Molecular BioSystems

RSC Publishing

PAPER

View Article Online
View Journal | View Issue

Cite this: *Mol. BioSyst.*, 2013, **9**, 3187

Synergetic regulatory networks mediated by oncogene-driven microRNAs and transcription factors in serous ovarian cancer†

Min Zhao,[a] Jingchun Sun[a] and Zhongming Zhao*[abcd]

Received 2nd May 2013,
Accepted 19th September 2013

DOI: 10.1039/c3mb70172g

www.rsc.org/molecularbiosystems

Although high-grade serous ovarian cancer (OVC) is the most lethal gynecologic malignancy in women, little is known about the regulatory mechanisms in the cellular processes that lead to this cancer. Recently, accumulated lines of evidence have shown that the interplay between transcription factors (TFs) and microRNAs (miRNAs) is critical in cellular regulation during tumorigenesis. A comprehensive investigation of TFs and miRNAs, and their target genes, may provide a deeper understanding of the regulatory mechanisms in the pathology of OVC. In this study, we have integrated three complementary algorithms into a framework, aiming to infer the regulation by miRNAs and TFs in conjunction with gene expression profiles. We demonstrated the utility of our framework by inferring 67 OVC-specific regulatory feed-forward loops (FFL) initiated by miRNAs or TFs in high-grade serous OVC. By analyzing these regulatory behaviors, we found that all the 67 FFLs are consistent in their regulatory effects on genes that are jointly targeted by miRNAs and TFs. Remarkably, we unveiled an unbalanced distribution of FFLs with different oncogenic effects. In total, 31 of the 67 coherent FFLs were mainly initiated by oncogenes. On the contrary, only 4 of the FFLs were initiated by tumor suppressor genes. These overwhelmingly observed oncogenic genes were further detected in a sub-network with 32 FFLs centered by miRNA let-7b and TF TCF7L1 to regulate cell differentiation. Closer inspection of 32 FFLs revealed that 75% of the miRNAs reportedly play functional roles in cell differentiation, especially when enriched in epithelial–mesenchymal transitions. This study provides a comprehensive pathophysiological overview of recurring coherent circuits in OVC that are co-regulated by miRNAs and TFs. The prevalence of oncogenic coherent FFLs in serous OVC suggests that oncogene-driven regulatory motifs could cooperatively act upon critical cellular processes such as cell differentiation in a highly efficient and consistent manner.

Introduction

Ovarian cancer (OVC) refers to heterogeneous cancers arising from the ovary. There are estimated to be 22 280 new cases and 15 500 deaths in the United States in 2012.[1] OVC is regarded as a ''silent killer'' due to its high mortality and low cure rates.[2] These facts are largely due to the absence of symptoms in this cancer's early stages. Patients are difficult to diagnose until the disease is in an advanced stage and has spread beyond the ovary. Most OVCs originate from ovarian surface epithelia, which can be classified into four major types in histology: serous (70%), endometrioid (10–15%), clear-cell (10%), and mucinous (3%) carcinomas.[3] According to the degree of differentiation, OVCs are grouped into well-differentiated low-grade and poorly differentiated high-grade. Additionally, it is known that serous OVCs account for 90% of high-grade tumors.[4] Despite that numerous genetic and pathogenic studies have been reported for OVC, the molecular mechanisms underlying this cancer, especially high-grade serous OVC, are largely unknown.

Like other types of tumors, OVC is characterized by uncontrolled cell growth, which is caused by the deregulated gene expression of tumor suppressors and oncogenes in controlling cell proliferation and apoptosis.[5,6] In these deregulated gene expression processes, two major groups of regulators affect cancer gene expression at the transcriptional and post-transcriptional levels. The first group is transcription factors (TFs), which operate through the transcription

[a] *Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN, USA. E-mail: zhongming.zhao@vanderbilt.edu*
[b] *Department of Cancer Biology, Vanderbilt University School of Medicine, Nashville, TN, USA*
[c] *Department of Psychiatry, Vanderbilt University School of Medicine, Nashville, TN, USA*
[d] *Center for Quantitative Sciences, Vanderbilt University Medical Center, Nashville, TN, USA*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c3mb70172g

activation or suppression of target genes with specific binding sites in regulatory regions.[7] The second group is microRNAs (miRNAs), which mediate degradation or translational repression of target genes by binding target genes with small complementary sequences.[8] In addition, these two types of regulatory mechanisms have reciprocal regulation and joint effects on their shared target genes, which form complex regulatory motifs such as feed-forward loops (FFLs) to influence gene expressions in cancer.[9–11] Recently, numerous individual identifications of transcriptional dysregulation of TFs and miRNAs in OVC have provided further implication of TFs and miRNAs in the etiology of OVC.[12,13] Though our previous TF–miRNA FFL study in GBM[10] and other studies of TF–miRNA FFLs in other types of cancers[11,14,15] highlight the interplay of miRNAs and TFs and their involvement in cancer development, the structure and function of the TF–miRNA regulatory FFLs based on genome-wide expression profiles in OVC have not been explored. Recent genome-wide studies performed by the Cancer Genome Atlas (TCGA) project provided vast quantities of gene expression profiling and other molecular profiling from hundreds of OVC samples, which might provide a promising opportunity to uncover the basic building blocks of regulatory networks in OVC.[16]

Usually, genome-wide data or *de novo* prediction can produce thousands of regulatory relationships. However, how does one compile a compendium of key regulators from mass data? To achieve this goal, most previous studies choose to focus on a specific biological question or use motif-finding algorithms to pinpoint the potential interesting regulators. To identify a relatively small subset of regulators to coordinate multiple cellular processes at an appropriate scale to generate reliable hypotheses, we combined the two best performing and complementary algorithms into a framework to infer regulation by miRNAs and TFs from collected co-expressed and reported OVC candidate genes. We applied this approach to identifying 67 significant co-regulating FFLs consisting of miRNAs, TFs, and OVC genes for in-depth analysis. Based on statistical tests, we discovered that all 67 FFL are coherent type, which means the regulators (miRNAs and TFs) have consistent regulatory effects on their joint target genes. In addition, these coherent FFLs are overwhelmingly initiated by oncogenic genes. Based on all the coherent significant FFLs with 25 TFs, 27 miRNAs and 23 target genes, we constructed a central regulatory network in OVC, which contained enriched genes involved in cell differentiation, cell cycle regulators, and OVC-specific signaling pathways. We have observed the prevalence of a coherent regulatory pattern of oncogenic miRNA and TF to their targets. This result may spotlight the importance of recurring coherent circuit elements, which have a tendency to initiate through oncogenic genes and promote oncogenic processes in a highly efficient and consistent style.

Methods

Expression profiles for TFs, miRNAs, and OVC-related genes

As shown in Fig. 1, our workflow started from TFs, miRNAs, and OVC-related genes using the gene expression data from TCGA.[16] The gene expression data were generated from three microarray gene expression platforms (Affymetrix Exon 1.0 array, Agilent 244K whole genome expression array, and Affymetrix HT-HG-U133A array) for the 489 high-grade serous OVC samples.[16] To provide unified expression measures, the TCGA investigators first normalized and calculated the expression values on each platform separately. Then, to obtain relative gene expression scores, they subtracted the mean value across samples for the same gene followed by dividing the expression value using standard deviation across samples. Finally, a factor analysis model was used to integrate the relative expression data from three platforms into a single data set of 11 864 mRNAs without batch effects.[16,17] The final unified expression data was downloaded from the TCGA website as a matrix, which was formatted as one row for each gene and one column for each sample (https://tcga-data.nci.nih.gov/docs/publications/ov_2011/).

To obtain the protein-coding transcriptional regulator information, we obtained 524 human TF genes from the TRANSFAC Professional database (release 2011.4) (Table S1, ESI†).[18] By overlapping the TF genes with the 11 864 genes with expression profile, we obtained 441 human TFs for follow-up analyses.

To focus on OVC-related genes, we utilized our previous OVC-specific gene set containing 1257 non-redundant OVC-related genes collected from fourteen data sources (Table S1, ESI†).[6] Among the 1257 OVC-related genes, 957 were found in the 11 864 genes with expression profiles from TCGA.

