Identification of genes universally differentially expressed in gastric cancer

CURRENT STATUS: POSTED

Haidan Yan
Fujian Medical University

Yidan Shi
Fujian Medical University

Jiahui Zhang
Fujian Medical University

Haifeng Chen
Fuzhou Second Hospital

Qingzhou Guan
Fujian Medical University

Jun He
Fujian Medical University

Meifeng Li
Fujian Medical University

Zheng Guo
Fujian Medical University

Ping Li
Fujian Medical University Union Hospital

DOI:
10.21203/rs.2.16787/v1

SUBJECT AREAS
Gastroenterology & Hepatology

KEYWORDS
Cancer genes, Differentially expressed genes, Frequency, Gastric cancer, Individual-
Abstract

Background Due to the remarkable heterogeneity of gastric cancer (GC), differentially expressed genes (DEGs) identified at the population-level by case-control comparison cannot afford dysregulation frequency of each DEG in GC.

Methods Firstly, the individual-level DEGs were identified for 1,090 GC tissues without paired normal tissues by the RankComp method. Secondly, we directly compared the gene expression in a cancer tissue with its paired normal tissue to identify individual-level DEGs for 448 paired cancer-normal gastric tissues.

Results We found 25 DEGs dysregulated not only in more than 90% of 1,090 GC tissues, but also in more than 90% of 448 GC tissues with paired normal tissues. The 25 genes were defined as universal DEGs for GC which were further validated in our additionally measured 24 paired cancer-normal gastric tissues. Among the universal DEGs, four up-regulation genes (BGN, E2F3, PLAU and SPP1) and one down-regulation gene (UBL3) were found to be cancer genes documented in the COSMIC or F-Census database. By analyzing protein-protein interaction network, we found 12 universal up-regulation genes and their 284 direct neighbor genes were significantly enriched with cancer genes and key biological pathways related to cancer, such as MAPK signaling pathway, Cell cycle and Focal adhesion. The 13 universal down-regulation genes and 16 direct neighbor genes were also significantly enriched with cancer genes and gastric acid secretion related pathway.

Conclusion These universal DEGs may be of special importance for GC diagnosis and treatment targets, and can help to study the molecular mechanism of GC.

Background

Gastric cancer (GC) is one of the most frequent malignant tumors with a high mortality rate[1-3]. The inter-individual molecular heterogeneity of GC is a big obstacle for its clinical diagnosis and treatment. It’s of special significance to find common molecular biomarkers for GC. Currently, various methods, such as SAM[4], limma[5, 6] and RankProd[7], have been used to identify the population-level differentially expressed genes (DEGs) between a group of GC tissues and a group of normal controls. However, the inter-individual heterogeneity of DEGs didn’t be considered in these methods.
Therefore, for a given population-level DEG, we cannot know whether it’s frequently dysregulated in a particular type of cancer.

Currently, three methods have been proposed to tackle this difficulty. First, the differential expression status of each gene was identified in a cancer tissue through comparing with the average expression level of the same gene in a group of normal tissues[8-10]. However, because the gene expression levels of normal tissues exist large inter-individual variations[11-13], this method may easily be affected by biological variations. Second, paired cancer-normal samples were used to identify individual-level DEGs by comparing the gene expression level of a gene in a cancer tissue with that in its paired normal tissue[14]. Because paired cancer-normal samples are relatively scarce, the application scope of this method is limited. In any case, paired cancer-normal samples are valuable and reliable specimens for individual analysis. Third, we previously developed an algorithm, named RankComp, to identify the DEGs in each cancer tissue by comparing the relative expression orderings (REOs) in the cancer tissue with the highly stable REOs predetermined in a large collection of normal tissues[15]. This method allows us to identify DEGs for each cancer tissue in the absence of its paired normal tissue. RankComp is proposed based on the findings that the relative expression orderings (REOs) of genes are highly stable in a type of normal tissues but widely disrupted in the corresponding cancer tissues[15]. In brief, the disrupted REOs in a cancer tissue caused by the dysregulation of its DEGs were compared with the stable REOs predetermined in accumulated normal tissues to identify the DEGs for the cancer tissue[15]. To identify robust common biomarkers for GC, we not only analyzed samples with paired cancer-normal gastric tissues, but also analyzed samples with only GC tissues by RankComp.

