HMG Advance Access published March 10, 2015

Genome-wide profiling of polyadenylation sites reveals a link between selective polyadenylation and cancer metastasis

Deng-Pan Lai[1,†], Sheng Tan[2,†], Ya-Ni Kang[1,†], Jun Wu[3], Hong-Sain Ooi[4], Jian Chen[5], Ting-Ting Shen[5], Ying Qi[5], Xiaoli Zhang[5], Yan Guo[1], Tao Zhu[2], Bingya Liu[1,5,6], Zhifeng Shao[1], Xiaodong Zhao[1,*]

[1]State Key Laboratory for Oncogenes & Related Genes, Bio-ID Center, School of Biomedical Engineering, Ruijin Hospital, Shanghai Jiao Tong University, Shanghai 200240, China

[2]Laboratory of Molecular Tumor Pathology, School of Life Sciences, University of Science and Technology of China, Hefei, Anhui 230027, China

[3]Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China

[4]Department of Biomedicine, Aarhus University, Wilhelm Meyers Allé 4, 8000 Aarhus C, Denmark

[5]Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, Shanghai 200240, China

[6]Shanghai Key Laboratory of Gastric Neoplasms, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 20025, China

[*]To whom correspondence should be addressed. Email: xiaodongzhao@sjtu.edu.cn, Tel: 0086-21-3420-6846, Fax: 0086-21-3420-6846

[†]These authors are first co-authors and equally contributed to the work.

Abstract

Alternative polyadenylation (APA) is an important post-transcriptional modification implicated in many diseases, including cancer. Although extensively characterized, the functional consequence of APA modulation on tumorigeneisis remains elusive. Here, we developed a deep sequencing-based approach that specifically profiles 3' termini of polyadenylated RNAs (herein termed 3T-seq) and analyzed APA events in two gastric cancer cell lines and one nontransformed counterpart. Overall, we identified more than 28,000 poly(A) sites, 70% of which are potentially novel. Further, we observed widespread APA-mediated 3'UTR shortening of 513 genes (false discovery rate < 0.05) across gastric cancer genome. We characterized one of these genes, *NET1*, in detail and found that the shortening of *NET1* 3'UTR significantly enhances transcriptional activity. Moreover, the *NET1* isoform with short 3'UTR promotes cellular migration and invasion *in vitro*. Collectively, our work provides an effective approach for genome-wide APA site profiling and reveals a link between APA modulation and gastric cancer metastasis.

Introduction

An increasing body of evidence has demonstrated that alternative polyadenylation (APA) is a widespread phenomenon in human genome (1, 2). APA contributes to the transcriptome complexity by generating transcript isoforms with diverse 3'UTRs of a given gene and provides an important layer of gene expression regulation. Through the alteration of 3'UTRs APA potentially affects the stability, translation efficiency and cellular localization of the transcript isoforms (3-5). APA events are tissue-specifically modulated (6, 7) and a recent report even demonstrated APA could be used to achieve a tissue-specific transcription grogram (8). In addition to regulation of gene expression, recent studies have shown that extensive APA modulation is associated with various biological processes and disease, such as proliferation, development, cellular differentiation, neuron activation and cancer (1).

An emerging theme of cancer biology is that APA modulation can activate some proto-oncogenes without genetic mutation and thereby promote oncogenic transformation (9). Transcript isoforms with short 3'UTRs generated by alternative cleavage and polyadenylation exhibit increased stability and produce more protein by loss of miRNA-mediated repression (10). A recent study reported the widespread preferential usage of proximal poly(A) sites in various cancers, including breast, colon, kidney, liver and lung tumors (11). These observations underscore a general role of APA modulation in tumorigenesis. However, the functional consequence of APA modulation on tumor formation and progression remains elusive.

Here, we developed a robust method to profile genome-wide 3' termini of transcripts and delineate APA sites in two gastric cancer cell lines, AGS and MKN28, and their non-transformed counterpart GES-1. AGS is non-metastatic which was derived from a poorly differentiated gastric adenocarcinoma. MKN28 is metastatic which was generated from a well differentiated gastric adenocarcinoma with intestinal type differentiation. We used this method to characterize the APA

sites in gastric cancer cells and explored the biological significance of APA modulation in gastric tumorigenesis.