To correlate miRNA gene expression with TF and OVC-related gene expression, we selected 267 Grade 3 OVC matched samples with miRNA and mRNA gene expression data. In these expression profiles, there were 11 864 coding genes and 799 miRNAs in total (Table S1, ESI†). For the miRNA set, we excluded 241 miRNAs from human virus or minor forms (miR*) and compiled 558 unique mature miRNAs from the TCGA Level 3 archive that contains normalized and processed miRNA expression data using Agilent miRNA microarray (8 × 15K human miRNA-specific microarrays).[16]

Expression correlation among miRNAs, TFs, and OVC-related genes

Based on the expression profiles of OVC-related genes, human TFs, and miRNAs, we first computed expression correlation scores for three types of pairs: TF ↔ miRNA, miRNA ⊣ OVC-related gene, and TF → OVC-related gene. Here, symbols ⊣ and → represent miRNA initiated repression and TF initiated regulation, respectively. The symbol ↔ represents the mutual regulation between TF and miRNA. We estimated the expression correlation among these regulatory pairs using the Spearman's correlation method that is implemented in the R language package (version 2.14.0) to calculate their expression correlation scores and corresponding *p*-values.[19] For all the *p*-values in each type of pair, a false discovery rate (FDR) was applied to correct the statistical significance of multiple testing. For all the pairs from both miRNA ⊣ OVC-related gene and miRNA ⊣ TF, we required their expression correlation scores be less than −0.2 and the FDR adjusted *p*-values be less than 0.01. For all the possible pairs of TF → miRNA and TF → OVC-related gene, we extracted the pairs with absolute expression correlation scores greater than 0.2 and corresponding FDR adjusted *p*-values less than 0.01. According to these criteria,

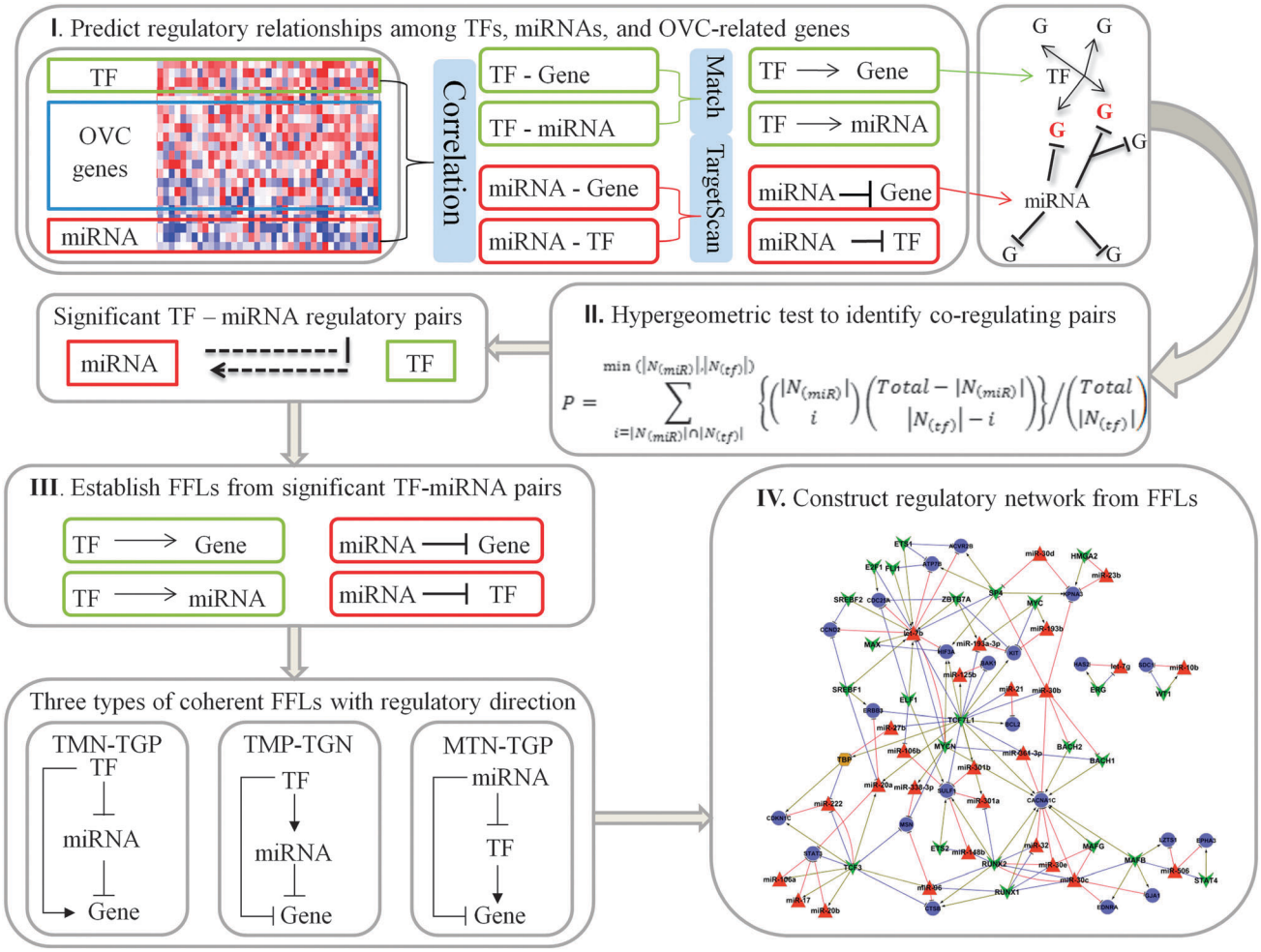

Fig. 1 Schematic of the discovery of microRNAs (miRNAs) and transcription factors (TFs) mediated feed-forward loops (FFLs) in serous ovarian cancer (OVC). The pipeline involves four main steps. (1) Identification of regulatory relationships in OVC. Based on the compiled 957 OVC-related genes, 799 miRNAs, and 264 known human TFs with expression profiles from the 267 tumor samples from The Cancer Genome Atlas (TCGA), we predicted four types of regulations (TF → gene, TF → miRNA, miRNA ⊣ gene, and miRNA ⊣ TF) by integrating miRNA-target data, TF-target data, and gene expression profiles (see Methods). (2) Detection of significant regulatory pairs between miRNAs and TFs using a hypergeometric test. (3) Identification of significant coherent FFLs by integrating regulatory directions with significant TF–miRNA pairs and their joint target genes. (4) Construction of the OVC-specific regulatory network based on significant FFLs. Other abbreviations in the figure: TMN-TGP: a coherent FFL initiated by a TF that negatively regulates a miRNA and positively regulates their joint target gene. TMP-TGN: a coherent FFL initiated by a TF that positively regulates a miRNA and negatively regulates their joint target gene. MTN-TGP: a coherent FFL initiated by a miRNA that negatively regulates a TF and their joint target genes, while the regulated TF positively regulates their joint target gene.

we have an appropriate number of pairs in each type of regulatory relationship for follow up analysis (data not shown).

Regulatory relationship of miRNA to gene/TF (miRNA ⊣ gene/TF)

The posttranscriptional repression of miRNAs to protein-coding genes can be typically inferred from information such as evolutionary conservation of seed sites on sequence.[20,21] Among those available prediction tools based on binding sites, TargetScan is the most popular and has the highest accuracy.[20,21] Therefore, we downloaded data from the TargetScan server (version 6.0, November 2011) to prepare the regulatory pairs between miRNAs and OVC-related genes/TFs.[22] Among the 558 human mature miRNAs with expression profiles we compiled above, 253 miRNAs have evolutionarily conserved targets in four organisms (human, mouse, rat, and dog). In total, 8162 target genes of the 253 miRNAs were collected after filtering based on their context score (higher than −0.30), which is a quantitative measure for the overall target binding efficacy.[20,21] After overlapping the predicted miRNA target relationships with the pairs of miRNAs and OVC-related genes with significantly negative correlation of expression profiles, we obtained 311 pairs with 103 miRNAs and 131 OVC-related genes in total (Table 1). Similarly, we obtained 222 pairs with 97 miRNAs and 80 human TFs after intersecting the results from TargetScan and TF–miRNA pairs with the significant correlation based on expression profiles (Table 1).

Regulatory relationship between TF and gene/miRNA (TF → gene/miRNA)

In our pipeline, we required any regulatory relationship between TF and gene/miRNA to have evidence from both computational

Table 1 Summary of relationships among OVC TFs, miRNAs, and OVC-related genes

Relationship	Number of the elements: Pairs	miRNAs	OVC genes	TFs	Method (+EXP Cor[e])
miRNA ⊣ gene[a]	311	103	131	—	TargetScan
miRNA ⊣ TF[b]	222	97	—	80	TargetScan
TF → gene[c]	2369	—	341	207	Match™
TF → miRNA[d]	1988	342	—	235	Match™

[a] miRNA ⊣ gene: miRNA repression of OVC-related gene expression. [b] miRNA ⊣ TF: miRNA repression of TF expression. [c] TF → gene: TF regulation of OVC-related gene expression. [d] TF → miRNA: TF regulation of miRNA expression. [e] EXP Cor: all the methods in the table were overlapped to expression correlation scores. For all the predicted pairs from both miRNA ⊣ gene and miRNA ⊣ TF, their expression correlation scores are less than −0.2 and FDR adjusted *p*-values < 0.01. For all the detected pairs from both TF → gene and TF → miRNA, the absolute values of their expression correlation scores are higher than 0.2 and FDR adjusted *p*-values < 0.01.

prediction and experimental gene expression. To predict reliable regulatory relationships, we only isolated the conserved TF → OVC-related gene regulations across human, mouse, and rat.[23] For all the 1257 OVC-related genes, we first extracted 1050 mouse homologous genes and 1019 rat homologous genes from the NCBI HomoloGene database (build 65, February 15, 2011).[24] Next, we inputted 1257 human OVC-related genes and their 1050 mouse and 1019 rat homologous genes to the UCSC Genome Browser[25] to obtain their sequence, separately. Then, we extracted promoter regions from −1500 to +500 around transcription start sites (TSS). Next, we performed a search of TF binding sites on these sequences using the TRANSFAC Match™ software (release 2011.4).[26] We searched the mouse and rat genes that were homologous to the 524 human TF genes using the HomoloGene database (build 65, February 15, 2011)[24] and then predicted the human, mouse, and rat TF-gene regulations separately. As a result, the 524 human TFs were linked to 2869 mouse and 2858 rat TRANSFAC sequence matrix IDs, which were used to predict TF targets. In our prediction of the targets of human, mouse and rat TFs, we applied stringent cut-off values to minimize false positive matches: a core similarity of 1.00 and a matrix similarity of 0.95. These are the two most important scores to evaluate the quality of a match based on its corresponding TF matrix from TRANSFAC. Finally, 16 953 conserved human TF–target gene pairs were identified using Match™ (conserved among human, mouse, and rat). These pairs contained 291 human TFs and 397 OVC related genes. Following our pipeline, the 16 953 TF → OVC-related gene pairs were intersected with TF-gene expression correlation scores from TCGA. Here, the correlation of expression profiles between TF and OVC-related genes could be negative or positive. An absolute value of the correlation score higher than 0.2 was used as a cut-off. This cut-off resulted in a total of 2369 TF → OVC-related gene pairs, which included 207 TFs and 341 OVC-related genes (Table 1).