In this study, the gene expression profiles of 448 paired cancer-normal gastric tissues and 1,090 GC tissues without paired normal tissues were collected from the Gene Expression Omnibus (GEO) database[16] and The Cancer Genome Atlas (TCGA) data portal[17]. Using 1,090 GC tissues without paired normal tissues, DEGs were identified in each cancer tissue by RankComp. For 448 paired cancer-normal gastric tissues, we directly compared the gene expression in a cancer tissue with its paired normal tissue to identify individual-level DEGs. Then, the up-regulation or down-regulation
frequency of each gene was calculated in both the 1,090 GC tissues and the 448 paired cancer-normal gastric tissues. The DEGs, dysregulated not only in at least 90% of the 1,090 GC tissues but also in 90% the 448 paired cancer-normal gastric tissues, were defined as universal DEGs for GC. These universal DEGs were further validated by 24 paired cancer-normal gastric tissues measured by us.

Methods

Samples and data measurement

We measured the gene expression profiles for 24 paired cancer-normal gastric tissues which were collected from the department of gastric surgery, Fujian Medical University Union Hospital. All gastric tissues were sampled from surgical resection and were fresh frozen. Trizol reagent (Invitrogen) was used to isolate total RNA from fresh frozen tissues according to the manufacture’s protocol. We used 1–2μg of total RNA for mRNA capture using NEBNextPolyA mRNA Magnetic Isolation Module. Using a NEBNext Ultra Directional RNA Library Prep Kit, the stranded RNA-seq libraries were constructed. The 2×150 paired-end sequencing was performed on an Illumina HiSeqXten (Illumina, US). We preprocessed the raw RNA-seq files (.fastq) by Trimmomatic[18], and aligned reads to the reference genome (GRCh37) by hisat2[19]. Finally, the reads per kilobase per million mapped reads (RPKM) values of genes were computed to represent the expression levels of genes using StringTie[20]. The study was approved by the Fujian Medical University Union Hospital Institutional Review Board of all participating institutions, and written consent forms were obtained from all participants.

Public data and preprocessing

Gene expression profiles of gastric tissues were downloaded from GEO[16] and TCGA (http://tcga-data.nci.nih.gov/tcga/)[17]. The datasets including paired cancer-normal gastric tissues or GC tissues without paired normal tissues were collected in this study, as described in Table 1.

For the gene expression profiles measured by the Affymetrix platform, we only downloaded the raw data (.CEL files). Then, the raw data was processed using the Robust Multi-array Average algorithm for background adjustment without quantile normalization[21]. For the datasets measured by the Illumina platform, the processed data was directly downloaded for the following analysis. Finally, we
mapped each probe ID to Entrez gene ID using the corresponding platform annotation file for the array-based data (Affymetrix and Illumina platforms). If a gene was mapped by multiple probes, the expression level of this gene was calculated as the arithmetic mean of these probes. For GC tissues measured by RNA-seq, the RPKM or FPKM values of genes were downloaded for the following analysis.

Table 1 about here

Identification of population-level DEGs

Student’s t-test was used to identify DEGs between GC tissues and GC-adjacent normal tissues. Here, we identified two lists of DEGs using two independent datasets, and evaluated the concordance score of the two lists of DEGs by the cumulative binomial distribution model[22]. The DEGs reproducibly detected by the two independent datasets were defined as population-level DEGs for GC.

Identification of individual-level DEGs

Here, we only analyzed the individual-level differential expression status for population-level DEGs to ensure that the DEGs were involved in GC. Firstly, a total of 74 normal gastric tissues measured by Affymetrix GPL570 and 32 normal gastric tissues measured by HiSeq_RNASeqV2 were used to identify significantly stable REOs for the two platforms, respectively. For normal tissues measured by a particular platform, a binomial distribution test was used to identify significantly stable REOs[23]. The consistent stable REOs in both Affymetrix GPL570 and HiSeq_RNASeqV2 were defined as cross-platform REOs. Based on the cross-platform stable REOs identified using normal gastric tissues, RankComp was used to identify DEGs in a cancer sample compared with its own previously normal state. The detailed description of the RankComp algorithm was described in Wang et al.[15]. Simultaneously, paired cancer-normal gastric tissues were also used to identify individual-level DEGs for each cancer tissue. A gene was defined as up-regulated if its expression level in the cancer tissue was higher than that in the paired normal tissue; Conversely, the gene was defined as down-regulated. Then, we could calculate the up- or down-regulation frequency for each DEG. The 95% confidence interval (CI) of the up- or down-regulation frequency of each DEG was also calculated by the binom.confint algorithm in R language.

