Published OnlineFirst March 20, 2012; DOI: 10.1158/0008-5472.CAN-11-3870

Cancer Research

AMPKα Modulation in Cancer Progression: Multilayer Integrative Analysis of the Whole Transcriptome in Asian Gastric Cancer

Yon Hui Kim, Han Liang, Xiuping Liu, et al.

Cancer Res 2012;72:2512-2521. Published OnlineFirst March 20, 2012.

Updated version	Access the most recent version of this article at: doi:10.1158/0008-5472.CAN-11-3870
Supplementary Material	Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2012/03/20/0008-5472.CAN-11-3870.DC1.html

Cited Articles	This article cites by 38 articles, 17 of which you can access for free at: http://cancerres.aacrjournals.org/content/72/10/2512.full.html#ref-list-1
Citing articles	This article has been cited by 3 HighWire-hosted articles. Access the articles at: http://cancerres.aacrjournals.org/content/72/10/2512.full.html#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.

Downloaded from cancerres.aacrjournals.org on July 31, 2013. © 2012 American Association for Cancer Research.

Published OnlineFirst March 20, 2012; DOI: 10.1158/0008-5472.CAN-11-3870

Integrated Systems and Technologies

Cancer Research

AMPKα Modulation in Cancer Progression: Multilayer Integrative Analysis of the Whole Transcriptome in Asian Gastric Cancer

Yon Hui Kim[1], Han Liang[2], Xiuping Liu[1], Ju-Seog Lee[3], Jae Yong Cho[4], Jae-Ho Cheong[5], Hoguen Kim[6], Min Li[7], Thomas J. Downey[7], Matthew D. Dyer[8], Yongming Sun[8], Jingtao Sun[8], Ellen M. Beasley[8], Hyun Cheol Chung[4], Sung Hoon Noh[5], John N. Weinstein[2], Chang-Gong Liu[1], and Garth Powis[1]

Abstract

Gastric cancer is the most common cancer in Asia and most developing countries. Despite the use of multimodality therapeutics, it remains the second leading cause of cancer death in the world. To identify the molecular underpinnings of gastric cancer in the Asian population, we applied an RNA-sequencing approach to gastric tumor and noncancerous specimens, generating 680 million informative short reads to quantitatively characterize the entire transcriptome of gastric cancer (including mRNAs and miRNAs). A multilayer analysis was then developed to identify multiple types of transcriptional aberrations associated with different stages of gastric cancer, including differentially expressed mRNAs, recurrent somatic mutations, and key differentially expressed miRNAs. Through this approach, we identified the central metabolic regulator AMP-activated protein kinase (AMPK)α as a potential functional target in Asian gastric cancer. Furthermore, we experimentally showed the translational relevance of this gene as a potential therapeutic target for early-stage gastric cancer in Asian patients. Together, our findings not only provide a valuable information resource for identifying and elucidating the molecular mechanisms of Asian gastric cancer, but also represent a general integrative framework to develop more effective therapeutic targets. *Cancer Res; 72(10); 2512–21. ©2012 AACR.*

Introduction

Gastric cancer is the fourth most common cancer and the second leading cause of cancer-related deaths in the world (1). The incidence of cancer affecting the distal stomach has significantly declined over the past 40 years, but the incidence of proximal gastric and gastroesophageal junction adenocarcinoma in the United States and Europe has increased at a rate substantially exceeding that of prostate cancer, brain cancer, or melanoma (2). Moreover, the 5-year relative survival rate of patients with gastric cancer has not improved significantly in recent decades, remaining at approximately 20% to 30%.

Authors' Affiliations: Departments of [1]Experimental Therapeutics, [2]Bioinformatics and Computational Biology, and [3]Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas; Departments of [4]Internal Medicine, [5]Surgery, and [6]Pathology, Yonsei University College of Medicine, Seoul, Korea; [7]Partek Inc., St. Louis, Missouri; and [8]Life Technologies, Foster City, California

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Y.H. Kim, H. Liang, X. Liu, and J.-S. Lee contributed equally to this work.

Corresponding Author: Yon Hui Kim, New Experimental Therapeutics Branch, National Cancer Center, Korea, 323 Ilsan-ro, Ilsandong-gu, Goyang-si, Gyeonggi-do 410-769, Korea; Phone: 82-31-920-2519; Fax: 82-31-920-2542; E-mail: yhkim@ncc.re.kr and Han Liang, Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, 1400 Pressler Street, Houston, TX 77030; Phone: 713-745-9815; Fax: 713-563-4242; E-mail: hliang1@mdanderson.org

doi: 10.1158/0008-5472.CAN-11-3870

©2012 American Association for Cancer Research.