To predict the TF → miRNA relationship, we first obtained genomic locations of the 558 mature miRNAs by searching miRBase (release 17).[27] The resulting genomic locations were used to extract their respective promoter regions around TSSs (−1500 to +500). Next, Match™ was applied to identify possible TF → miRNA regulatory pairs. In this procedure, we obtained 44 625 TF → miRNA pairs between 405 TFs and 542 miRNAs that satisfied the core score of 1.00 and matrix score over 0.95. After checking expression correlation scores between TF and miRNA pairs with an absolute value greater than 0.2, we obtained 1988 TF → miRNA regulatory pairs. Those pairs included 235 TFs and 342 miRNAs (Table 1).

Significant TF–miRNA co-regulation pairs

To improve the reliability of the predicted regulatory pairs, we adopted a cumulative hypergeometric test to sort out significant TF–miRNA co-regulation pairs with the same joint target genes. Using the function below, we calculated a *p*-value for each TF–miRNA pair in our data:[10,28]

$$P = \sum_{i=|N_{(\mathrm{miR})}|\cap|N_{(\mathrm{tf})}|}^{\min(|N_{(\mathrm{miR})}|,|N_{(\mathrm{tf})}|)} \times \left\{ \binom{|N_{(\mathrm{miR})}|}{i} \binom{\mathrm{Total} - |N_{(\mathrm{miR})}|}{|N_{(\mathrm{tf})}| - i} \Big/ \binom{\mathrm{Total}}{|N_{(\mathrm{tf})}|} \right\}$$

In this formula, $N_{(\mathrm{miR})}$ is the count of genes regulated by a miRNA, $N_{(\mathrm{tf})}$ is the number of genes regulated by a TF, and Total is the number of jointly regulated genes between all human genes regulated by all human miRNAs and all human genes regulated by all human TFs. The FDR was applied to adjust the false positive rate for multiple testing.[29] Finally, we selected the 335 significant TF–miRNA co-regulation pairs if their FDR corrected *p*-values were less than 0.01.

Coherent FFLs, network construction, and statistical functional evaluation

To reduce the complexity of regulatory networks and distill critical regulatory elements, we added the direction of the edges (regulation) using the expression correlation scores (*i.e.*, positive or negative co-expression). Using the significant TF–miRNA co-regulation pairs obtained above, we constructed a comprehensive TF–miRNA regulatory network, which was comprised of 67 FFLs with regulatory directions. This network included 27 miRNAs, 25 TFs, and 23 OVC-related genes (Table S2, ESI†).

We specifically examined the features of FFLs and OVC regulatory networks for those TFs and target genes that are categorized as tumor suppressors and oncogenes. We downloaded 196 protein-coding tumor suppressor genes with multiple sources of literature evidence from the tumor suppressor gene database TSGene (http://bioinfo.mc.vanderbilt.edu/TSGene/).[30] Additionally, we collected a list of 296 protein-coding oncogenes, each of which had evidence from both the UniProtKB keyword "Proto-oncogene"[31] and the Tumor Associated Gene (TAG) database (http://www.binfo.ncku.edu.tw/TAG/).

To detect biological pathways overrepresented in our constructed regulatory network, we performed a pathway enrichment analysis using the Ingenuity Pathway Analyses (IPA) tool from Ingenuity Systems.[32] Given a list of genes, Fisher's exact tests were conducted to detect the enrichment of these genes in

Ingenuity's manually curated canonical pathways. To control the error rate in the analysis results, we selected the significant pathways with a corrected *p*-value less than 0.05 (IPA provided this cut-off based on the Benjamini–Hochberg method[29]). We conducted functional enrichment tests using the online tool DAVID[33] to further assess the enrichment of interesting genes with Gene Ontology (GO) annotation and other functional terms enrichment. Only those functional terms were selected if their adjusted *p*-values were less than 0.0001 (hypergeometric test followed by the Benjamini–Hochberg correction).[29]

Randomization analyses

To evaluate the robustness of the final regulatory network in OVC, we applied an empirical re-sampling approach. Take the TF gene list as an example here. First, for 441 TFs mapped to the OVC gene expression data, we randomly selected 25 TFs. Similarly, we randomly selected 27 miRNAs from the 799 miRNAs and 23 OVC-related genes from the 957 genes with expression profile across the 267 high-grade serous OVC samples. We obtained all the regulatory pairs if their absolute expression correlation coefficients were less than 0.2 (this cutoff value was −0.2 for the regulatory pairs initiated by miRNAs) and the FDR adjusted *p*-values were less than 0.01. Based on the regulatory pairs TF → gene, TF → miRNA, miRNA ⊣ TF, and miRNA ⊣ gene, we examined how many FFLs could be formed. We repeated this randomization process 10 000 times. Next, we counted the number of randomly generated datasets (N) whose number of FFLs was greater than the observed number of FFLs (67 in our study). Lastly, we calculated an empirical *p*-value by N/10 000 for the observed number of FFLs being greater than 67. We obtained an empirical *p*-value 0.0017, indicating that the observed network with 67 FFLs was unlikely generated by chance. For the 17 datasets with more than 67 FFLs, we checked their number of coherent FFLs and the number of coherent FFLs initiated by oncogenes and tumor suppressors (TSGs). In our final regulatory network, there were 31 oncogene-initiated FFLs (46.3%) and 4 TSG-initiated FFLs (5.97%). As shown in Table S5 (ESI†), none of the 17 randomly selected datasets had a larger proportion of oncogene-initiated FFLs than the observed one.

Results

A computational pipeline to construct a TF–miRNA FFL regulatory network

To explore the regulatory relationships among miRNAs, TFs, and OVC-related genes, we applied a novel computational framework to our collected OVC-related genes by incorporating transcriptome data from TCGA. The pipeline has four critical steps to construct a final regulatory network (Fig. 1). First, we predicted four types of regulatory pairs, including miRNA ⊣ TF, miRNA ⊣ OVC-related gene, TF → miRNA and TF → OVC-related gene, which were combined using sequence-based regulatory predictions and gene expression data. In this step, the evolutionarily conserved miRNA targeting TFs and OVC-related genes were predicted using TargetScan;[22] meanwhile the evolutionarily conserved regulatory pairs initiated by TFs were predicted using the TRANSFAC Match™ software.[26] To reduce false positive possibilities and determine the positive or negative regulatory effects of miRNAs and TFs on their target genes, we further overlapped our predicted regulatory pairs to miRNA and mRNA expression from 267 Grade 3 OVC samples from TCGA. In the second step, significant TF–miRNA pairs were identified based on their shared target genes using a hypergeometric test. Thirdly, regulatory feed-forward loops (FFLs) between miRNAs, TFs and their joint target genes were formed based on the overlapping of these three elements to the significant co-occurring TF–miRNA pairs. In each loop, the three regulatory links with positive or negative regulatory effect were established to determine the joint regulatory direction. Finally, this extensive exploration of regulatory motifs initiated by TFs and miRNAs with regulatory flows resulted in a total of 67 FFLs with 25 TFs, 27 miRNAs, and 23 OVC-related genes (Fig. 2).

Complete coherent feed-forward regulatory loops

Previous studies showed that one of most abundant and significant regulatory network motifs in *E. coli* and Yeast is the FFL, which consists of two regulators and one joint target.[34] According to regulators' regulatory effects on their joint targets, the FFLs with two regulators can be divided into two main types. One is coherent FFLs, in which the two regulators have consistent effects on their joint target. The other is incoherent FFLs, which have inconsistent effects from two regulators to their joint target.[35] As shown in Fig. 2, all the 67 FFLs identified from OVC are coherent FFLs. More specifically, there are three consistent regulatory patterns. The first is TMN-TGP (TF–miRNA negative and TF–gene positive), which is initiated by a TF with a negatively regulating effect on a miRNA and positive regulatory effect on its joint target genes. As the miRNA represses the joint target, the negative control of the initial TF to the miRNA leads to an expression increase of their joint target genes. The second is TMP-TGN (TF–miRNA positive and TF–gene negative). In this type of FFL, the initial regulator is a TF to activate a miRNA and repress its joint target genes. Combined with the negative effect of the miRNA to the target gene, the activation of the initial TF to a miRNA will strength the reduction of their joint target. The third type of FFL is MTN-TGP (miRNA–TF negative and TF–gene positive), which is initiated by a miRNA with a negative effect on a TF while, at the same time, the repressed TF has a positive effect on its joint target gene. Thus, in this FFL, the joint target gene will be depressed by both the miRNA and the decreased TF.

