RESEARCH ARTICLE

Weighted Gene Co-expression Network Analysis in Identification of Endometrial Cancer Prognosis Markers

Xiao-Lu Zhu, Zhi-Hong Ai, Juan Wang, Yan-Li Xu, Yin-Cheng Teng*

Abstract

Objective: Endometrial cancer (EC) is the most common gynecologic malignancy. Identification of potential biomarkers of EC would be helpful for the detection and monitoring of malignancy, improving clinical outcomes. Methods: The Weighted Gene Co-expression Network Analysis method was used to identify prognostic markers for EC in this study. Moreover, underlying molecular mechanisms were characterized by KEGG pathway enrichment and transcriptional regulation analyses. Results: Seven gene co-expression modules were obtained, but only the turquoise module was positively related with EC stage. Among the genes in the turquoise module, COL5A2 (collagen, type V, alpha 2) could be regulated by PBX (pre-B-cell leukemia homeobox 1)1/2 and HOXB1 (homeobox B1) transcription factors to be involved in the focal adhesion pathway; CENP-E (centromere protein E, 312kDa) by E2F4 (E2F transcription factor 4, p107/p130-binding); MYCN (v-myc myelocytomatosis viral related oncogene, neuroblastoma derived [avian]) by PAX5 (paired box 5); and BCL-2 (B-cell CLL/lymphoma 2) and IGFBP-6 (insulin-like growth factor binding protein 6) by GLI1. They were predicted to be associated with EC progression via Hedgehog signaling and other cancer related-pathways. Conclusions: These data on transcriptional regulation may provide a better understanding of molecular mechanisms and clues to potential therapeutic targets in the treatment of EC.

Keywords: Endometrial cancer - Co-expression module - KEGG - transcriptional regulation

Introduction

Endometrial cancer (EC) is one of the most prevalent gynecologic malignancies and there are an estimated 40,000 new cases diagnosed in the United States (Jemal et al., 2009). EC is categorized into two subtypes, type I (the estrogen-related) and type II (the non-estrogen depended). Most patients (approximately 80%) are diagnosed with type I EC, low stage and grade, endometrioid histology, thus generally with a good prognosis (Rose, 1996; Tolentino Silva et al., 2012). In contrast, outcomes are extremely poor for patients with type II EC because of high stage and grade, non-endometrioid histology (Bokhman, 1983; Ray and Fleming, 2009). Therefore, there is an urgent need to explore some potential biomarkers of EC as a tool for the detection and monitoring of aggressive endometrial tumors. Meanwhile, a better understanding of the molecular mechanism involved in EC is beneficial for the development of novel and more selective molecular targeted approaches against EC.

Emerging evidence has suggested that type I shows specifically genetic alterations in microsatellite instability due to alterations of the mismatch repair genes and mutations in PTEN tumor suppressor gene, Pik3ca, K-Ras protooncogene, and CTNNB1 (beta-catenin). Type II exhibits p53 mutations, overexpression of Her2/neu oncogene, and chromosomal instability (Catasus et al., 2009; Dobrzycka and Terlikowski, 2010). Furthermore, Mig-6 ablation either singly or in combination with Pten ablation dramatically accelerates the development of EC through the down-regulation of the estrogen-induced apoptosis inhibitors Birc1 and the inhibition of ERK2 phosphorylation (Kim et al., 2010). Leptin promotes EC growth and invasiveness by up-regulation of cyclin D1 together with a down-regulation of cyclin-dependent kinase inhibitor p21WAF1/Cip1. However, cyclin D1 expression is depended on signal transducers and activators of transcription 3 (STAT3) and cyclic AMP-responsive element (CRE) binding protein, which can bind to the cyclin D1 promoter (Catalano et al., 2009; Liu et al., 2011). However, high-throughput screening of multiple EC related genes and investigation of their underlying molecular mechanisms remain still rare.

Network approaches provide a means to bridge the gap from individual genes to complex traits. Weighted Gene Co-expression Network Analysis (WGCNA) constructs gene sets (modules) from the observed gene expression data. These modules are then related to gene ontology information to study their biological plausibility and to eliminate spurious modules due to technical artifacts (Horvath and Dong, 2008). WGCNA has been successfully applied to identify the genes of brain cancer.
(Horvath et al., 2006), breast cancer (Shi et al., 2010), and glioblastoma (Xiang et al., 2012). Thus, we also applied the same strategy to identify more prognosis markers for EC. Moreover, we characterized their underlying molecular mechanisms by Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis and transcriptional regulation analysis.