The high mortality rate of gastric cancer is due largely to late-stage diagnosis of the cancer and a lack of effective medical treatment options. Treatment often consists of drug combinations that have provided survival advantages for patients with other cancer types (3, 4). Thus, there is clearly a need for new therapies specifically targeting gastric cancer. A comprehensive molecular profile of gastric cancer would provide important information about the disease pathways and targets that could facilitate the development of new therapeutic agents and strategies (5, 6).

So far, most transcriptional profiling studies in gastric cancer have used hybridization microarrays. For example, aberrant miRNA expression signatures in gastric tumor samples from Japanese and Italian patients have been reported using miRNA expression microarrays (7, 8), and mRNA expression signatures from Chinese patients have been recently reported using exon microarrays (9). RNA sequencing (RNA-seq) technology is rapidly supplanting hybridization-based approaches. This approach not only enables investigators to quantify gene expression levels but to simultaneously assess alternative splicing and gene fusion events and to detect nucleotide variations in transcribed regions (10, 11). Thus, multidimensional data sets from a single platform can generate a rich profile of cancer progression and development. In particular, whole-transcriptome RNA-seq provides a detailed and precise view of the entire spectrum of expressed transcripts for both mRNA and noncoding RNA.

Downloaded from cancerres.aacrjournals.org on July 31, 2013. © 2012 American Association for Cancer Research.

Published OnlineFirst March 20, 2012; DOI: 10.1158/0008-5472.CAN-11-3870

In this study, we generated comprehensive mRNA and miRNA profiles for Asian gastric tumors. First, we conducted transcriptome-wide, unbiased analyses of the RNA-seq data to identify different types of transcriptional aberrations (mRNA, miRNA, and somatic mutation candidates) to leverage the existing knowledge of the pathogenesis. Second, integrating the results of our multilayer analyses, we identified a potential role in cancer progression for *PRKAA2*, which encodes AMP-activated protein kinase (AMPK)α2, a subunit of the AMPK serine/threonine protein kinase complex involved in the regulation of cellular and organismal metabolism. We experimentally validated the expression changes of *PRKAA2* between early- and late-stage gastric cancers. Third, through functional studies in gastric cancer cell lines, we showed the translational relevance of *PRKAA2* as a potential therapeutic target. Our work provides a valuable information basis for elucidating the molecular mechanisms of gastric cancer progression and also represents a general framework for the more effective development of disease-focused therapeutic targets.

Materials and Methods

Sample collection and characterization

This is a retrospective study conducted in archival fresh frozen human tissue specimens obtained from the National Research Resource Bank Program of the Korea Science and Engineering Foundation in the Ministry of Science and Technology. Patients consented to the use of the tissue specimens for research purpose, and the Institutional Review Boards of the College of Medicine, Yonsei University (Seoul, Korea), and the University of Texas MD Anderson Cancer Center (Houston, TX) approved the use of the specimens. Histologic classification and tumor stage were reviewed by a pathologist at the Gene Bank at Yonsei University Severance Hospital. Among 82 initially enrolled gastric cancer cases, 24 tumors and 6 noncancerous gastric tissues that met the criteria (sufficient amount and quality of RNA) were included in this study. Clinical and histopathologic characteristics obtained from the patients are summarized in Supplementary Table S1.

RNA-seq library preparation and SOLiD sequencing

The Miravana Kit (Ambion/Applied Biosystems) was used to isolate total RNA according to the vendor's protocol. The whole-transcriptome sequencing (WT-seq) and small RNA-seq libraries were prepared by the small RNA expression kit (SREK, PN 4397682) of Applied Biosystems Inc. (ABI), based on SOLiD whole-transcriptome and small RNA standard protocols provided by ABI. The individual prepared "barcode" libraries were quantified and pooled equally together for multiplexing. The sequencing runs were carried out on SOLiD v 3.0 for both WT-seq and small RNA-seq. WT-seq samples were sequenced in 1/4 slide per sample using 50-nucleotide (nt) single tags; and small RNA-seq samples were sequenced in 1/10 slide per sample using 35-nt single tags. Detailed information is provided in the Supplementary Materials and Methods. The RNA-seq data were deposited in the National Centre for Biotechnology Information Gene Expression Omnibus database (accession number GSE36968).

