BIOINFORMATICS ORIGINAL PAPER

Vol. 31 no. 1 2015, pages 76–83
doi:10.1093/bioinformatics/btu597

Gene expression

Advance Access publication September 3, 2014

Gene regulatory networks by transcription factors and microRNAs in breast cancer

Sheng Qin[1,2], Fei Ma[2,*] and Liming Chen[1,*]

[1]The Key Laboratory of Developmental Genes and Human Disease, Ministry of Education, Institute of Life Science, Southeast University Nanjing 210096, China and [2]Laboratory for Comparative Genomics and Bioinformatics & Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Science, Nanjing Normal University, Nanjing 210046, China

Associate Editor: Ivo Hofacker

ABSTRACT

Motivation: Gene regulatory networks (GRNs) affect numerous cellular processes and every process of life, and abnormalities of GRN lead to breast cancer. Transcription factors (TFs) and microRNAs (miRNAs) are two of the best-studied gene regulatory mechanisms. However, the architecture and feature of GRNs by TFs and miRNAs in breast cancer and its subtypes were unknown. In this study, we investigated the GRNs by TFs and miRNAs with emphasis on breast cancer classifier genes at system level.

Contact: feima1968@gmail.com or chenliming1981@gmail.com

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on January 28, 2014; revised on May 30, 2014; accepted on August 29, 2014

1 INTRODUCTION

The products of genes encoded by the genome of organism enable cell survival and many cellular functions (Karlebach and Shamir, 2008). Regulation of gene expression determines the amount and the temporal pattern of these products in the cell that are crucial to the processes of life (Karlebach and Shamir, 2008). Although there are many layers of gene regulation, transcription factors (TFs) and microRNAs (miRNAs) are two of the best-studied gene regulatory mechanisms (Chen and Rajewsky, 2007). TFs regulate gene expression at the transcriptional level, whereas miRNAs, ~22-nuleotide long non-coding RNAs, regulate gene expression at the post-transcriptional level (Chen and Rajewsky, 2007). Expression abnormalities of both coding and non-coding genes were involved in cancer, which is a complex genetic disease (Calin and Croce, 2006).

Cancer is a leading cause of human death. TFs play important roles in expression regulation of both coding and non-coding genes, and at least three main groups of TFs are known to be important in human cancer (Darnell, 2002). Alterations of miRNAs are involved in both the initiation and progression of human cancer, and the link between deregulated miRNAs expression and cancer had been well established (Calin and Croce, 2006; Garzon *et al.*, 2009; Pencheva and Tavazoie, 2013). Understanding gene regulatory networks (GRNs) by TFs and miRNAs will shed light on the mechanisms of cancer, including breast cancer.

Breast cancer is the most common cancer in women, and ~1 million new breast cancer cases were diagnosed each year worldwide (Liu, 2012). Recent studies provided valuable information on the molecular mechanism of breast cancer via integrating information from genome transcriptome to proteome of representative numbers of breast cancer patient samples (Cancer Genome Atlas Network, 2012; Ellis *et al.*, 2012; Shah *et al.*, 2012; Yuan *et al.*, 2012). Gene regulations by TFs and miRNAs have been linked to breast cancer. Several TFs, e.g. oestrogen receptors, are known to be important in breast cancer (Darnell, 2002). MiRNAs, e.g. miR-200 family members, play important roles in breast cancer (Liu, 2012). However, the architecture and feature of GRNs by TFs and miRNAs in breast cancer and its subtypes were largely unknown.

Breast cancer is a remarkably heterogeneous disease. Breast cancer cell lines and primary breast tumors were found to be consistently classified into three major subtypes: luminal (Lu), basal A (BaA) and basal B (BaB) (Ellis *et al.*, 2012; Neve *et al.*, 2006). The breast cancer cell lines were found to compromise a system exhibiting the substantial genomic, transcriptional and biological heterogeneity properties of primary breast cancers (Neve *et al.*, 2006). Breast cancer classifier genes (BCCGs) involved in the differentiation status of the cell type and/or tumor biology were identified to be able to classify breast cancer cell lines into Lu, BaA and BaB groups (Neve *et al.*, 2006). Investigating the pattern and feature of GRNs composed of TFs, miRNAs and BCCGs in breast cancer cell lines will provide valuable information for understanding the molecular mechanism of gene regulation in breast cancer and the molecular basis of breast cancer subtypes.

In this study, we revealed the architecture and feature of GRNs by TFs and miRNAs in breast cancer via systematically investigating GRNs composed of TFs, miRNAs and BCCGs across 33 breast cancer cell lines representing three major breast cancer subtypes (Lu, BaA and BaB). First, TFs, miRNAs and BCCGs were collected from public database and published works, and the relationships among them were analyzed via computational methods combined with gene expression data. Second, for each breast cancer subtype, the correlation between TFs/miRNAs and their targets was calculated. After that, enrichment of miRNA target gene in BCCGs was analyzed,

*To whom correspondence should be addressed.