Results

Deep sequencing of mRNA 3' termini by 3T-seq

We aimed to profile the genome-wide APA patterns in gastric cancer genome. To achieve this, we developed a 3T-seq approach to assay APA events. This approach was outlined in Figure 1A. Briefly, total RNA was reversely transcribed with oligod(T) primer. The resulting DNA/RNA hybrids were treated with RNase H and DNA polymerase I for double strand DNA synthesis, which was immobilized on M280 streptavidin dynabeads (Invitrogen). Enzymatical fragmentation was then performed to break dsDNA to 200-500bp pieces with the 3' terminal fragments anchored on the beads. After GsuI digestion to remove poly(A) stretch, the 3' terminal fragments were released and subjected to deep sequencing by Illumina Hiseq2000.

We used this method to profile APA sites in two gastric cancer cell lines AGS and MKN28 and one gastric epithelial cell line GES-1. Totally, we generated approximate 22 million reads, of which nearly 80% are mappable. 9.3 million reads with unique genomic location were used to filter those from internal priming, yielding 9.1 million reads for the delineation of APA sites (Figure 1B). A statistical summary of the sequencing data sets was shown in Table S1.

We observed the majority of filtered reads were located in the 3' terminus of UCSC Refseq genes (Figure S1A). Particularly, nearly 90% reads were mapped to the 3' end of UCSC genes. We detected the expression of 66.7% of Refeseq genes (19,014) that are supported by at least two 3T-seq reads. These results suggest that our method effectively captures the 3' ends of transcripts.

In addition to determination of mRNA 3' ends, we examined the potential utility of 3T-seq approach in the measurement of gene expression levels. Currently, RNA-seq has been predominantly used in gene expression analysis. We compared our data in AGS with RNA-seq data under accession number GSE43093. The Pearson correlation is 0.59 (Figure 1C), which indicates that 3T-seq could be used for gene expression study.

Identification of polyadenylation sites

We utilized a similar strategy used in a previous report to infer the poly(A) sites (12). With the threshold that each poly(A) site is supported by at least 2 reads, totally we identified 28,588 poly(A) sites in three gastric cell lines. To estimate the reliability of our poly(A) site data, we firstly examined the presence of motifs in poly(A) sites reported in previous studies (12, 13). We found more than 80.5% of poly(A) sites in our analysis contain at least one of reported motifs within 100 nt upstream of the identified sites. Furthermore, we plotted the distance between the poly(A) sites in our study to the closest known poly(A) sites from the UCSC poly(A) database. The distribution of the 3T-seq defined poly(A) sites shows a sharp peak centered on known poly(A) sites (Figure S1B). These results indicates that the 3T-seq approach could accurately identify poly(A) sites.

Among the 28,588 identified poly(A) sites, 33.1% were mapped to UCSC transcription termination sites (TTS), 45% to 3'UTR regions (Figure S1C). 30% of these poly(A) sites have been annotated in the UCSC poly(A) database, and thus 70% are putatively novel. To test the authenticity of these novel APA sites, we randomly selected 10 of these poly(A) sites for verification with 3' RACE. All of 10 poly(A) sites were confirmed, 5 were located in 3'UTR regions, 3 in TTS, 1 in CDS and 1 in 1kb downstream TTS. This result suggests our method could effectively identify

unannotated poly(A) sites. Among the expressed genes detected by 3T-seq, we observed that 22.8% harbor three or more poly(A) sites (Figure S1D). Interestingly, we found the transcripts derived from genes with more than 3 poly(A) sites tend to be more differentially polyadenylated. In the following analysis, we focused on the genes with multiple poly(A) sites in our data sets.