In each FFL, there is one ''fast'' regulatory arm with direct regulation and a ''slow'' regulatory arm with two consecutive regulatory steps. The integration of the signals from both the fast and slow regulatory arms often emerges as a signal delay in each FFL. According to the theoretical analysis from ref. 34, all the coherent FFLs are sign-sensitive delays that contrast to the sign-sensitive acceleration of incoherent FFLs. The MTN-TGP and TMN-TGP show their delays on the initial sign-on. However, TMP-TGN will have a delay on the initial sign-off. These sign-sensitive delays can help the FFLs reject short initial pulses and assist them in responding only to persistent stimuli. Combined with their consistent effects, the three coherent FFLs

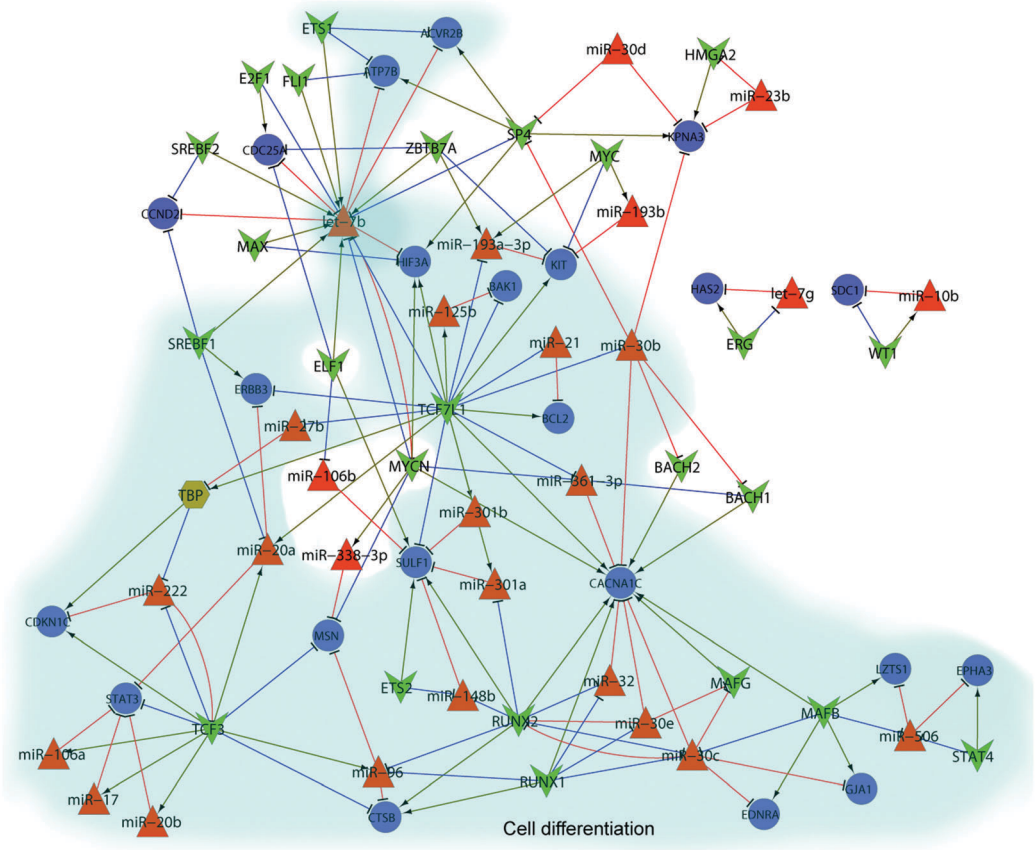

Fig. 2 MicroRNAs (miRNAs) and transcription factors (TFs) mediated feed-forward loops (FFLs) and regulatory networks in serous ovarian cancer (OVC). In the network, red nodes denote miRNAs, green nodes denote TFs, and blue nodes denote joint target genes. TBP is marked as an orange hexagon because it acts as both a TF and a target gene in this network. The links with green arrow ends are positive regulations from the initial TF to the end node. The links with blue ''⊣'' ends are negative regulations initiated by TFs. The lines with black ''⊣'' ends represent a negative regulatory relationship initiated by miRNAs. FFLs involved in cell differentiation are marked with a light navy background.

in OVC can respond to the initial cellular signals in a consistent and persistent style.

The previous study in *E. coli* and yeast revealed that the most common coherent FFLs with two TFs are formed with three positive regulations.[34] However, the inherent negative regulation on the miRNA targets causes a lack of the three positive links in coherent FFL co-regulated by TFs and miRNAs. Among the three types of coherent FFLs, TF-initiated TMN-TGP and TMP-TGN FFLs show relatively dominant numbers (83.58%) in our detected 67 FFLs.

Compared to MTN-TGP, both the TMN-TGP and TMP-TGN FFLs involve more unique miRNAs, TFs, and target genes (Fig. 3A–C), which means they have wider effects on cellular processes. In addition, most of target genes from MTN-TGP are covered by the other two types of FFLs. These comparisons might highlight the more important regulatory role of TFs as compared to miRNAs in serous OVC. Compared to the two TF-regulating FFLs, TMN-TGP and TMP-TGN in our network not only strengthen the regulatory effects at both the transcriptional and post-transcriptional level, but also consolidate the robustness with backup regulators. Compared to incoherent FFLs, even if the initial regulator in a coherent FFL is transiently ineffective, the secondary regulator is still able to confer fidelity to the regulation of the joint target gene.

The prevalence of oncogenes as initial regulators

The molecular mechanism underlying a coherent FFL shows that the final output of the joint target is influenced by both the initial regulator and its regulated regulator. In the three types of FFLs, the joint target genes from TMP-TGN and MTN-TGP show an inverted output compared to the input signals from its initial regulators. In contrast, TMN-TGP has the properties to stimulate the output of the joint targets. In this study, nearly half of the 67 FFLs belong to the TMN-TGP category in OVC. For the TMN-TGP FFLs, the final output of the joint targets is positively correlated with its initial TFs. By contrast, the final outputs of the joint targets from TMP-TGN and MTN-TGP are negatively correlated with its initial TFs or miRNAs. As shown in Fig. 3D, the expressions of seven targets are not inverted by their initial regulator since they only belong to TMN-TGP FFLs. However, the expression of another six targets will be inverted by their initial regulators, as they belong within TMP-TGN or MTN-TGP FFLs. Most interestingly, ten target genes can either be inverted or non-inverted, and their fate is determined by the

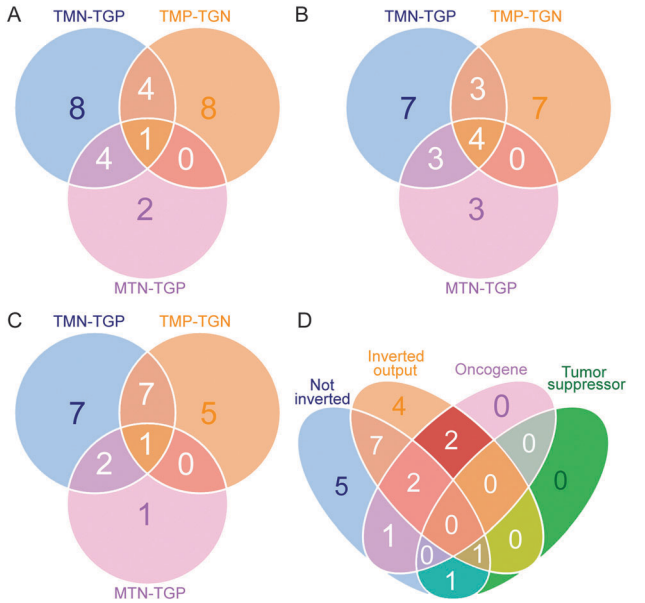

Fig. 3 The overlap of involved miRNAs, TFs and genes in three coherent FFLs. The overlap of all the involved miRNAs (A), TFs (B), and target genes (C) are shown for the three types of coherent FFLs. The definitions of the three types of coherent FFLs (TMN-TGP, TMP-TGN and MTN-TGP) are provided in Fig. 1. The output of the 23 target genes that overlap with tumor suppressor genes and oncogenes and their gene expression output status compared to their initial regulators (D). The changes of gene expression for target genes from TMN-TGP are consistent with those of their initial regulators (not inverted); however, the gene expression output of target genes from TMP-TGN and MTN-TGP were inverted by their initial regulators (inverted output).

strength of their initial regulators from different FFL types. Taking this information into consideration, the numbers of target genes with the inverted and non-inverted expression are finely balanced in the identified 67 FFLs in OVC.

