SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •
Special Focus on Biomolecular Network Analysis and Application

July 2016, Vol. 59 070105:1–070105:10
doi: 10.1007/s11432-016-5582-0

Understanding tissue-specificity with human tissue-specific regulatory networks

Weili GUO, Lin ZHU, Suping DENG, Xingming ZHAO* & Deshuang HUANG*

College of Electromics and Information Engineering, Tongji University, Shanghai 201804*, China*

Received March 28, 2016; accepted May 5, 2016; published online June 13, 2016

Abstract Tissue-specificity is important for the function of human body. However, it is still not clear how the functional diversity of different tissues is achieved. Here we construct gene regulatory networks in 13 human tissues by integrating large-scale transcription factor (TF)-gene regulations with gene and protein expression data. By comparing these regulatory networks, we find many tissue-specific regulations that are important for tissue identity. In particular, the tissue-specific TFs are found to regulate more genes than those expressed in multiple tissues, and the processes regulated by these tissue-specific TFs are closely related to tissue functions. Moreover, the regulations that are present in certain tissue are found to be enriched in the tissue associated disease genes, and these networks provide the molecular context of disease genes. Therefore, recognizing tissue-specific regulatory networks can help better understand the molecular mechanisms underlying diseases and identify new disease genes.

Keywords tissue-specificity, gene regulatory network, transcription factor, tissue-specific regulation, disease gene

Citation Guo W L, Zhu L, Deng S P, et al. Understanding tissue-specificity with human tissue-specific regulatory networks. Sci China Inf Sci, 2016, 59(7): 070105, doi: 10.1007/s11432-016-5582-0

1 Introduction

Cells in different tissues behave diversity in morphology and perform different functions despite sharing identical genetic information. For example, certain diseases are initiated in specific tissues and the corresponding treatments are generally designed for these tissues. In the literature, it has been assumed that tissue-specificity is determined by those genes that are specifically expressed in the tissue. However, to perform their functions within cells, genes should interact with each other. Therefore, the tissue-specificity is determined by gene regulations instead of certain genes [1,2], where dysregulations may lead to disease. Accordingly, target therapy could be designed based on tissue-specific regulations [1,3–6]. Pinpointing the regulatory circuits underlying tissue-specificity can help give insight into the developmental and pathological processes of tissues.

In recent years, many gene regulations have been identified and identification of gene regulatory networks has been extensively studied [2, 7–9]. Recently, many gene regulations have been determined experimentally. For example, the Encyclopedia of DNA elements (ENCODE) project provides large-scale

* Corresponding author (email: xm_zhao@tongji.edu.cn, dshuang@tongji.edu.cn)

binding data for transcription factors (TFs) based on chromatin immunoprecipitation combined with high-throughput sequencing (ChIP-Seq). In addition, curated TF database JASPAR and TRANSFAC provide regulatory motifs for TFs [10, 11], and other databases like TRED [12], ORegAnno [13] and TRRUST [14], collect TF-gene regulation interactions from the literature or experimental sources. Although large amounts of data have been accumulated in terms of gene regulation, these data are generated independently and distributed in different sources. Some computational approaches have also been proposed to construct gene regulatory networks. For instance, we have developed a new approach named NARROMI to infer gene regulatory networks based on gene expression data [15]. Gersten et al. [7] constructed a regulatory network including 119 TFs based on ChIP-seq data. The gene regulatory network provided a global view on the interactions among genes, which could help identify dysregulations responding to perturbations, such as the case in diseases. Despite the fact that these existing regulations and regulatory networks are useful, most can only be used as a reference since temporal and spatial information is not considered. Although some approaches have been proposed to construct tissue-specific gene regulatory networks, false positives are prevalent in these networks. For example, Li et al. [16] constructed gene regulatory networks for different tissues by predicting TF target genes based on their binding motifs from the TRANSFAC database, wherein a gene was regarded to be targeted by one TF if its binding motif occured in the promoter of the gene. Based on the gene expression data from GTEx, Emma et al. [2] built gene co-expression networks for different tissues, however co-expression does not necessarily mean regulation.

In this paper, we construct gene regulatory networks for 13 human tissues by integrating large scale known gene regulations with gene and protein expression profile across human tissues. We then conduct a detailed analysis of tissue specific regulatory networks across tissues, searching for regulatory principles underlying the diversity of tissue functions. We mainly focus on the properties and functions of tissue-specific transcription factors (TSTFs) and tissue-specific regulations in tissue-specificity. By investigating these tissue-specific gene regulatory networks, we find many tissue-specific regulations that are important for tissue identity. In particular, through analyzing the topological properties of TSTFs in the network as well as their target genes, we find that TSTFs regulate more genes compared with TFs expressed in multiple tissues, and the processes regulated by these TSTFs are closely related to tissue functions. Moreover, disease analysis for the regulations that are present in certain tissues reveals that they are enriched in the tissue associated disease genes, and the regulatory networks provide the molecular context of disease genes. Therefore, tissue-specific regulatory networks can help better understand tissue-specificity and the molecular mechanisms underlying diseases, and can identify new disease genes.