Shortened 3' UTRs are widespread in gastric cancer genome

A recent report observed that several well-chosen oncogenes preferentially generate short 3'UTR isoforms in cancer cells (9). We asked whether such a phenomenon is widespread in gastric cancer cells. Our 3T-seq approach could be able to distinguish short 3'UTR isoforms from long ones for a given gene, which enable us to address this question. To this end, we adopted the cancer 3' UTR length index (CULI) (14) to quantitatively characterize the 3'UTR alteration in gastric cancer genome. A positive CULI suggests that a gene harbors lengthened 3' UTR in a cancer cells compared with the nontransformed cells, and a negative CULI indicates the shortened one. With this criteria we identified 580 genes (FDR < 0.05) with a significant difference in 3' UTR length between AGS and GES-1 and 65% genes have shortened 3'UTRs in AGS (Figure 2A and Table S2). Among 248 genes (FDR < 0.05) with different 3'UTR length between MKN28 and GES-1, we found similar trend of switching that 80% genes have shortened 3'UTRs (Figure 2A and Table S3). These results suggest 3'UTR shortening is widespread in gastric cancer genome. Our finding is in agreement with previous reports in other cancer types (11, 14, 15). The observation from individual gene in a previous study (9) raised an assumption that 3'UTR length is inversely correlated with expression level. To gain the general view of the relationship between APA alteration and gene expression, we compared the APA change with mRNA abundance, and found no linear correlation on transcriptome scale (Figure 2B). These results imply APA could be involved in regulation of gene expression in a complicated manner.

Meanwhile, we also compared the 3'UTR alteration between AGS and MKN28, and identified 601 genes (FDR < 0.05) with altered 3'UTRs (Figure 2C and Table S4). Such a large fraction of genes with significantly altered 3'UTRs between the two gastric cancer cell lines indicated that there are enormous differences in the post-transcriptional regulation between AGS and MKN28. As they exhibit similar trend of 3'UTRs switching compared with the normal cell line, we speculated the APA-mediated 3'UTR differences between these two gastric cancer cell lines are potentially involved in their different malignant characteristics.

Functional enrichment of genes with switched APA sites

Since a considerable fraction of genes exhibited altered APA, next we attempted to understand the biological consequences of altered APA pattern in gastric cancer. We searched the overrepresented functional items in genes containing shortened 3'UTR with PANTHER (16). Analysis of 377 genes with 3'UTR shortening in AGS yielded proliferation-related biological processes that are statistically overrepresented, including metabolic process, cell cycle, mRNA processing, primary metabolic process, and RNA splicing (Table S5). Similar biological processes were also observed in MKN28. These results implicate a significant role of 3'UTR shortening in the unlimited proliferation of gastric cancer cells. In addition, in comparison with GES-1 or AGS we observed the biological processes of mRNA polyadenylation and mRNA 3'-end processing are overrepresented in MKN28 (Table S5), suggesting some components of 3-end-processing and polyadenylation machinery could be actively involved in this highly metastatic cell line.

3'UTR shortening promotes gastric cellular migration and invasion

Although the gene ontology analysis results suggest that the preferential usage of proximal poly(A)

sites contributes to the gastric cancer cell proliferation, the shortening of 3'UTR is also involved in T cell proliferation (17). As such, how the selective usage of poly(A) sites is particularly involved in gastric tumorigenesis remains elusive. To explore the functional relevance of 3'UTR shortening in gastric malignancy, we further analyzed the genes with switched APA patterns between two gastric cancer cell lines. We chose 313 genes that preferentially use proximal APA in MKN28 (Table S4) and searched the overrepresented pathways. The most statistically significant enriched item is cytoskeletal regulation by Rho GTPase pathway. Interestingly, a previous study demonstrated Rho GTPase pathway is involved in cell migration and invasion in gastric cancer and the activation of this pathway is mediated by *NET1* (18).

In our study we found *NET1* exhibits 3'UTR shortening in MKN28 (Figure 3A), a cell line more metastatic than AGS. To examine whether 3'UTR shortening of *NET1* increases transcriptional activity, a luciferase reporter assay was performed with short or long 3'UTR of *NET1*, respectively. As shown in Figure 3B, the short 3'UTR significantly promotes transcriptional activity of reporter gene. Interestingly, we observed such transcriptional activation occurs in both gastric cancer cell lines and does not occur in GES-1, which implies that the transcriptional activation induced by 3'UTR shortening is cellular context dependent.

Lastly, we evaluated the influence of APA-mediated 3'UTR alteration on gastric cellular migration and invasion. MKN28 cells were transfected with *NET1* containing short or long 3'UTR, respectively (Figure 4A). Their capability of migration and invasion was measured by wound-healing, migration and invasion assays. Compared with the isoform with long 3'UTR, *NET1* with short 3'UTR leads to a more rapid closing of the wound (Figure 4B). Concordantly, we observed the shortening of *NET1* 3'UTR significantly promotes gastric cellular migration (Figure

4C) and invasion (Figure 4D). These results indicate that APA-induced 3'UTR alteration facilitates gastric metastasis.