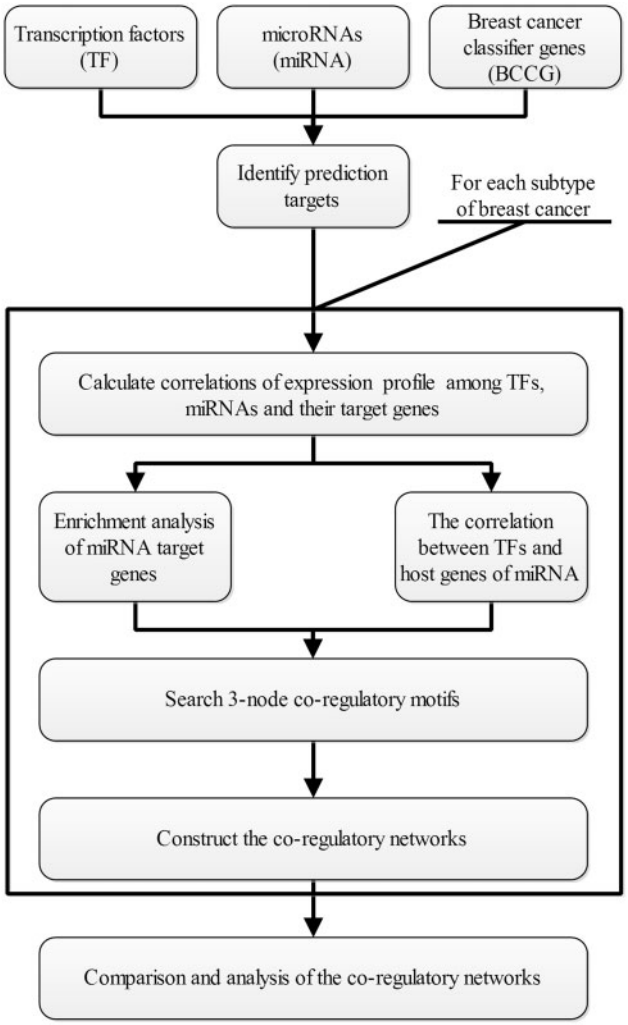

Fig. 1. Workflow. Workflow of the present study determining TF and miRNA co-regulatory network in breast cancer

and the correlation between TFs and miRNA host genes was calculated. Finally, 3-node co-regulatory motifs were identified and used to construct co-regulatory networks. The workflow was deciphered in Figure 1, and the detailed information was provided in Section 2. Our current work presented the architecture and feature of GRNs by TFs and miRNAs in breast cancer at system level.

2 METHODS

2.1 Data collection

All human miRNAs and their conserved predicted targets were downloaded from TargetScan Human V6.2 (http://www.targetscan.org) (Friedman *et al.*, 2009). The expression data of breast cancer cell lines were collected based on previous works, and the procedure was described as follows. mRNA expression and miRNA expression data of human breast cancer cell lines were collected from the studies by Richard M. Neve *et al.* (2006) and Muhammad Riaz *et al.* (2013), respectively. Thirty-three breast cancer cell lines with both mRNA and miRNA expression data were used for further analysis. These 33 cell lines cover three breast cancer subtypes: Lu, BaA and BaB (Supplementary Table S1). In all, 305 probes that were filtered by PAM analysis and used to classify subtypes of breast cancer were derived from the published data (Neve *et al.*, 2006; Tibshirani *et al.*, 2002). These probes were mapped to 264 genes as BCCGs by 'Gene ID Conversion' in the DAVID Web server (Huang *et al.*, 2008). The TFs, conserved TF binding sites (TFBSs) and transcription start sites (TSSs) were obtained from the UCSC Table Browser (Kent *et al.*, 2002). The TSSs of miRNAs were downloaded from miRstart database (Chien *et al.*, 2011).

2.2 MRNA and TF target prediction

MiRNA targets (including TFs and BCCGs with *P*ct $\geq$0.8) were generated using TargetScan (Friedman *et al.*, 2009). TF target genes were identified using the approach developed by Poos and his colleagues (Poos *et al.*, 2013), via defining the $\pm$2000 bp sequence around TSSs as the promoter region. The genes with promoter regions completely overlapped with TFBSs were considered as TF targets.

2.3 Correlation and enrichment analysis

Pearson correlation was used to estimate the relation between miRNAs, TFs and BCCGs. TFs were assumed to promote the expression of their targets, whereas miRNAs were assumed to inhibit the expression of their targets. Therefore, TFs expressions have positive correlations with their target genes ($r \geq 0.3$), and miRNAs expressions have negative correlations with their target genes ($r \leq -0.3$).

In miRstart, the TSS of an intragenic miRNA is the same as its host gene by default. However, not all of these miRNAs share their TSS with their host genes (Ozsolak *et al.*, 2008). In this study, a TF was considered to regulate an intragenic miRNA when the expression of TF has positive correlation with the host gene of the miRNA ($r \geq 0.3$).

Furthermore, we evaluated the significant relationship between the expression of miRNAs and BCCGs via enrichment analysis to restrict the target relations. The enrichment analysis was executed for each miRNA individually. Hyper geometric test was performed to assess the enrichment of miRNA target genes in BCCGs. Multiple-test correction was also performed to control the false discovery rate (FDR) (Benjamini and Hochberg, 1995). MiRNAs with significant enrichment in BCCGs were used for further analysis ($P \leq 0.05$ and FDR $\leq$0.05).

2.4 Co-regulatory motifs and co-regulatory networks analysis

TF-miRNA-BCCG GRNs were constructed for three subtypes of breast cancer based on correlated relationships among miRNAs, TFs and target genes. 3-node motifs were detected by FANMOD software, and two types of co-regulatory motif were filtered out from the dumping results (Wernicke and Rasche, 2006). The co-regulatory motif is defined as that in which a TF and a miRNA share a common target gene and regulate each other in 3-node motif. The GRNs for three breast cancer subtypes (Lu, BaA and BaB) were constructed separately.

To evaluate the significance of co-regulatory motifs in different GRNs, we compared them with those of random networks. To fully randomize a network, each edge in a network was exchanged five times via changing edges with same regulatory relationship among miRNAs, TFs and target genes. In a random network, the number of target genes by a TF or a miRNA and the number of TFs or miRNAs targeting a gene remains unchanged. The random networks were constructed 1000 times and compared with original networks. The *P*-values were computed to estimate the significance of co-regulatory motifs in real networks.

3-node loop motifs were merged to construct GRNs in breast cancer and breast cancer subtypes. In the GRNs, three nodes represent miRNA, TF and BCCG, while two edges represent activation and inhibition

Table 1. Summary of gene regulatory motifs in breast cancer subtypes

Subtype	miRNAs	TFs	Target genes	Interactions	Type I GRM (*P*-value)	Type II GRM (*P*-value)
BaA	19	20	19	140	38 (0.968)	19 (0.095)
BaB	23	30	17	171	42 (0.172)	23 (0.043)
Lu	14	20	7	85	15 (0.418)	17 (0.035)

Notes: The table lists the number of miRNAs, TFs, target genes and regulation interactions and GRM types in breast cancer subtypes.

