632	15q22	T
NFKB1/REL/RELA	GPX4	rs740495	19p13	G
REL/RELA	ICAM1	rs12720356	19p13	G
NFKB1	CREM	rs12242110	10p11	G
NFKB1/RELA	MUC1	rs3180018	1q22	A
NFKB1/REL/RELA	SCAMP3	rs3180018	1q22	A
NFKB1	ZPBP2	rs2872507	17q21	A
REL	FADS1	rs102275	11q12	C

networks is 0.703. We then focused on edges differently presented in these two networks to obtain a differentially network and performed the following analysis.

We ranked TFs based on the number of their target genes in the differential network in descending order and showed top 5 TFs in Additional file 1: Table S2. In this list, we found that SREBF1 and TWIST are androgen responsive in human cells, and SREBF1 and NFATC2 are androgen responsive in mouse cells, according to the Androgen Responsive Gene Database [51].

We collected target genes for TFs in treated and untreated LNCaP cell lines, respectively, and calculated a Jaccard distance to indicate the proportion of differential regulating edges for a TF. The top 5 TFs with the highest Jaccard distance are present in Table 4. In this list, NFIX is verified to be androgen responsive in human cells, and NFIX, NFATC2 and FOSL1 are verified in mouse cells by the Androgen Responsive Gene Database.

We finally ranked the target genes in the differential regulatory network according to the number of edges pointing to them, say, the number of TFs differentially regulating these genes. The top 5 genes having the largest number of regulation are shown in Table 5. In this list, WWTR1 is known as a downstream regulatory target in the Hippo signaling pathway that plays a key role in tumor suppression. CDKN1A encodes protein *p*21, which plays an important role in KEGG [52] prostate cancer pathway and is again verified by the Androgen Responsive Gene Database.

Table 4 Top-ranked TFs based on the Jaccard distance

TF	No. of differentially regulating edges	Jaccard distance	Rank
NFIX	3103	0.987902	1
NFATC2	6825	0.98004	2
BATF	2261	0.954814	3
FOSL1	1726	0.953591	4
HIVEP2	1697	0.920282	5

Table 5 Top-ranked genes based on their differential regulated edges

Gene	No. of differentially regulated edges	Rank
WWTR1	41	1
WWTR1-AS1	40	2
CDKN1A	39	3
PRRC2C	38	4
UBE2D3	38	4

Conclusions and discussion

In this paper, we have proposed a Markov random field model for integrating chromatin accessibility and gene expression data to construct regulatory networks specific to 110 cell lines and 13 tissues. We have demonstrated the high quality of these networks via comprehensive statistical analysis based on ChIP-seq experiments, functional annotations, taxonomic analysis, and literature surveys. Joint analysis of these networks with GWAS data provides results that are either consistent with literature or provided biological insights into regulatory mechanisms of human inherited diseases.

Main contributions of our work include the following aspects. First, we demonstrated the power of joint analysis on epigenomic and transcriptomic data towards the accurate construction of gene regulatory network. In the recent years, parallel assays of the epigenome and transcriptome has become popular, and computational methods for integrative analysis of such data are of urgent need, especially in single-cell multi-omics data analysis [53, 54]. Our work, as a beneficial attempt in this direction, can thus provide valuable experiences in methodology development for multi-omics data integration.

Second, our work provides a useful resource of regulatory networks to the community. Recently, Marbach et al. constructed 394 gene regulatory networks specific to human cell types and tissues by integrating TF motifs with CAGE data from the FANTOM5 project [12]. We compared their results with our networks on four shared cell lines involving 220 common TFs and found the proportion of shared edges ranging from 26 to 31%. On one hand, our regulatory networks consist with the networks constructed with CAGE data to some extent, indicating the rationality and robustness of our networks. On the other hand, the difference between our networks and theirs shows the complementarity and diversity of these two resources. In this sense, combined use of both resources may offer a complete landscape of human transcriptional regulatory networks.

Certainly, our work can further be improved from the following aspects. First, we construct cell line specific regulatory network based on multiple tissues and cell lines, and how to construct sample-specific regulatory networks is also very important [55–57]. Second, we only consider regulatory relationships between transcription factors and target genes in the current work. It is known that DNA regulatory elements are of great importance in gene regulation. Therefore, the incorporation of regulatory elements into a regulatory network is necessary [58]. Third, there have been great innovations in experimental technology for studying the epigenome in the recent years. For example, ATAC-seq [59] has been proposed as an more efficient alternative of DNase-seq. HiChIP [60] has been developed to directly assess enhancer activity and enhancer-promoter interactions. These techniques have provided a great opportunity to study gene regulatory networks towards the understanding of phenotypic consequences of human genetic variation on physiology traits or disease risks. How to bring the idea of integrative analysis in our work to facilitate deep analysis regarding multiple types of epigenomic and transcriptomic data will be a direction worth noting.