Computational analyses of RNA-seq data

WT-seq short reads were mapped to the human reference genome (hg19) and exon junctions (defined as RefSeq gene annotation) using the ABI Bioscope (version 1.21) WT-seq analysis pipeline with default parameters. The reads mapped to the sequences that were not of biologic interest, such as rRNAs, tRNAs, and repetitive elements, were first filtered. Then mapped reads with mapping quality 10 or more were defined as uniquely mapped reads and used in the downstream analysis. The SOLiD System Small RNA Analysis Pipeline Tool (corona RNA2MAP version 0.50) was used to analyze small RNA-seq reads: after filtering, the reads were mapped to mature miRNAs in miRBase (verson13.0; ref. 12) and the human reference genome, respectively.

To identify gastric cancer–related differentially expressed genes, the "reads per kilobase of exon per million mapped sequence reads" (RPKM; ref. 13) values of the human RefSeq genes were calculated using the RNA-seq flow in the Partek Genomics Suite (version 6.5 beta, Partek Inc.) and then log transformed. Single-factor ANOVA was used to detect differentially expressed genes among 18,890 protein-coding genes: $P < 9.5 \times 10^{-4}$ [false discovery rate (FDR) < 0.05] was used as the cutoff value in the 5-group comparison (normal, tumor stage I, II, III, or IV); and $P < 7\times10^{-4}$ was used in the 4-stage comparison (tumor stage I, II, III, or IV). A similar analysis was conducted on 2,569 long noncoding RNAs. Cross-platform gene expression comparison was conducted with a recent microarray study in gastric cancer (14). A gene ontology analysis was conducted using GoMiner (15) and a disease association analysis was conducted with Ingenuity Pathway Analysis software (version 7.0). Recurrent somatic mutations were identified on the basis of a recent exome-sequencing study on Asian patients with gastric cancer (16). miRNA expression was quantified as reads per million (RPM) of reads mapped to known miRNAs, and key differentially expressed miRNAs were defined on the basis of their differential expression and expression anticorrelation with their potential target genes. Detailed information is provided in the Supplementary Materials and Methods.

Experiments on biologic function

Cell culture. Human NCI-N87 and AGS gastric cancer cells were obtained from the American Type Culture Collection (ATCC; http://www.atcc.org/). The study was conducted within 6 months of resuscitation, and they were cultured in RPMI-1640 (CellGro) and 10% fetal calf serum (FCS; Hyclone) at 37°C in 5% CO_2. ATCC uses short tandem repeat (STR) profiling. For a hypoxia assay, culture flasks were incubated for various times at 37°C in humidified air, 5% (normoxia), or 1% O_2, 5% CO_2, and 94% N_2 (hypoxia) using an *in vivo* Hypoxia Workstation 500 with Ruskin hypoxic gas mixer (Biotrace International). Cells (2.5×10^5) were seeded and incubated under normoxic conditions to 70% confluence and then incubated under hypoxic conditions for 18 hours in the presence or absence of metformin at 10 mmol/L concentration. The NCI-N87-HRE cells were established according to the manufacturer's protocol, Cignal HIF Reporter (luc; SA Biosciences, QIAGEN Co.). Firefly

Downloaded from cancerres.aacrjournals.org on July 31, 2013. © 2012 American Association for Cancer Research.

Published OnlineFirst March 20, 2012; DOI: 10.1158/0008-5472.CAN-11-3870

luciferase activity was measured by the Dual Luciferase Reporter Assay System (Promega) according to the manufacturer's protocol.

Western blotting. Cells were grown under hypoxic conditions in the presence or absence of 10 mmol/L metformin. The cells were washed twice in a PBS solution and Western blotting was conducted, as previously described (17).

Real-time reverse transcriptase PCR. Total RNA was isolated from cell lysates by the PARIS Kit (Ambion/Applied Biosystems) according to the manufacturer's protocol. Next, TaqMan quantitative reverse transcriptase (RT)-PCR was carried out on the ABI 7300 system by the TaqMan one-step RT-PCR Master Mix kit and predesigned primer/probe pairs for *PRKAA1*, *PRKAA2*, *PRKAB1*, *PRKG1*, *STK11*, HNF4α, and β_2-microglobulin (Applied Biosystems). Normalization procedures and analyses were carried out with β_2-microglobulin using the $2(-\Delta\Delta C_t)$ method as the internal reference (18) and using Applied Biosystems GeneAmp 5700 SDS software. All measurements were carried out in triplicate.

