Pathol. Oncol. Res. (2018) 24:129–134
DOI 10.1007/s12253-017-0223-5

ORIGINAL ARTICLE

COL1A2 is a Novel Biomarker to Improve Clinical Prediction in Human Gastric Cancer: Integrating Bioinformatics and Meta-Analysis

Li Rong[1] · Wei Huang[2] · Shangkun Tian[1] · Xiangbo Chi[1] · Pan Zhao[1] · Fengfeng Liu[1]

Received: 12 October 2016 / Accepted: 3 April 2017 / Published online: 11 April 2017

Abstract Gastric cancer is the third most common cause of cancer-related death in worldwide. It is crucial to target the key genes controlling pathogenesis in the early stage of gastric cancer. This study describes an integrated bioinformatics to identify molecular biomarkers for gastric cancer in patients' cancer tissues. We reports differently expression genes in large gastric cancer cohorts from Gene Expression Ominus (GEO). Our findings revealed that 433 genes were significantly different expressed in human gastric cancer. Differently expression gene profile in gastric cancer was further validated by bioinformatic analyses, co-expression network construction. Based on the co-expression network and top-ranked genes, we identified collagen type I alpha 2 (COL1A2) which encodes the pro-alpha2 chain of type I collagen whose triple helix comprises two alpha1 chains and one alpha2 chain, was the key gene in a 37-gene network that modulates cell motility by interacting with the cytoskeleton. Furthermore, the prognostic role of COL1A2 was determined by use of immunohistochemistry on human gastric cancer tissue. COL1A2 was highly expressed in human gastric cancer as compared with normal gastric tissues. Statistical analysis showed COL1A2 expression level was significantly associated with histological type and lymph node status. However, there were no correlations between COL1A2 expression and age, lymph node numbers, tumor size, or clinical stage. In conclusion, the novel bioinformatics used in this study has led to identification of improving diagnostic biomarkers for human gastric cancer and could benefit further analyses of the key alteration during its progression.

Keywords gastric cancer · biomarker · bioinformatics · COL1A2

Introduction

Gastric cancer is the third most common cause of cancer-related death in the world, and it remains difficult to cure in worldwide, primarily because most patients present with advanced disease [1–3]. In the United States, stomach malignancy is currently the 15th most common cancer [4–6]. Gastric cancer may spread from the stomach to other parts of the body, particularly the liver,lungs, bones, lining of the abdomen and lymph nodes. The most common cause is infection by the bacterium Helicobacter pylori, which accounts for more than 60% of cases [7–10]. Early gastric cancer has no associated symptoms; however, some patients with incidental complaints are diagnosed with early gastric cancer. Gastric cancer is commonly diagnosed at an advanced stage and those patients with advanced disease 5-year survival rate are 24% [11–13]. All physical signs in gastric cancer are late events. By the time they develop, the disease is almost invariably too far advanced for curative procedures.

Recent gene expression microarray studies have identified a number of gene biomarkers of different types for human cancer. For example, researchers believe that mutations in the genes *BRCA1* and *BRCA2* cause as many as 60% of all

✉ Li Rong
lirongcq@126.com

[1] Department of gastroenterology, Chongqing Infectious Disease Medical Center, 1#Huangjuewan, Xiaolongkan, Shapingba District, Chongqing 400030, People's Republic of China

[2] Health Center of Southwest University of Political Science and Law, Chongqing 401120, People's Republic of China

cases of hereditary breast and ovarian cancers. In this study, we used gene expression microarray to detail the gastric cancer program of gene expression in different grade gastric cancer tissues and try to find out some genes associated with the tumorigenesis of gastric cancer [14–16]. Firstly, we mined published gene microarray data from the Gene Expression Omnibus (GEO), and explored differently expressed genes profiling. The use of this powerful technology has resulted in the identification of a novel biomarker for human gastric cancer, COL1A2, which plays an important role in the progression of human gastric cancer.

Materials and Methods

Microarray Data Processing and Gene Expression Profile Mining

The gastric cancer of microarray data and corresponding clinical data used were downloaded from the publicly GEO databases (http://www.ncbi.nlm.nih.gov/geo/): GSE27342 and GSE31789. The raw files of background were adjusted using Robust Multichip Average. Expression Console 1.4.1 and Transcriptome Analysis Console v3.0 was used for microarray analysis. Median levels of transcript expressions were calculated. Gene-level data was then filtered to include only those probe sets with annotations. Probe set annotation mainly reference the new version annotation files (*.transcript.csv) that can be download on affymetrix official website (http://www.affymetrix.com/support/technical/annotationfilesmain.affx) The CSV files contain both design-time data, such as the genomic location.

Cluster and co-Expression Network Analysis

Hierarchical Clustering tab was next performed as described in reference [17, 18]. We utilize the express profile of difference genes to construct the gene co-expression network. This network distinctly reveal the relation between the genes, find the key gene of regulation and "interaction venation" thoroughly. The co-expression network assimilate the scale-free property of the huge data, and to simulate the scale-free relation by the interaction between genes.

Immunohistochemistry Staining and Evaluation

Seventy-five paraffin-embedded human gastric cancer tissues were cut into 4 μm thick and 2 mm diameter sections to construct tissue microarrays, and then were subjected to heat-induced epitope retrieval in 0.01 mM citrate buffer (pH 6.0). Endogenous peroxidase activity was blocked in 3% H2O2 in citrate buffer for 20 min. The rabbit polyconal antibody to COL1A2 (14695–1-AP;Proteintech) was applied to the sections at a dilution of 1:200 and incubated at 4°Covernight. Immunohistochemical staining process was preformed using with 3, 3′-diaminobenzidine for 60s.The final immunoreactive scores were evaluated independently by two pathologists blinded to the clinical parameters, the scores were determined by the sum of intensity and extent of the staining. The intensity of staining was scored as 0(negative),1(weakly positive),2(moderately positive),and 3(strongly positive). According to the percentage of the positive staining area, the extent of staining was scored as 0(0–10%),1(11%–30%),2(31%–50%), 3(51%–70%), and 4(71%–100%). The

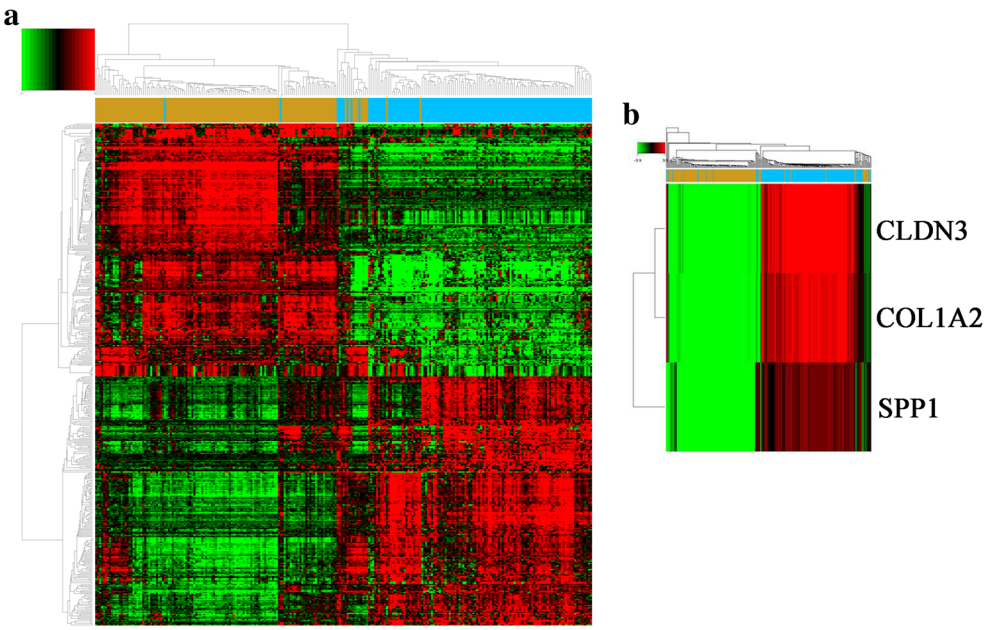

Fig. 1 Gene expression profile of human gastric cancer. (**a**) Heat map hierarchical clustering reveals 433 genes that were differentially expressed in gastric cancer tissues compared with stomach tissues. (**b**) Ranked the first three over-expressed genes in gastric cancer. Up-regulation or down-regulation gene is represented as red or green, respectively

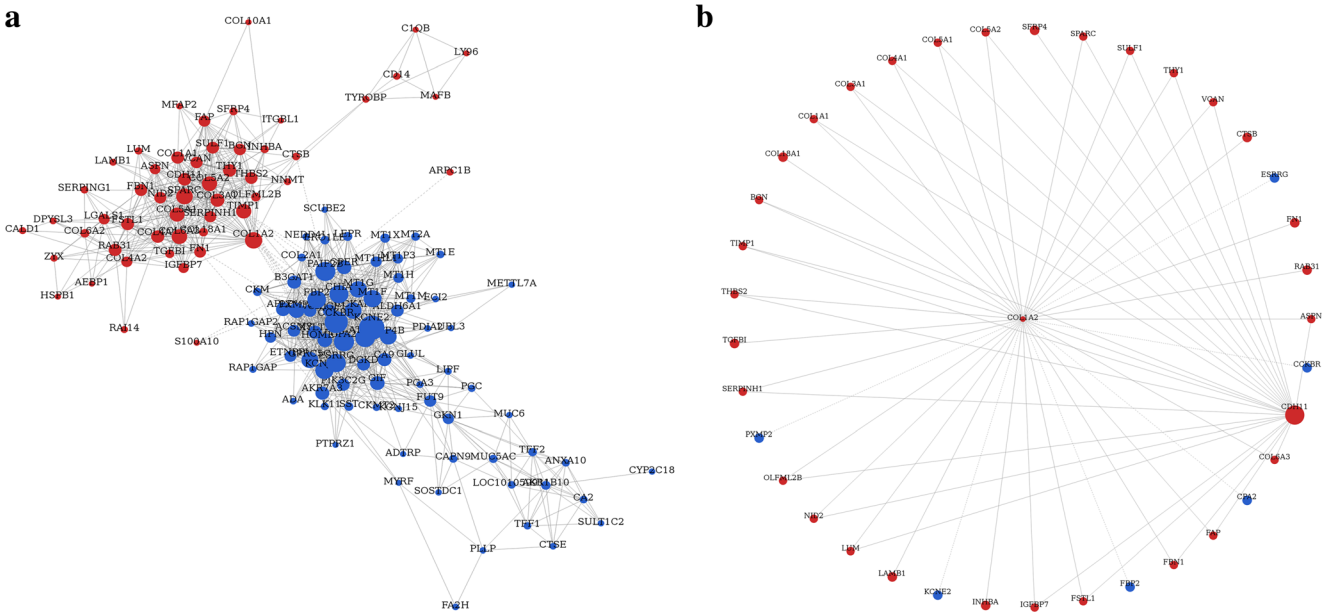

Fig. 2 Gastric cancer of gene co-expression network. (**a**) From the total differential genes, 211 candidate genes as function of collagen fibril organization by constructing a gene co-expression network with k-core algorithm. The COL1A2 gene localizes at the center of subnetwork in the co-expression network. (**b**) COL1A2 was the key gene in a 20-gene network and directly regulates 37 adjacent genes that network according to their degrees. Because the clustering coefficients of these genes are higher than other genes, these interactions rely strongly on COL1A2. Up-regulation gene is represented as red node; down-regulation gene is represented as green node

final staining scores (ranging from 0 to 7) of COL1A2 expression were divided into two groups: high expression groups (scores ≥ 3) and low expression groups (scores <3).

Statistical Analysis

For microarray analysis, differentially expressed genes were confirmed using a *p*-value threshold and FDR analysis. The threshold of truly significant genes was taken to be *p* value less than 0.05 and FDR value less than 0.05. IBM SPSS statistics 22 was used for statistical analysis. Chi-square was used to analyze the correlation between COL1A2 expression and clinicopathplogic parameters. $P \leq 0.05$ was considered as statistical significant.

Results

Identification of Differently Expressed Genes in Human Gastric Cancer

The human gastric cancer of Affymetrix Human Genome U133plus 2.0 microarray data and corresponding clinical data used were downloaded from the publicly GEO databases. Differently expressed genes of gastric cancer and paired non-tumoral were analyzed using Expression Console 1.4.1 and Transcriptome Analysis Console v3.0. Compared with gastric normal tissue, hierarchical clustering showed a total

Table 1 Correlation between the clinicopathologic features and expression of COL1A2

Characteristics	CAPG (%)		*P*-Value
	Low expression	High expression	
Age (y)			0.325
< 55	3(6.0)	47(94.0)	
≥ 55	7(28.0)	18(72.0)	
Tumor size(cm)			0.085
<3	4(7.5)	49(92.5)	
≥ 3	6(27.3)	16(72.7)	
Histological type			<0.0001
Malignant	10(13.4)	65(86.6)	
Normal	58(77.3)	17 (22.7)	
Pathological grading			0.016
I	1(5.3)	18(94.7)	
II	5(17.9)	23(82.1)	
III	2(12.5)	14(87.5)	
IV	2(17.6)	10(83.3)	
T classification			0.083
T1	3(13.7)	19(86.3)	
T2	5(17.6)	25(83.3)	
T3	2(8.7)	21(91.3)	
N classification			0.126
N0	6(13.3)	39(86.7)	
N1–3	4(13.4)	26(86.6)	

Chi-square test (*P<0.05)

of 433 genes were differentially expressed ($p < 0.01$) in invasive gastric cancer, as shown in Fig. 1.

Comprehensive co-Expression Network Analysis and Candidate Biomarker

In order to identify which gene may potentially play a biomarker in the development of human gastric cancer, all significantly different expressed genes then constructed to gene co-expression network, as shown in Fig. 2a. The degree of a node describes the number of links one gene has to others within the gene network. In the gastric cancer co-expression network, we founded the altered gene-expressions in gastric cancer are largely consistent with the previous study, such as the altered expression of CDH1, CKS2, CLDN1 et al. Some novel observations were also made in our study. Interestingly, Central to this network was COL1A2, which directly controlled 20 neighboring genes it interacted with (Fig. 2b).

Immunohistochemical Evaluation of COL1A2 Protein Expression as a Biomarker

In order to investigate the role of COL1A2 in human gastric cancer progression, the COL1A2 protein expression was examined in normal gastric and grades 2 to 4 gastric cancer by immunohistochemical analysis. We founded that COL1A2 expression is increased in 65 of 75(86.6%) gastric cancer tissues and 17 of 75(22.7%) (Table 1) normal stomach tissues. These data indicates that COL1A2 is highly expressed in gastric carcinoma tissues compared to adjacent non-tumor tissues (Fig. 3). We also found that COL1A2 is mainly expressed in cytoplasm in non-cancerous tissues, whereas COL1A2 is mainly expressed in nuclear in cervical carcinoma tissues, given that it encodes the pro-alpha2 chain of type I collagen, all this led us to hypothesis that COL1A2 may play an important role in tumorigenesis of gastric cancer.

Relationship between COL1A2 Expression and Clinical Features

To further determine COL1A2 expression to gastric cancer progression, relationship of COL1A2 in tumors tissues from 75 patients was analyzed. Each sample was assigned an immunoreactivity score ranging from 0 to 7. Statistical analysis showed COL1A2 expression level was significantly associated with histological type ($P<0.05$) and lymph node status. However, there was no correlations between COL1A2 expression and age, lymph node numbers, tumor size, or clinical

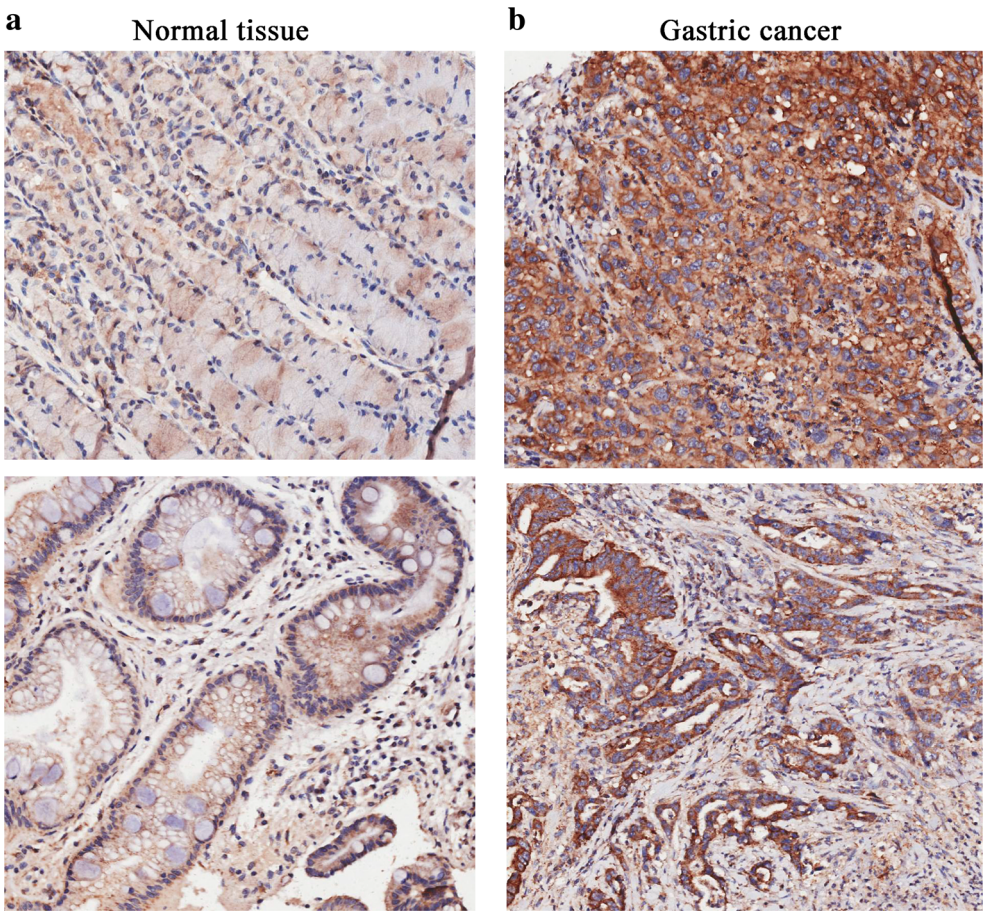

Fig. 3 Detection of COL1A2 expression in gastric caner and adjacent tissues. Representative images of (**a**) negative staining of COL1A2 in normal stomach tissues. (**b**) positive staining of COL1A2 in gastric cancer tissues

stage ($P > 0.05$), as shown in Table 1. It suggests that COL1A2 expression may associated with the lymphoid development in cervical carcinoma.

Discussion

Gastric cancer, also known as stomach cancer, is cancer developing from the lining of the stomach. Gastric cancer is the third most common cause of cancer-related death in the world, and it remains difficult to cure in Western countries, primarily because most patients present with advanced disease. In the United States, stomach malignancy is currently the 15th most common cancer. The most common cause is infection by the bacterium Helicobacter pylori, which accounts for more than 60% of cases. About 10% of cases run in families and between 1% and 3% of cases are due to genetic syndromes inherited from a person's parents such as hereditary diffuse gastric cancer. Most cases of stomach cancers are gastric carcinomas. This type can be divided into a number of subtypes [19–21]. Lymphomas and mesenchymal tumors may also develop in the stomach. Most of the time, stomach cancer develops in stages over years. Early gastric cancer has no associated symptoms; however, some patients with incidental complaints are diagnosed with early gastric cancer. Most symptoms of gastric cancer reflect advanced disease.