However, the same equilibrium does not exist in inverted and non-inverted effects if we explore the initial regulators and their roles in cancer development. Based on the known tumor suppressors and oncogenes, we classified the FFLs into three types: the oncogenic style, tumor suppressor style and unknown regulatory style (Table 3). In the tumor suppressor style, FFLs initiated by tumor suppressors are more likely to behave as the "the guardians of the cell" to establish cell-cycle checkpoints, induce apoptosis and control the cell to grow normally. For the same effect, the loops with the initial oncogene regulator can be classified as oncogenic FFLs. With their consistent regulatory effect on joint targets, the oncogenic FFLs can behave like oncogenes, which might promote cancer progression. As shown in Table 2, there are 12 oncogenes enriched in our final regulatory network of 67 FFLs (adjusted *p*-value, 1.60×10^{-10}). Ten of the oncogenes are the initial TFs in FFLs and result in 30 oncogenic FFLs. Adding in one additional oncogenic FFL initiated by an oncogenic miRNA miR-222, there are 31 oncogenic FFLs in total in our OVC regulatory network (Table 3). In contrast, there are only four tumor-suppressor-style FFLs in total. One is initiated by the protein-coding tumor suppressor WT1, while the other three are initiated by two miRNAs let-7b and miR-30c. Furthermore, in our constructed regulatory network consisting of 27 miRNAs, 25 TFs and 23 genes (Fig. 2), the WT1-governed FFL is

Table 2 Functional terms and pathways overrepresented in the 47 coding genes involved in the OVC-specific regulatory network with coherent feed-forward loops

Ingenuity canonical pathways	Adjusted *p*-value[a]
Acute myeloid leukemia signaling	8.0×10^{-3}
Cell cycle: G1/S checkpoint regulation	0.012
Ovarian cancer signaling	0.019
Molecular mechanisms of cancer	0.025
Telomerase signaling	0.028
ERK/MAPK signaling	0.033
Functional terms	**Adjusted *p*-value[b]**
Transcription regulator activity	3.1×10^{-15}
Proto-oncogene	1.6×10^{-10}
Positive regulation of macromolecule metabolic process	4.8×10^{-10}
Multicellular organismal development	2.7×10^{-7}
Regulation of cell differentiation	6.1×10^{-5}
Regulation of cell proliferation	9.7×10^{-5}
Regulation of cell cycle	1.8×10^{-4}
Cell differentiation	4.5×10^{-4}
Regulation of homeostatic process	5.5×10^{-4}

[a] Adjusted *p*-value was calculated by Fisher's exact test followed by Benjamini–Hochberg multiple testing correction using Ingenuity Pathway Analysis (IPA) tool (http://www.ingenuity.com/products/ipa). [b] Adjusted *p*-value was calculated by hypergeometric test followed by Benjamini–Hochberg multiple testing correction using David Classification Tool (http://david.abcc.ncifcrf.gov/).

Table 3 Summary of three types of coherent feed-forward loops

	Number of coherent FFLs			
Motif	All	TS[a]	OCG[b]	Unknown[c]
TMN-TGP[d]	33	0	18	15
TMP-TGN[d]	23	1	12	10
MTN-TGP[d]	11	3	1	7
Total	**67**	**4**	**31**	**32**

[a] The coherent FFL initiated by a TF or miRNA as tumor suppressors (TSs). [b] The coherent FFL initiated by a TF or miRNA as oncogenes (OCGs). [c] The coherent FFL initiated by a TF or miRNA without known oncogenic roles. [d] The definitions of the three types of coherent FFLs are provided in Fig. 1.

not connected with the other FFLs. In fact, 65 FFLs in our composite network are connected which shows that 97% of the FFLs can communicate with each other.

In sum, for the connected 65 FFLs, only three tumor suppressor FFLs (let-7b ↔ MYCN-HIF3A, miR-30c ↔ MAFG-CACNA1C, and miR-30c ↔ RUNX2-CACNA1C) are initiated by the miRNA tumor suppressors let-7b and miR-30c. All the remaining 31 FFLs with known roles on initial regulators belong to the oncogenic style category. This imbalanced distribution of FFLs is a striking feature in our constructed OVC regulatory network (Fig. 2 and Table 3), which might lead to a whole system collapse toward the oncogenic direction if the entire oncogenic initial regulators are able to play their roles as expected.

Nearly half of FFLs to regulate cell differentiation centered by TF TCF7L1

Network topological analyses (Fig. S1, ESI†) show that the constructed network with 67 FFLs is relatively compact, which

means the nodes easily communicate with each other in short steps. Among the network, there are 18 and 22 nodes without in-degree and out-degree, which correspond to 18 regulators and 22 target genes (Fig. S1a and b, ESI†). Aside from the nodes without in-degree or out-degree, the remaining nodes are highly connected. This feature made the shortest path distribution for the whole network skewed to a smaller number such as 1 and 2, which means most of the communications have occurred only by 1 or 2 steps (Fig. S1c, ESI†). This compact network can run with high efficiency if it is utilized for a specific cellular process.

In our network, the TF TCF7L1 has the highest out-degree, controlling 10 FFLs (Fig. 2). This gene belongs to the Wnt signaling pathway and antagonizes with the TGF-beta signaling pathway.[36] As is well-known, TGF-beta is a tumor suppressor in normal epithelial cells and anti-proliferative factor at the early stages of oncogenesis.[37] Not surprisingly, TCF7L1 was reported to regulate the cell cycle[38] in prostate cancer. In addition, TCF7L1 is also important to the regulation of cell differentiation.[39,40] Additionally, it is reportedly necessary to the terminal differentiation of epithelial cells.[40] Piecing together this evidence, TCF7L1 might be important in the epithelial cell differentiation process in ovarian cancer progression. Compared to TCF7L1, another most connected node, miRNA let-7b, is all but passively regulated in our constructed network. In total, let-7b only initiates a single FFL and is regulated by the other 14 TFs, including TCF7L1. let-7b is one of most important miRNA tumor suppressors.[41] In contrast, the TCF family was reported to play its oncogenic roles during cancer progression.[42,43] In summary, for the two most connected nodes in our network, the tumor suppressor let-7b is mostly passively regulated and can be regulated by another highly connected regulating oncogene, TCF7L1.

Systematic functional and pathway analyses on all 47 coding genes in the network revealed that the majority of these genes are enriched in cancer related pathways and cell growth (Table 2). The most interesting revelation is that 17 genes are related to cell differentiation (adjusted p-value, 4.53×10^{-4}). To create consistency, only the 267 samples belonging to the Grade 3 category of tumor were used to build the regulatory networks in this study. Grade 3 OVCs are characterized as poorly differentiated, which means that the tumor cells do not look like ovarian tissue. Thus, the enriched gene list in our coherent FFLs might provide a potential molecular mechanism for the dysregulation of cell differentiation in serous OVC.

In total, there are 13 more protein-coding genes that participate in the regulation of cell differentiation, including *TCF7L1*, *TCF3*, *RUNX1*, *MAFG*, *RUNX2*, *MAFB*, *ETS1*, *CCND2*, *BCL2*, *KIT*, *ACVR2B*, *LZTS1*, and *ZBTB7A*. Moreover, 11 of these 13 genes are oncogenes (*LZTS1* and *ZBTB7A* are the exceptions). For the 11 oncogenes, six play their roles as initial TFs in our network (*TCF7L1*, *TCF3*, *RUNX1*, *ETS1*, *RUNX2*, and *MAFB*) and form 32 oncogenic coherent FFLs (Fig. 2 and Table S3, ESI†). These 32 FFLs contain 6 TFs, 20 miRNAs and 17 joint target genes. Most interestingly, among the 20 miRNAs, 15 were recently reported to regulate cell differentiation, including let-7b,[44] miR-17,[45] miR-20a,[45] miR-21,[46–48] miR-27b,[49] miR-30b,[50] miR-30c,[51] miR-30e,[52] miR-96,[53] miR-125b,[54] miR-148b,[55] miR-222,[56] miR-301a,[57] miR-301b,[58] and miR-361-3p.[59] In summary, 75% of the miRNAs regulated by 6 oncogenic TFs had literature evidence to support their functions in cell differentiation. In addition, seven of the 15 reported miRNAs are related to epithelial differentiation or epithelial–mesenchymal transition, including miR-21,[46–48] miR-27b,[49] miR-30b,[50] miR-30c,[51] miR-30e,[52] miR-96,[53] and miR-361-3p.[59] Furthermore the other 6 miRNAs are reported to regulate cell differentiation in other cancers, including let-7b,[44] miR-17,[45] miR-20a,[45] miR-125b,[54] miR-148b,[55] and miR-301a.[57] Combining the FFLs' oncogenic behaviors, the sub-network in cell differentiation might terminate the epithelial cells development and change the cancer cell fate to poorly differentiated cells.