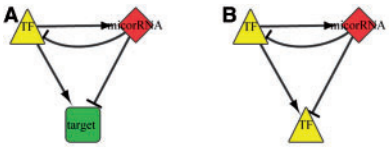

Fig. 2. 3-node co-regulatory motif of two breast cancer GRM types. (**A**) Type I GRM: A TF and a miRNA share a commonly regulated target BCCG. (**B**) Type II GRM: A TF and a miRNA regulate another TF. TFs are marked with triangle, miRNAs with diamond and BCCGs with round rectangle. Arrow indicates active relation, and 'T' indicates inhibitory relation

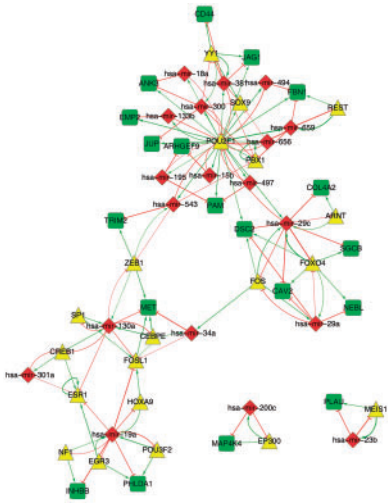

Fig. 3. The co-regulatory network of BaA. TFs are represented as triangles, miRNAs are represented as diamonds and BCCGs are represented as rectangles. Arrow with green edge indicates active relation, and 'T' with red edge indicates inhibitory relation

relation. The GRNs were visualized and analyzed by Cytoscape 3.0.2 (Smoot *et al.*, 2011). To evaluate the importance of nodes, we calculated the degrees and betweenness centrality of each node in each network by Cytoscape 3.0.2. The degree for a node is defined as the number of connections it has to other nodes. The betweenness centrality for a node is the number of shortest paths from all vertices to all others that pass through the node. The Venn and the Euler diagrams were used as a plug-in of Cytoscape to compare GRNs of three breast cancer subtypes (Wilkinson, 2012). The common and specific nodes and edges in networks among the three breast cancer subtypes were analyzed. To estimate whether the intersections of GRNs have biological relationship or just random overlap, significant tests were performed for the intersections among the three GRNs using network analysis tools (Brohée *et al.*, 2008).

3 RESULTS

3.1 Identification of miRNA and TF target genes in the BCCGs

There are 170 of 264 BCCGs targeted by 255 miRNAs and 146 TFs (Supplementary Table S2). In all, 4373 TF-regulating-BCCGs, 1352 miRNA-regulating-BCCGs, 3440 TF-regulating-miRNAs and 2357 miRNA-regulating-TFs pairs were identified (Supplementary Table S2).

The expression of 209 of 255 miRNAs was found to correlate with their target genes at $r \leq -0.3$ (BaA:143; BaB:161; Lu:123) (Supplementary Tables S3 and S4). In addition, 140 miRNAs are significantly enriched in the BCCGs ($P \leq 0.05$ & FDR ≤ 0.05) (BaA:104; BaB:111; Lu:91) (Supplementary Table S5).

In 146 TFs, 129 TFs targeting 160 BCCGs have expression data across the 33 breast cancer cell lines (Supplementary Table S3). Six genes belong to both TFs and BCCGs (*Sox9*, *Gata3*, *Esr1*, *Fosl1*, *Xbp1* and *Pbx1*). The expression of all 129 TFs was found to be associated to their target genes at $r \geq 0.3$ level (BaA: 129; BaB: 126; Lu: 125) (Supplementary Table S4). What's more, 129 TFs were associated with 78 miRNAs using 74 BCCGs as their host genes at $r \geq 0.3$ level (Supplementary Table S6).

3.2 The gene regulatory motifs feature of GRNs by TFs and miRNAs in breast cancer

The basal subtype (BaA and BaB) has significantly more nodes and edges in co-regulatory motif group than the luminal subtype (Lu). BaB contains 70 nodes and 171 edges, BaA contains 58 nodes and 140 edges, whereas Lu contains only 41 nodes and 85 edges (Table 1 and Supplementary Table S7). When TF-miRNA-BCCG GRNs were constructed using interaction data of the three subtypes of breast cancer, two types of gene regulatory motifs (GRMs) were detected (Fig. 2). Type I GRM is where a TF and a miRNA pair share a commonly regulated target gene which is a BCCG, and Type II GRM is where a TF and a miRNA pair commonly regulates another TF. Both Type I and II GRMs were observed in all breast cancer subtypes. Compared with random networks, type II GRM in BaB and Lu was significant ($P \leq 0.05$) (Table 1). This might be owing to regulatory motif that is usually apparent at the hierarchy of TFs rather than other genes.

3.3 The architecture and feature of GRNs by TF and miRNA in breast cancer

To study the architecture and feature of GRNs by TF and miRNA in breast cancer, for each breast cancer subtypes, GRNs were created via joining all the 3-node co-regulatory motifs (Figs 3–5) and merged the three co-regulatory networks into an entire network (Supplementary Figs S1 and S2).

