Supplementary File

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

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

Supplementary Methods and Materials

Pathological classification of tumor stage

Briefly, a pathologist at the Gene Bank at Yonsei University Severance Hospital accepted the resected specimens, and then harvested and stored a portion of tumor and normal tissues according to standard operating protocols. Tumor staging was based on the 6th Edition of the UICC/AJCC TNM classification system. Patients with advanced gastric cancer who underwent curative gastrectomy received a more extensive lymph node dissection. The Institutional Review Board of the Yonsei University Health System authorized the use of these tissues for research purposes. Eighty-two gastric cancer cases (tumor and normal pairs) were initially enrolled in the study, 24 tumors and 6 normal tissues that met the criteria (sufficient amount and quality of mRNA) were included in the study.

Library preparation and experimental procedure for SOLiD sequencing

The Miravana Kit (Ambion/Applied Biosystems, Foster City, CA, USA) was used to isolate total RNA according to the vendor's protocol. The isolated total RNA was then quantified using Nanodrop (Thermo Fisher Scientific, Waltham, MA, USA) and qualified using Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). The WT-seq and small RNA-seq libraries were prepared using the small RNA expression kit (SREK, PN 4397682) of Applied Biosystems Inc. (ABI), based on SOLiD WT and small RNA standard protocols provided by ABI.

Library preparation for WT-sequencing: The rRNA was depleted from the total RNA using the Invitrogen Ribominus Eukaryotic Kit (PN A1083708; Life Technologies Corp., Carlsbad, CA, USA); RNase III was used to fragment 0.5-1 ug of rRNA-depleted total RNA.

Then the fragmented rRNA-depleted total RNA was hybridized and ligated with adaptor mix A from the SOLiD small RNA Expression kit (SREK kit, PN4397682). Next, reverse transcription was performed to generate cDNA templates. Size-selection was used to obtain the cDNA in 100-200 nts from Novex 6% TBE-urea gel (Invitrogen). Then the excised gel piece containing 100-200 nt DNA was vertically split into four pieces using a clean razor blade. The selected cDNA was further amplified using the supplied primer sets containing a 6-bp "barcode" sequence on the 3' (reverse) primer and ~12-15 cycles of PCR. Next, the purified PCR products were arranged as a library in size ranges of 150-250 bp, containing 50-150 bp cDNA inserts, quantitated and qualitated by Agilent Bioanalyzer 2100. This readied the purified PCR products for the next steps of emulsion PCR in the preparation of template beads.

Library preparation for small RNA sequencing: The adaptor mix A provided in the small RNA expression Kit (SREK) was used to hybridize the sample containing small RNA. The adaptor mixes were sets of RNA/DNA oligonucleotides with a single-stranded degenerate sequence at one end and a defined sequence required for SOLiD sequencing at the other end. Hybridizing and ligating the sample with adaptor mix A sequentially yielded the template for SOLiD sequencing from the 5' ends of the small RNA. Next, the small RNA population was reversely transcribed with ligated adaptor to generate cDNA. To meet the sample quantity requirement for SOLiD sequencing and to append the required terminal sequences to each molecule, the cDNA library was amplified using one of the supplied primer sets containing a 6-bp "barcode" sequence on the 3' (reverse) primer and ~12-15 cycles of PCR. In preparation for the subsequent step of emulsion PCR, in which the molecules were attached to beads, the individual library PCR products were purified, and electrophoresis on 6% polyacrylamide gel

was used to size-select them in the range of 108-135 bp PCR products containing small RNA of 18-40 nt inserts.

Template beads preparation: The individual prepared "barcode" libraries were quantified and pooled equally together for multiplexing. Four barcoded individual WT-seq libraries were pooled together and ten individual barcoded small RNA-seq libraries were pooled together as templates for emulsion PCR; the template molecules were attached to beads, as described in the ABI SOLiD emulsion PCR protocol. After emulsion PCR, the template beads were enriched by adaptor P2-polysteryline beads in 60% glycerol gradient by centrifugation. The P2-enriched template beads were further modified for bead immobilization to the slide and prepared for deposit on substrate-coated glass slides, as described in the ABI protocol.

SOLiD sequencing: The sequencing runs were performed on SOLiD v 3.0 for both WT-seq and small RNA-seq. The library template beads were titrated by Work Flow Analysis run to determine the percentage of P2-positive beads in the total template of beads before they were deposited onto slides for sequencing. Each full sequencing slide contained a deposit of 300 million P2-positive template beads. Multiplex sequencing with four samples was performed in 50 nts of WT-seq and multiplex sequencing of 10 small RNA samples in 35 nts of small RNA-seq.

Computational analyses of RNA-seq data

Short read mapping: For the WT-seq dataset (30 samples), short reads were mapped to the human reference genome (hg19) and exon junctions (defined as RefSeq gene annotation) using the ABI Bioscope^{TM} (version 1.21) WT-seq analysis pipeline with default parameters. The reads mapped to the sequences that were not of biological interest, such as rRNAs, tRNAs, and

repetitive elements, were first filtered. Then mapped reads with mapping quality ≥10 were defined as uniquely mapped reads and used in the downstream analysis.

For the small RNA-seq dataset (25 samples), the SOLiD System Small RNA Analysis Pipeline Tool (corona RNA2MAP version 0.50) was used: first, the reads mapped to the sequences of no biological interest were filtered, and then the remaining reads were mapped to annotated mature miRNAs and the human reference genome (key parameters: seed length = 18, seed error = 2, and maximum number of errors = 4). The human mature miRNA sequences were obtained from miRBase (verson13.0) (1), including 698 non-redundant mature miRNAs. Because miRNAs are highly redundant in families, both uniquely mapped reads and multireads (the reads equally aligned to multiple miRNAs) were included in the analysis.