Discussion

With an increasing number of reports demonstrating APA events are actively involved in a wide variety of biological process, many efforts have been made to develop approaches to characterize the poly(A) sites in a genome-wide manner (1, 19). In this study, we developed a 3T-seq strategy to identify the APA sites in gastric cancer cells. Our method has advantages over some of those previously reported. 3T-seq could detect new poly(A) sites and quantitatively delineate the poly(A) usage preference between different samples. In contrast, microarray-based analysis of APA relies on the prior knowledge of probe sequence. RNA-seq is widely used for transcriptome analysis and has significantly advanced our understanding of alternative splicing events. Although it has been also used in APA analysis (7, 20), the relatively low fraction of reads located in the cleavage sites makes it inefficient in characterizing poly(A) sites. Recently, several 3'end-enriched approaches have been developed (1, 19), which significantly increase the efficiency in APA determination. However, in some methods the presence of untemplated poly(A) stretch in 3'end-enriched library affects the quality of deep sequencing (21, 22). To circumvent this limitation, a recent report introduced an amplification procedure to generate mutations in the poly(A) stretch (14). With a similar purpose, our method utilized restriction enzyme digestion to remove poly(A) stretch and release 3' terminal fragments for deep sequencing. As such, we provided a robust approach to perform genome-wide survey of APA events.

The functional coupling between 3'UTR alteration and transcription has been characterized. 3'UTR shortening could increase transcriptional activity, probably through the relief of

microRNA-mediated repression. Genome-wide APA analyses also seem to suggest a positive correlation between usage of proximal poly(A) sites and gene expression level (11, 23). However, in this study we observed a complicated trend when examining the relationship of 3'UTR change and gene expression. A large number of genes with both shortened and lengthened 3'UTR isoforms show increased transcription level (Figure 2B). Meanwhile, some affected genes have no obvious change in expression level or are even down-regulated. Such trend was also observed in other reports (14, 15). This discrepancy is not likely due to the experimental strategy in each study, as we used a similar strategy to enrich 3'end fragments as in a previous report (23). One consequence of 3'UTR shortening is the loss of binding sites of miRNAs. For the genes with shortened 3'UTR that are not up-regulated, it is possible that miRNAs targeting the affected genes are not actively transcribed in the examined cells. A comprehensive characterization of miRNA and APA modulation could facilitate the elucidation of the relation between 3'UTR alteration and transcription.

Cancer has been classically viewed as a neoplastic disease with the accumulation of genetic mutations which drive oncogene activation (24). However, a recent study advanced our understanding of tumorigenesis in which several well-chosen oncogenes were found to be activated by APA-mediated 3'UTR shortening rather than genetic mutation (9). Genome-wide analyses of APA sites have revealed the widespread shortening of 3'UTRs in both cancer cell lines and cancer tissues (11, 14). Although extensively characterized, how 3'UTR shortening contributes to tumor formation and progression remains elusive. In this study, we identified hundreds of genes with shortened 3'UTRs. Gene ontology analysis of these affected genes indicates Rho GTPase pathway is the most overrepresented ($p < 0.01$). Rho GTPase pathway is involved in gastric cell migration and invasion, whose activation is mediated by NET1 (18). Intriguingly, we observed the shortening of 3'UTR accompanies *NET1* transcriptional activation. We further demonstrated that 3'UTR

shortening could enhance transcriptional activity (Figure 3B) and promote gastric cellular migration and invasion *in vitro* (Figure 4). As mentioned above, oncogenic activation derived from selective polyadenylation is partially due to the escape of miRNA-mediated repression, it is likely that the APA-mediated *NET1* 3'UTR shortening activates Rho GTPase pathway and promotes gastric metastasis.

Briefly, in this study we developed a robust approach for global APA site profiling and observed that hundreds of genes exhibit the APA-mediated shortening of 3'UTR. Our results further indicate that 3'UTR shortening contributes to gastric cancer cellular migration and invasion and suggest a link between selective polyadenylation and cancer metastasis.