Pathway enrichment analysis
The 223 pathways covering 6290 unique genes were downloaded from the KEGG database on October 21, 2018[24]. Here, the human disease pathways were excluded from this study. The hypergeometric distribution model was used to determine biological pathways that were significantly enriched with genes of interest[25].

**Network analysis**

Protein-protein interaction (PPI) data was collected from the SiGNaling Network Open Resource[26] database and a literature-curated human signaling network (version 6, http://www.cancer-systemsbiology.org/HumanSignalingNet_v6.csv)[27]. A non-redundant PPI network containing 6,920 proteins and 72,840 interactions were further extracted for our analysis.

**Results**

**Limitations of the population-level DEGs**

Firstly, we used two independent datasets (GSE29272 and GSE29998) to identify the population-level DEGs. Using the paired Student’s t-test with 5% FDR control, 9,340 and 4,928 DEGs were identified from GSE29272 and GSE29998, respectively (Supplementary Table S1). The two lists of DEGs had 3,002 overlaps, and 93.23% (2,799) of the overlapped genes showed the same dysregulation directions in the GC tissues compared with the paired normal tissues (binomial test, \( p < 1.0 \times 10^{-16} \)). Then, the 2,799 reproducible DEGs were defined as the population-level DEGs for GC.

Due to the heterogeneity of cancer, a DEG, detected at the population-level, cannot tell us whether it is dysregulated in a particular cancer tissue or not. To tackle this difficulty, some studies compared the expression levels of genes in a cancer tissue with the average expression levels of the same genes in a set of normal tissues[8–10]. However, this method may easily be affected by biological variations because of the large inter-individual variations of gene expression in normal tissues[11–13]. For example, in 49 paired cancer-normal gastric tissues from GSE29998, gene \( \text{ACACB} \), a population-level down-regulated gene in GC, was identified as up-regulation in four of the 49 GC tissues when comparing with its average expression level in the 49 paired normal gastric tissues (Figure 1A). However, \( \text{ACACB} \) should be identified as down-regulation in two of the four GC tissues when comparing with the expression levels in the corresponding paired normal tissues. Besides, as
shown in the green box in Figure 1A, the expression level of ACACB was higher in the GC tissue than that in the paired normal tissue but below the average expression level in the normal tissues, and thus could not be identified as up-regulation in the GC tissue. The similar phenomenon was also observed in another example gene ACADVL (Figure 1B).

Figure 1 is about here

Universal up- or down-regulation genes in GC tissues

A total of 1,090 GC tissues without paired normal tissues and 448 GC tissues with paired normal tissues were collected from GEO and TCGA databases (Table 1). Here, we defined the DEGs, dysregulated not only in at least 90% of GC tissues without paired normal tissues, but also in 90% of paired cancer-normal tissues, as universal DEGs for GC. Considering the existence of measurement variations and low quality of tissue samples, we defined the genes, dysregulated in at least 90% of GC tissues rather than 100% of GC tissues, as universal DEGs. To ensure that the identified DEGs are involved in GC, we only analyzed the 2,799 population-level DEGs of GC.

Using 74 normal gastric tissues detected by Affymetrix GPL570 and 32 normal gastric tissues detected by HiSeq_RNASeqV2 (Table 1), 195,153,494 and 182,078,993 significantly stable REOs were identified for the two platforms (binomial test, FDR<0.05), respectively. Among the 105,408,410 overlapping REOs identified in both two platforms, all showed the same REO patterns (binomial test, p < 1.0x10^{-16}). Based on these 105,408,410 across-platform stable REOs, RankComp was used to identify the differential expression status of the 2,799 population-level DEGs in each of the 1,090 GC tissues (see Methods). Simultaneously, the 448 paired cancer-normal gastric tissues, pooled from 14 datasets measured by 10 different platforms, were used to identify individual-level DEGs by comparison of the gene expression levels between a cancer tissue and its paired normal tissue. The up- or down-regulation frequency in the 448 GC tissues and the 95% CI of the frequency were calculated for each of the 2,799 population-level DEGs. Finally, 25 DEGs, including 12 up-regulation and 13 down-regulation genes, were identified in at least 90% of the 1,090 GC tissues and in 90% of the 448 GC tissues (Figure 2). The 25 genes were defined as universal DEGs for GC.