The co-regulatory networks of each breast cancer subtypes clearly exhibit a small number of hubs (Figs 3–5). The most

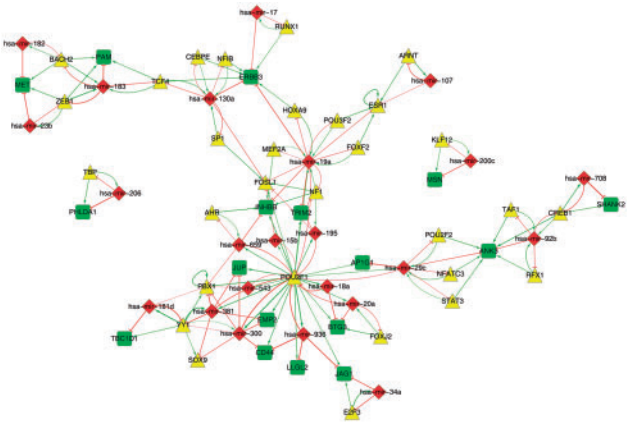

Fig. 4. The co-regulatory network of BaB. TFs are represented as triangles, miRNAs are represented as diamonds and BCCGs are represented as rectangles. Arrow with green edge indicates active relation, and 'T' with red edge indicates inhibitory relation

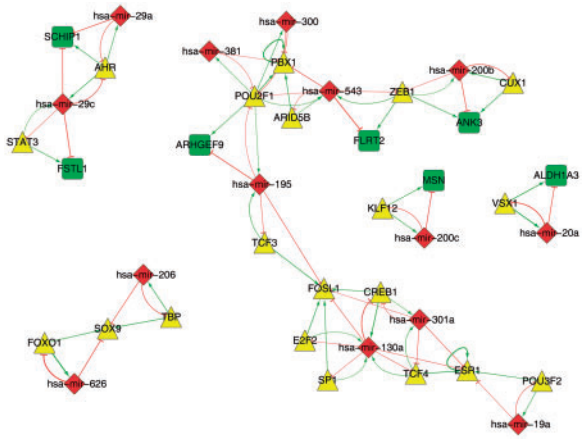

Fig. 5. The co-regulatory network of Lu. TFs are represented as triangles, miRNAs are represented as diamonds and BCCGs are represented as rectangles. Arrow with green edge indicates active relation, and 'T' with red edge indicates inhibitory relation

connected nodes with ≥10 connected genes are suggested to be hub genes in the co-regulatory networks. In the co-regulatory network of BaA, we identified four hub genes: POU2F1, hsa-mir-29c, hsa-mir-19a and hsa-mir-130a, which were connected to 34, 12, 12 and 11 genes, respectively (Fig. 3 and Supplementary Table S8). In comparison to the co-regulatory network of BaA, the co-regulatory network of BaB contains four identical hubgenes:POU2F1, hsa-mir-29c, hsa-mir-19a and hsa-mir-130a connected to 33, 17, 10 and 10 genes, respectively (Fig. 4 and Supplementary Table S8). Although POU2F1and hsa-mir-130a were also observed to be the only two hub genes in the co-regulatory network of Lu, the number of connected genes of POU2F1 is only 10, which is dramatically less than that in BaA (33) and BaB (34) (Fig. 5 and Supplementary Table S8). These results explain at least in part the less different biological and clinical behaviors between BaA and BaB than

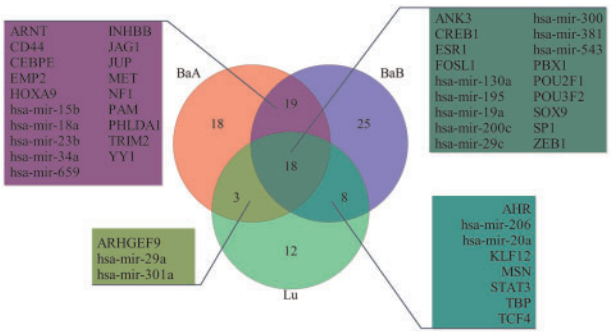

Fig. 6. Comparison of nodes in co-regulatory networks between BaA, BaB and Lu using Venn diagram

those between basal (BaA and BaB) and luminal (Lu) breast cancer.

Node degrees and betweenness centrality parameters were further calculated to evaluate the contribution of each node in three networks, respectively (Supplementary Table S8). Although the co-regulatory network of BaA and BaB shared the four hub genes, the connected genes of these hub genes were not the same. For example, 10 of 22 POU2F1 target genes were different between the co-regulatory networks of BaA and BaB (Figs 3 and 4, Supplementary Table S8). In addition, in BaA, POU2F1 were regulated by 12 genes including 4 genes only present in BaA and not in BaB, whereas in BaB, POU2F1 were regulated by 11 genes including 3 genes only present in BaB and not in BaA (Figs 3 and 4, Supplementary Table S8). These results together indicated that although BaA and BaB shared the common hub genes in their GRNs, they differ from each other via connecting to different genes to form subtype-specific GRNs, which are relevant to the different biological and clinical behaviors between BaA and BaB.

Beside the hub genes, we further explored the common gene nodes among breast cancer subtypes. Although 18 common nodes including POU2F1 and hsa-mir-130a are found among Lu, BaA and BaB, significant variations were found in the nodes among BaA, BaB and Lu (Fig. 6). There are 37 intersection nodes between two basal breast cancer subtypes (BaA and BaB), whereas only 21 intersection nodes were found between Lu and BaA and 26 intersection nodes were found between Lu and BaB (Fig. 6). The intersection nodes between basal subtypes are significantly higher than those found between basal subtypes and luminal subtype (Fig. 6). In addition, only 12 nodes were found in Lu GRN, only 18 nodes were found in BaA GRN and only 25 nodes were found in BaB GRN (Fig. 6 and Table 2).

When the edges in the GRNs of each breast cancer subtypes were compared, common and unique edges were identified (Fig. 7, Supplementary Figs S1 and S2 and Supplementary Table S10).Twenty-one common edges were found among BaA, BaB and Lu. There are 30 intersection edges between the two basal breast cancer subtypes (BaA and BaB), whereas only 10 intersection edges were found between Lu and BaA and 12 intersection edges were found between Lu and BaB (Fig. 7). BaA, BaB and Lu contain 79, 108, and 42 subtype-specific edges (Fig. 7, Supplementary Table S10). Compared with variation nodes of different breast cancer subtypes, the variation edges are much

