

## Original Article

# LncRNA expression profiles reveal the co-expression network in human colorectal carcinoma

Jun Liu\*, Shuangjie Wu\*, Mengjun Li, Xinhai Wang, Yifan Tang

Department of General Surgery, Huashan Hospital Affiliated to Fudan University, 12 Middle Urumqi Road, Shanghai 200040, P. R. China. \*Equal contributors.

Received October 16, 2015; Accepted November 28, 2015; Epub February 1, 2016; Published February 15, 2016

**Abstract:** Colorectal carcinoma (CRC) is the most common cancer of the gastrointestinal tract with high prevalence, and the poor prognosis of this disease lead to the high mortality rates. The aim of this study was to characterize the mechanism of CRC tumorigenesis and progression, and to screen the potential biomarker for CRC diagnosis. Expression profiles of mRNA and LncRNA of CRC tumor samples and its para-tumor samples were obtained with microarray technology. Co-expression network reveals several genes, such as CEACAM1, DHRS11, HNF4A, HGRLOS, MUC4 and TTN, were related to CRC, these genes play important roles in the progression of CRC and some have been suggested association with tumor cell transformation. Furthermore, significance analysis of all the gene expression profiling was performed. Functional annotation analysis reveals that differentially expressed genes were related to immune pathway and p53 signaling pathway which includes TP53 and CDKN2A. GO enrichment analysis reveals that DEGs were involved in kinase activity categories and epidermal growth factor. In total, our study described the transcriptome profile characteristic and providing clues for CRC researches.

**Keywords:** LncRNA, co-expression network, colorectal carcinoma

## Introduction

Colorectal carcinoma (CRC) is the third leading cause of cancer death worldwide [1, 2]. It accounts for about 80% of primary CRC cases and approximately two thirds of them are diagnosed at an advanced stage [2]. The poor prognosis of this disease is partially due to the lack of an effective means of early diagnosis. Discovery of an effective and reliable tool for early diagnosis of CRC would play a pivotal role in improving the prognosis of patients with CRC. However, the molecular pathogenesis of CRC remains to be elucidated.

Long non-coding RNAs (long ncRNAs, lncRNA) are non-protein coding transcripts longer than 200 nucleotides which can localize and function in the nucleolus, nucleus and in the cytoplasm [3]. It has been identified that the majority of protein-coding genes have antisense partners, including many tumour suppressor genes that are frequently silenced by epigenetic mechanisms in cancer [4]. A detailed analysis showed the p15 antisense ncRNA (CDKN2BAS) was able to induce changes to heterochromatin

and DNA methylation status of p15 by an unknown mechanism, thereby regulating p15 expression. Moreover, lncRNA dysregulation was involved in the development and progression of a variety of types, such as gastric cancer [5], breast cancer [6] and prostate cancer [7]. The lncRNAs are identified as tumor driving oncogenic lncRNAs and tumor suppressor. They play an important role in cell growth, apoptosis, cell migration and invasiveness as well as cancer cell stemness. In recent studies, lncRNAs could be used as a new prognostic marker to predict clinical outcomes of cancer patients [8]. Therefore, it is intriguing to analyze lncRNA expression profile, mRNA expression pattern and their regulation network in human colorectal carcinoma.

LncRNA and mRNA expression profile array is a powerful method to identify biological markers that has been applied in many cancer researches including lung cancer [9], glioma [10] and prostate cancer [11, 12]. It allows us to compare the whole-transcriptome expression profiles between tumor and normal sample pairs to reveal the alteration of lncRNA and mRNA pro-

files related to tumor genesis of CRC. Here we adopted this method to investigate lncRNA and mRNA expression profile in CRC, differentially expression genes and GO and pathway enrichment analysis was performed. Furthermore, lncRNA and mRNA co-expression network in CRC was constructed. Our study provided a clue for prognostic marker prediction research in CRCs.

## Materials and methods

### *Patients and total RNA preparation*

lnc-RNA and mRNA expression of tumor and adjacent non-tumorous tissues was investigated in 3 participants. Written informed consents were obtained from all subjects before recruitment. Patients were diagnosed based on the clinical manifestations, endoscopy, and pathology, confirmed by two gastroenterologists. Colonic biopsy specimens were obtained from all subjects during surgical operation. Total RNA from patients' colonic tumor and para-tumor mucosa were extracted by Total RNA Extraction Kit (SLNco, Cinoasia, China). All RNA samples were examined by Bioanalyzer 2100 QC (Agilent, USA) with RIN > 7.

### *lnc-RNA and mRNA expression profiling*

Three para-tumor/tumor pairs of CRC lnc-RNA and mRNA data were obtained by Human 8 × 60K lncRNA microarray (Arraystar, USA), raw data was normalized with robust Multi-array Analysis (RMA) algorithm for background correction and expression level summarisation, then normalized data was log<sub>2</sub> scaled which enable the comparison between para-tumor and tumor samples. Data pre-processing, Pearson correlation coefficient, and Hierarchical clustering analysis calculated by "ward" method were performed in the R ([www.r-project.org/](http://www.r-project.org/)) environment with its "base" function and "stat" packages.

### *Differential expression analysis*

Differential expression of genes between tumor samples and matched para-tumor samples were performed by calculating fold changes using the normalized value of each transcript between tumor and matched para-tumor samples, and statistical significance of differentially expressed genes was presented by calculating a t test *p*-value. Then, significance of a dif-

ferentially expressed lncRNA/mRNA between two samples was determined according the threshold of fold change larger than 2 and *p*-value was less than 0.01.

### *KEGG pathway enrichment analysis of differentially expressed genes*

Up-regulated genes and down-regulated genes identified by comparing gene expression between tumor and para-tumor samples were used to query the KEGG pathway database to determine the biological function of these DEGs. Enriched pathway was determined by both significant fisher exact test (*p*-value < 0.05), and at least 3 differentially expressed genes were involved in the pathway. The pathway enrichment analysis was performed by using "KEGG.db" and "KEGGprofile" packages in R project.

### *Co-expression gene network analysis*

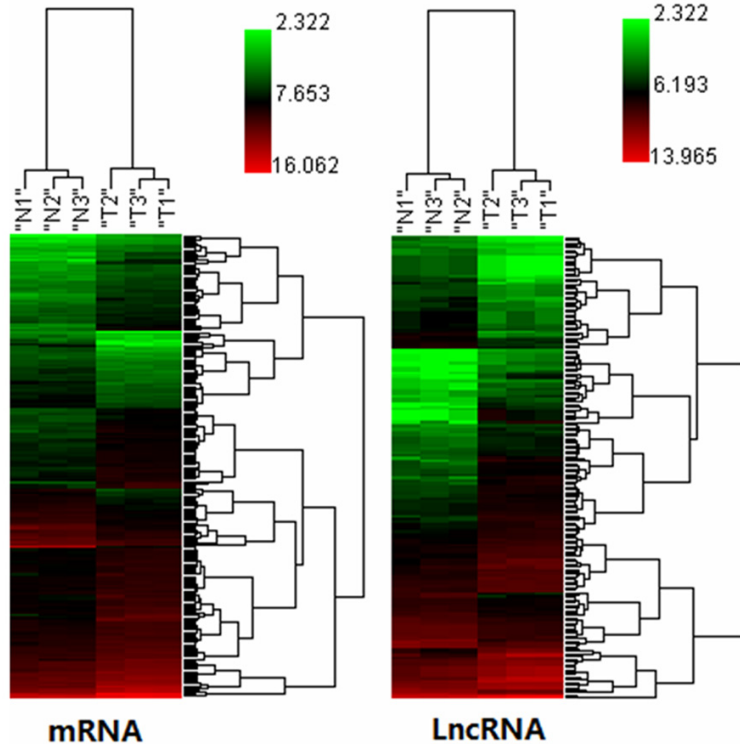
Gene co-expression networks were built according to the normalized signal intensity of specific expressed genes. In order to get rid of genes with low information content a combined filter based on between-sample variability and gene minimal signal was used. The filter leaves out only those gene probe sets that fulfilled both of the two following conditions: 1<sup>st</sup>: Genes which have an expression difference or variability between samples (DExp<sub>g</sub> highest-lowest) lower than the median of all the expression differences calculated for each gene; 2<sup>nd</sup>: Genes which have a mean expression signal between samples (mean Exp samples) lower than the median of all the expression signals calculated for each gene. We constructed the network adjacency between two genes, *i* and *j*, defined as a power of the Pearson correlation between the corresponding gene expression profiles. By computing the correlation coefficient of these genes, we obtained the gene-gene co-expression adjacency matrix, *M* (*i*, *j*) that only genes with the strongest correlations (0.99 or greater) were selected to drawn in co-expression network graphs.

## Results

### *Gene expression profiling of colon cancer*

The expression of 26,109 protein coding transcripts and 30,586 lncRNAs of colon cancer were detected with Human lncRNA Microarray

## LncRNA and mRNA co-expression network in human colorectal carcinoma



**Figure 1.** Differential expression of lncRNAs and mRNAs in CRC and hierarchical clustering analysis for of 254 mRNAs (left) and 174 lncRNAs (right) that were differentially expressed between CRC samples (ca, cancer tissues) and nontumor samples (greater than 2.0-fold;  $P < 0.01$ ). Expression values are represented in colors of red and green which indicates expression above and below the median expression value across all samples (log scale 2), respectively.

(V3.0). Gene expression fold change and t test were performed by comparing cancer group with the control group. Differentially expressed transcripts (DET) were selected according to the cutoff of t test  $p$  value less than 0.05 combined with fold change larger than 1.5. Totally, 224 mRNA and 180 lncRNA were identified as differentially expressed transcripts. The transcripts expression change range from 2.0 to 13.0 folds for mRNA and change from 2.0 to 13.4 folds for lncRNA expression. Both the number and range of change of mRNA and lncRNA expression between tumor and normal are very similar. The heat map in **Figure 1** shows the hierarchical clustering analysis of differentially expression genes in tumor and normal samples.

### Pathway enrichment analysis and GO annotation with DETS

In order to reveal the biological process changes between tumor and normal samples, the

pathway enrichment analysis and GO annotation were performed. The 224 differentially expressed transcripts were using to query the KEGG database and GO annotation database, significantly enriched KEGG database with fisher exact  $p$  value  $< 0.05$  were listed in **Table 1**. The top enriched pathway is primary immunodeficiency, and T cell B cell receptor signally pathways were also enriched which may indicates the immunology changes are related to colon tumorigenesis. CD30, CD3E, CD72, CD2 and IL7R genes were involved in these immune pathways. In addition, TP53 and CDKN2A involved in TP53 signaling pathway were significantly affected. Cell adhesion molecules signaling pathway play a pivotal role in the development of recurrent invasive, and distant metastasis of cancer was significantly affected, HLA-DOB and HLA-F were involved in the pathway changes. The results of molecular function of GO annotation

were listed in **Table 2**. Most of differentially expressed mRNAs are belong to the kinase activity categories. Several genes also related to epidermal growth factor (GO:0005154) and transcription factor (GO:0003700). The function annotation analysis reveals the relationship between gene expression alteration of immune and cancer pathway and tumorigenesis.

### Co-expression analysis reveal gene interaction in colon cancer

Gene co-expression analysis assumes that genes with similar expression patterns are hypothesized to have similar functions or involved in same biological process. In order to reveal the interaction between mRNA and lncRNA in colon cancer, we constructed three co-expression networks. Co-expression networks could cluster thousands of transcript into a functional modules based on the correlation of gene expression. Co-expression modules

**Table 1.** Pathway enrichment analysis of mRNA

Pathway	Genes	P Value
Primary immunodeficiency	DCLRE1C, CD3D, CD3E, IL7R, BTK	2.25E-03
Hematopoietic cell lineage	CD3D, CD3E, IL4R, CD2, ITGA4, IL7R	1.21E-02
Cell adhesion molecules (CAMs)	CD2, CTLA4, CLDN2, ITGA4, SELPLG, HLA-DOB, HLA-F	1.91E-02
p53 signaling pathway	TP53I3, CDKN2A, RPRM, TP53, SFN	2.34E-02
T cell receptor signaling pathway	PTPN6, CD3D, CD3E, CTLA4, MALT1, VAV2	2.95E-02
B cell receptor signaling pathway	PTPN6, MALT1, CD72, VAV2, BTK	3.22E-02

may relate to biological function. As shown in **Figure 2**, each gene corresponds to a node. Two genes are connected by an edge, indicating a strong correlation (i.e., either positive or negative). Within the network analysis, we focused on the genes that are interacted with 5 more other genes. CEACAM1, DHRS11, HNF4A, HGRLOS, FAM13C and UTY are hub genes in the network.

**Discussion**

The abnormal expression of mRNA and lncRNA is associated with many cancer types, and in this study, we investigated the significantly differential expressed mRNA and lncRNA in human colon cancer, compared to the corresponding normal sample. Our results reveal the DEGs are related to immune system and cancer bioprocess. System biology approaches allow us to investigate the co-expression networks between lncRNA and mRNA. A small group of lncRNA and mRNA are aberrantly expressed in human CRC.

For CRC, the dysfunction of immune system and activation of cancer related pathways may contribute to the tumorigenesis and progression. We found that hundreds of genes are abnormal expression in tumor samples, which includes many immune factors, likes CD3D, DCLRE1C, IL7R, BTK and CD3E which are involved in primary immunodeficiency. T cell and B cell signaling pathways are also significantly affected. Our results were in agreement with recent studies that the alteration of immune system may play an important role in CRC [13] and immune factors may promote tumor growth, progression and metastasis [14]. In addition, TP53, CDKN2A are key elements of TP53 signaling pathway are also suggested to be differentially expressed in CRC tissues, which indicates the activation of cancer pathways in CRC. TP53 has been reported as

an independent and strong predictor of survival [15].

Previous study shown the lncRNA high expression in many cancer types, and several lncRNA-mRNA co-expression networks were identified [16]. HNF4α is a transcription factor that plays a key role in cancer differentiation, our study show it is coexpressed with DHRS11 and CEACAM1. The association of CEACAM1 expression and tumor cell transformation has been reported by several studies [17-20]. It creates a pro-angiogenic tumor microenvironment that supports tumor vessel maturation. We also found the MUC4 and MUC12 genes were coexpressed with many other genes. Several studies have shown the aberrant expression profile of mucins in different malignancies, suggesting that mucins have a great potential to be used as a diagnostic and prognostic marker in gastric cancer. In addition, GHRLOS is a non-protein coding RNA; it is coexpressed with TTN and FAM13C. Finally, we found the UTY is coexpressed with many genes in CRC tissues. UTY gene codes for an HLA-B60-restricted human male-specific minor histocompatibility antigen involved in stem cell graft rejection, characterization of the critical polymorphic amino acid residues for T-cell recognition, it is important to cell immune [21-23]. The function of these co-expression networks will be further studied in our future experiment researches.

Overall, we provided an insight of lncRNA expression associate with complexes to regulate gene expression in CRC. The progression of CRC is associated with immune system and lncRNA and mRNA dysregulation.

**Acknowledgements**

This project is supported by 2013 Science Foundation from Science and Technology

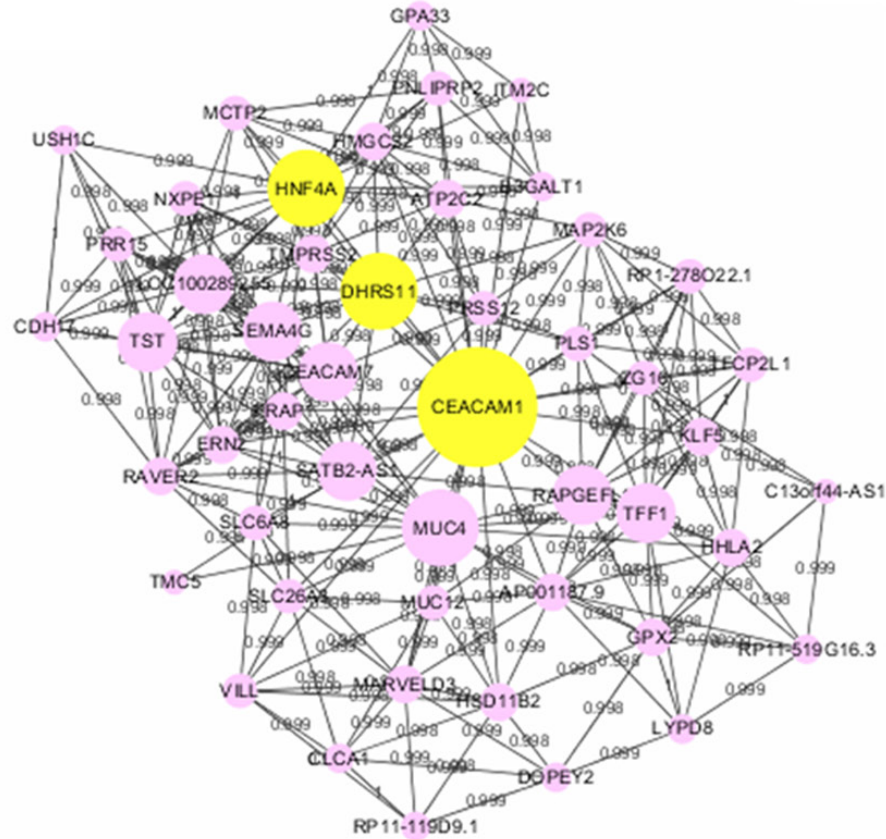
## LncRNA and mRNA co-expression network in human colorectal carcinoma

**Table 2.** GO annotation of mRNA

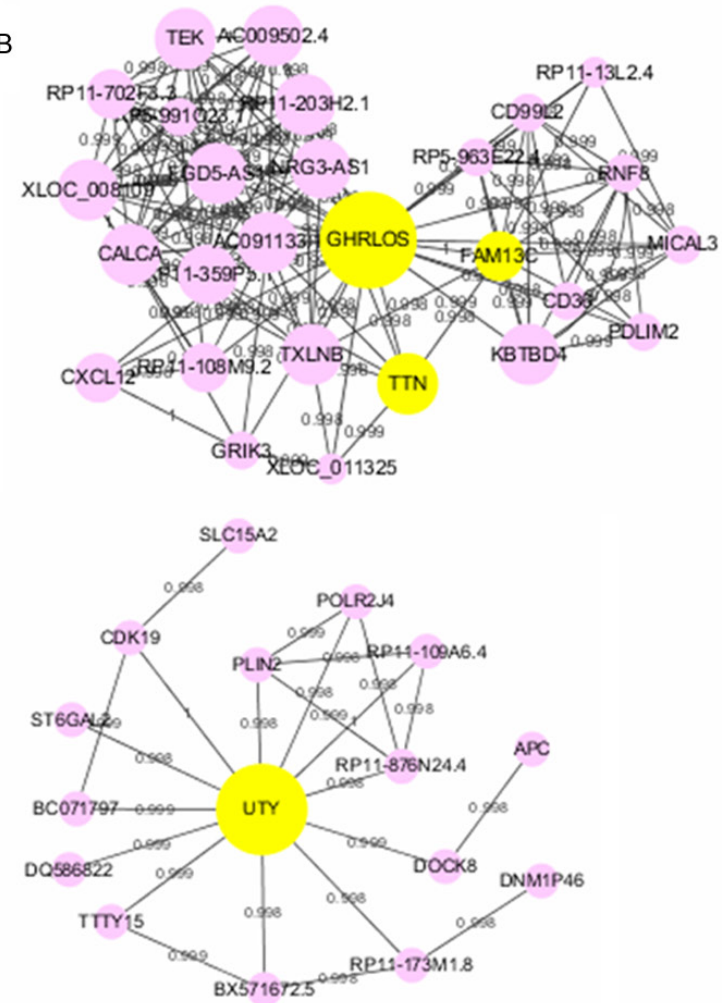
Term	Genes	P value
GO:0019207~kinase regulator activity	CDKN2A, TAOK3, MALT1, SFN, EGF, NRG1, RHOH	7.70E-04
GO:0046983~protein dimerization activity	THRA, CD3D, GZMA, CD3E, TP53, ASNS, APOBEC3F, FOXP1, TP53I3, ZNF862, ATP2A1, JUND, CD2, CDA, NKX2-5	4.55E-03
GO:0019210~kinase inhibitor activity	CDKN2A, TAOK3, SFN, RHOH	9.14E-03
GO:0005154~epidermal growth factor receptor binding	HBEGF, EGF, VAV2	1.13E-02
GO:0019887~protein kinase regulator activity	CDKN2A, TAOK3, SFN, EGF, NRG1	1.50E-02
GO:0042803~protein homodimerization activity	TP53I3, THRA, GZMA, ATP2A1, CD2, CDA, ASNS, APOBEC3F, NKX2-5, FOXP1	1.68E-02
GO:0042802~identical protein binding	UPF2, THRA, GZMA, ASNS, ITGA4, KCNIP2, APOBEC3F, FOXP1, BTK, TP53I3, ATP2A1, CD2, CLDN2, CDA, NKX2-5	1.80E-02
GO:0003700~transcription factor activity	PPARD, ERG, THRA, TBX20, TP53, FOXP1, ASCL1, CDKN2A, MEIS2, REL, TFEC, JUND, IRF1, IRF2, ZNF394, BHLHE41, TBX18, NKX2-5, PITX2, ETV3	1.94E-02
GO:0003950~NAD+ ADP-ribosyltransferase activity	ART3, PARP15, PARP8	3.43E-02
GO:0019209~kinase activator activity	MALT1, EGF, NRG1	3.43E-02

LncRNA and mRNA co-expression network in human colorectal carcinoma

A



B



**Figure 2.** Co-expression network of mRNA and lncRNA expression. Node with interaction larger than 10 was colored in yellow. The node size was corresponding to the number of interaction.

Commission of Shanghai Municipality (13ZR1-405400).

**Disclosure of conflict of interest**

None.

**Address correspondence to:** Dr. Yifan Tang, Department of General Surgery, Huashan Hospital Affiliated to Fudan University, 12 Middle Urumqi Road, Shanghai 200040, P. R. China. Tel: +86-181-2118-6477; E-mail: tangyifan618@126.com

**References**

- [1] Baum B and Hewicker-Trautwein M. Classification and Epidemiology of Mammary Tumours in Pet Rabbits (*Oryctolagus cuniculus*). *J Comp Pathol* 2015; 152: 291.
- [2] Everhart JE and Ruhl CE. Burden of digestive diseases in the United States part II: lower gastrointestinal diseases. *Gastroenterology* 2009; 136: 741.
- [3] Zhang K, Shi ZM, Chang YN, Hu ZM, Qi HX, Hong W. The ways of action of long non-coding RNAs in cytoplasm and nucleus. *Gene* 2014; 547: 1.
- [4] Zhao Q, Shi X, Xie Y, Huang J, Shia B, Ma S. Combining multidimensional genomic measurements for predicting cancer prognosis: observations from TCGA. *Brief Bioinform* 2015; 16: 291.
- [5] Sun M, Nie FQ, Wang ZX and De W. Involvement of lncRNA dysregulation in gastric cancer. *Histol Histopathol* 2016; 31: 33-9.
- [6] Liu Y, Sharma S and Watabe K. Roles of lncRNA in breast cancer. *Front Biosci (Schol Ed)* 2015; 7: 94.
- [7] Malik R, Patel L, Prensner JR, Shi Y, Iyer MK, Subramaniyan S, Carley A, Niknafs YS, Sahu A, Han S, Ma T, Liu M, Asangani IA, Jing X, Cao X, Dhanasekaran SM, Robinson DR, Feng FY, Chinnaiyan AM. The lncRNA PCAT29 inhibits oncogenic phenotypes in prostate cancer. *Mol Cancer Res* 2014; 12: 1081.
- [8] Tian X and Xu G. Clinical value of lncRNA MALAT1 as a prognostic marker in human cancer: systematic review and meta-analysis. *BMJ Open* 2015; 5: e8653.
- [9] Wang J, Zhou Y, Lu J, Sun Y, Xiao H, Liu M, Tian L. Combined detection of serum exosomal miR-21 and HOTAIR as diagnostic and prognostic biomarkers for laryngeal squamous cell carcinoma. *Med Oncol* 2014; 31: 148.
- [10] Li R, Qian J, Wang YY, Zhang JX and You YP. Long noncoding RNA profiles reveal three molecular subtypes in glioma. *CNS Neurosci Ther* 2014; 20: 339.
- [11] Malik R, Patel L, Prensner JR, Shi Y, Iyer MK, Subramaniyan S, Carley A, Niknafs YS, Sahu A, Han S, Ma T, Liu M, Asangani IA, Jing X, Cao X, Dhanasekaran SM, Robinson DR, Feng FY, Chinnaiyan AM. The lncRNA PCAT29 inhibits oncogenic phenotypes in prostate cancer. *Mol Cancer Res* 2014; 12: 1081.
- [12] Xue Y, Wang M, Kang M, Wang Q, Wu B, Chu H, Zhong D, Qin C, Yin C, Zhang Z, Wu D. Association between lncrna PCGEM1 polymorphisms and prostate cancer risk. *Prostate Cancer Prostatic Dis* 2013; 16: 139-44, S1.
- [13] Cui G, Shi Y, Cui J, Tang F and Florholmen J. Immune microenvironmental shift along human colorectal adenoma-carcinoma sequence: is it relevant to tumor development, biomarkers and biotherapeutic targets? *Scand J Gastroenterol* 2012; 47: 367.
- [14] Liu H, Zhang Z, Tabuchi T, Wang S and Wang J. The role of pro-inflammatory cytokines and immune cells in colorectal carcinoma progression. *Oncol Lett* 2013; 5: 1177.
- [15] Rechsteiner M, von Teichman A, Rüschoff JH, Fankhauser N, Pestalozzi B, Schraml P, Weber A, Wild P, Zimmermann D, Moch H. KRAS, BRAF, and TP53 deep sequencing for colorectal carcinoma patient diagnostics. *J Mol Diagn* 2013; 15: 299.
- [16] Yang F, Zhang L, Huo XS, Yuan JH, Xu D, Yuan SX, Zhu N, Zhou WP, Yang GS, Wang YZ, Shang JL, Gao CF, Zhang FR, Wang F, Sun SH. Long noncoding RNA high expression in hepatocellular carcinoma facilitates tumor growth through enhancer of zeste homolog 2 in humans. *Hepatology* 2011; 54: 1679.
- [17] Yang C, He P, Liu Y, He Y, Yang C, Du Y, Zhou M, Wang W, Zhang G, Wu M, Gao F. Assay of serum CEACAM1 as a potential biomarker for breast cancer. *Clin Chim Acta* 2015; 450: 277.
- [18] Yamamoto N, Yokoyama S, Ieda J, Mitani Y, Yamaguchi S, Takifuji K, Hotta T, Matsuda K, Watanabe T, Shively JE, Yamaue H. CEACAM1 and hollow spheroid formation modulate the chemosensitivity of colorectal cancer to 5-fluorouracil. *Cancer Chemother Pharmacol* 2015; 75: 421.
- [19] Shi JF, Xu SX, He P and Xi ZH. Expression of carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1) and its correlation with angiogenesis in gastric cancer. *Pathol Res Pract* 2014; 210: 473.
- [20] Arabzadeh A and Beauchemin N. Stromal CEACAM1 expression regulates colorectal cancer metastasis. *Oncoimmunology* 2012; 1: 1205.
- [21] Warren EH, Gavin MA, Simpson E, Chandler P, Page DC, Distechi C, Stankey KA, Greenberg PD, Riddell SR. The human UTY gene encodes

## LncRNA and mRNA co-expression network in human colorectal carcinoma

- a novel HLA-B8-restricted H-Y antigen. *J Immunol* 2000; 164: 2807.
- [22] Vogt MH, Goulmy E, Kloosterboer FM, Blokland E, de Paus RA, Willemze R, Falkenburg JH. UTY gene codes for an HLA-B60-restricted human male-specific minor histocompatibility antigen involved in stem cell graft rejection: characterization of the critical polymorphic amino acid residues for T-cell recognition. *Blood* 2000; 96: 3126.
- [23] Ivanov R, Hol S, Aarts T, Hagenbeek A, Slager EH, Ebeling S. UTY-specific TCR-transfer generates potential graft-versus-leukaemia effector T cells. *Br J Haematol* 2005; 129: 392.