Discussion

To facilitate an understanding of the cancer at the systems biology level, numerous studies have successfully utilized genome-wide data to uncover thousands of regulatory motifs and modules involved in tumor development. In this study, we developed a novel computational framework to construct and analyze the regulatory networks consisting of feed-forward loops, and we successfully demonstrated the power of recruiting critical cellular circuits in ovarian cancer. The key aspect of our approach is the utilization of gene expression profile correlations to infer the regulatory direction of miRNAs and TFs to their joint target genes, which is expected to generate an interlocked network with complex behaviors. To improve the robustness of our approach to identify conserved regulatory modules, we required all the predicted regulatory relationships to be conserved across human, mouse, rat and dog in our pipeline. This generalizable FFL and network analysis pipeline is highly useful to discover critical regulators and their downstream targets in tumorigenesis, and this approach can be applied to explore the regulation systems among miRNAs, TFs, tumor suppressors, oncogenes and other cancer related genes in other tumor types or other complex diseases. However, this process involves a series of computational methods, such as regulatory prediction using predefined gene sets, gene expression correlations, the identification of statistical co-regulation pairs, and functional enrichment analyses of interlocked networks. Each step, which can produce significant results for interpretation, should therefore be approached with due caution.

Most of the prior approaches to construct gene regulatory networks were based on co-expression data or sequence-based prediction results individually, which often generate undirected networks with a large number of genes.[10,11,15] Our results are preliminary, benefiting from simultaneous genomic measurements of expression in miRNAs and mRNAs in the same tumor samples. From our data flow, the correlations with gene-expression change can significantly narrow down co-expression networks and discovered functionary modules. To improve sensitivity and specificity for an in-depth analysis, we employed both co-expression and sequence-based predictions to characterize the complex

regulatory relationships with associated directions rather than the undirected protein–protein interaction network. In our results, TF related regulatory pairs were still too large and consisted of thousands of regulatory pairs (Table 1). However, benefiting from the small number of regulatory pairs related to miRNAs, we further adopted a co-regulating pair analysis to reduce the number of regulatory pairs to an appropriate scale (Table 1). This finding implied that the bottleneck regulators, participating in co-regulation, may assist in excluding inappropriate molecules by employing co-occurring enrichment techniques and following false positive controls in the particular biological processes involved in the tumor formation and development.

Our pipeline provides a framework to generate an appropriate scaled network from mass genomic data. However, we lack the ability to detect other genes that may be outside the network, yet still play important roles in the structure of the networks and cancer development. Our results showed that 41 of the 47 protein-coding genes (87.23%) detected by our method were previously reported to associate with cancer and other related diseases. In addition, the 47 protein-coding genes are mutated in 63.3% of all OVC cases from TCGA using the cBio Cancer genomics Portal.[60] Those protein-coding genes were also categorized by the Gene Ontology to have functions in many fundamental cellular processes such as regulation of cell cycles. Due to these genes' important roles, it is easy to expand the key regulatory network by incorporating direct or long-range protein–protein interaction networks or pathway regulatory information. For example, for the hub node TCF7L1 in our network, we identified five additional interesting OVC-related genes that interact with TCF7L1 including AXIN2, CCND1, CTNNB1, SMARCA4, and TERT.

Besides TF–miRNA FFLs, we may apply similar computational approaches to investigate two or multiple TFs, or two or multiple miRNAs, regulate gene(s), which has been often found in cellular systems. So far, the role of FFLs with two TFs in cancer has not been extensively explored by computational approaches. There are FFLs with two TFs found by experimental work[61] and theoretical analyses.[34,62] In addition to TF–TF FFLs, two TF regulators may form regulatory TF dimers[63] or more complex regulatory units to affect the downstream gene expression.[64]

Compared to a single oncogene analysis, oncogenic FFLs with miRNAs and TFs provide more robust tools to uncover the regulatory mechanisms and biomarkers in cancer.[34] By introducing the coherent FFL concept to cancer regulatory network analyses, we have demonstrated their power to generate unanticipated hypotheses for oncogenic regulatory networks. It is worth noting that incoherent FFLs exist too. In fact, incoherent FFLs are common in bacteria and yeast and have been found in animals.[35] So far, there has been no systematic comparison of coherent FFLs and incoherent FFLs in animals yet, especially in cancer cells. Our study of the complete coherent FFLs in ovarian cancer does not conflict with those previous studies of incoherent FFLs; rather, we focused on co-regulation mechanisms observed in cancer.[35] The systematic expansion of the properties and functions related to these recurrent coherent FFLs in cancer provide numerous experimentally testable regulatory relationships between 27 miRNAs, 25 TFs and 23 critical OVC-related genes. In turn, these tested regulatory relationships may explain the cellular wiring patterns for OVC. Among these relationships were the six genes enriched in ovarian cancer signaling from IPA pathway analysis, including GJA1, E2F1, EDNRA, TCF7L1, TCF3, and BCL2 (Table S4, ESI†). These six coding genes are connected with 16 miRNAs in our final network. In summary, these enrichments of ovarian cancer signaling, cell cycle, and cell differentiation not only demonstrate the power of our pipeline to identify critical cellular events, but also may provide a clearer picture of and urge further experimental testing for the OVC cellular signaling pathway as relates to non-coding molecules.

Among the 65 connected FFLs, a single gene *TBP* (TATA box binding protein) has roles as both a TF and a target gene. Because of its two-tier role, it can interlock two FFLs (TCF7L1 ↔ miR27b – TBP and TBP ↔ miR222 – CDKN1C) rigidly. TBP and TBP-associated factors (TAFs) belong to the transcription factor IID, which is the trigger of transcription by RNA polymerase II.[65] Though TBP was reported as widely expressed, it showed its highest expression in the testis and ovary.[66] In addition, TAFs reportedly promote colorectal cancer cell transformation of epithelial–mesenchymal transitions.[67] The loss of the epithelial features and change to mesenchymal phenotypes are the characteristic features in acquiring invasiveness in ovarian carcinoma. Combining these clues together, TBP may play important roles during epithelial–mesenchymal transitions in OVC.

In addition, there are three TF–miRNA pairs with joint targets that can regulate each other reciprocally, including miR-30c ↔ RUNX2, miR-222 ↔ TCF3, and let-7b ↔ MYCN. This reciprocal regulation might create multiple outputs of three joint target genes (*CACNA1C*, *CDKN1C*, and *HIF3A*). If the miRNA is the initiator, it will decrease the gene expression of the TF and joint targets in the two reciprocal loops dominantly. Reciprocal results will increase joint targeting gene expression and decrease miRNA expression if the TF is the dominant initiator. More interestingly, miR-30c and let-7b both belong to tumor suppressors,[68] and their TF competitors RUNX2 and MYCN are both oncogenes.

Our final network not only revealed the interlocked and reciprocal coherent FFLs, but also delivered novel topological and functional insights into the OVC cellular processes. For all 27 miRNAs, 17 (63%) have demonstrated effects on ovarian cancers.[13] The remaining 10 are all reported to have roles in cancer development, including: miR-17,[69] miR-193b,[70] miR-301a,[71] miR-301b,[72] miR-106b,[73] miR-27b,[74] miR-30b,[75] miR-32,[76] miR-361-3p,[59] and miR-506.[77] Among these, let-7b is one of the hubs and has 15 connectivities. However, only one FFL is initiated by let-7b; the remaining 14 connections are all regulations from TF to let-7b. Among the 14 FFLs, 4 are oncogenic style, and the remaining 10 are unknown. The other tumor suppressor, miR-30c, is the second highly-connected miRNA, with 7 regulatory relationships. Two of the FFLs are self-initiated. All the other 5 FFLs stem from oncogenic TFs. These findings

may highlight the importance of let-7b and miR-30c as critical tumor suppressors in OVC, which is competed by a wide range of TFs. Though these significant regulatory relationships have been discovered, more experimental examinations are needed to confirm the complex output of let-7b and miR-30c under the inferred oncogenic FFLs.

To better understand the molecular activities in the important biological processes in OVC, we generated and analyzed the sub-networks related to cell differentiation. For this purpose, we extracted all the FFLs initiated by six oncogenes involved in cell differentiation. Three-fourths of the miRNAs in this sub-network were reported to have functions in cell differentiation. Particularly, several miRNAs are related to epithelial–mesenchymal transition. Of note, three miRNAs from the miR-30 family were detected in a wide range of tissues and cells, including miR-30b in mouse mammary glands,[50] miR-30c in human adipocytes,[51] and miR-30e in intestinal cells.[52] In addition, the abundance of miRNAs from miR-30 was reduced during epithelial–mesenchymal transition in human pancreatic cells.[78] However, the role of cell differentiation of the miR-30 family in OVC has not been well studied and our results provide a new hypothesis that miR-30 family dysregulation in OVC might be related to cancer cell differentiation. These regulatory circuits discovered in OVC might expand our understanding of the function of oncogenic TFs–miRNAs through the characterization of their direct reciprocal regulatory relationship. Additionally, these intact regulatory networks not only systematically illuminate the gene regulation building blocks in the underlying genetics in cancer, but also present potential therapeutic and diagnostic protein-coding and non-coding molecules for further evaluation.

Conclusions

In summary, this study provides a comprehensive patho-physiological overview of coherent circuits in OVC that are co-regulated by miRNAs and TFs. Our finding of the strong prevalence of oncogenic coherent FFLs in serous OVC suggested an oncogene-driven regulatory machine that acts upon critical cellular processes, such as cell differentiation, in a highly efficient and cooperative way.