Table 2. The special nodes in each breast cancer subtype

Type of BC	Special nodes
BaA	CAV2, COL4A2, DSC2, EGR3, EP300, FBN1, FOS, FOXO4, hsa-mir-133b, hsa-mir-494, hsa-mir-497, hsa-mir-656, MAP4K4, MEIS1, NEBL, PLAU, REST, SGCB
BaB	AP1G1, BACH2, BTG3, E2F3, ERBB3, FOXF2, FOXJ2, hsa-mir-107, hsa-mir-17, hsa-mir-181d, hsa-mir-182, hsa-mir-183, hsa-mir-708, hsa-mir-92 b, hsa-mir-936, LLGL2, MEF2A, NFATC3, NFIB, POU2F2, RFX1, RUNX1, SHANK2, TAF1,TBC1D1
Lu	ALDH1A3, ARID5B, CUX1, E2F2, FLRT2, FOXO1, FSTL1, hsa-mir-200 b, has-mir-626, SCHIP1, TCF3, VSX1

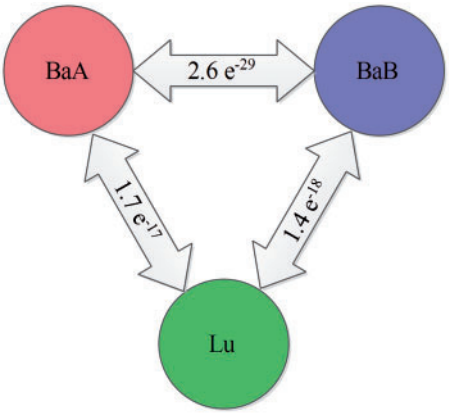

Fig. 8. The statistical test of intersection edges among three breast cancer subtypes

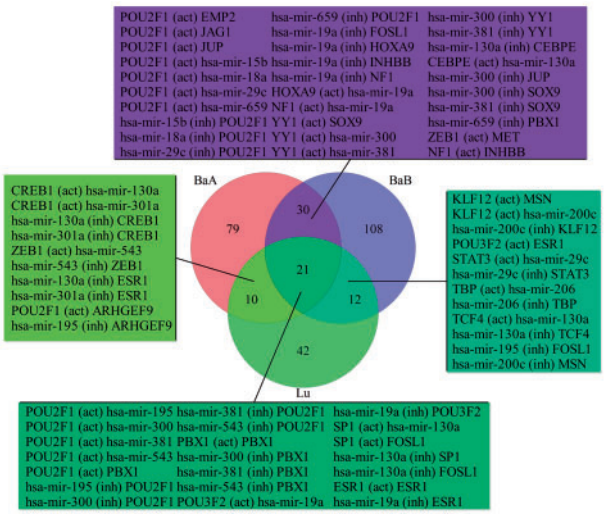

Fig. 7. Comparison of edges in co-regulatory networks between BaA, BaB and Lu using Venn diagram. 'act' stands for activation. 'inh' stands for inhibition

bigger (Figs 6 and 7, Table 2, Supplementary Table S10). These results suggested that both nodes and edges contribute to the different biological and clinical behaviors of breast cancer subtypes.

The intersection edges among GRNs are statistically significantly ($P<0.01$) (Fig. 8). This result indicated that the differences on nodes and edges among GRNs of three breast cancer subtypes reflect the molecular basis under breast tumorigenesis and breast cancer subtyping.

4 DISCUSSION

Numerous cellular processes and every process of life are affected by GRNs (Karlebach and Shamir, 2008). Precise and coordinated control of gene expression is required for an organism to grow, develop and function normally.(Cox and Goding, 1991). It is known for decades that a failure of normal regulation of expression of genes in cell growth and differentiation leads to the appearance and behavior of cancer cells (Cox and Goding, 1991). Abnormities in GRNs causing gene expression abnormities play important roles in human cancer including breast cancer. Breast cancer is one of the most commonly diagnosed cancer types in the world and is one of the most investigated areas of oncology (Demicheli and Coradini, 2011). Although there are >100 000 published papers on breast cancer in the past decade and GRNs were claimed to be a new conceptual framework to analyze breast cancer behaviors (Demicheli and Coradini, 2011), the study on the architecture and the feature of the GRNs in breast cancer is still limited.

GRNs can provide explanations of developmental and physiological functions at system level (Davidson, 2010). GRNs contain many layers of gene regulation, including cell signaling; mRNA splicing, polyadenylation and localization; chromatin modifications; and mechanisms of protein localization, modification and degradation (Chen and Rajewsky, 2007). TFs and miRNAs are two of the best-studied gene regulatory mechanism (Chen and Rajewsky, 2007). TFs are considered as key intrinsic regulators of cell fate via controlling gene expression directly (Demicheli and Coradini, 2011). Compared with TFs that regulate gene expression at the level of transcriptional regulation, miRNAs regulate gene expression at post-transcriptional level and raised entirely new mechanisms of gene regulation. More than 60% of human protein-coding genes were predicted to be regulated by miRNAs (Friedman *et al.*, 2009). Thus, investigation on GRNs by TFs and miRNAs will provide valuable information for understanding the gene regulation mechanism of cancer, including breast cancer.

In breast cancer, GRNs by TFs have been the subject of intense scientific investigation over the past several decades, and the related studies also lead to the improvement of breast cancer treatment. Take oestrogen receptor-α (ESR1), which is one of the most extensively studied TFs in breast cancer, as an example. ESR1 is found to be the driving TF in majority of breast cancer and to target genes dictating cell growth and endocrine responses

(Ross-Innes *et al.*, 2012). The studies on ESR1 and its GRNs lead to the development of hormone therapy, which is the mainly current systemic therapy (Bedard *et al.*, 2013). Recent technological advances improve the genome-wide map of GRN by ESR1 from model systems to breast cancer (Carroll *et al.*, 2006; Lin *et al.*, 2007; Ross-Innes *et al.*, 2012; Welboren *et al.*, 2009). Compared with TFs, miRNAs, which are short (20–23 nucleotides) regulatory RNAs that cause destabilization or translational repression of target mRNAs, are rising to be important new mechanisms of gene regulation in breast cancer (Le Quesne and Caldas, 2010). It has been mostly investigated on the role of individual miRNAs in breast cancer, and a new framework for studying the biology of miRNAs in breast cancer was presented recently via analyzing the miRNA expression profiles of 1302 breast tumors with matching detailed clinical annotation, long-term follow-up and genomic and miRNA expression data (Dvinge *et al.*, 2013).