2 Methods

2.1 Construction of gene regulatory networks

Before constructing tissue-specific regulatory networks, a background gene regulatory network was first constructed by integrating the experimentally determined regulations between TFs and genes and those reported in literature. We collected experimentally determined regulations from the Encyclopedia of DNA elements (ENCODE) ChIP-seq data [17] and JASPAR [10]. Because the regulations between TFs and genes are not directly available, we used the regulatory DNA elements of transcriptional factor binding sites (TFBS) from ENCODE and JASPAR to determine the regulations. For the TFBSs in ENCODE, we obtained the human transcription factor binding sites from the table Unique TFBS of UCSC Genome Browser [17], and identified the potential regulations between TFs and their target genes with TIP [18] as described in [7]. For the TFBS data in JASPAR, we obtained the TF binding regions using the MotifFeatures and AnnotatedFeatures tables from Ensembl [19]. After mapping the binding sites to human genome (UCSC hg19 human genome assembly), a TF was regarded to regulate one gene if its TFBS lies in the region 1-kb upstream and 500-bp downstream around the transcription start site (TSS) of the gene. In addition, we downloaded gene regulation data from 7 public databases, including TRED [12], ORegAnno [13], BCI [20], PAZAR [21], TRRUST [14], TFactS [22] and FANTOM4 [23].

Finally we built a background human gene regulatory network consisting of 925 TFs and 19855 genes with 326986 regulations.

The protein expression data across 30 human tissues were obtained from the Human Proteome Map [24]. In particular, a protein was regarded as present in one tissue if its expression was captured in the particular tissue (those with an expression value above zero were considered as present). In a similar way, the gene expression profiles across 36 tissues were obtained from the GEO database (GSE2361 [25]), and the data were processed using MAS5. For each tissue, only the TFs that were expressed at both the gene and protein level, and the genes expressed at gene level were considered, and accordingly the regulations between those TFs and genes were extracted from the background regulatory network to construct the tissue-specific regulatory network. As a result, 13 gene regulatory networks were constructed with each network for one tissue.

2.2 Identification of tissue-specific genes, housekeeping genes and disease genes

Generally, tissue-specific genes are considered as particularly expressed and function in one or several tissues. With the gene expression profiles of 36 human tissues, 3378 tissue-specific genes were identified using the method in [26]. We validated the extracted TSGs with the TiGER database, the result showed that on average 90% of the TSGs in each tissue overlaped with those of TiGER, illustrating the reliability of the TSGs. One TF was called a tissue-specific TF if its expression was dependent at protein level, and tissue-specific TFs were identified using a similar method to that with TSGs. TF was considered as a general TF if it was expressed in at least 90% of tissues. As a result, 150 tissue-specific, and 119 general TFs were obtained for 13 tissues.

The housekeeping (HK) genes are essential for biological systems and are widely expressed in various tissues. Here, the 2064 HK genes were collected from [26]. In general, each disease only occurs in certain tissues. To verify the tissue-specific gene regulatory networks, for each disease, we investigated if the tissue-specific regulations were involved in the disease. The tissue-associated diseases were extracted from the high-confident disease-tissue covariation matrix as described in [4]. The disease-gene associations were obtained from the Online Mendelian Inheritance in Man (OMIM) database [27]. For the 13 tissues considered here, 698 tissue-specific disease genes were extracted.

3 Results

3.1 Gene regulatory networks for 13 tissues

In this paper, 13 tissues (heart, ovary, kidney, pancreas, prostate, colon, placenta, fetal brain, spinal cord, testis, liver, lung and fetal liver) were considered since both gene and protein expression data are not available for other tissues. For each tissue, a specific gene regulatory network was constructed based on the background regulatory network, wherein a TF was required to be expressed at both protein and transcriptional level, and genes were required to be expressed at transcriptional level in the tissue. Table 1 shows the statistics of the gene regulatory networks for the 13 tissues. It can be seen that the network sizes are quite different for distinct tissues, wherein the number of TFs varies from 104 to 276 and the number of genes varies from 3737 to 5972.

3.2 Tissue-specific TFs play important roles in tissue-specificity

Since transcription factors regulate the expression of genes, we first investigated the regulatory features of TFs in the networks and the role TSTFs play in defining tissue identity, and how TFs cooperate to regulate gene expression in different tissues. By examining the tissues in which the TFs are expressed at the protein level, we found that TFs were expressed in various numbers of tissues as shown in Figure 1(a), wherein most TFs were expressed in fewer than three tissues and a small number of TFs were expressed in all 13 tissues. By grouping the TFs into tissue-specific and general TFs as described in Methods, we investigated the protein expression levels of the two groups of TFs. As shown in Figure 1(b), general

Table 1 Number of TFs, genes, and regulations of gene regulatory network in each tissue

Tissue	TF	Gene	Edge	Tissue	TF	Gene	Edge
Heart	104	3737	13414	Fetal brain	175	4863	32868
Ovary	276	5404	56182	Spinal cord	147	5392	24143
Kidney	114	4766	18206	Testis	255	5804	47892
Pancreas	195	4363	29361	Liver	165	4169	24232
Prostate	233	5972	46498	Lung	192	5826	43099
Colon	150	4790	22978	Fetal liver	174	4749	31131
Placenta	147	5229	27323				