Materials and Methods

Cell line and growth condition

Gastric epithelial cell line GES-1 and two cancer cell lines AGS and MKN28 were reported in our previous study (25) and grown in RPMI-1640 medium (Gibco, USA) with 10% fetal bovine serum (Gibco, USA) at 37℃ in a humidified atmosphere with 5% CO2.

3T-seq library preparation

Total RNA was extract with TRIzol Reagent (Invitrogen, USA) and treated with TURBO DNA-free™ Kit (Life technologies) to remove genomic contamination. 50ug total RNA was incubated with streptavidin dynabeads coated with biotinylated GsuI-oligo(dT) primer (Table S6). The first strand cDNA was synthesized with SuperScript III (Invitrogen, USA) and dNTP mix containing 5-methylated-dCTP instead of dCTP. Then dsDNA was generated with the method reported previously (26). Next, the resulting dsDNA was fragmentated to approximately 200-500bp

with fragmentase (NEB). 3' terminal fragments were released from beads through Gsu I digestion, which were subjected to deep sequencing by Illumina HiSeq2000. All primers used in this study were listed in Table S6. A detailed 3T-seq library construction method was provided in Supplemental Information.

Data analysis

The processing pipelines were similar to a previous report (14). We screened and trimmed the raw read data with our C++ scripts and the processed reads were mapped to the human genome (hg19) with bowtie2 (27). The poly(A) sites were constructed by iteratively clustering the mapping results as described previously (12). The 3'UTR switching for each gene among the three samples was detected by measuring the linear trend alternative to independence. The false discovery rate was estimated with R software. Functional analysis of the genes with altered 3' UTR was performed the online software PANTHER (16).

3' RACE validation of novel poly(A) sites

3' RACE was performed on AGS and MKN28 according to the reported protocol (28). Primers for 3' RACE were listed in Table S7. The PCR fragments were purified with polyacrylamide gel electrophoresis and ligated with pEASY-T1 vector (Transgen). Ten cloned fragments were randomly chosen for capillary sequencing.

Luciferase assay

The short and long 3URTs of *NET1* were amplified and cloned downstream Renilla luciferase in a psiCHECK2 vector (Promega). Cells plated on 24-well plates were transfected with 100ng of psiCHECK2-3'UTR_short, psiCHECK2-3'UTR_long and psiCHECK2, respectively. After 48 h,

cells were lysed and assayed with Dual Luciferase Assay (Promega) according to the manufacturer's instructions. Three independent experiments were performed in triplicate.

Wound-healing assay

Wound-healing assay was performed as we previously reported (25). Briefly, the ORF of *NET1* with short and long 3'UTRs were amplified and cloned into pIRESneo3 (Clontech), respectively. MKN28 cells were seeded in six-well plates, cultured overnight, and transfected with pIRESneo3, *NET1*-3'UTR-short and *NET1*-3'UTR-long, respectively. Upon reaching confluence, cells were serum starved for 24 h, and then the cell layer was scratched with a sterile plastic tip and washed twice with PBS and cultured in RPMI-1640 (serum free) at 37 °C in a humidified incubator with 5% CO2. At different time points, photo images of the plates were taken under a microscope.

Cell migration and invasion assay

To estimate the influence on cell motility mediated by *NET1* with altered 3'UTRs, assays were performed in BioCoat transwell chambers (Corning Costar) with uncoated porous filters (pore size 8 μm), as described in our previous report (29). Filters were precoated with matrigel for the examination of cell invasion. Cells were serum-deprived for 24 h before experiment and then seeded on the upper chamber in serum-free medium with 0.2% BSA. Filters were fixed in 90% ethanol and stained with 0.1% crystal violet (Sigma) for counting. Values for cell migration or invasion were expressed as the average number of cells per microscopic field over three fields per one filter for triplicate experiments. Experiments were repeated at least three times.

Acknowledgements

The authors would like to thank Daniel M Czajkowsky for helpful discussion and careful reading.

This work was supported by Longhua Medical Project of State Clinical Research Center of TCM in Longhua Hospital [LYTD-21 and JDZX2012123], State Key Development Program for Basic Research of China [2013CB967402] and National Natural Science Foundation of China [91019004 and 91229123].

Conflict of Interest statement. None declared.

Accession number

Raw sequencing data have been submitted to the NCBI Sequence Read Archive (SRA; http://www.ncbi.nlm.nih.gov/sra) under accession number SRP048647.