Acknowledgements

We thank Ms Rebecca Hiller Posey for polishing an earlier draft of the manuscript. This work was partially supported by National Institutes of Health grants (R01LM011177, R03CA167695, P30CA68485, and P50CA095103), The Robert J. Kleberg, Jr and Helen C. Kleberg Foundation (to Z.Z.), and Ingram Professorship Funds (to Z.Z.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1 R. Siegel, D. Naishadham and A. Jemal, *Ca-Cancer J. Clin.*, 2012, **62**, 10–29.
2 E. Lengyel, *Am. J. Pathol.*, 2010, **177**, 1053–1064.
3 K. R. Cho and M. Shih Ie, *Annu. Rev. Pathol.: Mech. Dis.*, 2009, **4**, 287–313.
4 J. D. Seidman, I. Horkayne-Szakaly, M. Haiba, C. R. Boice, R. J. Kurman and B. M. Ronnett, *Int. J. Gynecol. Pathol.*, 2004, **23**, 41–44.
5 A. Balmain, J. Gray and B. Ponder, *Nat. Genet.*, 2003, **33**(Suppl), 238–244.
6 M. Zhao, J. Sun and Z. Zhao, *PLoS One*, 2012, **7**, e44175.
7 J. E. Darnell, *Nat. Rev. Cancer*, 2002, **2**, 740–749.
8 T. A. Farazi, J. I. Spitzer, P. Morozov and T. Tuschl, *J. Pathol.*, 2011, **223**, 102–115.
9 H. Hermeking, *Nat. Rev. Cancer*, 2012, **12**, 613–626.
10 J. Sun, X. Gong, B. Purow and Z. Zhao, *PLoS Comput. Biol.*, 2012, **8**, e1002488.
11 Z. Yan, P. K. Shah, S. B. Amin, M. K. Samur, N. Huang, X. Wang, V. Misra, H. Ji, D. Gabuzda and C. Li, *Nucleic Acids Res.*, 2012, **40**, e135.
12 A. P. Crijns, R. S. Fehrmann, S. de Jong, F. Gerbens, G. J. Meersma, H. G. Klip, H. Hollema, R. M. Hofstra, G. J. te Meerman, E. G. de Vries and A. G. van der Zee, *PLoS Med.*, 2009, **6**, e24.
13 M. T. van Jaarsveld, J. Helleman, E. M. Berns and E. A. Wiemer, *Int. J. Biochem. Cell Biol.*, 2010, **42**, 1282–1290.
14 P. V. Nazarov, S. E. Reinsbach, A. Muller, N. Nicot, D. Philippidou, L. Vallar and S. Kreis, *Nucleic Acids Res.*, 2013, **41**(5), 2817–2831, DOI: 10.1093/nar/gks1471.
15 M. B. Gerstein, A. Kundaje, M. Hariharan, S. G. Landt, K. K. Yan, C. Cheng, X. J. Mu, E. Khurana, J. Rozowsky, R. Alexander, R. Min, P. Alves, A. Abyzov, N. Addleman, N. Bhardwaj, A. P. Boyle, P. Cayting, A. Charos, D. Z. Chen, Y. Cheng, D. Clarke, C. Eastman, G. Euskirchen, S. Frietze, Y. Fu, J. Gertz, F. Grubert, A. Harmanci, P. Jain, M. Kasowski, P. Lacroute, J. Leng, J. Lian, H. Monahan, H. O'Geen, Z. Ouyang, E. C. Partridge, D. Patacsil, F. Pauli, D. Raha, L. Ramirez, T. E. Reddy, B. Reed, M. Shi, T. Slifer, J. Wang, L. Wu, X. Yang, K. Y. Yip, G. Zilberman-Schapira, S. Batzoglou, A. Sidow, P. J. Farnham, R. M. Myers, S. M. Weissman and M. Snyder, *Nature*, 2012, **489**, 91–100.
16 The Cancer Genome Atlas Research Network, *Nature*, 2011, **474**, 609–615.
17 R. G. Verhaak, K. A. Hoadley, E. Purdom, V. Wang, Y. Qi, M. D. Wilkerson, C. R. Miller, L. Ding, T. Golub, J. P. Mesirov, G. Alexe, M. Lawrence, M. O'Kelly, P. Tamayo, B. A. Weir, S. Gabriel, W. Winckler, S. Gupta, L. Jakkula, H. S. Feiler, J. G. Hodgson, C. D. James, J. N. Sarkaria, C. Brennan, A. Kahn, P. T. Spellman, R. K. Wilson, T. P. Speed, J. W. Gray, M. Meyerson, G. Getz, C. M. Perou and D. N. Hayes, *Cancer Cell*, 2010, **17**, 98–110.
18 V. Matys, O. V. Kel-Margoulis, E. Fricke, I. Liebich, S. Land, A. Barre-Dirrie, I. Reuter, D. Chekmenev, M. Krull, K. Hornischer, N. Voss, P. Stegmaier, B. Lewicki-Potapov, H. Saxel, A. E. Kel and E. Wingender, *Nucleic Acids Res.*, 2006, **34**, D108–D110.
19 R Development Core Team, 2008.
20 M. Selbach, B. Schwanhausser, N. Thierfelder, Z. Fang, R. Khanin and N. Rajewsky, *Nature*, 2008, **455**, 58–63.