Additional file

Additional file 1: Figure S1. The directed acyclic subgraph of the human cell hierarchical taxonomy graph. **Figure S2** The similarity of Th2 cell and Crohn's disease. **Figure S3** The enrichment degree for target set of NFKB1 in 110 cell lines. **Table S1** TFs and corresponding GO terms that alter between normal and cancer cell lines. **TableS2** Top ranked TFs based on their differential regulating edges. (DOC 501 kb)

Acknowledgements
Rui Jiang is a RONG professor at the Institute for Data Science, Tsinghua University. We acknowledge the authors of PIQ, who provide us valuable software.

Funding
Publication costs are funded by the National Key Research and Development Program of China (No. 2018YFC0910404), the National Natural Science Foundation of China (Nos. 61873141, 71871019, 61721003, 61573207, 61772197 and 71471016), the US National Science Foundation (No. IIS-1646333) and the Tsinghua-Fuzhou Institute for Data Technology.

Availability of data and materials
The data of CSNets could be found at http://bioinfo.au.tsinghua.edu.cn/jianglab/CSNETS/.

About this supplement
This article has been published as part of *BMC Genomics* Volume 19 Supplement 10, 2018: Proceedings of the 29th International Conference on Genome Informatics (GIW 2018): genomics. The full contents of the supplement are available online at https://bmcgenomics.biomedcentral.com/articles/supplements/volume-19-supplement-10.

Authors' contributions
TJ and RJ provide guidance and planning for this project. SM produced programs, analysed results and wrote the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

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

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

Author details
[1]Ministry of Education Key Laboratory of Bioinformatics; Bioinformatics Division, Beijing National Research Center for Information Science and Technology; Department of Automation, Tsinghua University, Beijing 100084, China. [2]Department of Statistics, Department of Biomedical Data Science, Bio-X Program Stanford University, Stanford, CA 94305, USA. [3]Department of Computer Science and Engineering, University of California, Riverside, CA 92521, USA.