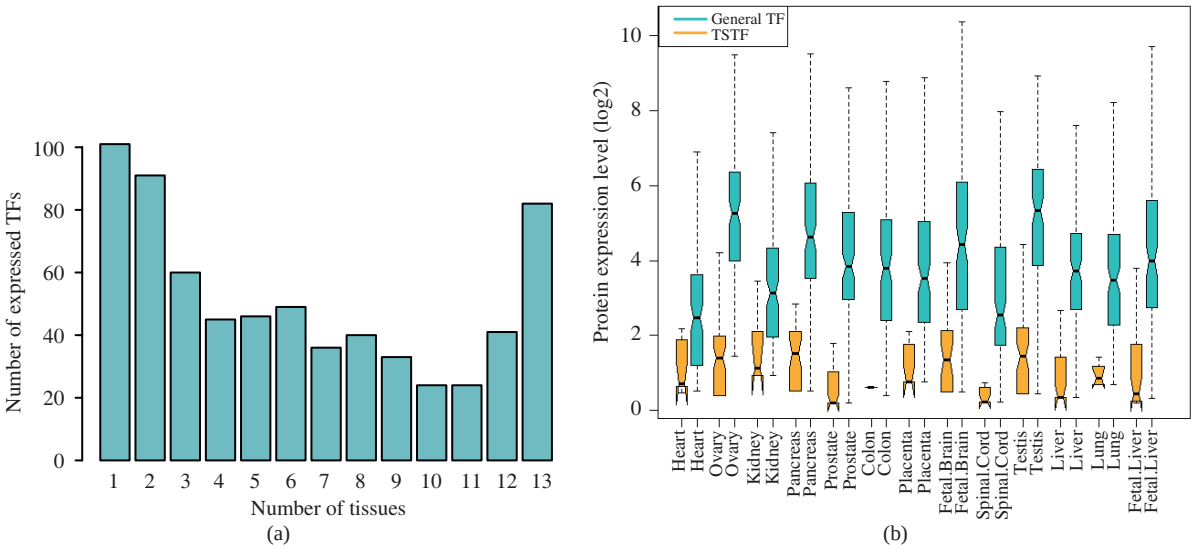

Figure 1 Expression level of TFs in 13 tissues. (a) Number of tissues in which TFs are expressed. TFs are likely to be expressed either specifically or generally; (b) Distribution of protein expression level for TSTFs (yellow) and general TFs (cyan) in each tissue, shown as a boxplot. In all tissues, TSTFs have lower average expression than general TFs.

TFs were found to have significantly higher expression level than TSTFs across all 13 tissues (p-value $< 1 \times 10^{-8}$, Wilcoxon rank-sum test), which is consistent with the previous conclusion that universally expressed genes are likely to have higher expression levels [28].

By investigating the out-degrees of TFs in the tissue-specific gene regulatory networks, we found that TSTFs tended to have higher out-degrees than general TFs (p-value < 0.001; Wilcoxon rank-sum test), which indicated that on average TSTFs regulate more genes than general TFs. Figure 2 shows the out-degree distributions of both TSTFs and general TFs in 13 tissues, from which we can see that the out-degrees of TSTFs vary among different tissues whereas those of general TFs are similar across tissues, indicating the diversity of TSTFs activities and the stability of general TFs across tissues.

Proteins are translated from genes, and the expression of proteins should be proportional to that of their coding genes. Therefore, it is expected that those TSTFs should also be TSGs. However, we found that the TSTFs have no tendency to be TSGs in tissues (p-value $< 1 \times 10^{-3}$, Fisher exact test). Furthermore, we investigated the target genes of TSTFs to determine if their target genes tend to be TSGs that are generally regarded as playing important roles in tissues-specific function. Surprisingly, it was discovered that target genes regulated by TSTFs were not enriched in TSGs (Fisher exact test, p-vlaue < 0.05). Despite both TSTFs and their target genes are not enriched in TSGs, we doubt whether they still have important roles in the determination of tissue-specificity. We performed functional enrichment analysis on the genes regulated by TSTFs using DAVID [29], a functional annotation tool. The functional enrichment analysis results showed that these genes were enriched in tissue-specific functions. For example, in fetal liver (Table 2), the enriched functions included liver development (GO:0001889), regulation of