To further validate these universal DEGs, we measured gene expression profiles for 24 paired cancer-
normal gastric tissues using RNA-seq. Compared with paired normal gastric tissues, nine of the 12 universal up-regulation genes were up-regulated in more than 90% of the 24 GC tissues. The other three genes, CEMIP, OLFML2B and SPP1, were also up-regulated in 88% (21) of the 24 GC tissues (Figure 2). For the 13 universal down-regulated genes, nine genes were validated in more than 90% of the 24 paired cancer-normal tissues. The remained three genes were also validated in 88% (21) of the 24 paired cancer-normal tissues (Figure 2).

Figure 2 is about here

The functional analysis of universal genes

To further analyze the biological functions of common DEGs in GC, we mapped the 25 universal DEGs to PPI network (see methods). The 25 universal DEGs and corresponding 300 direct neighbor genes with interaction relations were extracted as a new sub-network for the following analysis.

For the 12 up-regulation genes, 284 genes were directly linked to them. The 296 genes were significantly enriched in 56 pathways (Supplementary Table S2), such as MAPK signaling pathway, Cell cycle and Focal adhesion (Figure 3A, hypergeometric distribution model, FDR < 0.05). Moreover, the 296 genes were significantly enriched with cancer genes documented in COSMIC[28] or F-Census database[29] (hypergeometric distribution model, p<0.05). Four of the 12 up-regulation genes, BGN, E2F3, PLAU and SPP1, were cancer genes. Among the the top four genes (NEK2, SPP1, PLAU, E2F3) with the largest degrees (Figure 3C), three genes were cancer genes except for the gene NEK2. Gene NEK2 had 106 direct neighbor genes, among which 30 genes were cancer genes (Figure 3E).

Moreover, NEK2 was found to be frequently up-regulated in pan-cancer[14]. NEK2, as a serine/threonine-protein kinase, is involved in mitotic regulation[30]. The over-expression of NEK2 can enhance the ability of Mad2 to cause delays in cell division[31] The dysregulation of NEK2 may play important roles in tumorigenesis and maybe an effective target for cancer treatment[32–34].

For the 13 down-regulation genes, 16 genes were directly linked to them. The 29 genes were significantly enriched in 15 pathways (Supplementary Table S2), such as Calcium signaling pathway and Gastric acid secretion (Figure 3B, hypergeometric distribution model, FDR < 0.05). Eight of the 29 genes were cancer genes documented in COSMIC[28] or F-Census[29], which can’t happen by chance
(hypergeometric distribution model, p<0.05). As shown in Figure 3, CCKBR was the gene with the largest degree among the universal down-regulation genes and directly interacted with two cancer genes (Figure 3D, F). CCKBR has been reported to play important roles in gastric carcinogenesis by regulating stem cell function and epithelial homeostasis[35].

Hypermethylation CpG sites within promoter of a gene is an important event leading to the down-regulation of this gene[36]. Here, we collected DNA methylation profiles of 115 paired cancer-normal gastric tissues from GSE30601 and GSE25869 datasets. A gene was defined as hypermethylated when more than 50% of the 115 GC tissues had higher methylation level than that in the paired normal tissues. For the 13 down-regulation genes, 10 universal down-regulated genes were hypermethylated, and two genes, ESRRG and AQP4, showed higher methylation level in 49% and 41% of the 115 GC tissues, respectively (Supplementary Table S3). The result showed that hypermethylation of gene promoter maybe driver events for the down-regulation of the 13 universal down-regulated genes.