Analysis of gastric-cancer-related differentially expressed genes: The RefSeq gene annotation file (hg19) was downloaded from the UCSC Genome Browser. Only genes in the main assembly were included (i.e., chr1-22, X, Y), and transcripts with multiple genomic loci were excluded. In total, there were 18,890 annotated protein-coding genes and 2,569 long non-coding RNA genes. The RPKM values of the human genes were calculated using the RNA-seq flow in the Partek® Genomics Suite™ (version 6.5 beta, Partek Inc., St. Louis, MO, USA).

Differentially expressed coding genes were detected using the Partek® Genomics Suite ANOVA. To stabilize the variance of RPKM, log transformation was applied to the raw values (0 was set to 0.001). In a single-factor ANOVA, tissue type was considered as the independent variable. As a result, 356 out of 18,890 genes were identified as gastric-cancer-related differentially expressed genes at $P < 9.5\times10^{-4}$ and $FDR < 0.05$ (2) in the five-group comparison (normal, tumor stage I, II, III or IV); and 28 genes were identified as stage-specific differentially expressed genes at $P < 7\times10^{-4}$ and $FDR < 0.35$ in the four-stage comparison (tumor stage I, II, III

or IV). To independently validate our results, the same five-group ANOVA was applied to a published microarray dataset (83 samples, including normal and tumor stages I, II, III and IV) (see figure legends in Supplementary Fig. 1). A similar analysis was performed on the 2,569 long non-coding RNAs, and $FDR < 0.1$ was used.

To identify the common theme of the 356 gastric-cancer-related differentially expressed genes, a gene ontology analysis was performed using GOminer (3) and a disease association analysis was performed using Ingenuity Pathway Analysis software (version 7.0). To avoid potential bias due to higher expression of differential expressed genes in the above enrichment analyses, a set of 12,213 genes with the same expression distribution (indexing the expression level of a gene by the median RPKM value across 30 samples) as that of the differentially expressed genes was used as the reference gene set.

Analysis of recurrent somatic mutation candidates in gastric cancer: A list of somatic mutations in Asian patients of gastric cancer was obtained from a recent exome-sequencing study (4). Only mutations with a potential functional effect (nonsynonymous/nonsense mutations and those at splicing sites) were retained. Based on WT-seq read mapping, nucleotide positions with variant alleles (relative to the human reference genome) were identified for each sample using pileup command in SAMtools (5). Given the reported somatic mutation positions, potential recurrent somatic mutations were identified if the exact mutant alleles were observed. We also performed a similar analysis based on the somatic mutations present in the COSMIC database (6).

Analysis of key differentially expressed miRNAs related to gastric cancer: Through the read mapping step, short reads of small RNA data were mapped to mature miRNA sequences. To count the reads mapped to each miRNA in a sample, uniquely mapped reads were first counted and then the multireads were split to the potential loci according to the amount of uniquely

mapped reads in these loci. In view of many annotated miRNAs with zero or extremely low expression level, miRNAs were ranked based on their maximal expression level across the 25 samples and focused on the top 402 miRNAs in the subsequent analysis. To identify differentially expressed miRNAs, a single-factor ANOVA was performed on the log-transformed RPM of miRNAs, with the tissue type as the independent factor (normal, tumor stages I, II, III or IV) using the Partek® Genomics Suite. At $P < 0.01$, 26 miRNAs were identified. To identify the miRNAs with detectable repression effect on their potential target genes, for each pair of miRNA and protein-coding genes with 3' UTR (16,882 genes), Spearman's rank correlation (*R*s) was used to quantify their expression correlations across the 25 samples with both the coding gene and the miRNA expression data. For each of the 26 miRNAs, the 16,882 protein-coding genes were classified as its potential target genes or not based on whether the 3' UTR contained the corresponding seed-matched 7-mers, as proposed by Farh et al. (7). Then Wilcoxon's rank sum test was used to determine whether the *R*s values of potential target genes of a miRNA were significantly lower than those of other genes using. At $P < 0.01$, six miRNAs were defined as key differentially expressed miRNAs.

Cells and reagents for functional experiments

Human NCI-N87 and AGS gastric cancer cells were obtained from the American Type Culture Collection (http://www.atcc.org/). Monoclonal antibodies to human HIF-1α were purchased from BD Transduction Laboratories™ (BD Biosciences, San Jose, CA, USA); AMPK-α, LKB1, and HNF4α were purchased from Cell Signaling (Cell Signaling Tech. Inc., Boston, MA, USA); and β-actin was purchased from Santa Cruz Biotechnology (Santa Cruz, CA, USA). Metformin was purchased from Sigma-Aldrich (St. Louis, MO, USA).

Supplementary References

1. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res 2006; 34: D140-4.

2. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological 1995; 57: 289-300.

3. Zeeberg BR, Feng W, Wang G, et al. GoMiner: a resource for biological interpretation of genomic and proteomic data. Genome Biol 2003; 4: R28.

4. Wang K, Kan J, Yuen ST, et al. Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. Nat Genet 2011.

5. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009; 25: 2078-9.

6. Forbes SA, Bhamra G, Bamford S, et al. The Catalogue of Somatic Mutations in Cancer (COSMIC). Curr Protoc Hum Genet 2008; Chapter 10: Unit 10 1.

7. Farh KK, Grimson A, Jan C, et al. The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. Science 2005; 310: 1817-21.