Materials and Methods

EC gene expression profile data were accessible at National Center for Biotechnology Information (NCBI) Gene Expression Omnibus data repository (http://www.ncbi.nlm.nih.gov/geo/) using the series accession number GSE36389. The base data are based on the Affymetrix Human Genome U133 Plus 2.0 Array. The 20 tissue samples for the microarray study consisted of 7 control samples, 3- G1 samples, 8- G2 samples and 2 - G3 samples.

A total of 7234 pairs of regulatory relationship between 376 transcription factors (TFs) and 2653 target genes were collected from TRANSFAC (Matys et al., 2003) (http://www.gene-regulation.com/pub/databases.html) and TRED (Jiang et al., 2007) (http://rulai.cshl.edu/TRED/) database.

KEGG pathway database is a collection of manually drawn pathway maps of the molecular interaction and reaction networks. KEGG datasets were downloaded on 2011-12 that contained 211 pathways.

Raw gene expression data files were read into R statistical package using the Affy package. The probes were eliminated if they mapped several or without genes. As for the genes mapping for several probes, we used an average to calculate, resulting in 12633 genes. The genes only with p-value less than 0.05 were selected as differentially expressed gene (DEGs) by analysis of variance. Finally, a total of 1055 genes were obtained, indicating these genes play important roles in the development of EC at various stages.

To explore the interactions between these genes, we applied a systems biology approach using WGCNA (Horvath and Dong, 2008). Unlike other gene co-expression networks using a binary variable to encode gene co-expression (connected, 1; unconnected, 0), the WGCNA converts co-expression measure into connection weights or topology overlap measures (TOM). We inputted expression profiles of 1055 DEGs to construct weighted gene co-expression modules using 1-TOM as the distance measure with a height cutoff of 0.95 and a minimum size cutoff of 40. A total of 7 gene co-expression modules were obtained and the enriched genes of these gene co-expression modules were 193 (turquoise), 192 (blue), 185 (brown), 99 (yellow), 94 (green), 89 (red), and 67 (black). A total of 136 genes (grey) were not included in co-expression modules.

Eigengenes were calculated to display the correlation between each module by using the function module eigengenes within the WGCNA package. The dendrogram was constructed by hierarchical clustering of module eigengenes (Figure 1). We found there was high similarity between brown and black module, green and blue module, red and yellow module.

Based on the clinical information of samples, we attempted to confirm the EC related co-expression modules (that is, association between each module and sample label). The correlation coefficient was presented in Figure 2. We found the turquoise module was positively correlated with EC development(r=0.49, p=4.7e-13). These findings showed turquoise module was highly correlated with EC grade. Further, this turquoise module was annotated in KEGG pathway with the P<0.05. Finally, a total of 11 pathways were enriched (Table 1).

In addition, we mined the relationship between transcription factors and target genes within the turquoise module. We found COL5A2 (collagen, type V, alpha 2) was annotated in KEGG pathway with the P<0.05. The resulting dendrogram. If two genes have a high TOM, they are correlated with the same set of genes.

To explore whether genes in each module share a common biological function, we searched for overrepresentation in KEGG pathway. The enriched genes in KEGG pathway were statistically calculated by hypergeometric distribution. If KEGG pathway has M total genes, the number of DEGs is K, and M genes are involved in one specific pathway, the probability of generating i overlapping genes at least is calculated as follows:

\[ P = 1 - \sum_{i=0}^{K} \binom{K}{i} \binom{M-K}{N-i} \]

### Results

A total of 1055 genes were identified by using the analysis of variance with the p<0.05. Therefore, we suggest these genes may play important roles in EC development at different stages. Further, these genes were enriched into pathway. A total of 6 pathways were detected (Table 1).

Expression profiles of 1055 DEGs were used to construct weighted gene co-expression modules using 1-TOM as the distance measure with a height cutoff of 0.95 and a minimum size cutoff of 40. A total of 7 gene co-expression modules were obtained and the enriched genes of these gene co-expression modules were 193 (turquoise), 192 (blue), 185 (brown), 99 (yellow), 94 (green), 89 (red), and 67 (black). A total of 136 genes (grey) were not included in co-expression modules.