Previous gene expression microarray studies have identified a number of gene biomarkers of different types for human cancer. For example, researchers believe that mutations in the genes *BRCA1* and *BRCA2* cause as many as 60% of all cases of hereditary breast and ovarian cancers [8, 10]. In this study, we describe an integrated bioinformatics to identify molecular biomarkers for gastric cancer in patients' cancer tissues. We reports differently expression genes in large gastric cancer cohorts from Gene Expression Ominus (GEO). Firstly, we founded that 433 genes were significantly different expressed in human gastric cancer. Differently expression gene profile in gastric cancer was further validated by bioinformatic analyses, co-expression network construction. In the gastric cancer co-expression network, we founded the altered gene-expressions in gastric cancer are largely consistent with the previous study, such as the altered expression of CDH1, CKS2, CLDN1. Some novel observations were also made in our study. Interestingly, Central to this network was COL1A2, which directly controlled 20 neighboring genes it interacted with.

Furthermore, the COL1A2 protein expression was examined in normal gastric and grades 2 to 4 gastric cancer by immunohistochemical analysis. We founded that COL1A2 expression is increased in 65 of 75(86.6%) gastric cancer tissues and 17 of 75(22.7%) normal stomach tissues. These data indicates that COL1A2 is highly expressed in gastric carcinoma tissues compared to adjacent non-tumor tissues. We also found that COL1A2 is mainly expressed in cytoplasm in non-cancerous tissues, whereas COL1A2 is mainly expressed in nuclear in cervical carcinoma tissues, given that COL1A2 encodes the pro-alpha2 chain of type I collagen, all this led us to hypothesis that COL1A2 may play an important role in tumorigenesis of gastric cancer. Statistical analysis showed COL1A2 expression level was significantly associated with histological type and lymph node status.

In conclusion, the novel bioinformatics used in this study has led to identification of improving diagnostic biomarkers for human gastric cancer and could benefit further analyses of the key alteration during its progression.

Acknowledgments We thank Department of gastroenterology, Chongqing Infectious Disease Medical Center for providing study data.

Compliance with Ethical Standards

Conflict of Interest No potential conflicts of interest were disclosed.

References

1. Orditura M, Galizia G, Sforza V, Gambardella V, Fabozzi A, Laterza MM, Andreozzi F, Ventriglia J, Savastano B, Mabilia A, Lieto E, Ciardiello F, De Vita F (2014) Treatment of gastric cancer. World J Gastroenterol 20:1635–1649
2. Lee YY, Derakhshan MH (2013) Environmental and lifestyle risk factors of gastric cancer. Arch Iran Med 16:358–365
3. Chandanos E, Lagergren J (2008) Oestrogen and the enigmatic male predominance of gastric cancer. Eur J Cancer 44:2397–2403
4. Qin J, Liu M, Ding Q, Ji X, Hao Y, Wu X, Xiong J (2014) The direct effect of estrogen on cell viability and apoptosis in human gastric cancer cells. Mol Cell Biochem 395:99–107
5. Gonzalez CA, Sala N, Rokkas T (2013) Gastric cancer: epidemiologic aspects. Helicobacter 18(Suppl 1):34–38
6. Hatakeyama M, Higashi H (2005) Helicobacter pylori CagA: a new paradigm for bacterial carcinogenesis. Cancer Sci 96:835–843
7. Thrumurthy SG, Chaudry MA, Hochhauser D, Mughal M (2013) The diagnosis and management of gastric cancer. BMJ 347:f6367
8. Chan TH, Qamra A, Tan KT, Guo J, Yang H, Qi L, Lin JS, Ng VH, Song Y, Hong H et al: ADARMediated RNA Editing Predicts Progression and Prognosis of Gastric Cancer. Gastroenterology 2016; 151(4):637–650 e610
9. Kataoka K, Shiraishi Y, Takeda Y, Sakata S, Matsumoto M, Nagano S, Maeda T, Nagata Y, Kitanaka A, Mizuno S, Tanaka H, Chiba K, Ito S, Watatani Y, Kakiuchi N, Suzuki H, Yoshizato T, Yoshida K, Sanada M, Itonaga H, Imaizumi Y, Totoki Y, Munakata W, Nakamura H, Hama N, Shide K, Kubuki Y, Hidaka T, Kameda T, Masuda K, Minato N, Kashiwase K, Izutsu K, Takaori-Kondo A, Miyazaki Y, Takahashi S, Shibata T, Kawamoto H, Akatsuka Y, Shimoda K, Takeuchi K, Seya T, Miyano S and Ogawa S. Aberrant PD-L1 expression through 3′-UTR disruption in multiple cancers. Nature 2016; 534: 402–406.
10. Burki TK: Progression-free survival with regorafenib in gastric cancer. The Lancet Oncology 2016; 17(8):e323
11. Venturi S, Donati FM, Venturi A, Venturi M (2000) Environmental iodine deficiency: a challenge to the evolution of terrestrial life? Thyroid 10:727–729
12. Jakszyn P, Gonzalez CA (2006) Nitrosamine and related food intake and gastric and oesophageal cancer risk: a systematic review of the epidemiological evidence. World J Gastroenterol 12:4296–4303