Published: 31 December 2018

References
1. Wang Y, Jiang R, Wong WH. Modeling the causal regulatory network by integrating chromatin accessibility and transcriptome data. NATL SCI REV. 2016;3(2):240–51.
2. Macneil LT, Walhout AJ. Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. Genome Res. 2011;21(5):645–57.
3. Wu M, Lin Z, Ma S, Chen T, Jiang R, Wong WH. Simultaneous inference of phenotype-associated genes and relevant tissues from GWAS data via Bayesian integration of multiple tissue-specific gene networks. J Mol Cell Biol. 2017;9(6):436–52.
4. Wu M, Zeng W, Liu W, Lv H, Chen T, Jiang R. Leveraging multiple gene networks to prioritize GWAS candidate genes via network representation learning. METHODS. 2018.
5. Ma S, Jiang T, Jiang R. Differential regulation enrichment analysis via the integration of transcriptional regulatory network and gene expression data. BIOINFORMATICS. 2015;31(4):563–71.
6. Zhao XM, Li S. HISP: a hybrid intelligent approach for identifying directed signaling pathways. J Mol Cell Biol. 2017;9(6):453–62.
7. Chen JC, Alvarez MJ, Talos F, Dhruv H, Rieckhof GE, Iyer A, Diefes KL, Aldape K, Berens M, Shen MM, et al. Identification of causal genetic drivers of human disease through systems-level analysis of regulatory networks. CELL. 2016;166(4):1055.
8. Zickenrott S, Angarica VE, Upadhyaya BB, Del SA. Prediction of disease-gene-drug relationships following a differential network analysis. Cell Death Dis. 2016;7:e2040.
9. Jiang R. Walking on multiple disease-gene networks to prioritize candidate genes. J Mol Cell Biol. 2015;7(3):214–30.
10. Hu G, Chen J. A genome-wide regulatory network identifies key transcription factors for memory CD8(+) T-cell development. Nat Commun. 2013;4:2830.
11. Li Y, Shao T, Jiang C, Bai J, Wang Z, Zhang J, Zhang L, Zhao Z, Xu J, Li X. Construction and analysis of dynamic transcription factor regulatory networks in the progression of glioma. Sci Rep. 2015;5:15953.
12. Marbach D, Lamparter D, Quon G, Kellis M, Kutalik Z, Bergmann S. Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. Nat Methods. 2016;13(4):366–70.
13. Roy S, Ernst J, Kharchenko PV, Kheradpour P, Negre N, Eaton ML, Landolin JM, Bristow CA, Ma L, Lin MF, et al. Identification of functional elements and regulatory circuits by Drosophila modENCODE. SCIENCE. 2010;330(6012):1787–97.
14. Chen X, Gu J, Wang X, Jung JG, Wang TL, Hilakivi-Clarke L, Clarke R, Xuan J. CRNET: an efficient sampling approach to infer functional regulatory networks by integrating large-scale ChIP-seq and time-course RNA-seq data. BIOINFORMATICS. 2017.
15. Cheng C, Alexander R, Min R, Leng J, Yip KY, Rozowsky J, Yan KK, Dong X, Djebali S, Ruan Y, et al. Understanding transcriptional regulation by integrative analysis of transcription factor binding data. Genome Res. 2012;22(9):1658–67.
16. Neph S, Stergachis AB, Reynolds A, Sandstrom R, Borenstein E, Stamatoyannopoulos JA. Circuitry and dynamics of human transcription factor regulatory networks. CELL. 2012;150(6):1274–86.
17. Verfaillie A, Imrichova H, Atak ZK, Dewaele M, Rambow F, Hulselmans G, Christiaens V, Svetlichnyy D, Luciani F, Van den Mooter L, et al. Decoding the regulatory landscape of melanoma reveals TEADS as regulators of the invasive cell state. Nat Commun. 2015;6:6683.
18. Tang B, Hsu HK, Hsu PY, Bonneville R, Chen SS, Huang TH, Jin VX. Hierarchical modularity in ERalpha transcriptional network is associated with distinct functions and implicates clinical outcomes. Sci Rep. 2012;2:875.
19. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE. High-resolution mapping and characterization of open chromatin across the genome. CELL. 2008;132(2):311–22.
20. Min X, Zeng W, Chen N, Chen T, Jiang R. Chromatin accessibility prediction via convolutional long short-term memory networks with k-mer embedding. BIOINFORMATICS. 2017;33(14):i92–i101.
21. Neph S, Vierstra J, Stergachis AB, Reynolds AP, Haugen E, Vernot B, Thurman RE, John S, Sandstrom R, Johnson AK, et al. An expansive human regulatory lexicon encoded in transcription factor footprints. NATURE. 2012;489(7414):83–90.
22. Pique-Regi R, Degner JF, Pai AA, Gaffney DJ, Gilad Y, Pritchard JK. Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. Genome Res. 2011;21(3):447–55.
23. Sherwood RI, Hashimoto T, O'Donnell CW, Lewis S, Barkal AA, van Hoff JP, Karun V, Jaakkola T, Gifford DK. Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. Nat Biotechnol. 2014;32(2):171–8.
24. Sheffield NC, Thurman RE, Song L, Safi A, Stamatoyannopoulos JA, Lenhard B, Crawford GE, Furey TS. Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. Genome Res. 2013;23(5):777–88.

25. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, et al. Integrative analysis of 111 reference human epigenomes. NATURE. 2015;518(7539):317–30.
26. Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan KK, 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.
27. Mathelier A, Fornes O, Arenillas DJ, Chen CY, Denay G, Lee J, Shi W, Shyr C, Tan G, Worsley-Hunt R, et al. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 2015.
28. Wingender E, Chen X, Hehl R, Karas H, Liebich I, Matys V, Meinhardt T, Pruss M, Reuter I, Schacherer F. TRANSFAC: an integrated system for gene expression regulation. Nucleic Acids Res. 2000;28(1):316–9.
29. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. BIOINFORMATICS. 2011;27(12):1739–40.
30. Minka TP. Expectation propagation for approximate Bayesian inference. In: *UAI'01*. San Francisco, CA, USA; 2001. p. 362–9.
31. Sandberg R, Ernberg I. Assessment of tumor characteristic gene expression in cell lines using a tissue similarity index (TSI). Proc Natl Acad Sci U S A. 2005;102(6):2052–7.
32. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A. 2001;98(9): 5116–21.
33. Besag J. On the statistical analysis of dirty pictures. J R Stat Soc Ser B Methodol. 1986:259–302.
34. Yang EW, Girke T, Jiang T. Differential gene expression analysis using coexpression and RNA-Seq data. BIOINFORMATICS. 2013;29(17):2153–61.
35. Wei Z, Li H. A Markov random field model for network-based analysis of genomic data. BIOINFORMATICS. 2007;23(12):1537–44.
36. Kolmogorov V, Zabih R. What energy functions can be minimized via graph cuts? IEEE Trans Pattern Anal Mach Intell. 2004;26(2):147–59.
37. Weiss Y. Correctness of local probability in graphical models with loops. Neural Comput. 2000;12(1):1–41.
38. Chavez L, Bais AS, Vingron M, Lehrach H, Adjaye J, Herwig R. In silico identification of a core regulatory network of OCT4 in human embryonic stem cells using an integrated approach. BMC Genomics. 2009;10:314.
39. Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. CELL. 2005;122(6):947–56.
40. Rosse C, Mejino JJ. A reference ontology for biomedical informatics: the foundational model of anatomy. J Biomed Inform. 2003;36(6):478–500.
41. Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. Nucleic Acids Res. 2004;32(Database issue):D267–70.
42. Peluso I, Pallone F, Monteleone G. Interleukin-12 and Th1 immune response in Crohn's disease: pathogenetic relevance and therapeutic implication. World J Gastroenterol. 2006;12(35):5606–10.
43. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. A map of human genome variation from population-scale sequencing. NATURE. 2010;467(7319):1061–73.
44. Franke A, McGovern DP, Barrett JC, Wang K, Radford-Smith GL, Ahmad T, Lees CW, Balschun T, Lee J, Roberts R, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat Genet. 2010;42(12):1118–25.
45. Lamparter D, Marbach D, Rueedi R, Kutalik Z, Bergmann S. Fast and rigorous computation of gene and Pathway scores from SNP-based summary statistics. PLoS Comput Biol. 2016;12(1):e1004714.
46. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545–50.
47. Schreiber S, Nikolaus S, Hampe J. Activation of nuclear factor kappa B inflammatory bowel disease. GUT. 1998;42(4):477–84.
48. Ellis RD, Goodlad JR, Limb GA, Powell JJ, Thompson RP, Punchard NA. Activation of nuclear factor kappa B in Crohn's disease. Inflamm Res. 1998;47(11):440–5.
49. Visekruna A, Joeris T, Seidel D, Kroesen A, Loddenkemper C, Zeitz M, Kaufmann SH, Schmidt-Ullrich R, Steinhoff U. Proteasome-mediated degradation of IkappaBalpha and processing of p105 in Crohn disease and ulcerative colitis. J Clin Invest. 2006;116(12):3195–203.
50. Atreya I, Atreya R, Neurath MF. NF-kappaB in inflammatory bowel disease. J Intern Med. 2008;263(6):591–6.
51. Jiang M, Ma Y, Chen C, Fu X, Yang S, Li X, Yu G, Mao Y, Xie Y, Li Y. Androgen-responsive gene database: integrated knowledge on androgen-responsive genes. Mol Endocrinol. 2009;23(11):1927–33.
52. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30.
53. Kelsey G, Stegle O, Reik W. Single-cell epigenomics: recording the past and predicting the future. SCIENCE. 2017;358(6359):69–75.
54. Liu Z, Lou H, Xie K, Wang H, Chen N, Aparicio OM, Zhang MQ, Jiang R, Chen T. Reconstructing cell cycle pseudo time-series via single-cell transcriptome data. Nat Commun. 2017;8(1):22.
55. Liu C, Louhimo R, Laakso M, Lehtonen R, Hautaniemi S. Identification of sample-specific regulations using integrative network level analysis. BMC Cancer. 2015;15:319.
56. Liu X, Wang Y, Ji H, Aihara K, Chen L. Personalized characterization of diseases using sample-specific networks. Nucleic Acids Res. 2016;44(22): e164.
57. Kuijjer ML, et al. Estimating sample-specific regulatory networks. arXiv; 2015. preprint 1505.06440.
58. Duren Z, Chen X, Jiang R, Wang Y, Wong WH. Modeling gene regulation from paired expression and chromatin accessibility data. Proc Natl Acad Sci U S A. 2017;114(25):E4914–23.
59. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: a method for assaying chromatin accessibility genome-wide. Curr Protoc Mol Biol. 2015;109:21–9.
60. Mumbach MR, Rubin AJ, Flynn RA, Dai C, Khavari PA, Greenleaf WJ, Chang HY. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. Nat Methods. 2016;13(11):919–22.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions