EJGO European Journal of Gynaecological Oncology

Identification of potential miRNAs and candidate genes of cervical intraepithelial neoplasia by bioinformatic analysis

C. Yang[1], X. Xu[2], H. Jin[1]

[1] *Department of Gynecology, Women's Hospital, School of Medicine, Zhejiang University, Hangzhou (China)*

[2] *Sansom Institute for Health Research, School of Pharmacy and Medical Sciences, University of South Australia, Adelaide (Australia)*

Summary

Purpose: The objective of this study was to predict potential target genes and key miRNAs for cervical intraepithelial neoplasia (CIN) by bioinformatics analysis. *Materials and Methods:* The microarray data of GSE51993 were downloaded from Gene Expression Omnibus (GEO) database. Total 30 chips data from two platforms (each platform including eight CIN III samples data and seven normal cervix samples data) were used to identify the feature miRNAs and genes between CIN III and normal samples, respectively. Then the miRNA-mRNA regulatory network was constructed using Cytoscape software. Gene Ontology (GO) function and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were performed for all target genes with the Database for Annotation, Visualization and Integrated Discovery (DAVID) online tool. Transcription factors (TFs) and cancer-related genes were analyzed. *Results:* Total 21 putative target miRNAs and 361 putative target mRNAs were gained. The miRNA-mRNA regulatory network results showed that miR-338-5p, miR-193a-5p, and miR-216b were top three hub nodes. GO terms significantly enriched were extracellular region ($p = 0.004191$) and embryonic skeletal system ($p = 0.004742$). No significantly enriched KEGG pathway term was found in this study. *PBX1* (pre-B-cell leukemia transcription factor 1) and *LAMC2* (laminin subunit gamma-2) were cancer-promoting genes and also, *PBX1* was TF. *Conclusions: PBX1* and *LAMC2* may be target genes for CIN. MiR-338, and miR-216 may be key miRNAs in CIN development.

Key words: Cervical intraepithelial neoplasia; Bioinformatics; miRNA-mRNA regulatory network.

Introduction

Cervical cancer, the malignant neoplasm of the cervix uteri, is the second most common cancer among women worldwide [1]. Cervical intraepithelial neoplasia (CIN) is considered a precursor of cervical cancer. Development of cervical cancer goes through several premalignant stages, from low-grade CIN (CIN I) through high-grade CIN (CIN II/ III) to cervical cancer [2]. These lesions have the ability to progress from hyperplasia to cervix, to preinvasive carcinoma, and ultimately to invasive carcinoma [3]. There is a general agreement that either ablation or excision of CIN-II, III could reduce the incidence and mortality caused by invasive cervical cancer in women with these lesions [4].

Numerous studies have been obtained in exploring the pathological mechanism underlying CIN development. MicroRNA (miRNA) has been found to play a key role in the CIN progression. MiRNAs are a class of 19 to 23 nucleotide single-stranded RNA molecules [5] and are epigenetic factors that regulate cell proliferation, tumor cell growth, cancer formation, and metastasis by regulating tumor suppressor genes or oncogenes [6]. Martinez *et al.* reported that miR-218 levels in patients with high-risk CIN were lower than those with low-risk CIN and therefore, downregulated miR-218 may be involved in the pathogenesis of cervical cancer [7]. In addition, tumor suppressor genes such as p16 and retinoblastoma proteins played roles in the neoplastic changes of CIN [8]. Previous studies showed that vascular endothelial growth factor (VEGF) expression was associated with the progression of CIN [9]. Using microarray gene expression data and bioinformatic analysis, Prashant *et al.* suggested that transcription factor (TF) family E2F played an important role in cervical carcinogenesis [10]. Moreover, apolipoprotein A1, a truncated form of transthyretin and a cleavage fragment of inter-α-trypsin inhibitor heavy chain H4 were identified to be potential biomarker [11]. Although many factors have been found in CIN development, the pathogenic mechanisms of CIN are still not clearly demonstrated, and it is lack of effective markers for CIN treatment.

In this study, the authors applied biological informatics methods to identify the feature genes and miRNAs between CIN III and normal samples. Additionally, the miRNA-mRNA regulatory network was constructed and the functional enrichment analysis was performed. The authors aimed to explore the molecular mechanism and discover the potential target genes and key miRNAs of CIN.