Figure 3 is about here

Discussion

In this study, we aimed to find genes universally dysregulated in all or almost all patients of GC to deal with inter-individual heterogeneity. To find reliable universal DEGs, we not only analyzed samples with only GC tissues but also analyzed samples with paired cancer-normal gastric tissues. For samples with only GC tissues, RankComp was used to identify the DEGs in each cancer tissue. For samples with paired cancer-normal gastric tissues, we directly compared the gene expression in a cancer tissue with its paired normal tissue. Through calculating the dysregulation frequencies of genes, we found 25 DEGs not only dysregulated in more than 90% of the 1,090 GC tissues but also in more than 90% of the 448 GC tissues with paired normal tissues. Furthermore, the 24 paired cancer-normal gastric tissues additionally measured by us validated these 25 universal DEGs. Among the 25 universal DEGs, 12 genes were universally up-regulated in GC tissues and 13 genes were universally down-regulated in GC tissues.

Notably, only one gene (NEK2) of the 25 universal DEGs was found to be up-regulated in at least 90%
of 649 cancer tissues in pan-cancer analysis across 23 cancer types[14]. The pan-cancer analysis may miss GC-specific universal aberrations. For example, SNX10 was universally up-regulated not only in 93% of the 1,090 GC tissues but also in 91% of the 448 paired cancer-normal tissues. However, it was only highly expressed in 61% of the 649 cancer tissues in the pan-cancer analysis[14]. The up-regulation of SNX10 may lead to gastric acidification defects[37]. For another example, SLC7A8 was down-regulated in 98% of the 1,090 GC tissues and in 90% of the 448 paired cancer-normal tissues but only in 55% of the 649 cancer tissues in the pan-cancer analysis[14]. SLC7A8, also known as LAT2, is a member of the L-type amino acid transporter (LAT) family. Although the LAT family has been reported to play important roles in cancer development[38–41], few studies have validated the role of LAT2 in cancer cell growth. The biological functions of the universal DEGs should be further studied in the future, which may help to study the mechanism of GC.

In summary, identification of individual-level DEGs for GC tissues with and without paired normal tissues may make full use of all gastric cancer data to find reliable common biomarkers for GC. These common biomarkers may be of special importance for GC diagnosis and treatment targets.

Availability Of Data And Materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

List Of Abbreviations

GC: gastric cancer

DEGs: differentially expressed genes

REOs: relative expression orderings

GEO: Gene Expression Omnibus

TCGA: The Cancer Genome Atlas

RPKM: the reads per kilobase per million mapped reads

CI: confidence interval

PPI: protein-protein interaction

Declarations

Ethics approval and consent to participate
The study was approved by the Fujian Medical University Union Hospital Institutional Review Board.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests

Author Contributions
HY contributed to design, conceptualization, methodology and writing the original draft. YS contributed to acquisition of data, formal analysis, validation and revising the original draft. JZ contributed to analyze parts of the results and revise the manuscript. HC contributed to supervision, writing revising and editing. QG, JH and ML contributed to revise the manuscript. ZG and PL contributed to conceptualization, methodology and writing revising. All authors have read and approved the manuscript.

Funding
This work was supported by the National Natural Science Foundation of China [grant numbers. 61801118]. The education research project for young and middle-aged teachers in Fujian province [grant number. JAT170214] and the Fujian natural science foundation [grant number. 2019J01678]. These funds do not have their role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Acknowledgements
Not Applicable.

References
1. Chen WQ, Zhang SW, Zou XN et al. Cancer incidence and mortality in china, 2006, Chin J Cancer Res 2011;23:3–9.
2. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013, CA Cancer J Clin 2013;63:11–30.
3. Yang L, Parkin DM, Ferlay J et al. Estimates of cancer incidence in China for 2000 and projections for 2005, Cancer Epidemiol Biomarkers Prev 2005;14:243–250.
4. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation
response, Proc Natl Acad Sci U S A 2001;98:5116–5121.

5. Ritchie ME, Phipson B, Wu D et al. limma powers differential expression analyses for RNA-sequencing and microarray studies, Nucleic Acids Res 2015;43:e47.

6. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments, Stat Appl Genet Mol Biol 2004;3:Article3.

7. Breitling R, Armengaud P, Amtmann A et al. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments, FEBS Lett 2004;573:83–92.

8. Li J, Han L, Roebuck P et al. TANRIC: An Interactive Open Platform to Explore the Function of IncRNAs in Cancer, Cancer Res 2015;75:3728–3737.

9. Warnecke-Eberz U, Metzger R, Holscher AH et al. Diagnostic marker signature for esophageal cancer from transcriptome analysis, Tumour Biol 2016;37:6349–6358.