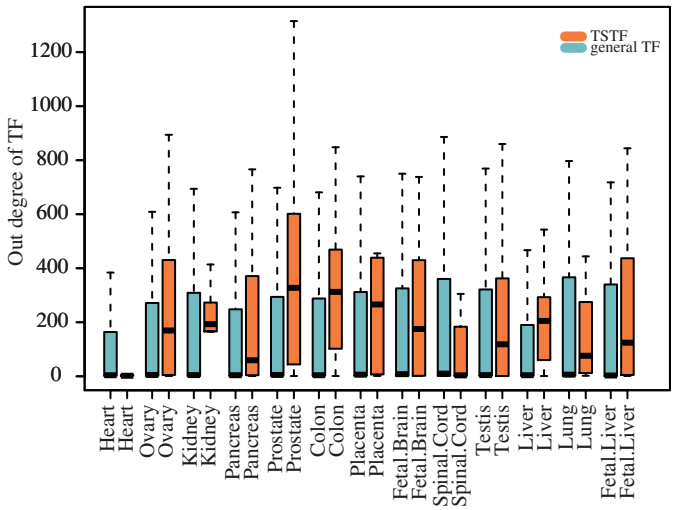

Figure 2 Out-degrees of TSTFs and general TFs in 13 tissues.

cellular protein metabolic process (GO:0032268), sterol metabolic process (GO:0016125) and cholesterol metabolic process (GO:0008203), which are well known liver functions. In the fetal brain (Table 2), the enriched functions, including regulation of neuron apoptosis (GO:0043523), gliogenesis (GO:0042063), neuron projection development (GO:0031175) and in utero embryonic development (GO:0001701), are essential biological processes for neuron and fetal brain development. From the function of the target genes of TSTFs, we can see that these TFs play key roles in determining tissue identity, although the regulators themselves are not TSGs. The analysis on both TSTFs and their target genes indicated that the tissue-specific regulations could provide complementary information with TSGs about tissue-specificity. Despite target genes not being enriched in TSGs, we noticed that the TSGs tended to be regulated by TSTFs in most tissues. In 6 out of 13 tissues, TSTFs were significantly enriched in the regulators of TSGs (p-value < 0.05, Fishers exact test), which implies that TSTFs indeed play important roles in determining tissue-specificity.

Based on the above interesting findings, we further explored the target genes of those TFs in each tissue-specific gene regulatory network. It was discovered that the target genes of general TFs were more likely to be TFs (43% on average) compared to those of TSTFs (p-value < 0.001, Wilcoxon rank-sum test). The findings indicated that the TSTFs tended to directly regulate the expression of genes, whereas general TFs tend to cooperate with each other in regulating gene expression. Furthermore, the target genes of general TFs were enriched in housekeeping genes (p-value $< 1 \times 10^{-5}$ in 11 out of 13 tissues, Fisher exact test), which means that general TFs are likely involved in the essential functions of biological systems by regulating housekeeping genes. We speculated that general TFs and TSTFs have different roles in the biological systems. General TFs tend to be involved in complex regulations and essential biological processes, whereas the TSTFs tend to directly regulate those genes related to tissue-specific functions, such as the morphogenesis and differentiation of tissues.

3.3 Tissue-specific regulations can help understand tissue-specific diseases

Except for TSTFs, we also noticed some gene regulatory interactions appeared in only one tissue, and these regulations were called tissue-specific regulations hereafter. Note that the tissue-specific regulations may or may not consist of TSTFs or TSGs. Table 3 shows the statistics of the tissue-specific regulations for 13 tissues. Since these regulations only occur in certain tissues, we believe they are important for the function of tissues. We investigated the functions of genes involved in tissue-specific regulations by performing functional enrichment analysis, and found that those genes are related to tissue-specific functions. For example, the genes from heart-specific regulations were found to be enriched in heart specific functions, such as heart development (GO: 0007507), regulation of muscle contraction (GO: 0006937), blood circulation (GO: 0008015), and cardiac muscle contraction (KEGG pathway: hsa04260).

Table 2 The functions and pathways enriched in the target genes of tissue-specific TFs in Fetal liver and Fetal brain respectively (p-value < 0.05)

Tissue	Category	Term	p-value
Fetal liver	GOTERM_BP	GO:0007049 cell cycle	2.2×10^{-30}
	GOTERM_BP	GO:0032268 regulation of cellular protein metabolic process	1.5×10^{-18}
	GOTERM_BP	GO:0001889 liver development	6.0×10^{-8}
	GOTERM_BP	GO:0002520 immune system development	3.7×10^{-7}
	GOTERM_BP	GO:0016125 sterol metabolic process	4.8×10^{-6}
	GOTERM_BP	GO:0008203 cholesterol metabolic process	6.4×10^{-6}
	GOTERM_BP	GO:0034101 erythrocyte homeostasis	1.0×10^{-5}
	GOTERM_BP	GO:0034381 lipoprotein particle clearance	3.5×10^{-4}
Fetal brain	GOTERM_BP	GO:0043523 regulation of neuron apoptosis	2.4×10^{-4}
	GOTERM_BP	GO:0042063 gliogenesis	2.5×10^{-4}
	GOTERM_BP	GO:0031175 neuron projection development	4.6×10^{-4}
	GOTERM_BP	GO:0001701 in utero embryonic development	8.4×10^{-3}
	GOTERM_BP	GO:0048666 neuron development	1.8×10^{-2}
	GOTERM_BP	GO:0050770 regulation of axonogenesis	2.2×10^{-2}
	KEGG_PATHWAY	hsa04520 : Adherens junction	5.9×10^{-6}
	KEGG_PATHWAY	hsa05016 : Huntington's disease	1.3×10^{-5}
	KEGG_PATHWAY	hsa05214 : Glioma	1.5×10^{-5}
	KEGG_PATHWAY	hsa05010 : Alzheimer's disease	1.8×10^{-5}