Eigengenes were calculated to display the correlation between each module by using the function module eigengenes within the WGCNA package. The dendrogram was constructed by hierarchical clustering of module eigengenes (Figure 1). We found there was high similarity between brown and black module, green and blue module, red and yellow module.

Based on the clinical information of samples, we attempted to confirm the EC related co-expression modules (that is, association between each module and sample label). The correlation coefficient was presented in Figure 2. We found the turquoise module was positively correlated with EC development(r=0.49, p=4.7e-13). These findings showed turquoise module was highly correlated with EC grade. Further, this turquoise module was annotated in KEGG pathway with the P<0.05. Finally, a total of 11 pathways were enriched (Table 2).

In addition, we mined the relationship between transcription factors and target genes within the turquoise module. We found COL5A2 (collagen, type V, alpha 2) was annotated in KEGG pathway with the P<0.05. The resulting dendrogram. If two genes have a high TOM, they are correlated with the same set of genes.

To explore whether genes in each module share a common biological function, we searched for overrepresentation in KEGG pathway. The enriched genes in KEGG pathway were statistically calculated by hypergeometric distribution. If KEGG pathway has M total genes, the number of DEGs is K, and M genes are involved in one specific pathway, the probability of generating i overlapping genes at least is calculated as follows:

\[ P = 1 - \sum_{i=0}^{K} \binom{K}{i} \binom{M-K}{N-i} \]

### Table 1. KEGG Pathway Enrichment

<table>
<thead>
<tr>
<th>Path ID</th>
<th>Path name</th>
<th>Total gene</th>
<th>Count gene</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hsa00630</td>
<td>Glyoxylate and dicarboxylate metabolism</td>
<td>15</td>
<td>5</td>
<td>0.0057</td>
</tr>
<tr>
<td>Hsa04110</td>
<td>Cell cycle</td>
<td>115</td>
<td>16</td>
<td>0.0263</td>
</tr>
<tr>
<td>Hsa04141</td>
<td>Protein processing in endoplasmic reticulum</td>
<td>148</td>
<td>19</td>
<td>0.0349</td>
</tr>
<tr>
<td>Hsa04640</td>
<td>Hematopoietic cell lineage</td>
<td>84</td>
<td>12</td>
<td>0.0423</td>
</tr>
<tr>
<td>Hsa05410</td>
<td>Hypertrophic cardiomyopathy (HCM)</td>
<td>76</td>
<td>12</td>
<td>0.0210</td>
</tr>
<tr>
<td>Hsa05414</td>
<td>Dilated cardiomyopathy</td>
<td>81</td>
<td>13</td>
<td>0.0147</td>
</tr>
</tbody>
</table>
Table 2. Turquoise Module Enriched Pathway

<table>
<thead>
<tr>
<th>Path ID</th>
<th>Path name</th>
<th>Total gene</th>
<th>Count gene</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hsa02008</td>
<td>Ribosome biogenesis in eukaryotes</td>
<td>58</td>
<td>5</td>
<td>0.0022</td>
</tr>
<tr>
<td>Hsa04510</td>
<td>Focal adhesion</td>
<td>189</td>
<td>8</td>
<td>0.0102</td>
</tr>
<tr>
<td>Hsa04640</td>
<td>Hematopoietic cell lineage</td>
<td>84</td>
<td>5</td>
<td>0.0107</td>
</tr>
<tr>
<td>Hsa00564</td>
<td>Glycerophospholipid metabolism</td>
<td>66</td>
<td>4</td>
<td>0.0208</td>
</tr>
<tr>
<td>Hsa05200</td>
<td>Pathways in cancer</td>
<td>312</td>
<td>10</td>
<td>0.0261</td>
</tr>
<tr>
<td>Hsa04141</td>
<td>Protein processing in endoplasmic reticulum</td>
<td>148</td>
<td>6</td>
<td>0.0304</td>
</tr>
<tr>
<td>Hsa04810</td>
<td>Regulation of actin cytoskeleton</td>
<td>189</td>
<td>7</td>
<td>0.0308</td>
</tr>
<tr>
<td>Hsa04340</td>
<td>Hedgehog signaling pathway</td>
<td>45</td>
<td>3</td>
<td>0.0347</td>
</tr>
<tr>
<td>Hsa04530</td>
<td>Tight junction</td>
<td>115</td>
<td>5</td>
<td>0.0364</td>
</tr>
<tr>
<td>Hsa05164</td>
<td>Influenza A</td>
<td>159</td>
<td>6</td>
<td>0.0411</td>
</tr>
<tr>
<td>Hsa05160</td>
<td>Hepatitis C</td>
<td>122</td>
<td>5</td>
<td>0.0451</td>
</tr>
</tbody>
</table>