10. Shen X, Li S, Zhang L et al. An integrated approach to uncover driver genes in breast cancer methylation genomes, PLoS One 2013;8:e61214.

11. Andreassen CN, Alsner J, Overgaard J. Does variability in normal tissue reactions after radiotherapy have a genetic basis—where and how to look for it?, Radiother Oncol 2002;64:131–140.

12. Hsiao LL, Dangond F, Yoshida T et al. A compendium of gene expression in normal human tissues, Physiol Genomics 2001;7:97–104.

13. Hughes DA, Kircher M, He Z et al. Evaluating intra- and inter-individual variation in the human placental transcriptome, Genome Biol 2015;16:54.

14. Gross AM, Kreisberg JF, Ideker T. Analysis of Matched Tumor and Normal Profiles Reveals Common Transcriptional and Epigenetic Signals Shared across Cancer Types, PLoS One 2015;10:e0142618.

15. Wang H, Sun Q, Zhao W et al. Individual-level analysis of differential expression of genes and pathways for personalized medicine, Bioinformatics 2015;31:62–68.

16. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository, Nucleic Acids Res 2002;30:207–210.

17. Hudson TJ, Anderson W, Artez A et al. International network of cancer genome projects, Nature 2010;464:993–998.
18. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 2014;30:2114–2120.

19. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements, Nat Methods 2015;12:357–360.

20. Pertea M, Pertea GM, Antonescu CM et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads, Nat Biotechnol 2015;33:290–295.

21. Irizarry RA, Hobbs B, Collin F et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data, Biostatistics 2003;4:249–264.

22. Bahn AK. Application of binomial distribution to medicine: comparison of one sample proportion to an expected proportion (for small samples). Evaluation of a new treatment. Evaluation of a risk factor, J Am Med Womens Assoc 1969;24:957–966.

23. Guan Q, Chen R, Yan H et al. Differential expression analysis for individual cancer samples based on robust within-sample relative gene expression orderings across multiple profiling platforms, Oncotarget 2016;7:68909–68920.

24. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes, Nucleic Acids Res 2000;28:27–30.

25. Hong G, Zhang W, Li H et al. Separate enrichment analysis of pathways for up- and downregulated genes, J R Soc Interface 2014;11:20130950.

26. Perfetto L, Briganti L, Calderone A et al. SIGNOR: a database of causal relationships between biological entities, Nucleic Acids Res 2016;44:D548–554.

27. Wang E, Zaman N, McGee S et al. Predictive genomics: a cancer hallmark network framework for predicting tumor clinical phenotypes using genome sequencing data, Semin Cancer Biol 2015;30:4–12.

28. Futreal PA, Coin L, Marshall M et al. A census of human cancer genes, Nat Rev Cancer 2004;4:177–183.

29. Gong X, Wu R, Zhang Y et al. Extracting consistent knowledge from highly inconsistent cancer gene data sources, BMC Bioinformatics 2010;11:76.
30. Chen Y, Riley DJ, Zheng L et al. Phosphorylation of the mitotic regulator protein Hec1 by Nek2 kinase is essential for faithful chromosome segregation, J Biol Chem 2002;277:49408-49416.

31. Liu Q, Hirohashi Y, Du X et al. Nek2 targets the mitotic checkpoint proteins Mad2 and Cdc20: a mechanism for aneuploidy in cancer, Exp Mol Pathol 2010;88:225-233.

32. Kokuryo T, Senga T, Yokoyama Y et al. Nek2 as an effective target for inhibition of tumorigenic growth and peritoneal dissemination of cholangiocarcinoma, Cancer Res 2007;67:9637-9642.

33. Wu G, Qiu XL, Zhou L et al. Small molecule targeting the Hec1/Nek2 mitotic pathway suppresses tumor cell growth in culture and in animal, Cancer Res 2008;68:8393-8399.

34. Hu CM, Zhu J, Guo XE et al. Novel small molecules disrupting Hec1/Nek2 interaction ablate tumor progression by triggering Nek2 degradation through a death-trap mechanism, Oncogene 2015;34:1220-1230.

35. Hayakawa Y, Jin G, Wang H et al. CCK2R identifies and regulates gastric antral stem cell states and carcinogenesis, Gut 2015;64:544-553.