Table 3 Number of tissue-specific regulations

Tissue	Tissue-specific regulation	Tissue	Tissue-specific regulation
Heart	110	Fetal Brain	3093
Ovary	5274	Spinal Cord	2303
Kidney	533	Testis	5042
Pancreas	304	Liver	1697
Prostate	4164	Lung	2218
Colon	218	Fetal Liver	1306
Placenta	4222		

Since TSGs are generally thought to be important for tissue-specific functions, we compared the enriched functions of tissue-specific regulations with those of TSGs, and found that specific regulations can provide additional and complementary information about tissue functions, especially pathway information. For instance, the brain specific regulations were enriched in biological processes, such as neuron differentiation, gliogenesis, and forebrain development (Table 4), which are crucial for fetal brain development. Moreover, the tissue-specific regulations provided additional information about the molecular pathways underlying tissue functions compared with tissue-specific genes, whereas the pathways can help better understand the functions and development of tissues [1, 3, 8]. For example, in the fetal brain, both tissue-specific genes and tissue-specific regulations were enriched in brain-related biological processes (Table 4). However, the regulations provided more information about brain-related pathways that cannot be provided by TSGs. Pathways like Wnt signaling pathway, TGF-beta signaling pathway, and Adherens junction are enriched in the brain-specific regulations, and have been shown to play important roles in brain functions [30, 31]. For instance, recent evidence indicates that the Wnt signaling and TGF-beta signaling pathways play important roles in the proper function of brain neural circuitry as well as the brain vascular network [30, 32]. Through VEGF and Wnt signaling, neural progenitors facilitate the ingression of blood vessels from the neural tube and communicate with endothelial cells to stabilize

Table 4 The functions, pathways and disease processes enriched in fetal brain-specific regulations and genes (p-value < 0.05)

Tissue-specific regulations		Tissue-specific genes	
Category	Term	Category	Term
GOTERM_BP	Regulation of gene-specific transcription	GOTERM_BP	Neuron differentiation
GOTERM_BP	Chromatin organization	GOTERM_BP	Axonogenesis
GOTERM_BP	Neuron differentiation	GOTERM_BP	Neuron projection development
GOTERM_BP	Gliogenesis	GOTERM_BP	Cell morphogenesis involved in neuron differentiation
GOTERM_BP	Embryonic development ending in birth or egg hatching	GOTERM_BP	Neuron development
GOTERM_BP	Neuron projection development	GOTERM_BP	Axon guidance
GOTERM_BP	Pallium development	GOTERM_BP	Neuron migration
GOTERM_BP	Forebrain development	GOTERM_BP	Forebrain development
KEGG_PATHWAY	Pathways in cancer	GOTERM_BP	Central nervous system neuron development
KEGG_PATHWAY	Wnt signaling pathway	GOTERM_BP	Memory
KEGG_PATHWAY	TGF-beta signaling pathway	KEGG_PATHWAY	Gap junction
KEGG_PATHWAY	Axon guidance	KEGG_PATHWAY	Axon guidance
KEGG_PATHWAY	Alzheimer's disease	GENETIC_ASSOCIATION_DB_DISEASE	Huntington disease-like
KEGG_PATHWAY	Glioma		
GENETIC_ASSOCIATION_DB_DISEASE	Huntington disease-like		
GENETIC_ASSOCIATION_DB_DISEASE	Schizophrenia		
GENETIC_ASSOCIATION_DB_DISEASE	Alzheimer's disease cognitive function		

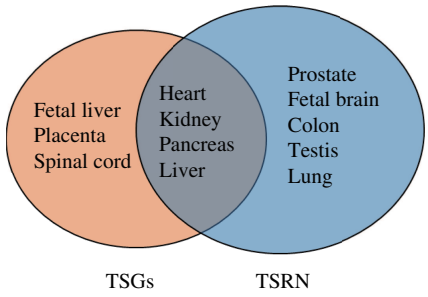

Figure 3 (Color online) Tissues in which tissue-specific genes and tissue-specific regulations are enriched.

nascent brain vessels by down-regulating Wnt pathway. Furthermore, the nascent brain vessel integrity was promoted through integrin v8-dependent TGF signaling crucial processes for establishing the brain vascular network [30, 32].