Discussion

Our study explored several gene co-expression modules in EC progression. Among them, turquoise module was predicted positively correlated. Further, we observed several target genes were regulated by transcription factors to involve in various pathways. For example, COL5A2 may be involved in Focal adhesion pathway and regulated by PBX1/2 and HOXB1 transcription factors; CENP-E, MYCN, BCL-2, and IGFBP-6 may be involved in pathway in cancer. Their expressions were regulated by E2F4, PAX5, and GLI1 transcription factors, respectively. Interestingly, GLI1 is one component involved in Hedgehog signaling pathway.

Extracellular matrix (ECM) has been demonstrated to play a crucial role in cell proliferation, migration, differentiation, apoptosis as well as carcinogenesis.

Collagen is the major constituent of the tumor ECM components and several types of collagens have been identified to participate in focal adhesion pathway. The collagenous matrix is degraded in malignant tissue, due to the up-regulation of several degradative enzymes in tumor. For example, a specific up-regulation of COL5A2 and COL11A1 was observed in colon carcinoma (Fischer et al., 2001). An increase of type V collagen expression has also been observed in human breast carcinoma, which can up-regulate Bcl-xS, Bad, Dap kinase, hsf-1, mthsp75, caspase-1,-5,-8,-9, and -14, whilst down-regulate Bcl-2, Bcl-xβ, and hsp60 (Luparello and Sirchia, 2005). Recent study indicates components of the extracellular matrix (SPARC, COL5A1, COL5A2, COL5A1, COL6A3 and COL15A1) are involved in EC development and progression (Yeramian et al., 2012). Similarly, we also found COL5A2 was also involved in EC progression in this study.

Cell type-specific expression of the human COL5A2 gene depends on a cis-acting element that consists of two contiguous protein binding sites (FPA and FPB) located between nucleotides -149 and -95, relative to the transcription start site. Three nuclear factors, PBX1/2 and HOXB1, can bind the bipartite core sequence of FPB and this in turn leads to promoter transactivation (Penkov et al., 2000; Nagato et al., 2004).

CENP-E, a kinesin-like motor protein, is required for efficient capture and attachment of spindle microtubules.
by kinetochores, a necessary step in chromosome alignment during metaphase (McEwen et al., 2001; Sardar and Gilbert, 2012). Microinjection or the antisense approach shows that cells with CENP-E defects have prolonged mitotic arrest, not progressed into anaphase, and even initiated apoptosis (Yao et al., 2000). Both CENP-E mRNA and protein levels are significantly reduced in HCC tissues and HepG2 cells, respectively (Liu et al., 2009). After treatment with paclitaxel, the growth inhibition rate is obviously increased and the apoptotic ratio is obviously increased in human cervix cancer HeLa cells with down-regulated CENP-E. This study shows that down-regulation of CENP-E gene expression can enhance the sensitivity of cervical cancer cells to paclitaxel (SuYan et al., 2009). In this study, we found CENP-E was associated with EC progression. Importantly, CENP-E has been demonstrated to be regulated by E2F4, whose activity helps to sustain G2 arrest and prevent cell death (Plesca et al., 2007).

The BDII rat is genetically predisposed to estrogen-dependent endometrial adenocarcinoma and has been recognized as a model system to study genes involved in this tumor. Several studies indicate that extra copies of the proximal region of rat chromosome (RNO) 6 commonly could be detected in endometrial adenocarcinoma. Cancer-related genes were located in RNO6q14-q16, such as MYCN (Carlsson et al., 2001; Adamovic et al., 2005). MYCN is a member of the MYC family of oncogenes which code for transcription factors containing a basic-helix-loop-helix-leucine zipper domain. MYCN is proved high amplification in endometrial adenocarcinoma. This suggests overexpression of MYCN contributes to the development of EC in the BDII rat (Carlsson et al., 2001). The PAX5 protein targets and activates the expression of MYCN (Nutt et al., 1998; Bougel et al., 2010).