Supplementary Material

Supplementary Material is available at HMG online

References

1 Elkon, R., Ugalde, A.P. and Agami, R. (2013) Alternative cleavage and polyadenylation: extent, regulation and function. *Nat Rev Genet*, **14**, 496-506.

2 Shi, Y. (2012) Alternative polyadenylation: new insights from global analyses. *RNA*, **18**, 2105-2117.

3 Lewis, J.D., Gunderson, S.I. and Mattaj, I.W. (1995) The influence of 5' and 3' end structures on pre-mRNA metabolism. *J Cell Sci Suppl*, **19**, 13-19.

4 Jacobson, A. and Peltz, S.W. (1996) Interrelationships of the pathways of mRNA decay and translation in eukaryotic cells. *Annu Rev Biochem*, **65**, 693-739.

5 Wickens, M., Anderson, P. and Jackson, R.J. (1997) Life and death in the cytoplasm: messages from the 3' end. *Curr Opin Genet Dev*, **7**, 220-232.

6 Zhang, H., Lee, J.Y. and Tian, B. (2005) Biased alternative polyadenylation in human tissues. *Genome Biol*, **6**, R100.

7 Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P. and Burge, C.B. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature*, **456**, 470-476.

8 Lianoglou, S., Garg, V., Yang, J.L., Leslie, C.S. and Mayr, C. (2013) Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes & development*, **27**, 2380-2396.

9 Mayr, C. and Bartel, D.P. (2009) Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell*, **138**, 673-684.

10 Di Giammartino, D.C., Nishida, K. and Manley, J.L. (2011) Mechanisms and consequences of alternative polyadenylation. *Mol Cell*, **43**, 853-866.

11 Lin, Y., Li, Z., Ozsolak, F., Kim, S.W., Arango-Argoty, G., Liu, T.T., Tenenbaum, S.A., Bailey, T., Monaghan, A.P., Milos, P.M. *et al.* (2012) An in-depth map of polyadenylation sites in cancer. *Nucleic Acids Res*, **40**, 8460-8471.

12 Tian, B., Hu, J., Zhang, H. and Lutz, C.S. (2005) A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res*, **33**, 201-212.

13 Beaudoing, E., Freier, S., Wyatt, J.R., Claverie, J.M. and Gautheret, D. (2000) Patterns of variant polyadenylation signal usage in human genes. *Genome Res*, **10**, 1001-1010.

14 Fu, Y., Sun, Y., Li, Y., Li, J., Rao, X., Chen, C. and Xu, A. (2011) Differential genome-wide profiling of tandem 3' UTRs among human breast cancer and normal cells by high-throughput sequencing. *Genome Res*, **21**, 741-747.

15 Morris, A.R., Bos, A., Diosdado, B., Rooijers, K., Elkon, R., Bolijn, A.S., Carvalho, B., Meijer, G.A. and Agami, R. (2012) Alternative cleavage and polyadenylation during colorectal cancer development. *Clinical cancer research : an official journal of the American Association for Cancer Research*, **18**, 5256-5266.

16 Mi, H., Muruganujan, A., Casagrande, J.T. and Thomas, P.D. (2013) Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc*, **8**, 1551-1566.

17 Sandberg, R., Neilson, J.R., Sarma, A., Sharp, P.A. and Burge, C.B. (2008) Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science*, **320**, 1643-1647.

18 Murray, D., Horgan, G., Macmathuna, P. and Doran, P. (2008) NET1-mediated RhoA activation facilitates lysophosphatidic acid-induced cell migration and invasion in gastric cancer. *Br J Cancer*, **99**, 1322-1329.

19 Sun, Y., Fu, Y., Li, Y. and Xu, A. (2012) Genome-wide alternative polyadenylation in animals: insights from high-throughput technologies. *J Mol Cell Biol*, **4**, 352-361.

20 Pickrell, J.K., Marioni, J.C., Pai, A.A., Degner, J.F., Engelhardt, B.E., Nkadori, E., Veyrieras, J.B., Stephens, M., Gilad, Y. and Pritchard, J.K. (2010) Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature*, **464**, 768-772.

21 Yoon, O.K. and Brem, R.B. (2010) Noncanonical transcript forms in yeast and their regulation during environmental stress. *RNA*, **16**, 1256-1267.