21 D. Baek, J. Villen, C. Shin, F. D. Camargo, S. P. Gygi and D. P. Bartel, *Nature*, 2008, **455**, 64–71.
22 B. P. Lewis, I. H. Shih, M. W. Jones-Rhoades, D. P. Bartel and C. B. Burge, *Cell*, 2003, **115**, 787–798.
23 A. Y. Guo, J. Sun, P. Jia and Z. Zhao, *BMC Syst. Biol.*, 2010, **4**, 10.
24 E. W. Sayers, T. Barrett, D. A. Benson, E. Bolton, S. H. Bryant, K. Canese, V. Chetvernin, D. M. Church, M. Dicuccio, S. Federhen, M. Feolo, I. M. Fingerman, L. Y. Geer, W. Helmberg, Y. Kapustin, S. Krasnov, D. Landsman, D. J. Lipman, Z. Lu, T. L. Madden, T. Madej, D. R. Maglott, A. Marchler-Bauer, V. Miller, I. Karsch-Mizrachi, J. Ostell, A. Panchenko, L. Phan, K. D. Pruitt, G. D. Schuler, E. Sequeira, S. T. Sherry, M. Shumway, K. Sirotkin, D. Slotta, A. Souvorov, G. Starchenko, T. A. Tatusova, L. Wagner, Y. Wang, W. J. Wilbur, E. Yaschenko and J. Ye, *Nucleic Acids Res.*, 2012, **40**, D13–D25.
25 L. R. Meyer, A. S. Zweig, A. S. Hinrichs, D. Karolchik, R. M. Kuhn, M. Wong, C. A. Sloan, K. R. Rosenbloom, G. Roe, B. Rhead, B. J. Raney, A. Pohl, V. S. Malladi, C. H. Li, B. T. Lee, K. Learned, V. Kirkup, F. Hsu, S. Heitner, R. A. Harte, M. Haeussler, L. Guruvadoo, M. Goldman, B. M. Giardine, P. A. Fujita, T. R. Dreszer, M. Diekhans, M. S. Cline, H. Clawson, G. P. Barber, D. Haussler and W. J. Kent, *Nucleic Acids Res.*, 2013, **41**, D64–D69.
26 A. E. Kel, E. Gossling, I. Reuter, E. Cheremushkin, O. V. Kel-Margoulis and E. Wingender, *Nucleic Acids Res.*, 2003, **31**, 3576–3579.
27 A. Kozomara and S. Griffiths-Jones, *Nucleic Acids Res.*, 2011, **39**, D152–D157.
28 D. S. Goldberg and F. P. Roth, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 4372–4376.
29 Y. Benjamini and Y. Hochberg, *J. R. Stat. Soc. Ser. B (Methodological)*, 1995, **57**, 289–300.
30 M. Zhao, J. Sun and Z. Zhao, *Nucleic Acids Res.*, 2013, **41**, D970–D976.
31 The UniProt Consortium, *Nucleic Acids Res.*, 2013, **41**, D43–D47.
32 D. Wodarz, *Oncogene*, 2004, **23**, 7799–7809.
33 D. W. Huang, B. T. Sherman, Q. Tan, J. R. Collins, W. G. Alvord, J. Roayaei, R. Stephens, M. W. Baseler, H. C. Lane and R. A. Lempicki, *Genome Biol.*, 2007, **8**, R183.
34 S. Mangan and U. Alon, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 11980–11985.
35 L. Goentoro, O. Shoval, M. W. Kirschner and U. Alon, *Mol. Cell*, 2009, **36**, 894–899.
36 C. Chesnutt, L. W. Burrus, A. M. Brown and L. Niswander, *Dev. Biol.*, 2004, **274**, 334–347.
37 J. J. Hill, T. L. Tremblay, C. Cantin, M. O'Connor-McCourt, J. F. Kelly and A. E. Lenferink, *Proteome Sci.*, 2009, **7**, 2.
38 S. Filleur, J. Hirsch, A. Wille, M. Schon, C. Sell, M. H. Shearer, T. Nelius and I. Wieland, *Cancer Cell Int.*, 2009, **9**, 28.
39 L. Ombrato, F. Lluis and M. P. Cosma, *Cell Cycle*, 2012, **11**, 39–47.
40 C. E. Barton, K. N. Johnson, D. M. Mays, K. Boehnke, Y. Shyr, P. Boukamp and J. A. Pietenpol, *Cell Death Dis.*, 2010, **1**, e74.
41 Y. S. Lee and A. Dutta, *Genes Dev.*, 2007, **21**, 1025–1030.
42 M. Idogawa, T. Yamada, K. Honda, S. Sato, K. Imai and S. Hirohashi, *Gastroenterology*, 2005, **128**, 1919–1936.
43 M. Lepourcelet, Y. N. Chen, D. S. France, H. Wang, P. Crews, F. Petersen, C. Bruseo, A. W. Wood and R. A. Shivdasani, *Cancer Cell*, 2004, **5**, 91–102.
44 B. Boyerinas, S. M. Park, A. Hau, A. E. Murmann and M. E. Peter, *Endocr.-Relat. Cancer*, 2010, **17**, F19–F36.
45 M. He, Q. Y. Wang, Q. Q. Yin, J. Tang, Y. Lu, C. X. Zhou, C. W. Duan, D. L. Hong, T. Tanaka, G. Q. Chen and Q. Zhao, *Cell Death Differ.*, 2013, **20**(3), 408–418, DOI: 10.1038/cdd.2012.130.
46 P. Bhat-Nakshatri, G. Wang, N. R. Collins, M. J. Thomson, T. R. Geistlinger, J. S. Carroll, M. Brown, S. Hammond, E. F. Srour, Y. Liu and H. Nakshatri, *Nucleic Acids Res.*, 2009, **37**, 4850–4861.
47 Y. J. Kim, S. J. Hwang, Y. C. Bae and J. S. Jung, *Stem Cells*, 2009, **27**, 3093–3102.
48 Z. Niu, S. M. Goodyear, S. Rao, X. Wu, J. W. Tobias, M. R. Avarbock and R. L. Brinster, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 12740–12745.
49 S. A. Melo and R. Kalluri, *Blood*, 2012, **119**, 2439–2440.
50 S. Le Guillou, N. Sdassi, J. Laubier, B. Passet, M. Vilotte, J. Castille, D. Laloe, J. Polyte, S. Bouet, F. Jaffrezic, E. P. Cribiu, J. L. Vilotte and F. Le Provost, *PLoS One*, 2012, **7**, e45727.
51 M. Karbiener, C. Neuhold, P. Opriessnig, A. Prokesch, J. G. Bogner-Strauss and M. Scheideler, *RNA Biol.*, 2011, **8**, 850–860.
52 Y. Liao and B. Lonnerdal, *Cell. Mol. Life Sci.*, 2010, **67**, 2969–2978.
53 S. K. Laine, J. J. Alm, S. P. Virtanen, H. T. Aro and T. K. Laitala-Leinonen, *J. Cell. Biochem.*, 2012, **113**, 2687–2695.
54 S. Bhattacharjya, S. Nath, J. Ghose, G. P. Maiti, N. Biswas, S. Bandyopadhyay, C. K. Panda, N. P. Bhattacharyya and S. Roychoudhury, *Cell Death Differ.*, 2013, **20**(3), 430–442, DOI: 10.1038/cdd.2012.135.
55 Y. Song, Y. Xu, Z. Wang, Y. Chen, Z. Yue, P. Gao, C. Xing and H. Xu, *Int. J. Cancer*, 2012, **131**, 1042–1051.
56 K. Qian, L. L. Hu, H. Chen, H. X. Li, N. Liu, Y. F. Li, J. H. Ai, G. J. Zhu, Z. P. Tang and H. W. Zhang, *Endocrinology*, 2009, **150**, 4734–4743.
57 M. Wang, C. Li, B. Yu, L. Su, J. Li, J. Ju, Y. Yu, Q. Gu, Z. Zhu and B. Liu, *J. Gastroenterol.*, 2013, **48**(9), 1023–1033, DOI: 10.1007/s00535-012-0733-6.
58 M. R. Suh, Y. Lee, J. Y. Kim, S. K. Kim, S. H. Moon, J. Y. Lee, K. Y. Cha, H. M. Chung, H. S. Yoon, S. Y. Moon, V. N. Kim and K. S. Kim, *Dev. Biol.*, 2004, **270**, 488–498.
59 C. Roth, I. Stuckrath, K. Pantel, J. R. Izbicki, M. Tachezy and H. Schwarzenbach, *PLoS One*, 2012, **7**, e38248.
60 E. Cerami, J. Gao, U. Dogrusoz, B. E. Gross, S. O. Sumer, B. A. Aksoy, A. Jacobsen, C. J. Byrne, M. L. Heuer, E. Larsson,

Y. Antipin, B. Reva, A. P. Goldberg, C. Sander and N. Schultz, *Cancer Discovery*, 2012, **2**, 401–404.

61 H. Xin, R. Z. Lu, H. Lee, W. Zhang, C. Y. Zhang, J. H. Deng, Y. Liu, S. D. Shen, K. U. Wagner, S. Forman, R. Jove and H. Yu, *J. Biol. Chem.*, 2013, **288**, 13842–13849.

62 R. Murugan, *PLoS One*, 2012, **7**, e41027.

63 A. Jankowski, E. Szczurek, R. Jauch, J. Tiuryn and S. Prabhakar, *Genome Res.*, 2013, **23**(8), 1307–1318, DOI: 10.1101/gr.154922.113.

64 B. C. Haynes, E. J. Maier, M. H. Kramer, P. I. Wang, H. Brown and M. R. Brent, *Genome Res.*, 2013, **23**(8), 1319–1328, DOI: 10.1101/gr.150904.112.

65 R. D. Kornberg, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 12955–12961.

66 C. Di Pietro, M. Ragusa, L. Duro, M. R. Guglielmino, D. Barbagallo, A. Carnemolla, A. Lagana, P. Buffa, R. Angelica, A. Rinaldi, M. S. Calafato, I. Milicia, C. Caserta, R. Giugno, A. Pulvirenti, V. Giunta, A. Rapisarda, V. Di Pietro, A. Grillo, A. Messina, A. Ferro, K. H. Grzeschik and M. Purrello, *DNA Cell Biol.*, 2007, **26**, 369–385.

67 A. Voulgari, S. Voskou, L. Tora, I. Davidson, T. Sasazuki, S. Shirasawa and A. Pintzas, *Mol. Cancer Res.*, 2008, **6**, 1071–1083.

68 J. Shen, C. B. Ambrosone and H. Zhao, *Int. J. Cancer*, 2009, **124**, 1178–1182.

69 H. Luo, J. Zou, Z. Dong, Q. Zeng, D. Wu and L. Liu, *Biochem. J.*, 2012, **442**, 311–321.

70 C. Xie, X. H. Jiang, J. T. Zhang, T. T. Sun, J. D. Dong, A. J. Sanders, R. Y. Diao, Y. Wang, K. L. Fok, L. L. Tsang, M. K. Yu, X. H. Zhang, Y. W. Chung, L. Ye, M. Y. Zhao, J. H. Guo, Z. J. Xiao, H. Y. Lan, C. F. Ng, K. M. Lau, Z. M. Cai, W. G. Jiang and H. C. Chan, *Oncogene*, 2013, **32**, 2282–2291, DOI: 10.1038/onc.2012.251.

71 Z. X. Lu, Y. Li, A. Takwi, B. H. Li, J. W. Zhang, D. J. Conklin, K. H. Young, R. Martin and Y. Li, *EMBO J.*, 2011, **30**, 57–67.

72 W. Shi, K. Gerster, N. M. Alajez, J. Tsang, L. Waldron, M. Pintilie, A. B. Hui, J. Sykes, C. P'ng, N. Miller, D. McCready, A. Fyles and F. F. Liu, *Cancer Res.*, 2011, **71**, 2926–2937.

73 F. Petrocca, A. Vecchione and C. M. Croce, *Cancer Res.*, 2008, **68**, 8191–8194.

74 Y. F. Wang, R. Rathinam, A. Walch and S. K. Alahari, *J. Biol. Chem.*, 2009, **284**, 23094–23106.

75 A. Gaziel-Sovran, M. F. Segura, R. Di Micco, M. K. Collins, D. Hanniford, E. V.-S. de Miera, J. F. Rakus, J. F. Dankert, S. Shang, R. S. Kerbel, N. Bhardwaj, Y. Shao, F. Darvishian, J. Zavadil, A. Erlebacher, L. K. Mahal, I. Osman and E. Hernando, *Cancer Cell*, 2011, **20**, 104–118.

76 S. E. Jalava, A. Urbanucci, L. Latonen, K. K. Waltering, B. Sahu, O. A. Janne, J. Seppala, H. Lahdesmaki, T. L. J. Tammela and T. Visakorpi, *Oncogene*, 2012, **31**, 4460–4471.

77 Y. Zhao, H. Liu, Y. Li, J. Wu, A. R. Greenlee, C. Yang and Y. Jiang, *Toxicol. Lett.*, 2011, **205**, 320–326.

78 M. V. Joglekar, D. Patil, V. M. Joglekar, G. V. Rao, D. N. Reddy, S. Mitnala, Y. Shouche and A. A. Hardikar, *Islets*, 2009, **1**, 137–147.