BCL-2 gene family genes are the most crucial regulators of apoptosis. Bcl-2 has the ability to block a wide variety of apoptotic signals (Adams and Cory, 1998), inhibit apoptosis, reduce the requirement for growth factors, and thereby extend the survival of cells. Down-regulated expression of Bcl-2 has been reported in a large number of human cancer, including endometrial carcinoma (Fyallah et al., 2011). There is a negative correlation between BCL-2 expression and histologic type and tumor grade (Erdem et al., 2003). A high Bcl-2/Bax ratio makes cells resistant to apoptotic stimuli, while a low ratio induces cell death. Bcl-2 and Bax are found expressed in normal and hyperplastic endometrium, and the Bcl-2/Bax ratio is lower in endometrial carcinoma (Vaskivuo et al., 2002). Interestingly, the angiogenic factor adrenomedullin treated or stably transfected Ishikawa cells show increased resistance to hypoxia induced apoptosis via upregulation of Bcl-2 in an autocrine/paracrine mechanism (Oehler et al., 2001). Bcl-2 overexpression in Ishikawa also leads to decreased cisplatin-induced apoptosis (Rouette et al., 2012). In addition, Mitf, a basic helix-loop-helix leucine zipper (b-HLH-Zip) transcription factor, occupies the BCL2 promoter, regulates endogenous levels of BCL2 in cancer (McGill et al., 2002). Insulin-like growth factors (IGF-I and IGF-II) are potent mitogens and postulated to exert autocrine, and paracrine effects on proliferation and differentiation in many cell types. Their mitogenic effects are tightly regulated by the IGF binding proteins (IGFBPs) stimulate. For example, overexpression of IGF-I and IGF-II is observed in gastric cancer cell line and tissues concomitant with highly expressed IGFBP-6 (Yi et al., 2001). IGFBP-6 can completely inhibit IGF-II-induced proliferation and adhesion effects in colon cancer cells (Leng et al., 2001). Similarly, over-expression of IGFBP-6 significantly suppresses the proliferation, invasion and metastastic activity of nasopharyngeal carcinoma cells and increases their apoptosis by regulating the expression of EGR-1 (Kuo et al., 2010). In EC tissues, the mean IGFBP-6 mRNA level is similar to that in cycling endometrium during the peri-ovulatory period (Rutanen et al., 1994).

Aberrant activation of Hedgehog signaling pathway has been linked to multiple types of human cancer. Transcriptional activation occurs through the GLI family of proteins resulting in activation of target genes. The expression of Gli1 is very weak in normal endometrium, but is increased in endometrial carcinoma. The cycloamine/siGli1-induced growth suppression is associated with the down-regulation of cyclins D1 and A and N-myc (Feng et al., 2007). Ectopic overexpression of Gli1 into EC cells can also lead to reduced expression of β-catenin in cell cytoplasm and increased expression of β-catenin in the nuclei (Liao et al., 2009). In addition, Gli1 can bound to promoter regions of Bcl-2 and IGFBP6 genes and induce their expression (Harris et al., 2011). The levels of IGFBP6 and Bcl-2 mRNA are decreased as Gli1 mRNA significantly by RNAi in cultured pancreatic cancer cells (Xu et al., 2009). In this study, we also found Gli1 could specifically regulate Bcl-2 and IGFBP6 expression in endometrial carcinoma.

In conclusion, our present findings shed new light on the EC progression, such as COL5A2-PBX1/2 and COL5A2-HOXB1 interaction via Focal adhesion pathway; CENPE -E2F4, MYCN-PAX5, BCL2-GLI1, and IGFBP6-GLI1 interaction via pathway in cancer. However, further experiments are still needed to confirm our conclusion in EC.

References


Weighted Gene Co-expression Network Analysis in Identification of Endometrial Cancer Prognosis Markers


DOI:http://dx.doi.org/10.7314/APJCP.2012.13.9.4607