Materials and Methods

Microarray data

The data of GSE51993, which was deposited by Mo *et al*. on 2013, were downloaded from GEO (Gene Expression Omnibus) (http://www.ncbi.nlm.nih.gov/geo/) database based on the plat-

Revised manuscript accepted for publication May 28, 2015

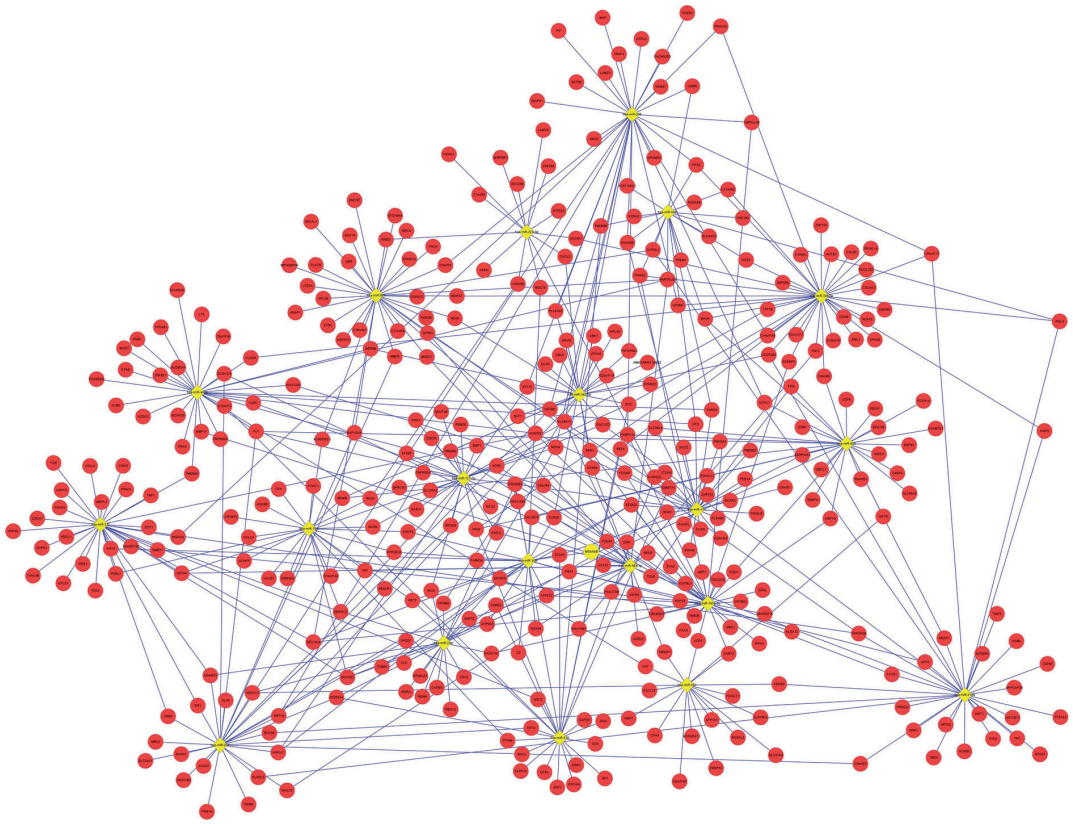

Figure 1. — Regulatory network of miRNAs and their target genes. Yellow nodes represent miRNAs; red nodes represent target genes related to miRNA.

form of GPL8179 (Human v2 microRNA expression beadchip) and GPL10558 (Human HT-12 V4.0 expression beadchip). Total 30 chips data from platforms of GPL8179 (15 chips data) and GPL10558 (15 chips data) were used for the development of the genome wide expression profiles of both miRNAs and mRNAs data, including eight CIN III samples data and seven normal cervix samples data, respectively. The data were normalized with robust multi-array average (RMA) [12] algorithm and subjected to logarithmic transformation.

Feature miRNAs and genes analysis

Probe sets were mapped to miRNAs and genes, respectively. Nonspecific probes were filtered. When multiple probe sets were mapped to the same miRNA or gene, the average expression value was calculated to represent the miRNA or gene expression level.

The feature miRNAs and genes between CIN samples and normal samples were analyzed. At first, interquartile range (IQR) [13] was used to filter miRNAs or genes based on gene expression levels distribution. All miRNAs or genes whose variability less than 1/5 overall IQR were eliminated. Then, ANOVA analysis was performed, and feature miRNAs or genes were filtered based on random forest [14].

TaLasso online analysis and miRNA-mRNA regulatory network

The TaLasso is both website (http://talasso.cnb.csic.es/) and the algorithm [15]. It has been tested with two datasets with matched miRNA and mRNA expression data.

