Comprehensive Genomic And Epigenomic Analyses on Transcriptomic Regulation In Stomach Adenocarcinoma

Junxing Chen

Second Affiliated Hospital of Fujian Medical University

Weinan Liu

Second Affiliated Hospital of Fujian Medical University

Jiabin Du

Second Affiliated Hospital of Fujian Medical University

Pengcheng Wang

Second Affiliated Hospital of Fujian Medical University

Jintian Wang

Second Affiliated Hospital of Fujian Medical University

kai Ye (✉ Yekai1972@fjmu.edu.cn)

Second Affiliated Hospital of Fujian Medical University <https://orcid.org/0000-0001-7743-9992>

Primary research

Keywords: stomach adenocarcinoma, integrative analysis of genomic and epigenomic regulation, DNA copy-number-correlated genes, methylation-correlated genes

Posted Date: August 24th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-292053/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Frontiers in Genetics on February 11th, 2022. See the published version at <https://doi.org/10.3389/fgene.2021.778095>.

Abstract

DNA copy number variations (CNVs) and alterations in methylation (MET)-induced transcriptomic dysregulation plays an important role in promoting heterogeneous Stomach adenocarcinoma (STAD) progression. However, the association between DNA MET and CNVs, and the effect of this association during cancer development remain unknown. Altogether 313 STAD specimens had been extracted to examine DNA MET and copy numbers, and to measure messenger RNA (mRNA) expression (EXP). According to our results, copy-number-correlated (CNVcor) genes were remarkably co-regulated with MET-correlated (METcor) ones. Additionally, CNVcor and METcor genes were carried out multi-omics integration, and 3 STAD prognostic subtypes had been identified and verified using the independent data. Typically, the subtype that had higher aggressiveness was found to be associated with reduced PIK3CA, ARID1A, SPECC1, ZFH3, KMT2D, OBSCN, ZBTB20, FSIP2, RANBP2 and TTN mutation rates, as well as increased JPH3, PLCXD3, and KCNB1 expression. Our results suggested that above mentioned genes played vital parts during invasive STAD development. And results of this comprehensive analyses on transcriptomic regulation at genomic and epigenetic levels facilitate to understand STAD pathology at various points of view, and assist in developing the effective STAD treatment.

Background

Gastric cancer (GC), a highly malignant tumor, accounts for about 10% of cancer-related deaths in the world in 2015. Related treatments have been developed for GC; however, it is still the fatal malignant tumor that ranks the second place in terms of its mortality in 2012 in East Asia, which is ascribed to its advanced diagnosis and great morbidity[1]. Typically, the overall survival (OS) rate of GC at 5 years is less than 25%, especially for the recurrent and advanced types[2]. Stomach adenocarcinoma (STAD) takes up about 90% of global GC cases[3, 4]. Most STAD patients among western countries are diagnosed at the metastatic or advanced stage[5]. Despite improvements in surgery, radiotherapy, and chemotherapy, the prognosis for STAD are still poor, and some patients with GC, even those with the same TNM (Tumor Node Metastasis) stage, have different prognoses and treatment responses[6]. STAD diagnosis at an early stage can markedly improve patient outcomes. As a result, the biomarkers for diagnosis and prognosis are in urgent need to improve the STAD diagnosis and to predict the outcomes for patients.

Markers for diagnosis and prognosis can be used to closely monitor and treat the high-risk patients, so as to extend their OS time. The common STAD biomarkers applied in clinic are mainly the clinicopathological parameters, such as tumor stage, age, chemotherapy response, and infection with *Helicobacter pylori*[7]. Developing new biomarkers can introduce and design new therapeutic strategies for improving the survival for patients. Therefore, comprehensive examination of molecular characteristics of disease is of crucial necessity.