13. Barker N, Huch M, Kujala P, van de Wetering M, Snippert HJ, van Es JH, Sato T, Stange DE, Begthel H, van den Born M, Danenberg E, van den Brink S, Korving J, Abo A, Peters PJ, Wright N, Poulsom R, Clevers H (2010) Lgr5(+ve) stem cells drive self-renewal in the stomach and build long-lived gastric units in vitro. Cell Stem Cell 6:25–36
14. Ohtsu A, Ajani JA, Bai YX, Bang YJ, Chung HC, Pan HM, Sahmoud T, Shen L, Yeh KH, Chin K, Muro K, Kim YH, Ferry D, Tebbutt NC, Al-Batran SE, Smith H, Costantini C, Rizvi S, Lebwohl D, Van Cutsem E (2013) Everolimus for previously treated advanced gastric cancer: results of the randomized, double-blind, phase III GRANITE-1 study. J Clin Oncol 31:3935–3943
15. Wang K, Yuen ST, Xu J, Lee SP, Yan HH, Shi ST, Siu HC, Deng S, Chu KM, Law S, Chan KH, Chan AS, Tsui WY, Ho SL, Chan AK, Man JL, Foglizzo V, Ng MK, Chan AS, Ching YP, Cheng GH, Xie T, Fernandez J, Li VS, Clevers H, Rejto PA, Mao M, Leung SY (2014) Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. Nat Genet 46:573–582
16. Ramucirumab extends gastric cancer survival. Cancer Discov 2014; 4: OF3.
17. Garay J, Piazuelo MB, Majumdar S, Li L, Trillo-Tinoco J, Del Valle L, Schneider BG, Delgado AG, Wilson KT, Correa P et al: The homing receptor CD44 is involved in the progression of precancerous gastric lesions in patients infected with Helicobacter pylori and in development of mucous metaplasia in mice. Cancer letters 2016; 371(1):90–98
18. Cristescu R, Lee J, Nebozhyn M, Kim KM, Ting JC, Wong SS, Liu J, Yue YG, Wang J, Yu K, Ye XS, Do IG, Liu S, Gong L, Fu J, Jin JG, Choi MG, Sohn TS, Lee JH, Bae JM, Kim ST, Park SH, Sohn I, Jung SH, Tan P, Chen R, Hardwick J, Kang WK, Ayers M, Hongyue D, Reinhard C, Loboda A, Kim S, Aggarwal A (2015) Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. Nat Med 21:449–456
19. Schlaermann P, Toelle B, Berger H, Schmidt SC, Glanemann M, Ordemann J, Bartfeld S, Mollenkopf HJ, Meyer TF (2016) A novel human gastric primary cell culture system for modelling helicobacter pylori infection in vitro. Gut 65:202–213
20. Tabata S, Ikeda R, Yamamoto M, Shimaoka S, Mukaida N, Takeda Y, Yamada K, Soga T, Furukawa T, Akiyama S (2014) Thymidine phosphorylase activates NFκB and stimulates the expression of angiogenic and metastatic factors in human cancer cells. Oncotarget 5:10473–10485
21. Kim SY, Park C, Kim HJ, Park J, Hwang J, Kim JI, Choi MG, Kim S, Kim KM, Kang MS (2015) Deregulation of immune response genes in patients with Epstein-Barr virus-associated gastric cancer and outcomes. Gastroenterology 148:137–147 e139