In general, disease occurs in specific tissues, and the development of disease involves the dysregulation of many genes. Therefore, we hypothesized that tissue-specific regulatory networks could help identify disease-related processes. For example, brain-specific regulations were enriched in several well-known brain diseases, such as Alzheimer's disease, schizophrenia, and glioma, whereas this information could not be provided by genes specifically expressed in brain. It has been discovered that pathways enriched in brain-specific regulations are important for vascular function. Increasing evidence indicates that vascular dysfunction, such as vessel integrity, plays an important role in the pathogenesis of many brain-associated diseases such as stroke, Alzheimer's disease, and Huntington disease [33]. To further investigate the associations between tissue-specificity and diseases, we collected tissue-specific disease genes, wherein the tissue-disease associations were obtained from [4] and disease associated genes were retrieved from

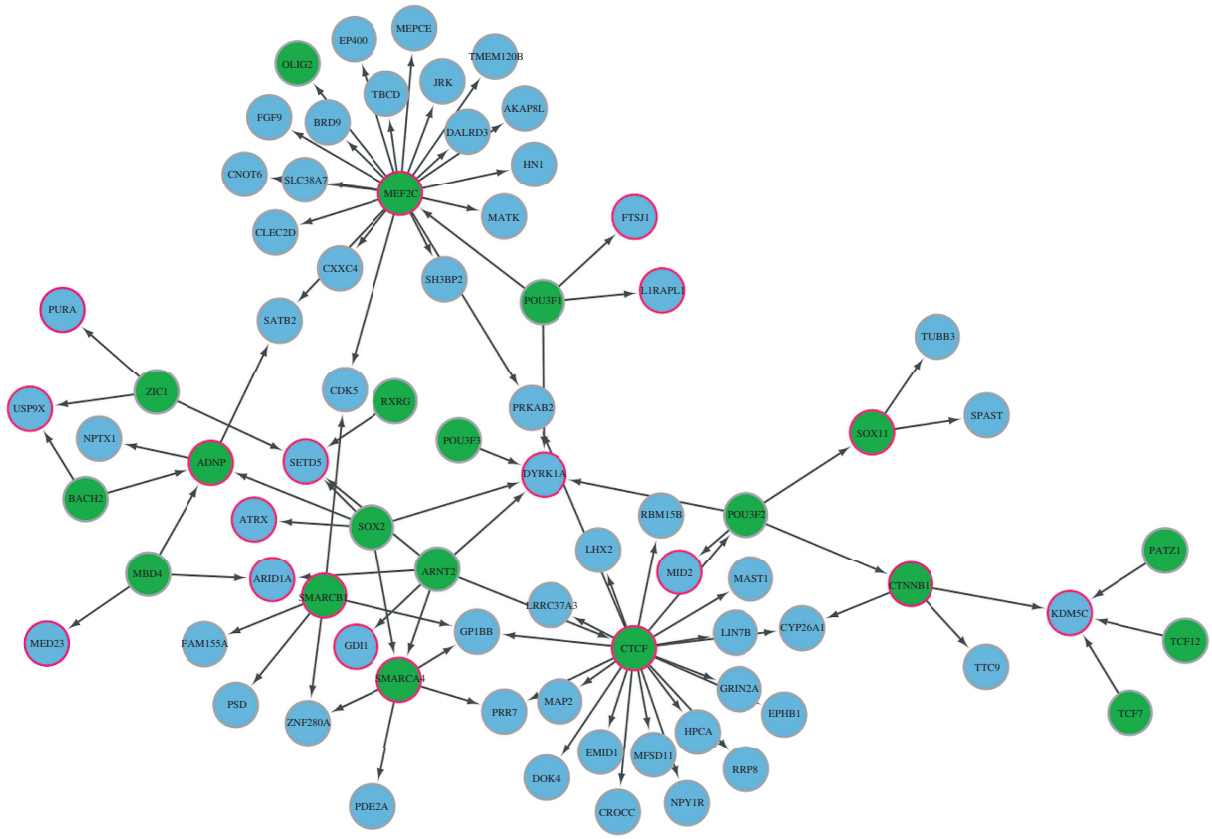

Figure 4 The gene regulatory network of mental retardation associated genes. The TFs and their target genes are marked in green and blue, and disease genes are marked with a red circle.

OMIM. Figure 3 shows the tissues in which both TSGs and tissue-specific regulations were enriched in the corresponding tissue-specific disease genes (p-value < 0.05, Fishers exact test). From the results, we can see that TSGs were enriched in tissue-specific disease genes in seven tissues while tissue-specific regulations were enriched in tissue-specific disease genes in nine tissues, including four common tissues. These findings indicate that the dysfunction of some tissue-specific regulations can lead to some diseases, which is consistent with the previous conclusion that some diseases happen due to the dysregulation of molecular networks instead of single genes [1, 34, 35].

3.4 Case study: the regulatory network of intellectual disability associated genes

As mentioned above, tissue-specific regulations are enriched in tissue-specific disease genes; tissue-specific regulatory networks can also help to understand how these disease genes are regulated. For example, Figure 4 shows the regulatory network consists of 19 mental retardation (also known as intellectual disability) disease genes from OMIM as well as their related regulations from the fetal brain-specific regulatory network. In the network, KDM5C is a well-known disease gene for X-linked intellectual disability [36], and it is regulated by four TFs, including PATZ1, TCF12, TCF7, and CTNNB1. Among those TFs, CTNNB1 is already known as a mental retardation disease gene, and the mutation of TCF12 was reported in patients with developmental delay or learning disability. TCF12 microdeletion was supposed to be responsible for intellectual deficiency [37, 38]. It can be seen that the regulatory network can tell how the disease genes are regulated and may also help identify new disease genes. It was determined that one disease gene, SETD5, was co-regulated by transcription factors ZIC1 and SOX2, whereas SOX2 regulates four disease genes (ATRX, ADNP, DYRK1A, and SMARCA) and ZIC regulates two other disease genes (PURA and USP9X). We suspected that these two TFs might play important roles in intellectual deficiency. SOX2 is known to be a key factor in the regulation of pluripotency and neural