Short interfering RNA transfection. Short interfering RNA (siRNA) SMARTpool sequences were obtained from Dharmacon/Thermo Fisher Scientific. The cells were transfected with 25 nmol/L siRNA-*PRKAA2*, siRNA-*PRKAB1*, siRNA-*PRKAG1*, siRNA-*PRKAG2*, or a siRNA nontargeting control with Dharma-FECT 1 lipid transfection reagent. The transfection medium was removed after 24 hours and replaced with fresh medium, and the cells were grown in 5% CO_2 at 37°C for an additional 48 to 72 hours. RT-PCR and/or Western blot analyses were conducted to confirm target knockdown by siRNA. The transfected cells were treated with metformin and cultured under hypoxic conditions for an additional 18 hours.

Statistical analysis. For the experiments on individual genes in this section, statistical significance ($P < 0.05$) was determined using the Student t test to compare data points with control data.

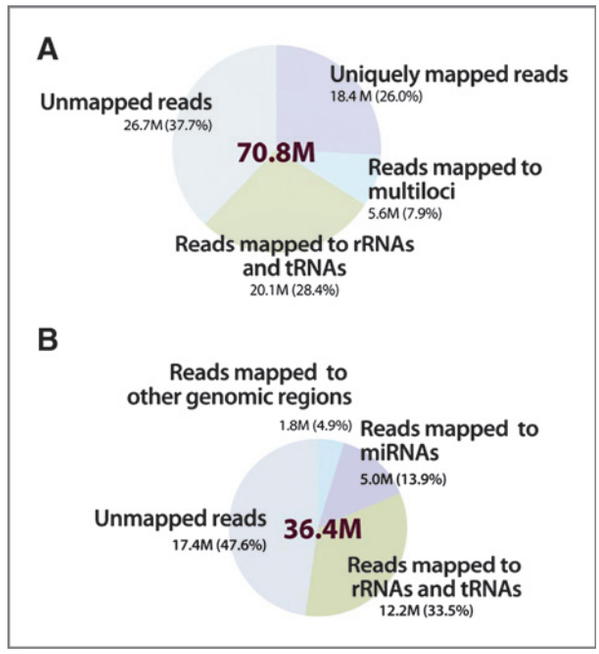

Figure 1. Distribution of short reads in the WT-seq and small RNA-seq data sets when mapped to the human reference genome. A, WT-seq data set of 2.1 billion 50-nt short reads from the 30 samples. Using the ABI Bioscope (version 1.21) WT-seq analysis pipeline, 62.3% of the short reads were mappable: 28.4% were mapped to the sequences of no biologic interest (e.g., rRNAs and tRNAs) and were therefore filtered; and 26.0% (18.4 million per sample) were uniquely mapped to the human reference genome (hg19) or exon junctions. B, Small RNA-seq data set of 894 million 35-nt short reads, mean of 36.4 million reads from the 25 samples. Using the SOLiD System Small RNA Analysis Pipeline Tool, 52.4% of the short reads were mappable; 33.6% mapped to sequences of no biologic interest were removed from further analysis. On average 13.9% [5.1 million (M) per sample] were mapped to known miRNAs in the miRBase database (version 13.0).

Results

Overview of the gastric cancer RNA-seq data

Using Life Technologies SOLiD sequencing platform, we conducted transcriptome-wide profiling of gastric cancer samples from 30 anonymous, unrelated Asians of both sexes. Included were 6 noncancerous gastric tissue samples and 24 gastric tumor samples that represented stages I through IV of tumor development (Clinical characteristics of patients are shown in Supplementary Table S1). Applying 2 protocols that complementarily cover RNA fragments of different sizes from each sample resulted in 2 parallel large-scale data sets that allowed us to simultaneously measure mRNA and miRNA expression.

From the WT-seq protocol we generated a WT-seq data set of 2.1 billion 50-nt short reads from the 30 samples (mean of 70.8 million; Supplementary Table S1). Using the ABI Bioscope WT-seq analysis pipeline, 62.3% of the short reads, on average, were mappable: 28.4% were mapped to sequences of no biologic interest for this study (e.g., rRNAs and tRNAs), so we removed them from further analysis. On average, 26.0% (18.4 million per sample) of the short reads were uniquely mapped to the human reference genome (hg19) or exon-junction sequences (0.5 million per sample); we used these in the downstream analysis (Fig. 1A and Supplementary Table S1).