The authors converted names of mRNAs and miRNAs which were screened from gene expression data into ensemble names. Then, data without ensemble name were deleted and the remaining data of mRNAs and miRNAs were put into TaLasso. The union of TarBase, miRecords, and miRWalk databases [15] were used for putative target genes. Cytoscape [16] is a software for visualizing complex networks and integrating networks between genes. With the application of Cytoscape software, the miRNA-mRNA regulatory network was constructed with these target genes.

Functional enrichment analysis

Gene Ontology (GO) database [17] is a collection of a large number of gene annotation terms. Kyoto Encyclopedia of Genes and Genomes (KEGG) knowledge database [18] is applied to identify the functional and metabolic pathway. Database for Annotation, Visualization and Integrated Discovery (DAVID) [19] is a gene functional enrichment analysis tool to understand the biological meaning for investigators. GO and KEGG pathway enrichment analysis were conducted with DAVID. A p-value < 0.05 was the cut-off criterion for the gene enrichment analysis.

Transcription factors and cancer related genes analysis

The Transcription Factors Database (TFD) [20] is a specialized database focusing on TFs and their properties. TFs, the target genes, were selected and identified from target genes based on TFD.

The tumor-associated gene (TAG) database [21] (http://www.binfo.ncku.edu.tw/TAG/) is designed to utilize information from well-characterized oncogenes and tumor suppressor genes to facilitate cancer research. Tumor suppressor genes (TSG) database [22] is a literature-based resource of tumor suppressor by integrating genomic data of mutations, gene expressions, regulations, methylations, and interactions. The cancer-promoting genes were extracted from TAG database, and tumor suppressor genes were extracted from TAG database and TSG database.

Table 1. — *Gene ontology analysis.*

GO category	GO ontology	GO term	Observed	*p*-value
GO:0005576	CC	Extracellular region	57	0.004191
GO:0048706	BP	Embryonic skeletal system development	7	0.004742
GO:0006959	BP	Humoral immune response	7	0.005376
GO:0048562	BP	Embryonic organ morphogenesis	9	0.005796
GO:0045087	BP	Innate immune response	9	0.0072

Note: CC: cellular component; BP: biological process.
Observed: number of the observed target genes in the category.

Results

Feature miRNAs and genes selection

Total 25 feature miRNAs were selected, including hsa-miR-338-5p, hsa-miR-193a-5p, hsa-miR-216b, hsa-miR-204, hsa-miR-21*, hsa-miR-887, hsa-miR-323-3p, hsa-miR-887, hsa-miR-1294, and so on.

Total 1,143 feature mRNAs were selected, including pre-B-cell leukemia transcription factor 1 (PBX1), laminin subunit gamma-2 (LAMC2), FBJ murine osteosarcoma viral oncogene homolog (FOS), chromodomain helicase DNA binding protein 1-like (CHD1L), chemokine (C-C motif) ligand 28 (CCL28), nicotinamide nucleotide transhydrogenase (NNT), nuclear receptor subfamily 0, group B, member 1 (NROB1), and so on.

TaLasso online analysis and miRNA-mRNA regulatory network

For putative target miRNAs and genes, 21 miRNAs and 361 mRNAs were gained. The miRNA-mRNA regulatory network was established with the 596 unique miRNA-target gene pairs (Figure 1). The degree of each miRNA in the network was then calculated and the top six were hsa-miR-338-5p (42), hsa-miR-193a-5p (38), hsa-miR-216b (37), hsa-miR-887 (35), hsa-miR-204 (35), and hsa-miR-21* (35).

Functional enrichment analysis

GO enrichment analysis were carried out for all target genes. The top five GO terms are shown in Table 1. The most significant term of cellular component (CC) was extracellular region (p = 0.004191), and that of biological process (BP) was embryonic skeletal system (p = 0.004742).

Transcription factors and cancer related genes analysis

In this study, 15 target genes, such as PBX1, homeobox D9 (HOXD9), homeobox B8 (HOXB8), and hepatocyte nuclear factor 4, gamma (HNF4G), were TFs. Five cancer-promoting genes including PBX1, LAMC2, FOS, CHD1L, and spalt-like transcription factor 4 (SALL4), as well as 14 tumor suppressor genes such as protein tyrosine phosphatase, non-receptor type 6 (PTPN6), microseminoprotein (MSMB), lactotransferrin (LTF), and histidine triad nucleotide binding protein 1 (HINT1) were detected.