differentiation, and was found to regulate PQBP1, a mental retardation gene in neural stem progenitor cells [39, 40]. ZIC1 is implicated in vertebrate brain development. The heterozygous loss of ZIC1 and ZIC4 has been suspected to lead to Dandy-Walker along with mental retardation, and mutations in ZIC1 were reported to be associated with learning disability [41, 42]. From the analysis, it can be observed that gene regulatory networks provide rich information about how disease genes are regulated and may help predict new disease genes.

4 Conclusion

Tissue-specificity is important to understand developmental and pathological processes of tissues, which can in turn help design better therapies for the diseases happening in certain tissues. In this paper, we constructed tissue-specific gene regulatory networks for 13 tissues. These regulatory networks provide information on how tissue-specific and disease genes are regulated in different tissues. We also investigate how TSTFs and tissue specific regulatory interactions contribute to tissue specificity. General TFs tended to regulate housekeeping genes and other TFs, whereas tissue-specific TFs preferred to directly regulate gene expression. Although target genes of TSTFs were not enriched in TSGs, TSTFs regulate biological processes important for tissue-specificity, indicating that the regulations can provide complementary information compared with tissue-specific genes. Furthermore, we found many tissue-specific regulations that are important for tissue identity. In particular, the tissue-specific regulations were enriched in tissue-specific disease genes, implying these regulations might be involved in the pathogenesis process. The regulatory networks of disease genes not only provide the regulatory information on disease genes but also can help predict new disease genes. In summary, tissue-specific gene regulatory networks constructed here provide new insights into tissue-specificity, which can help facilitate understanding of the pathological and developmental processes of tissues.

Acknowledgements This work was partly supported by National Natural Science Foundation of China (Grant Nos. 61520106006, 31571364, 61532008, 61411140249, 61133010, 91529303, 61572363), Innovation Program of Shanghai Municipal Education Commission (Grant No. 13ZZ072), Shanghai Pujiang Program (Grant No. 13PJD032), Ph.D. Programs Foundation of Ministry of Education of China (Grant No. 20120072110040), Fundamental Research Funds for the Central Universities, and Special Program for Applied Research on Super Computation of the NSFC-Guangdong Joint Fund (the second phase).

Conflict of interest The authors declare that they have no conflict of interest.

References

1 Greene C S, Krishnan A, Wong A K, et al. Understanding multicellular function and disease with human tissue-specific networks. Nat Genet, 2015, 47: 569–576
2 Pierson E, Koller D, Battle A, et al. Sharing and specificity of co-expression networks across 35 human tissues. Plos Comput Biol, 2015, 11: e1004220
3 Zhao X M, Chen L. Network-based biomarkers for complex diseases. J Theor Biol, 2014, 362: 1–2
4 Lage K, Hansen N T, Karlberg E O, et al. A large-scale analysis of tissue-specific pathology and gene expression of human disease genes and complexes. Proc Nat Acad Sci USA, 2008, 105: 20870–20875
5 Zheng C H, Huang D S, Zhang L, et al. Tumor clustering using non-negative matrix factorization with gene selection. IEEE Trans Inf Technol Biomed, 2009, 13: 599–607
6 Deng S P, Zhu L, Huang D S. Mining the bladder cancer-associated genes by an integrated strategy for the construction and analysis of differential co-expression networks. BMC Genom, 2015, 16: S4
7 Gerstein M B, Kundaje A, Hariharan M, et al. Architecture of the human regulatory network derived from ENCODE data. Nature, 2012, 489: 91–100
8 Ji Z W, Wu D, Zhao W, et al. Systemic modeling myeloma-osteoclast interactions under normoxic/hypoxic condition using a novel computational approach. Sci Rep, 2015, 5: 13291
9 Deng S P, Zhu L, Huang D S. Predicting hub genes associated with cervical cancer through gene co-expression networks. IEEE/ACM Trans Comput Biol Bioinform, 2016, 13: 27–35
10 Mathelier A, Zhao X, Zhang A W, et al. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucl Acid Res, 2013, gkt997