Applying the second small RNA-seq protocol to 19 gastric tumor samples (5 of the original 24 yielded insufficient sample amounts) and 6 noncancerous gastric tissue samples resulted in a small RNA-seq data set of 894 million 35-nt short reads (mean of 36.4 million; Supplementary Table S1). Using the SOLiD System Small RNA Analysis Pipeline Tool, 52.4% of the short reads, on average, were mappable: 33.6% were mapped to sequences of no biologic interest, as in the whole-transcriptome analysis, and we removed them from further analysis. On average, 13.9% (5.1 million per sample) were mapped to known miRNAs in the miRBase database (Fig. 1B and Supplementary Table S1; version 13.0; ref. 12). To our knowledge, this is one of the largest RNA-seq data sets on human cancer available to date.

Identification of gastric cancer–related differentially expressed genes

With millions of short reads mapped to the human genome/transcriptome, we quantified the expression levels of known

Downloaded from cancerres.aacrjournals.org on July 31, 2013. © 2012 American Association for Cancer Research.

Published OnlineFirst March 20, 2012; DOI: 10.1158/0008-5472.CAN-11-3870

Table 1. Gene ontology analysis of gastric cancer–related differentially expressed genes

GO term	Description	Enrichment fold	Log_{10} (*P*)	FDR
Biologic process				
GO:0007586	Digestion	6.82	−6.39	0.002
GO:0006599	Phosphagen metabolism	19.35	−3.55	0.059
GO:0006599	Ion transport	2.02	−3.80	0.071
Molecular function				
GO:0022891	Substrate-specific transmembrane transporter	1.98	−3.66	0.052
GO:0022804	Transmembrane transporter	2.63	−3.75	0.055
GO:0031420	Metal ion binding	3.11	−3.79	0.061
GO:0015662	ATPase activity	6.45	−3.57	0.059

NOTE: Biologic characteristics of the 356 differentially expressed genes identified using GoMiner. To adjust the bias toward highly expressed genes in detecting differential expression analysis, we used as the reference set 12,213 genes with the same expression distribution as that of the 356 differentially expressed genes.
Abbreviation: GO, Gene ontology.

genes in each sample using the conventional parameter, RPKM (13). Among 18,890 annotated RefSeq coding genes, 15,421 genes, on average, per sample had detectable expression (RPKM > 0.05, about one mapped read for a gene with 1kb exons; Supplementary Table S1). To evaluate the reproducibility of our RNA-seq approach on expression quantification, for a subset of samples with available microarray data from a previous study (14), we compared the gene expression data from the 2 platforms (microarray and RNA-seq) and found sample-by-sample correlations, $R_S = 0.73 \pm 0.04$ (Spearman rank correlation, Supplementary Fig. S1A) that are similar to those reported in the literature (11, 13). The clustering pattern from an unsupervised analysis largely reflected the disease/staging status of the samples under survey (Supplementary Fig. S2). Moreover, a principal component analysis of the global gene expression profiles showed that gastric cancer samples differed from gastric intestinal stromal tumor samples, thereby confirming the pathologic classification of our tumor samples (Supplementary Fig. S3).

To identify gastric cancer–related differentially expressed genes, we conducted a single-factor ANOVA on the RPKM (log-scale transformed) data of 18,890 protein-coding genes across 5 groups (tissue type = normal or stage I, II, III, or IV as the independent variable). At FDR < 0.05 (ref. 19; $P < 9.5 \times 10^{-4}$), we identified 356 differentially expressed genes (Supplementary Table S2). As an independent validation, we conducted the same analysis on a published microarray data set (14) consisting of 83 Asian gastric samples taken from normal tissue and 4 tumor stages. In that analysis, 78% of the differentially expressed genes identified by our RNA-seq analysis showed significant expression changes in the same direction (tumor vs. normal) based on the microarray data; whereas only 17% would be expected if a same-size gene set were randomly chosen ($P < 1 \times 10^{-4}$; Supplementary Fig. S1B). In addition, we conducted a similar analysis on 2,569 long noncoding RNA genes and identified 7 differentially expressed genes (Supplementary Table S2).