22 Shepard, P.J., Choi, E.A., Lu, J., Flanagan, L.A., Hertel, K.J. and Shi, Y. (2011) Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA*, **17**, 761-772.

23 de Klerk, E., Venema, A., Anvar, S.Y., Goeman, J.J., Hu, O., Trollet, C., Dickson, G., den Dunnen, J.T., van der Maarel, S.M., Raz, V. *et al.* (2012) Poly(A) binding protein nuclear 1 levels affect alternative polyadenylation. *Nucleic Acids Res*, **40**, 9089-9101.

24 Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N. and Stratton, M.R. (2004) A census of human cancer genes. *Nature reviews. Cancer*, **4**, 177-183.

25 Wang, M., Li, C., Nie, H., Lv, X., Qu, Y., Yu, B., Su, L., Li, J., Chen, X., Ju, J. *et al.* (2012) Down-regulated miR-625 suppresses invasion and metastasis of gastric cancer by targeting ILK. *FEBS letters*, **586**, 2382-2388.

26 Wu, Z.J., Meyer, C.A., Choudhury, S., Shipitsin, M., Maruyama, R., Bessarabova, M., Nikolskaya, T., Sukumar, S., Schwartzman, A., Liu, J.S. *et al.* (2010) Gene expression profiling of human breast tissue samples using SAGE-Seq. *Genome Res*, **20**, 1730-1739.

27 Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, **10**, R25.

28 Scotto-Lavino, E., Du, G. and Frohman, M.A. (2006) 3' end cDNA amplification using classic RACE. *Nat Protoc*, **1**,

2742-2745.

29 Tan, S., Li, R., Ding, K., Lobie, P.E. and Zhu, T. (2011) miR-198 inhibits migration and invasion of hepatocellular carcinoma cells by targeting the HGF/c-MET pathway. *FEBS letters*, **585**, 2229-2234.

Legends to Figures, Tables

Figure 1. Identification of APA sites by 3T-seq. (A) Schematic view of 3T-seq. The detailed procedure was described in text. (B) A genomic view of poly(A) sites defined by 3T-seq in IGV genome browser. (C) Scatter plot of gene expressions generated with two approaches, where RPKM (reads per kilobase of exon models per million) was used for RNA-seq and RPM (reads per million) for 3T-seq. The Pearson's correlation coefficient was shown for comparison. SMB, streptavidin magnetic beads; black bar attached on beads, biotinylated GsuI-oligod(T) primer.

Figure 2. APA-mediated 3'UTR alteration and the transcriptional activity of the affected genes in gastric cancer cells compared with the normal counterpart. (A) Cancer 3' UTR length index (CULI) was used for the quantitative measurement of 3'UTR alteration in cancer cell lines compared with the nontransformed counterpart (FDR < 0.05). (B) Violin plot showing transcription level change of genes with shortened and lengthened 3'UTR in cancer vs normal pair comparison. (C) Genomic view of APA site usage dynamics of GES-1, AGS and MKN28 in IGV browser.

Figure 3. APA enhances transcriptional activity by 3'UTR shortening. (A) *NET1* transcript isoforms with alternative poly(A) sites in GES-1 (blue), AGS (brown) and MKN28 (red). (B) GES-1, AGS and MKN28 were transfected with plasmid reporter containing long or short 3'UTR of *NET1*, respectively. After 48 h, luciferase assay was performed and normalized to the internal firefly luciferase activity. (**) $P < 0.01$ and (*) $P < 0.05$, two-tailed permutation test.

Figure 4. 3'UTR shortening of *NET1* promotes gastric cellular migration and invasion. (A, B) *NET1* isoforms with long or short 3'UTR were transfected in MKN28 and then subjected to wound-healing assay. The expression levels were measured by RT-qPCR. (C, D) The capability of cellular migration and invasion mediated by *NET1* isoforms with short or long 3'UTR was examined by trans-well assay. (*) $P < 0.05$, two-tailed permutation test.

Abbreviations

APA: alternative polyadenylation; CDS: coding DNA sequence; CULI: cancer 3' UTR length index; FDR: false discovery rate; PAS: poly(A) signal; UTR: untranslated region; 3T-seq: 3' termini sequencing.