11 Matys V, Fricke E, Geffers R, et al. TRANSFAC: transcriptional regulation, from patterns to profiles. Nucl Acid Res, 2003, 31: 374–378
12 Jiang C, Xuan Z, Zhao F, et al. TRED: a transcriptional regulatory element database, new entries and other development. Nucl Acid Res, 2007, 35: D137–D140
13 Griffith O L, Montgomery S B, Bernier B, et al. ORegAnno: an open-access community-driven resource for regulatory annotation. Nucl Acid Res, 2008, 36: D107–D113
14 Han H, Shim H, Shin D, et al. TRRUST: a reference database of human transcriptional regulatory interactions. Sci Rep, 2015, 5: 11432
15 Zhang X, Liu K, Liu Z P, et al. NARROMI: a noise and redundancy reduction technique improves accuracy of gene regulatory network inference. Bioinformatics, 2013, 29: 106–113
16 Li J, Hua X, Haubrock M, et al. The architecture of the gene regulatory networks of different tissues. Bioinformatics, 2012, 28: i509–i514
17 Consortium E P. An integrated encyclopedia of DNA elements in the human genome. Nature, 2012, 489: 57–74
18 Cheng C, Min R, Gerstein M. TIP: a probabilistic method for identifying transcription factor target genes from ChIP-seq binding profiles. Bioinformatics, 2011, 27: 3221–3227
19 Flicek P, Amode M R, Barrell D, et al. Ensembl 2014. Nucl Acid Res, 2013, gkt1196
20 Lefebvre C, Lim W K, Basso K, et al. A context-specific network of protein-DNA and protein-protein interactions reveals new regulatory motifs in human B cells. Syst Biol Comput Proteom, 2007, 4532: 42–56
21 Portales-Casamar E, Arenillas D, Lim J, et al. The PAZAR database of gene regulatory information coupled to the ORCA toolkit for the study of regulatory sequences. Nucl Acid Res, 2009, 37: D54–D60
22 Essaghir A, Toffalini F, Knoops L, et al. Transcription factor regulation can be accurately predicted from the presence of target gene signatures in microarray gene expression data. Nucl Acid Res, 2010, 38: e120
23 Severin J, Waterhouse A M, Kawaji H, et al. FANTOM4 EdgeExpressDB: an integrated database of promoters, genes, microRNAs, expression dynamics and regulatory interactions. Genome Biol, 2009, 10: R39
24 Kim M S, Pinto S M, Getnet D, et al. A draft map of the human proteome. Nature, 2014, 509: 575–581
25 Ge X, Yamamoto S, Tsutsumi S, et al. Interpreting expression profiles of cancers by genome-wide survey of breadth of expression in normal tissues. Genomics, 2005, 86: 127–141
26 Chang C W, Cheng W C, Chen C R, et al. Identification of human housekeeping genes and tissue-selective genes by microarray meta-analysis. PLoS ONE, 2011, 6: e22859
27 Hamosh A, Scott A F, Amberger J S, et al. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucl Acid Res, 2005, 33: D514–D517
28 Su A I, Wiltshire T, Batalov S, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Nat Acad Sci Usa, 2004, 101: 6062–6067
29 Dennis Jr G, Sherman B T, Hosack D A, et al. DAVID: database for annotation, visualization, and integrated discovery. Genome Biol, 2003, 4: P3
30 Santhosh D, Huang Z. Regulation of the nascent brain vascular network by neural progenitors. Mech Develop, 2015, 138: 37–42
31 Zlokovic B V. The blood-brain barrier in health and chronic neurodegenerative disorders. Neuron, 2008, 57: 178–201
32 Posokhova E, Shukla A, Seaman S, et al. GPR124 functions as a WNT7-specific coactivator of canonical β-catenin signaling. Cell Rep, 2015, 10: 123–130
33 Lynch J K. Epidemiology and classification of perinatal stroke. Semin Fetal Neonatal Med, 2009, 14: 245–249
34 Liu X, Liu Z P, Zhao X M, et al. Identifying disease genes and module biomarkers by differential interactions. J Amer Med Inform Assoc, 2012, 19: 241–248
35 Qin G, Zhao X M. A survey on computational approaches to identifying disease biomarkers based on molecular networks. J Theor Biol, 2014, 362: 9–16
36 Brookes E, Laurent B, Ounap K, et al. Mutations in the intellectual disability gene KDM5C reduce protein stability and demethylase activity. Hum Mol Genet, 2015, ddv046
37 Piard J, Roze V, Gzorny A, et al. TCF12 microdeletion in a 72-year-old woman with intellectual disability. Amer J Med Genet Part A, 2015, 167: 1897–1901
38 Kuechler A, Willemsen M H, Albrecht B, et al. De novo mutations in beta-catenin (CTNNB1) appear to be a frequent cause of intellectual disability: expanding the mutational and clinical spectrum. Hum Genet, 2015, 134: 97–109
39 Li C, Ito H, Fujita K, et al. Sox2 transcriptionally regulates PQBP1, an intellectual disability-microcephaly causative gene, in neural stem progenitor cells. PLoS ONE, 2013, 8: e68627
40 Zhang S C, Cui W. Sox2, a key factor in the regulation of pluripotency and neural differentiation. World J Stem Cells, 2014, 6: 305–311
41 Tohyama J, Kato M, Kawasaki S, et al. Dandy-Walker malformation associated with heterozygous ZIC1 and ZIC4 deletion: report of a new patient. Amer J Med Genet Part A, 2011, 155: 130–133
42 Twigg S R, Forecki J, Goos J A, et al. Gain-of-function mutations in ZIC1 are associated with coronal craniosynostosis and learning disability. Amer J Hum Genet, 2015, 97: 378–388