To identify biologic characteristics of the differentially expressed genes, we conducted a gene ontology analysis using GoMiner (15). We found that the 356-gene data set showed high enrichment of genes involved in the biologic processes of digestion and phosphagen metabolism and the molecular functions of transmembrane transport and ATPase activity (Table 1). Using the Ingenuity Pathway Analysis software, we identified 101 of the 356 genes (28.4%) as associated with cancer (the Fisher exact test, $P = 2.8 \times 10^{-4}$), and 46 of them (12.9%) as related to gastrointestinal disease (the Fisher exact test, $P = 1.1 \times 10^{-4}$). Figure 2A categorizes these 356 genes by biologic and molecular functions (Supplementary Table S2). Because highly expressed genes tend to be identified in RNA-seq–based differential analysis (20), we used 12,213 genes with the same expression distribution as that of the 356 genes, rather than the whole gene set, as a reference set in the above analyses (see Supplementary Methods and Materials). As shown in Fig. 2B, these differentially expressed genes provided substantial power for classifying normal versus tumor tissue, as well as for distinguishing different clinical stages of the gastric tumors (although the distinction between stages III and IV became a little fuzzy). In comparison with the expression levels in the normal tissue, genes related to gastrointestinal disease were downregulated in 4 tumor stages (Fig. 2B). For example, loss of expression of gastrokines *GKN1* and *GKN2* occurs frequently in gastric adenocarcinoma, which is associated with shorter overall survival in the intestinal subtype of distal gastric cancer (21). Our results showed a dramatic tumor-related loss of expression level in both *GKN1/2* and trefoil factor family peptides 1/2 (*TFF1/2*), supporting their potential use as predictive biomarkers (9).

In addition, we conducted a similar analysis on the 24 tumor samples using single-factor ANOVA (tissue type = stage I, II, III, or IV). In general, the expression variation among tumor samples was much less than that between normal and tumor samples. At $P < 7 \times 10^{-4}$, we identified 28 genes with significant stage-specific expression change. On the basis of these genes,

Downloaded from cancerres.aacrjournals.org on July 31, 2013. © 2012 American Association for Cancer Research.