In recent years, the huge multi-omics greatly contributes to comprehensively displaying disease dysregulation at genomic and epigenetic levels[8]. Genomic alternations induced by cancer filing, such as copy number variations (CNVs) and DNA mutations, are common in the process of tumor genesis, which

promote cancer progression[9]. Additionally, cancer genomic regulation via DNA methylation (MET) at epigenetic level exerts a crucial part in the behaviors of various cancers, including STAD[10, 11]. Genomic profiling research suggests that genomic and epigenomic dysregulation is highly heterogeneous. DNA CNVs exert a vital part in STAD for regulating STAD development; in addition, the resultant transcriptional dysregulation is potentially a driving event during the progression of STAD[12, 13]. Besides, research on DNA MET profiling suggests that, epigenetic regulation is highly significant in the development of cancer from the points of view of biology and clinic[14, 15]. At the same time, some critical tumor-associated genes, such as RASSF1A, SAMD14 and SOCS3, can modulate DNA MET, thereby regulating their functions[16–18].

Nonetheless, the association between DNA MET and CNVs is still unknown, even though DNA MET alternations and CNVs affect cancer whole genome. Besides, the effect of this association during cancer development remains unknown. In the current study, STAD samples were extracted for analyzing DNA MET and copy numbers, as well as the expression of messenger RNA (mRNA, EXP). In addition, DNA copy-number-correlated (CNVcor) and MET-correlated (METcor) genes were defined following related gene expression in the above-mentioned samples, separately, so as to recognize genes with genomically and/or epigenetically modulated expression. Of them, CNVcor genes suggest transcriptional dysregulation based on DNA copy numbers, whereas METcor ones represent transcriptional dysregulation according to DNA MET. Expression of CNVcor genes was evidently correlated with that of METcor genes, which indicated co-regulate of cancer transcriptome via genomic DNA CNVs and epigenetic DNA MET abnormalities. Additionally, CNVcor and METcor genes were performed multi-omics integration, so as to examine the typical molecular subtypes to predict the prognosis for STAD. Novel precise targets and biomarkers to distinguish different subtypes of cancer were identified through intensive systematic analysis.

Methods

mRNA expression patterns, DNA MET and DNA copy numbers

A STAD data set was downloaded based on TCGA for analyzing related results. The mRNA expression profiles, DNA MET and DNA copy numbers, together with STAD mutation data, were collected from the TCGA official data portal. Each sample mark was matched in every platform; afterwards, altogether 313 data sets had been utilized (Table S1), and they had corresponding data sets for mRNA expression profiles, DNA MET and DNA copy numbers. Additionally, gene expression patterns were standardized through log2 transformation and quantile normalization, followed by HUGO official symbol aggregation. The expression profiles were subsequently standardized through eliminating the average values from each probe within non-cancer tissues, which stood for the fold change in cancer tissues compared with that in normal tissues. Thereafter, the circular binary segmentation algorithm from “DNACopy.” library of R package was utilized for mapping DNA CNV data at genetic level for all samples based on the segmented

CNV data[19]. With regard to the DNA MET profiles, the β -values of the probe were filtered for eliminating probes positioned on the sex chromosomes. Later, probes located in the regions associated with the CpG islands had been mapped to associated genes, such as Shore regions, CpG islands, Shelf, differentially methylated regions, first-exon regions, 5'UTR, as well as the gene promoter areas that contained 2500 upstream basic groups from TSS. Probes that had over 30% missing values across the samples had been eliminated out of every processed pattern.

The “liftOver,” library of R package was utilized for every data set, so as to restore the probe genomic coordinates to the human reference genome hg38. Afterwards, probes had been matched to corresponding ones in the EXP profiles. Cancer-specific alterations were computed through eliminating the mean probe intensity within non-cancer tissues. Then, probes that had over 50% missing values, and those on the sex chromosomes were eliminated, and related information was input using the sk-nearest neighbor algorithm. Thereafter, the pairwise Pearson’s correlation coefficients of all genes were computed across the matched EXP and CNV profiles, as well as EXP and MET, respectively. If over one probe was mapped to one gene, the probe that had the average or the smallest correlation coefficient would be utilized to be the typical pair-matched probe for CNV and MET profiles, separately.

Cluster analysis on the genomic patterns at different levels

Stable clusters of samples were recognized using negative matrix factorization (NMF) cluster analysis based on 50 iterations and the “brunet” method through CNVcor and METcor genes, separately[20]. Notably, the cluster number k was set as 2-10, and the best cluster number was calculated according to the observed consensus map, together with the cluster cophenetic correlations. At the same time, the mean silhouette width was computed to examine the consensus membership matrix using “NMF.” of R package. For each member, the lowest cluster number was set as 10. The R package “iCluster” method was utilized to comprehensive analyze the EXP, MET and DNA CNV profiles; besides, 20 iterations and the default parameters were utilized[21].

Results

Transcriptomic alternations in DNA copy number or DNA MET

DNA CNVs and MET at genetic and epigenomic levels, as well as gene EXP patterns, were collected from 313 STAD specimens. Then, the raw materials were preprocessed according to the method described previously in “Methods.” Later, the correlation coefficients of DNA CNV or MET profiles with corresponding mRNA EXP data were computed, so as to evaluate the influence of genomic and/or epigenomic abnormalities. The correlation coefficient r had been standardized according to the Fisher’s Z-transformation for variance stabilization.

Consistent with prior study, overall distribution of correlation coefficients between DNA CNV and related EXP data markedly skewed to the right (skewness = 1.2352, $p < 1e^{-5}$). In comparison, the correlation

coefficients between DNA MET and related EXP data skewed to the left (skewness = -0.37363, $p < 1e^{-5}$) (Fig.1A), which indicated that DNA CNVs and MET abnormalities positively and negatively modulated transcription, respectively.

A great number of genes participated in the above 2 gene sets (CNVcor and METcor genes, Table S2-3); as a result, genes that were markedly correlated with OS were selected in later analysis (log rank $p < 0.05$). Gene signatures with positive correlation were screened to examine the DNA copy number (CNVcor, $n = 368$), and while those with negative correlation were screened for DNA MET (METcor, $n = 348$). CNVcor genes indicated DNA CNVs-dependent transcriptional dysregulation, but METcor genes suggested MET-dependent transcriptional dysregulation. CNVcor genes were not overlapped with METcor ones, of them, 116 overlapped genes had been identified, which suggested that CNVcor and METcor genes could specifically regulate transcriptional dysregulation (Fig.1B).

CNVcor genes preferred to DNA CNVs within some genomic regions, especially on chromosomes 10, 17, 18 and 21 (Fig.1C and Table S4). Consistent with prior research, CNVcor genes were abundant on chromosomes 10 and 17, which suggested that gene expression was sensitive to DNA dose at certain regions[22]. Additionally, METcor genes could be found within desired chromosomal regions, like chromosome 19 (Fig.1D and Table S5), most of which were genes that encoded proteins (Fig.1E) and distributed in the CpG islands (Fig.1F). This study suggested that, CNVcor and METcor genes were remarkably effective on transcriptional dysregulation of STAD, which required to be further examined in future studies.

Different CNVcor and METcor genes-dependent molecular subtypes.

Subsequently, the effects of CNVcor and METcor gene expression on predicting prognostic subgroups were explored. NMF cluster analysis was performed for all gene set data, and the cluster number k was set at 2-10; later, the value of k was calculated for each profile ($k=3$ for CNV and MET, respectively) (Fig.2A-B). Surprisingly, CNVcor genes-identified subtypes were remarkably overlapped with the METcor genes-identified ones ($p < 1e^{-5}$ upon χ^2 -test), and such results were consistent with CNVcor and METcor gene regulation in STAD (Fig.2E-F). Additionally, Kaplan–Meier (KM) curve results suggested that, CNVcor or METcor genes-identified subtypes predicted the OS of patients (Fig.2C-D), separately ($p < 0.05$).

Molecular subtypes associated with CNVcor and METcor gene expression were identified at various aspects. iCluster, the integrated cluster method, was employed to integrate genomic data on mRNA EXP, DNA CNVs and MET. Afterwards, cluster analysis would be carried out, with the cluster number k of 2-4. 20 cluster iterations at $K=4$ (category 5), $K=3$ (category 4) and $K=2$ (category 3) were carried out separately to assess the best iCluster cluster results. According to our findings, stable cluster results were obtained at $K=2$ relative to those at $K=3$ or 4 (Fig.S1-2). As a result, all samples were clustered as 3 subclasses of iC1-iC3 ($n=93$, 100 and 120, respectively). Fig.S3A-B display the cluster results of those 3 subclasses, and Table S6 presents those of all samples.

KM analysis results suggested that, iC1 could attain the optimal OS across those 3 subtypes ($p < 0.05$, Fig.3A). The OS of patients in iC1 subgroup was compared with that of the other 2 subgroups (Fig.3B and Fig.S4), and the results suggested that, the difference in prognosis between iC1 and iC2 subgroups was statistically significant ($p < 0.001$). Notably, the subtypes identified by icluster remarkably overlapped with those identified through CNVcor and METcor genes ($p < 1e-5$, χ^2 -test, Fig.3C-D). According to the above findings, comprehensive analysis on CNVcor and METcor genes facilitated to identify various molecular subtypes, and all of them showed heterogenous combinations of genomic and epigenomic features associated with prognosis and transcriptional dysregulation.

Coordinated DNA CNVs and METs abnormalities.

The frequencies between genome-wide DNA METs abnormalities and DNA CNVs were compared following batch effect correction. Additionally, DNA copy-number gain (CNVgain, $\beta > 0.3$) and loss (CNVloss, $\beta < -0.3$), together with DNA hypermethylation (METHyper, $\beta > 0.8$) and hypomethylation (METHypo, $\beta < 0.2$) were computed according to the determined threshold (fold change of 0.3), which were subsequently compared with the level of every probe. Our results indicated that (Table S7), CNVgain frequency showed marked correlation with CNVloss frequency ($p < 1e^{-5}$, Fig.4A). Besides, METHyper frequency displayed evident correlation with METHypo frequency ($p < 1e^{-5}$, Fig.4F). Directional METHyper and CNVgain were tightly correlated with CNVloss, which indicated that all correlations were directional abnormality-free (Fig.4B-E, $p < 0.05$). In conclusion, our results suggested that, STAD patients with increased frequency of DNA CNVs had elevated frequency of DNA MET abnormality. Correlations in the frequency of abnormal METcor and CNVcor genes indicated the close correlation of DNA MET with DNA CNVs.

Identification of STAD subtype key features

Firstly, clinical features (including TNM, Stage, Gender and Primary Site) were compared across those 3 subtypes. Fig.S5 and Table 1 indicate that, the differences in clinical characteristics are not statistically significant among 3 STAD subtypes. Nonetheless, as for stage distribution, III and IV samples were mainly the iC2 subtype that had the poorest prognosis. In addition, samples across these 3 subtypes were calculated and determined based on the tumor immune estimation resource (TIMER) method (Table S8) [23]. As could be seen, differences in the 5/6 immune cell scores (CD8⁺T, CD4⁺T, macrophages, neutrophils, and dendritic cells) of samples across these 3 subtypes were statistically significant (Fig.5A and B, $p < 1e^{-5}$). The above results indicated that, immunocyte infiltration degree or the immune microenvironment of STAD was correlated with DNA CNV or MET level.

Table 1 Clinical informations of STAD patients across the 3 subtypes were compared.

Event	Total	iC1	iC2	iC3
Alive	188	69	47	68
Dead	125	22	49	49
T				
T1	14	7	5	2
T2	63	17	26	20
T3	146	41	47	58
T4	81	26	18	37
N				
N0	90	35	24	31
N1	80	20	25	35
N2	65	20	22	23
N3	62	14	23	25
NX	7	2	2	3
M				
M0	275	87	86	102
M1	18	2	5	11
MX	11	2	5	4
Stage				
I	39	14	14	11
II	99	34	27	38
III	129	36	40	53
IV	28	3	13	12
Un	9	4	2	3
Grade				
G1	8	2	3	3
G2	107	37	49	21
G3	180	51	40	89
GX	9	1	4	4

Age				
0~50	27	9	7	11
50~60	72	15	20	37
60~70	91	23	33	35
70~80	99	38	32	29
80~100	15	6	4	5
Gender				
FEMALE	101	29	27	45
MALE	203	62	69	72

Besides, differences in gene EXP, DNA CNVs and MET across iC2 and iC1 subtype samples were compared. Later, DNA MET and CNVs levels had been classified as 3 types, including Normal, Gain and Loss (for CNV), as well as Normal, HyperMethy and HypoMethy (for MET). DNA CNVcor or METcor genes with marked difference between iC2 and iC1 subtypes were obtained through Fisher's exact test. These results are presented in Table S9-10. As for EXP patterns, differentially expressed genes (DEGs) between iC1 and iC2 subtypes were acquired through DESeq2[24] ($p < 0.05$), and the results are presented in Table S11. Fig. S7 displays EXP, CNV and METcor genes that were markedly different across samples of 3 subtypes. In addition, to illustrate the crucial prognostic features across various subtypes, altogether 5 genes with distinct difference between iC1 and iC2 samples at all the three (MET, CNV and EXP) levels had been screened to carry out univariate survival analysis. According to our findings, 2 genes (PLCXD3 and KCNB1) showed correlation with prognosis (log-rank $p < 0.05$), suggesting that the above-mentioned two genes in iC2 subtype (with poorer prognosis) had greater CNV and hypermethylation levels than iC1 subtype (with positive prognosis), and the expression of the 2 genes in iC1 subtype were down-regulated compared with ones in the iC2 subtype. Afterwards, PLCXD3 and KCNB1 expression was divided as low (L1), medium (L2) or high (L3). Our findings suggested that, L1, L2 and L3 groups in terms of the PLCXD3 and KCNB1 expression levels had positive correlation with prognosis (Fig.6A-B). The above findings indicated that, the distinct PLCXD3 and KCNB1 expression was related to DNA MET or CNV level as well as the patients prognostic outcome. Moreover, the GEO GSE62254[25] STAD data set ($n=266$) had been utilized to analyze the relationship between those of 5 genes and the prognosis for patients. The expression levels of JPH3 and KCNB1 were available, which suggested that only JPH3 in GSE62254 STAD data set[26] showed marked correlation with prognosis. Fig.S6 displays these results.

Eventually, the STAD mutation profiles had been determined for exploring the associations with the sub-classification. Specifically, the synonymous mutations were eliminated to obtain the missense and nonsense mutations. Overall, the mutation frequency of genes in different subtypes showed statistically significant differences. In addition, some distinctly different mutant genes (PIK3CA, ARID1A, SPECC1, ZFHX3, KMT2D, OBSCN, ZBTB20, FSIP2, RANBP2 and TTN) between iC1 and iC2 subtype were selected

upon fisher's test, as presented in Fig.7 and Table S12 ($\text{FDR} < 0.1$ upon fisher's test). The different mutational spectra were analyzed, which suggested that the above 10 genes in iC1 subtype had remarkably increased mutation frequencies compared with those in iC2 and iC3 subtypes ($p < 0.01$). Interestingly, mutation frequencies of these 10 genes in iC3 subtype showing common prognosis were also increased compared with that of iC2 subtype ($p < 0.05$). In conclusion, the above findings indicated that, the STAD molecular subtypes related to DNA copy numbers and DNA MET showed correlation with mutations in the above 10 genes, and they might modulate the progression of STAD subtypes.