Discussion

CIN is the leading cause of death among gynecological malignancies and represents the second-leading cause of cancer-related deaths in women worldwide [23]. Regardless of the fact that some genes have been reported in the progression of CIN or cervical cancer, there is a lack of detailed molecular pathogenesis mechanism. In this study, the authors identified the feature miRNAs and mRNAs between normal cervix samples and CIN III samples using bioinformatics analysis. Total 21 putative target miRNAs and 361 putative target genes were identified. TF analysis results showed that 15 TFs were associated with the regulation of CIN. Among these factors, *PBX1* and *LAMC2* were cancer-promoting genes. The miRNA-mRNA network was constructed, and miR-338-5p, miR-193a-5p, and miR-216b were hub nodes in this regulatory network.

PBX1 encodes a nuclear protein that belongs to the *PBX* homeobox family of transcriptional factors. *PBX* was a cofactor for HOX-class homeobox proteins [24]. Previous studies had shown that *HOX* and *PBX* genes were involved in oncogenic processes, such as chromatin binding, cell cycle control, proliferation, apoptosis, angiogenesis, and cell-cell communications [25-28]. Richard *et al.* reported that disrupting the interaction between HOX proteins and their co-factor PBX will retard tumour growth *in vivo* [24]. Thus, *HOX/PBX* interaction may be a potential target in cervical cancer therapy. In this study, *PBX1* was a feature gene and cancer-promoting gene, suggesting that it was possibly involved in oncogenic processes in the CIN. The miRNA-mRNA regulatory network results also showed that *PBX1* was regulated by miR-193a-5p which was the second hub node in the network. Previous study found that miR-193 regulated cell growth through the transforming growth factor-β (TGF-β) pathway by regulating Smad3 in glioma [29]. In addition, Chen *et al.* reported that miR-193b, another member of the miR-193 family, repressed cell proliferation and regulates cyclin D1 expression in melanoma [30]. Taking these factors into account, the present authors suggested that *PBX1* may play an important role in CIN development and miR-193a-5p targeting *PBX1* expression may be a critical event in CIN development. As well as *PBX1, LAMC2* is a cancer-promoting gene. It belongs to the laminin family, which is an epithelial basement membrane protein. It has been shown that LAMC2 was involved in a wide variety of biological processes including cell adhesion, differentiation, migration, and tumor invasion [31-33]. Immunohistochemical analysis revealed that

LAMC2 protein was highly expressed in cervix carcinomas [34] and was a marker to predict the risk of progression of CIN lesions [35]. Therefore, LAMC2 may be a key gene in CIN development.

Aside from miR-193a-5p, several other miRNAs such as miR-338 and miR-216b may be important in the development of CIN. It was reported that miR-338 suppressed the gastric cancer progression through PTEN-AKT signaling by targeting phosphatidylinositol-3,4,5-trisphosphat e-dependent Rac exchange factor 2 (P-REX2a) [36], as well as its role as a suppressor of the Smoothened-independent signaling pathways [37]. Furthermore, miR-216b was reported to suppress tumor growth and invasion by targeting KRAS (kirsten rat sarcoma viral oncogene homolog) in nasopharyngeal carcinoma [38]. Kim *et al.* showed that miR-216b promoted cellular senescence through the p53–$p21^{Cip1/WAF1}$ pathway in colon cancer [39]. In this study, miR-338 and miR-216b were hub nodes in miRNA-mRNA regulatory network. Thus, the present authors speculated that miR-338 and miR-216b may be key regulators in CIN development.

In order to predict potential target genes for CIN treatment, only CIN III and normal samples, not CIN II or CIN I, were analyzed. In conclusion, the genes of *PBX1* and *LAMC2* may play an important role in CIN development. These genes showed potential perspective in treatment of CIN. MiR-338 and miR-216 may be the key miRNAs in CIN development. They may be used to predict the risk of progression of CIN lesions. However, further experiments are still needed to confirm the present results.

Acknowledgements

This project was supported by Department of Education of Zhejiang Province, China (Grant No. Y201224994).

References

[1] Arbyn M., Castellsague X., De Sanjose S., Bruni L., Saraiya M., Bray F., Ferlay J.: "Worldwide burden of cervical cancer in 2008". *Ann. Oncol.,* 2011, *22*, 2675.

[2] Wentzensen N., Klug S.J.: "Early detection of cervical carcinomas". *Review Article DA Int.,* 2008, *105*, 617.

[3] Östör A.G.: "Natural history of cervical intraepithelial neoplasia: a critical review". *Int. J. Gynecol. Pathol.,* 1993, *12*, 186.