Published OnlineFirst March 20, 2012; DOI: 10.1158/0008-5472.CAN-11-3870

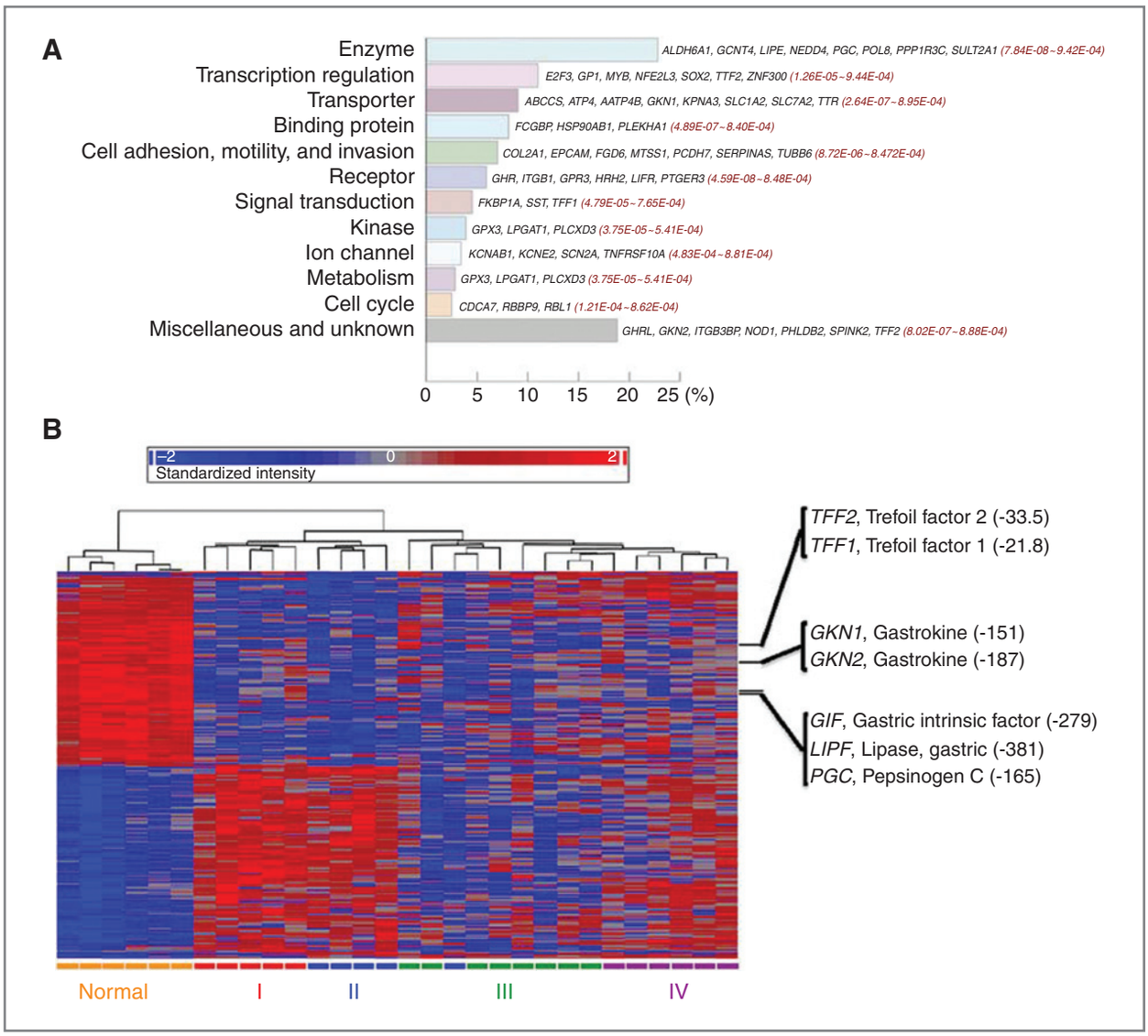

Figure 2. Biologic and molecular characteristics of 356 differentially expressed genes related to gastric cancer. These genes were identified with a single-factor ANOVA on RPKM (log-scale transformed) across 5 groups (tissue type = normal or stage I, II, III, or IV as the independent variable; at FDR < 0.05 and $P < 9.5 \times 10^{-4}$). A, biologic and molecular functions of representative genes known to be related to gastric cancer. Their raw P values in the ANOVA are shown in red. B, the clustering heatmap of 30 samples based on the 356 differentially expressed genes, generated with Partek Genomics Suite v 6.5. Each column is labeled with different colors according to the sample type; several key genes related to gastrointestinal disease are highlighted with their P values and fold changes (FC) for differential expression.

the tumor samples clearly clustered according to their stage, with the largest distinction between stages I, II versus III, IV (Supplementary Fig. S4).

Identification of recurrent somatic mutation candidates in gastric cancer

To take full use of our RNA-seq data, we also made efforts to identify somatic mutation candidates in gastric cancer. Because we did not sequence the normal DNA from the same patients, we had a very limited power to infer somatic mutations based on our RNA-seq data alone. Instead, we took advantage of a recent exome-sequencing study on Asian patients with gastric cancer (16) and obtained a list of 2,651 somatic mutations with a potential functional effect (nonsynonymous/nonsense mutations and those at splicing sites). Among these reported somatic mutation positions, we detected the exact mutant alleles at 14 mutation positions in our WT-seq data, suggesting that they are recurrent somatic mutations (Supplementary Table S3). *TP53* is the only gene with multiple recurrent mutation candidates (four mutations), consistent with its known high mutation frequency in gastric cancer (22). In addition, we detected 92 potential recurrent coding somatic mutations based on the COSMIC database (ref. 23; although

Downloaded from cancerres.aacrjournals.org on July 31, 2013. © 2012 American Association for Cancer Research.

Published OnlineFirst March 20, 2012; DOI: 10.1158/0008-5472.CAN-11-3870

these recurrent mutations may not be specific to gastric cancer, Supplementary Table S4).

Identification of key differentially expressed miRNAs related to gastric cancer

In parallel with the analysis of gene expression using the WT-seq data set, we used the small RNA-seq data set to quantify the expression levels of known miRNAs in each sample using RPM, a measure analogous to RPKM for coding genes. Among 698 annotated, nonredundant mature miRNAs in the miRBase (version 13.0), on average, approximately 60% of the miRNAs had mapped reads. The observation that a large proportion of annotated miRNAs (40%) have zero or very low expression levels may be partially due to the inflation of the current miRBase annotation, as suggested by recent studies (24, 25). Therefore, we focused on 402 miRNAs with reliable expression (max RPM > 4 in 25 samples) in the subsequent analysis (see Supplementary Methods and Materials).

To identify the miRNAs that play a key role in gastric tumor development, we reasoned that (i) the key miRNAs themselves should show significant expression variations across different sample groups; and (ii) they should have detectable repression effects on the expression of their target genes (26). Therefore, we conducted a 2-step analysis. First, we conducted single-factor ANOVA (tissue type as the independent variable) on the RPM data (log-scale transformed) across 5 groups (normal tissue and four tumor stages). At $P < 0.01$ (FDR < 0.15), 26 miRNA genes showed significant differential expression: 9 upregulated and 17 downregulated (Fig. 3A and Supplementary Table S5). One of the upregulated miRNAs, miR-21 is the most commonly upregulated miRNA in both solid and hematologic tumors (8). Seven of the 26 differentially expressed miRNAs (Supplementary Table S5) were also identified by a recent microarray study on the miRNA biomarkers for the progression/prognosis of gastric cancer using samples from Japanese patients (7).