36. Yao C, Li H, Shen X et al. Reproducibility and concordance of differential DNA methylation and gene expression in cancer, PLoS One 2012;7:e29686.

37. Ye L, Morse LR, Zhang L et al. Osteopetrorickets due to Snx10 deficiency in mice results from both failed osteoclast activity and loss of gastric acid-dependent calcium absorption, PLoS Genet 2015;11:e1005057.

38. Pritchard C, Mecham B, Dumpit R et al. Conserved gene expression programs integrate mammalian prostate development and tumorigenesis, Cancer Res 2009;69:1739-1747.

39. Kaira K, Oriuchi N, Imai H et al. Expression of L-type amino acid transporter 1 (LAT1) in neuroendocrine tumors of the lung, Pathol Res Pract 2008;204:553-561.

40. Wang Q, Holst J. L-type amino acid transport and cancer: targeting the mTORC1 pathway to inhibit neoplasia, Am J Cancer Res 2015;5:1281-1294.

41. Kobayashi K, Ohnishi A, Promsuk J et al. Enhanced tumor growth elicited by L-type amino acid transporter 1 in human malignant glioma cells, Neurosurgery 2008;62:493-503; discussion 503-494.
Table 1. The public datasets used in this study.

| Datasets       | Platforms          | The number of normal samples | The number of cancer samples |
|----------------|--------------------|------------------------------|------------------------------|
| **Datasets with paired cancer-normal samples** |                    |                              |                              |
| GSE29272       | Affymetrix GPL96   | 134                          | 134                          |
| GSE2685        | Affymetrix GPL80   | 6                            | 6                            |
| GSE63089       | Affymetrix GPL5175 | 45                           | 45                           |
| GSE13195       | Affymetrix GPL5175 | 25                           | 25                           |
| GSE56807       | Affymetrix GPL5175 | 5                            | 5                            |
| GSE13911       | Affymetrix GPL570  | 31                           | 31                           |
| GSE19826       | Affymetrix GPL570  | 12                           | 12                           |
| GSE79973       | Affymetrix GPL570  | 10                           | 10                           |
| GSE65801       | Agilent GPL14550   | 32                           | 32                           |
| GSE51575       | Agilent GPL13607   | 26                           | 26                           |
| GSE29998       | Illumina GPL6947   | 49                           | 49                           |
| GSE13861       | Illumina GPL6884   | 19                           | 19                           |
| GSE63288       | SOLiD GPL13393     | 22                           | 22                           |
| TCGA            | HiSeq_RNASeqV2     | 32                           | 32                           |
| **Datasets without paired cancer-normal samples** |                    |                              |                              |
| GSE54129       | Affymetrix GPL570  | 21                           | 111                          |
| GSE38749       | Affymetrix GPL570  | -                            | 15                           |
| GSE57303       | Affymetrix GPL570  | -                            | 70                           |
| GSE34942       | Affymetrix GPL570  | -                            | 56                           |
| GSE22377       | Affymetrix GPL570  | -                            | 43                           |
| GSE35809       | Affymetrix GPL570  | -                            | 70                           |
| GSE51105       | Affymetrix GPL570  | -                            | 94                           |
| GSE15459       | Affymetrix GPL570  | -                            | 200                          |
| GSE34942       | Affymetrix GPL570  | -                            | 56                           |
| TCGA            | HiSeq_RNASeqV2     | -                            | 375                          |

Figures
The expression levels of ACACB (A) and COL4A5 (B) in the 49 paired cancer-normal gastric tissues from GSE29998.

The dysregulation frequencies of the 25 universal DEGs in 1,990 GC tissues, 448 paired cancer-normal tissues and 24 validated cancer-normal tissues, respectively. (A) The up-regulation frequencies. (B) The down-regulation frequencies.
Figure 3

Functional analysis of up-regulation and down-regulation genes, respectively. The top ten pathways enriched with up-regulation (A) and down-regulation (B) genes. (C) The degrees of up-regulation genes. (D) The degrees of down-regulation genes. (E) The sub-network contains NEK2 (or CCKBR) and its direct neighbor genes.

Supplementary Files
This is a list of supplementary files associated with this preprint. Click to download.
Table S1.docx
Table S2.docx