[4] Bjerre B., Eliasson G., Linell F., Soderberg H., Sjoberg N.: "Conization as only treatment of carcinoma in situ of the uterine cervix". *Am. J. Obstet. Gynecol.,* 1976, *125*, 143.

[5] Felekkis K., Touvana E., Stefanou C., Deltas C.: "microRNAs: a newly described class of encoded molecules that play a role in health and disease". *Hippokratia,* 2010, *14*, 236.

[6] Ju J.A., Huang Y.C., Lan S.H., Wang T. H., Lin P. C., Lee J. C., *et al.:* "Identification of colorectal cancer recurrence-related microRNAs". *Genomic Med. Biomark Health Sci.,* 2012, *4*, 19.

[7] Martinez I., Gardiner A., Board K., Monzon F., Edwards R., Khan S.: "Human papillomavirus type 16 reduces the expression of microRNA-218 in cervical carcinoma cells". *Oncogene,* 2007, *27*, 2575.

[8] Sano T., Oyama T., Kashiwabara K., Fukuda T., Nakajima T.: "Immunohistochemical overexpression of p16 protein associated with intact retinoblastoma protein expression in cervical cancer and cervical intraepithelial neoplasia". *Pathol Int.,* 1998, *48*, 580.

[9] Dobbs S., Hewett P., Johnson I., Carmichael J., Murray J.: "Angiogenesis is associated with vascular endothelial growth factor expression in cervical intraepithelial neoplasia". *Br. J. Cancer,* 1997, *76*, 1410.

[10] Srivastava P., Mangal M., Agarwal S.M.: "Understanding the transcriptional regulation of cervix cancer using microarray gene expression data and promoter sequence analysis of a curated gene set". *Gene,* 2014, *535*, 233.

[11] Zhang Z., Bast R.C., Yu Y., Li J., Sokoll L.J., Rai A.J., *et al.:* "Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer". *Cancer Res*., 2004, *64*, 5882.

[12] Irizarry R.A., Hobbs B., Collin F., Beazer-Barclay Y.D., Antonellis K.J., Scherf U., Speed T.P.: "Exploration, normalization, and summaries of high density oligonucleotide array probe level data". *Biostatistics,* 2003, *4*, 249.

[13] Bernardet U., Blanchard M., Verschure P.F.: "IQR: a distributed system for real-time real-world neuronal simulation". *Neurocomputing,* 2002, 44, 1043.

[14] Svetnik V., Liaw A., Tong C., Culberson J.C., Sheridan R.P., Feuston B.P.: "Random forest: a classification and regression tool for compound classification and QSAR modeling". *J. Chem. Inf. Comput. Sci*., 2003, *43*, 1947.

[15] Muniategui A., Nogales-Cadenas R., Vazquez M., Aranguren X.L., Agirre X., Luttun A., *et al.:* "Quantification of miRNA-mRNA interactions". *PLoS One,* 2012, *7*, e30766.

[16] Shannon P, Markiel A, Ozier O, Baliga N.S., Wang J.T., Ramage D., *et al.*: "Cytoscape: a software environment for integrated models of biomolecular interaction networks". *Genome Res* 2003;13:2498-504.

[17] Ashburner M, Ball CA, Blake J.A, *et al.:* "Gene Ontology: tool for the unification of biology". *Nat. Genet.,* 2000, *25*, 25.

[18] Altermann E., Klaenhammer T.R.: "PathwayVoyager: pathway mapping using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database". *BMC Genomics,* 2005, *6*, 60.

[19] Huang D.W., Sherman B.T., Tan Q., Kir J., Liu D., Bryant D., *et al.:* "DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists". *Nucleic Acids Res.,* 2007, *35*, W169.

[20] Ghosh D.: "Status of the transcription factors database (TFD)". *Nucleic Acids Res.,* 1993, *21*, 3117.

[21] Chen J.S., Hung W.S., Chan H.H., Tsai S.J., Sun H.S.: "In silico identification of oncogenic potential of fyn-related kinase in hepatocellular carcinoma". *Bioinformatics*, 2013, *29*, 420.

[22] Zhao M., Sun J., Zhao Z.: "TSGene: a web resource for tumor suppressor genes". *Nucleic Acids Res.,* 2013, *41*, D970.

[23] Brisson J., Morin C., Fortier M.: "Risk factors for cervical intraepithelial neoplasia: differences between low- and high-grade lesions". *Am. J. Epidemiol.,* 1994, *140*, 700.