Second, for each of the 26 miRNAs, we used the Spearman rank correlation (*R*s) to quantify its expression correlations with protein-coding genes across 25 samples with both the available coding gene and the miRNA expression data. We then tested whether the *R*s values of its potential target genes were significantly lower than those of other genes using the Wilcoxon rank sum test, resulting in 6 miRNAs showing significant anticorrelation with their potential targets (the Wilcoxon rank sum test $P < 0.01$, see Supplementary Methods and Materials). We defined these 6 miRNAs as key differentially expressed miRNAs of Asian gastric cancer (Fig. 3B).

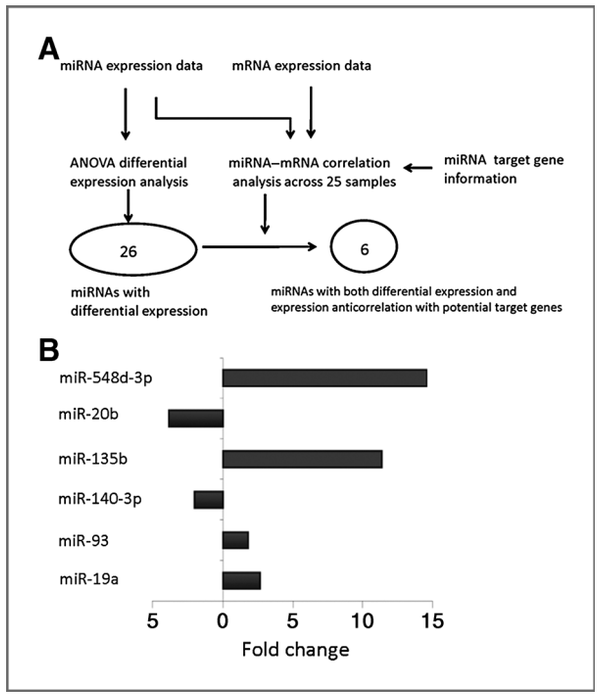

Figure 3. Integrative approach to identify key differentially expressed miRNAs related to gastric cancer. A, two criteria were used to identify key differentially expressed miRNAs: based on the ANOVA on miRNA expression data, at $P < 0.01$, 26 miRNA genes showed significant differential expression; and integrating miRNA expression, mRNA expression and miRNA target information, at $P < 0.01$, 6 of the 26 miRNAs showed significant anticorrelation with their potential target genes. As a result, the 6 miRNAs were defined as key differentially expressed miRNAs related to gastric cancer. B, the expression fold change of the 6 miRNAs in the tumor samples related to normal samples.

Integrative analysis suggests a potential role of *PRKAA2*, an AMPK activator in early-stage gastric cancer

To identify candidate genes with the highest potential functional impact in gastric tumorigenesis, we surveyed all 3 types of transcriptional aberrations: (i) differentially expressed genes; (ii) genes related to recurrent somatic mutation candidates; and (iii) potential target genes of key differentially expressed miRNAs. Through a simple scoring analysis, *PRKAA2* (AMPKα2) was the only gene identified by all the 3 criteria, suggesting that it is a potential key modulator in gastric cancer progression (Supplementary Fig. S5). *PRKAA2* is the catalytic subunit of AMPK, which is a heterotrimer consisting of a catalytic α-subunit and regulatory noncatalytic β- and γ-subunits, each with 2 or 3 isoforms (27). AMPK is an energy-sensing enzyme and plays a central role in the regulation of energy homeostasis (28).

The most compelling evidence is that *PRKAA2* is among the most highly differentially expressed genes identified in both 5-group and 4-stage expression analyses (5-group ANOVA, $P = 4.65 \times 10^{-6}$; 4-stage ANOVA, $P = 4.7 \times 10^{-4}$; only 22 genes were identified by both analyses). *PRKAA2* showed a differential loss of mRNA level in tumor stage I/II relative to normal or tumor stage III/IV (Supplementary Fig. S6). Using qRT-PCR on an independent set of Asian gastric cancer samples (including normal, stage I through IV cases), we validated our findings by comparing the expression patterns of *PRKAA2* along with the other 3 subunits of AMPK, *PRKAA1*, *PRKAB1*, and *PRKAG1*, for which we had not observed changes in the RNA-seq experiment (Fig. 4).

Moreover, through our mRNA:miRNA integrative analysis (Fig. 4), *PRKAA2* mRNA expression is inversely correlated with miR-19a expression (Pearson correlation test, $P < 0.02$), one of

Downloaded from cancerres.aacrjournals.org on July 31, 2013. © 2012 American Association for Cancer Research.