Several studies have demonstrated the regulation between TFs and miRNAs in breast cancer. For examples, c-MYC directly transacted the *miR-17-92* cluster to modulate *E2f1* expression in breast cancer (O'Donnell *et al.*, 2005), and *miR-221/222* directly target *Esr1* mRNA to function in tamoxifen resistance in breast cancer (Zhao *et al.*, 2008). However, little is known on the pattern and features of GRNs by TFs and miRNA in breast cancer. Besides TargetScan, >10 integrative approaches have been developed for miRNA target prediction (Naifang *et al.*, 2013). In these approaches, approaches including MMIA (Nam *et al.*, 2009), mirConnX (Huang *et al.*, 2011) and MAGIA (Sales *et al.*, 2010) were correlation-based approaches, which directly consider the correlation between miRNA and mRNA expression. Besides correlation-based approaches, linear mode approaches, including GenmiR++ (Huang *et al.*, 2007) and Bayesian network approach (Liu *et al.*, 2009) predict miRNA target via formulating mRNA and miRNA expression with linear model with latent variables, and via using Bayesian network to model the miRNA–mRNA regulatory network, respectively. Regarding to the TFs and their targets, many approaches and databases, including UCSC database (Kent *et al.*, 2002), have been developed. However, none of these was able to be directly applied for analyzing GRN by miRNA and TF. To investigate the GRNs by the miRNA and the TF in breast cancer and among breast cancer subtypes, a pipeline shown in Figure 1 was designed to integrate the established approaches from miRNA target prediction to transcription targets analysis to decipher the architecture and features of GRNs consisted of miRNAs, TF and BCCGs in breast cancer and among breast cancer subtypes. There are two types of GRMs in breast cancer: Type I and Type II GRMs (Fig. 2). All the three breast cancer subtypes (BaA, BaB and Lu) have the two GRM types with significant variation number of nodes and edges (Fig. 2 and Supplementary Table S7). Basal subtype (BaA andBaB) has significantly more nodes and edges in co-regulatory motif group than luminal subtype (Lu) (Table 1, Fig. 2 and Supplementary Table S7). To study the combinatorial regulation pattern of GRNs by miRNAs and TFs in breast cancer, for each breast cancer subtypes, the GRNs by miRNAs and TFs were created via joining all 3-node co-regulatory motifs and merged three co-regulatory networks into the entire GRNs (Figs 3–5 and Supplementary Figs S1 and S2). The number of intersection nodes and edges between BaA and BaB is much higher than that between Lu and BaA and between Lu and BaB (Figs 6 and 7). Different breast cancer subtypes have their unique nodes and edges (Table 2 and Supplementary Tables S9 and S10). These features support the significant differences between breast cancer subtypes on many aspects from molecule to prognosis to some extent. Interestingly, a much smaller gene list containing only protein coding gene nodes from the GRNs is able to classify breast cancer subtypes as efficient as the gene list containing all BCCGs (Supplementary Figs S3 and S4). This result indicates that it might be potentially used analysis of the architecture and features of the GRNs in breast cancer and among breast cancer subtypes to build a more accurate and efficient breast subtype discriminator. Interestingly, most common nodes are TFs or miRNAs. This might be the consequence of restriction to BCCGs as TFs and miRNA targets in current study. Further studies would be required to confirm whether this phenomenon is a feature of breast cancer.

Identification of breast cancer driver genes will lay the ground for developing new diagnostics and individualizing cancer treatment. In the GRNs by TFs and miRNAs, 18 genes were common in GRNs across all breast cancer subtypes and designated non–subtype-specific genes (NSG) (Fig. 6). Nine of these 18 NSGs are TFs. In the nine TFs, eight (CREB1, ESR1, FOSL1, PBX1, POU2F1, SOX9, SP1 and ZEB1) were reported to play roles in breast cancer. CREB1 was reported to be a positive transcription regulator of aromatase (Sofi *et al.*, 2003), and its expression was correlated with the prognosis of breast cancer (Chhabra *et al.*, 2007). ESR1 roles in breast cancer were well established (Dumas and Diorio, 2011). FOSL1 was found to modulate growth, vacuolization and death of ant estrogen-resistant MCF-7 cells (Pennanen *et al.*, 2011). PBX1, Pre-B-cell leukemia homeobox 1, was reported to act as a TF to function as a pioneer factor defining aggressive Erα-positive breast cancer (Magnani *et al.*, 2011). POU2F1 was found to be a hub gene in GRNs across all breast cancer subtypes (Figs 3–5 and Supplementary Table S8). POU2F1, octamer TF 1 (OCT1), is reported to be associated with epithelial-mesenchymal transition (EMT) and breast tumor malignancy (Hwang-Verslues *et al.*, 2013). SOX9 was recently reported to cooperate with SLUG to determine the mammary stem cell state (Guo *et al.*, 2012). SP1 was reported to interact with ESR1 and play important roles in breast cancer (Ando and Catalano, 2012; Kim *et al.*, 2005). ZEB1 was reported to control proliferation and epithelial-to-mesenchymal transition of breast cancer cells (Hugo *et al.*, 2013). Besides the eight TFs in NSGs as discussed above, little is known about roles of ANK3 and POU3F2 in breast cancer, and *Ank3*and *Pou3f2* are considered to be new breast cancer gene candidates. In comparison, the eight miRNAs in NSGs are much less studied. Only two miRNAs in NSGs were reported to play roles in breast cancer: *has-mir-130a* was recently reported to modulate *Hoxa5*, which is a critical mediator of a breast cancer chemoprevention agent retinoic acid-induced cell growth inhibition (Yang *et al.*, 2013a), and *has-mir-19a* was found to target *Pten* to play a role in breast cancer chemoresistance (Liang *et al.*, 2011) and regulate the expression of tissue factor, which is an important regulator of tumor angiogenesis and metastasis in both luminal and basal breast cancer cells (Tsopanoglou and Maragoudakis, 2007; Zhang *et al.*, 2011).