[24] Morgan R., Plowright L., Harrington K.J., Michael A., Pandha H.S.: "Targeting HOX and PBX transcription factors in ovarian cancer". *BMC Cancer,* 2010, *10*, 89.

[25] Cillo C., Cantile M., Faiella A., Boncinelli E.: "Homeobox genes in normal and malignant cells". *J. Cell. Physiol.,* 2001, *188*, 161.

[26] Pando S.M., Taylor H.S.: "Homeobox gene expression in ovarian cancer". *In:* Sharon Stack M., Fishman D.A. (eds). *Ovarian cancer*. New York: Springer, 2002, 231.

[27] Sarno J.L., Kliman H.J., Taylor H.S.: "HOXA10, Pbx2, and Meis1 protein expression in the human endometrium: formation of multimeric complexes on HOXA10 target genes". *J. Clin. Endocr. Metab.,* 2005, *90*, 522.

[28] Bhatlekar S., Fields J.Z., Boman B.M.: "HOX genes and their role in the development of human cancers". *J. Mol. Med. (Berl.),* 2014, *92*, 811.

[29] Zhong Q., Wang T., Lu P., Zhang R., Zou J., Yuan S.: "miR-193b promotes cell proliferation by targeting Smad3 in human glioma". *J. Neurosci. Res.,* 2014, *92*, 619.

[30] Chen J., Feilotter H.E., Paré G.C., Zhang X., Pemberton J.G., Garady C., *et al.:* "MicroRNA-193b represses cell proliferation and regulates cyclin D1 in melanoma". *Am. J. Pathol.,* 2010, *176*, 2520.
[31] Garg M., Braunstein G., Koeffler H.P.: "LAMC2 as a therapeutic target for cancers". *Expert Opin. Ther. Targets,* 2014, *18*, 979.
[32] Kosanam H., Prassas I., Chrystoja C.C., Soleas I., Chan A., Dimitromanolakis A., *et al.:* "Laminin, gamma 2 (LAMC2): a promising new putative pancreatic cancer biomarker identified by proteomic analysis of pancreatic adenocarcinoma tissues". *Mol. Cell. Proteomics,* 2013, *12*, 2820.
[33] Garg M., Kanojia D., Okamoto R., Jain S., Madan V., Chien W., *et al.:* "Laminin-5γ-2 (LAMC2) is highly expressed in anaplastic thyroid carcinoma and is associated with tumor progression, migration, and invasion by modulating signaling of EGFR". *J. Clin. Endocr. Metab.,* 2013, *99*, E62.
[34] Skyldberg B., Salo S., Eriksson E., Aspenblad U., Moberger B., Tryggvason K., Auer G.: "Laminin-5 as a marker of invasiveness in cervical lesions". *J. Natl. Cancer Inst.,* 1999, *91*, 1882.
[35] Kohlberger P., Beneder C., Horvat R., Leodolter S., Breitenecker G.: "Immunohistochemical expression of laminin-5 in cervical intraepithelial neoplasia". *Gynecol. Oncol.,* 2003, *89*, 391.
[36] Guo B., Liu L., Yao J., Ma R., Chang D., Li Z., *et al.:* "miR-338-3p Suppresses Gastric Cancer Progression through a PTEN-AKT Axis by Targeting P-REX2a". *Mol. Cancer Res.,* 2014, *12*, 313.
[37] Xue Q., Sun K., Deng H.J., Lei S.T., Dong J.Q., Li G.X.: "MicroRNA-338-3p Inhibits Colorectal Carcinoma Cell Invasion and Migration by Targeting Smoothened". *Jpn. J. Clin. Oncol.,* 2014, *44*, 13.
[38] Deng M., Tang H., Zhou Y., Zhou M., Xiong W., Zheng Y., *et al.:* "miR-216b suppresses tumor growth and invasion by targeting KRAS in nasopharyngeal carcinoma". *J. Cell. Sci.,* 2011, *124*, 2997.
[39] Kim S.Y., Lee Y.H., Bae Y.S.: "miR-186, miR-216b, miR-337-3p, and miR-760 cooperatively induce cellular senescence by targeting subunit of protein kinase CKII in human colorectal cancer cells". *Biochem. Biophys. Res. Commun.,* 2012, *429*, 173.

Address reprint requests to:
H. JIN, M.D.
Department of Gynecology
Women's Hospital, School of Medicine
Zhejiang University
No.1 Xueshi Road
Hangzhou 310006 (China)
e-mail: hangmeijinhmj@163.com