The remaining six miRNAs (*has-mir-195*, *has-mir-200c*, *hsa-mir-543*, *hsa-mir-300*, *hsa-mir-381* and *hsa-mir-29c*) in NSGs are new miRNAs that could be play roles in breast cancer. Besides NSGs, the remaining genes were designated as breast cancer subtype-related genes (SRGs) (Fig. 3 and Table 2). Although SRGs cannot be simply interpreted as breast cancer subtype-specific genes whose functions are limited to specific breast cancer subtypes, SRGs would also play roles in breast cancer, and further studies are needed to validate whether the functions of SRGs in breast cancer are highly depended on the context of breast cancer subtypes. MiRNA regulatory networks were found to alter their configuration to conform to the stable TF regulatory networks with an increased circuit redundancy, and a marked reduction in the repertoire of composite feed-forward circuits targeted genes and this redundancy-adding role is preferentially attributable (Iwama *et al*., 2011). Although it is difficult to compile human breast cancers to other mammals including mice to study evolution of related GRNs, our GRNs by TFs and miRNAs in breast cancer and subtypes might be likely to retain similar properties. Further investigations are required to verify whether the redundancy-adding role would serve as a niche for many miRNA connections to survive avoiding conflicts with the stable TF regulatory networks in breast cancer.

More importantly, the GRNs raise important clues for further functional validation the GRNs by certain TFs and miRNAs and facilitate profound understanding of the functions of breast cancer genes. Several regulatory relationships have been experimentally validated. For example, a recent study showed that *microRNA-34* directly targets *Fosl1* to suppress breast cancer invasion and metastasis (Yang *et al*., 2013b). Thus, our results give important clue for the experimental validation of GRNs in breast cancer.

Our results reveal that the architecture and feature of GRNs by TFs and miRNAs in breast cancer have both distinct and common features among breast cancer subtypes. In addition, our results raise abundant clues for further experimental investigation of gene regulation by TFs and miRNAs in breast cancer via providing the new candidate breast cancer genes and regulatory relationship. Taken together, our results shed light on the mechanism of breast cancer development and subtyping.

ACKNOWLEDGEMENTS

The authors thank the members of the Li Fei laboratory for scientific discussions.

Funding: This work was supported by National Natural Science Foundation of China [81272260]; Natural Science Foundation of Jiangsu Province [SBK201241448]; National Basic Research Program of China (973 Program) [2015CB965000]; and Program for New Century Excellent Talents in University [NCET-12-0112].

Conflict of interest: none declared.

REFERENCES

Ando,S. and Catalano,S. (2012) The multifactorial role of leptin in driving the breast cancer microenvironment. *Nat. Rev. Endocrinol.*, **8**, 263–275.

Bedard,P.L. *et al.* (2013) Tumour heterogeneity in the clinic. *Nature*, **501**, 355–364.

Benjamini,Y. and Hochberg,Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.*, **10**, 289–300.

Brohée,S. *et al.* (2008) NeAT: a toolbox for the analysis of biological networks, clusters, classes and pathways. *Nucleic Acids Res.*, **36**, W444–W451.

Calin,G.A. and Croce,C.M. (2006) MicroRNA signatures in human cancers. *Nat. Rev. Cancer*, **6**, 857–866.

Cancer Genome Atlas Network. (2012) Comprehensive molecular portraits of human breast tumours. *Nature*, **490**, 61–70.

Carroll,J.S. *et al.* (2006) Genome-wide analysis of estrogen receptor binding sites. *Nat. Genet.*, **38**, 1289–1297.

Chen,K. and Rajewsky,N. (2007) The evolution of gene regulation by transcription factors and microRNAs. *Nat. Rev. Genet.*, **8**, 93–103.

Chhabra,A. *et al.* (2007) Expression of transcription factor CREB1 in human breast cancer and its correlation with prognosis. *Oncol. Rep.*, **18**, 953–958.

Chien,C.H. *et al.* (2011) Identifying transcriptional start sites of human microRNAs based on high-throughput sequencing data. *Nucleic Acids Res.*, **39**, 9345–9356.

Cox,P.M. and Goding,C.R. (1991) Transcription and cancer. *Br. J. Cancer*, **63**, 651–662.

Darnell,J.E. Jr (2002) Transcription factors as targets for cancer therapy. *Nat. Rev. Cancer*, **2**, 740–749.

Davidson,E.H. (2010) Emerging properties of animal gene regulatory networks. *Nature*, **468**, 911–920.

Demicheli,R. and Coradini,D. (2011) Gene regulatory networks: a new conceptual framework to analyse breast cancer behaviour. *Ann. Oncol.*, **22**, 1259–1265.

Dumas,I. and Diorio,C. (2011) Estrogen pathway polymorphisms and mammographic density. *Anticancer Res.*, **31**, 4369–4386.

Dvinge,H. *et al.* (2013) The shaping and functional consequences of the microRNA landscape in breast cancer. *Nature*, **497**, 378–382.

Ellis,M.J. *et al.* (2012) Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature*, **486**, 353–360.

Friedman,R.C. *et al.* (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, **19**, 92–105.

Garzon,R. *et al.* (2009) MicroRNAs in cancer. *Annu. Rev. Med.*, **60**, 167–179.

Guo,W. *et al.* (2012) Slug and Sox9 cooperatively determine the mammary stem cell state. *Cell*, **148**, 1015–1028.

Huang,D.W. *et al.* (2008) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, **4**, 44–57.

Huang,G.T. *et al.* (2011) mirConnX: condition-specific mRNA-microRNA network integrator. *Nucleic Acids Res.*, **39**, W416–W423.

Huang,J.C. *et al.* (2007) Using expression profiling data to identify human microRNA targets. *Nat. Methods*, **4**, 1045–1049.

Hugo,H.J. *et al.* (2013) Direct repression of MYB by ZEB1 suppresses proliferation and epithelial gene expression during epithelial-to-mesenchymal transition of breast cancer cells. *Breast Cancer Res.*, **15**, R113.

Hwang-Verslues,W.W. *et al.* (2013) Loss of corepressor PER2 under hypoxia up-regulates OCT1-mediated EMT gene expression and enhances tumor malignancy. *Proc. Natl Acad. Sci. USA*, **110**, 12331–12336.

Iwama,H. *et al.* (2011) MicroRNA networks alter to conform to transcription factor networks adding redundancy and reducing the repertoire of target genes for coordinated regulation. *Mol. Biol. Evol.*, **28**, 639–646.

Karlebach,G. and Shamir,R. (2008) Modelling and analysis of gene regulatory networks. *Nat. Rev. Mol. Cell Biol.*, **9**, 770–780.

Kent,W.J. *et al.* (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.

Kim,K. *et al.* (2005) Analysis of estrogen receptor alpha-Sp1 interactions in breast cancer cells by fluorescence resonance energy transfer. *Mol. Endocrinol.*, **19**, 843–854.

Le Quesne,J. and Caldas,C. (2010) Micro-RNAs and breast cancer. *Mol. Oncol.*, **4**, 230–241.

Liang,Z. *et al.* (2011) Regulation of miR-19 to breast cancer chemoresistance through targeting PTEN. *Pharm. Res.*, **28**, 3091–3100.

Lin,C.Y. *et al.* (2007) Whole-genome cartography of estrogen receptor alpha binding sites. *PLoS Genet.*, **3**, e87.

Liu,B. *et al.* (2009) Exploring complex miRNA-mRNA interactions with Bayesian networks by splitting-averaging strategy. *BMC Bioinformatics*, **10**, 408.

Liu,H. (2012) MicroRNAs in breast cancer initiation and progression. *Cell Mol. Life Sci.*, **69**, 3587–3599.

Magnani,L. *et al.* (2011) PBX1 genomic pioneer function drives ERalpha signaling underlying progression in breast cancer. *PLoS Genet.*, **7**, e1002368.

Naifang,S. *et al.* (2013) Integrative approaches for microRNA target prediction: combining sequence information and the paired mRNA and miRNA expression profiles. *Curr. Bioinform.*, **8**, 37–45.
Nam,S. *et al.* (2009) MicroRNA and mRNA integrated analysis (MMIA): a web tool for examining biological functions of microRNA expression. *Nucleic Acids Res.*, **37**, W356–W362.
Neve,R.M. *et al.* (2006) A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell*, **10**, 515–527.
O'Donnell,K.A. *et al.* (2005) c-Myc-regulated microRNAs modulate E2F1 expression. *Nature*, **435**, 839–843.
Ozsolak,F. *et al.* (2008) Chromatin structure analyses identify miRNA promoters. *Genes Dev.*, **22**, 3172–3183.
Pencheva,N. and Tavazoie,S.F. (2013) Control of metastatic progression by microRNA regulatory networks. *Nat. Cell. Biol.*, **15**, 546–554.
Pennanen,P.T. *et al.* (2011) Inhibition of FOSL1 overexpression in antiestrogen-resistant MCF-7 cells decreases cell growth and increases vacuolization and cell death. *Steroids*, **76**, 1063–1068.
Poos,K. *et al.* (2013) How microRNA and transcription factor co-regulatory networks affect osteosarcoma cell proliferation. *PLoS Comput. Biol.*, **9**, e1003210.
Riaz,M. *et al.* (2013) miRNA expression profiling of 51 human breast cancer cell lines reveals subtype and driver mutation-specific miRNAs. *Breast Cancer Res.*, **15**, R33.
Ross-Innes,C.S. *et al.* (2012) Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature*, **481**, 389–393.
Sales,G. *et al.* (2010) MAGIA, a web-based tool for miRNA and genes integrated analysis. *Nucleic Acids Res.*, **38**, W352–W359.
Shah,S.P. *et al.* (2012) The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature*, **486**, 395–399.
Smoot,M.E. *et al.* (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, **27**, 431–432.
Sofi,M. *et al.* (2003) Role of CRE-binding protein (CREB) in aromatase expression in breast adipose. *Breast Cancer Res. Treat.*, **79**, 399–407.
Tibshirani,R. *et al.* (2002) Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc. Natl Acad. Sci. USA*, **99**, 6567–6572.
Tsopanoglou,N.E. and Maragoudakis,M.E. (2007) Inhibition of angiogenesis by small-molecule antagonists of protease-activated receptor-1. *Semin. Thromb. Hemost.*, **33**, 680–687.
Welboren,W.J. *et al.* (2009) ChIP-Seq of ERalpha and RNA polymerase II defines genes differentially responding to ligands. *EMBO J.*, **28**, 1418–1428.
Wernicke,S. and Rasche,F. (2006) FANMOD: a tool for fast network motif detection. *Bioinformatics*, **22**, 1152–1153.
Wilkinson,L. (2012) Exact and approximate area-proportional circular Venn and Euler diagrams. *IEEE Trans. Vis. Comput. Graph.*, **18**, 321–331.
Yang,F. *et al.* (2013a) Retinoic acid-induced HOXA5 expression is co-regulated by HuR and miR-130a. *Cell Signal.*, **25**, 1476–1485.
Yang,S. *et al.* (2013b) MicroRNA-34 suppresses breast cancer invasion and metastasis by directly targeting Fra-1. *Oncogene*, **32**, 4294–4303.
Yuan,Y. *et al.* (2012) Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling. *Sci. Transl. Med.*, **4**, 157ra143.
Zhang,X. *et al.* (2011) MicroRNA-19 (miR-19) regulates tissue factor expression in breast cancer cells. *J. Biol. Chem.*, **286**, 1429–1435.
Zhao,J.J. *et al.* (2008) MicroRNA-221/222 negatively regulates estrogen receptor alpha and is associated with tamoxifen resistance in breast cancer. *J. Biol. Chem.*, **283**, 31079–31086.