Discussion

Prior studies indicate that, comprehensive analysis on tumor genomic characteristics at different levels contributes to identifying various molecular subtypes, which can thereby shed new lights on tumor clinical and mechanical differences, and reveal the potential biomarkers and therapeutic targets[8]. Nonetheless, illustrating the complex STAD genomic data is still challenging. In this study, the genomic and epigenomic dysregulations of CNVcor and METcor genes had been determined through the TCGA database. This was a strategy to integrate genomic and epigenomic data at different levels. According to our results, the above correlated genes contributed to identifying STAD subtypes, which reflected different molecular characteristics and genes correlated with immunity at different aspects, and prognostic outcomes. Additionally, STADs that had increased abnormal CNVcor frequency were associated with increased abnormal METcor frequency, indicating that patients with increased DNA CNVs frequency had elevated abnormal DNA MET frequency. The above results indicated that it was necessary to consider the aberrant DNA CNVs and MET during data analysis.

Fortunately, the current comprehensive analysis using CNVcor and METcor genes had identified the new critical molecular features indicating novel STAD biomarkers and therapeutic targets. Analysis results of mutation types across different subtypes indicated that, the differences in these 10 genes were significant in terms of the mutation frequency. Particularly, genes in iC1 subtype with optimal prognosis were associated with the greatest mutation frequency. 5 of the above 10 genes had been reported to participate in STAD progression, pathogenesis, immune microenvironment and malignant transformation, which were PIK3CA, SPECC1, ARID1A, ZBTB20, and KMT2D[27–32], and showed correlation with the prognosis and survival for patients. The above results suggested that, the bioinformatic mining results were highly reliable and accurate. Nonetheless, the associations of other 5 genes with STAD have not been verified in fundamental or clinical study, which is the point that we are interested in. Among these 5 genes, ZFH3 and RANBP2 are associated with an alternative mutation frequency among various malignant tumors (such as endometrial cancer and prostate cancer)[33–35]. Nonetheless, it is still unknown about the underlying mechanism for modulating cancer genesis and development. Besides, the association of gene expression and mutation frequencies with STAD remains unknown. In this study, our results indicated that mutations in RANBP2 and PIK3CA partially mediated the favourable prognosis of STAD patients in iC1 subtype. Nonetheless, further protein-protein interaction (PPI) analysis and molecular biological experiments are needed to verify these results.

Further, JPH3, PLCXD3 and KCNB1 are the possible crucial regulating factors in the initiation and development of STAD. KCNB1 (Kv2.1), the major voltage-gated potassium channel (Kv), is recognized to be the new prognostic factor to predict the survival for some cancers, including glioma, and it plays a tumor suppressing role via inducing autophagy[36, 37]. JPH3 belongs to the junctional membrane complex (JMC) protein family member, which serves for stabilizing the JMC between plasma membrane (PM) and endoplasmic reticulum (ER), and maintaining the cellular ultrastructure between intracellular ionic channels and cell surface. JPH3 is recognized to be a new tumor suppressor gene that is methylated in colorectal cancer, which enhances apoptosis mediated by mitochondria; in addition, it is the candidate biomarker for survival and metastasis for digestive tract cancers[38, 39]. Yet, further studies are needed to illustrate PLCXD3 role in STAD. Our results suggested that, expression of JPH3, PLCXD3 and KCNB1 showed marked correlation with DNA MET.

Conclusion

To sum up, the current comprehensive analysis on the basis of gene expression at genomic and epigenomic levels reveal coordinated STAD genomic alternations at distinct points of view. Our results facilitate to determine the STAD molecular subtypes and to reveal accurate STAD clinical and mechanical diagnosis and treatments.

Abbreviations

stomach adenocarcinoma

STAD; copy number variations:CNVs; methylation:MET; messenger RNA:mRNA; CNVs-correlated:CNVcor; MET-correlated:METcor; Gastric cancer:GC; TNM:Tumor Node Metastasis; negative matrix factorization:NMF; junctional membrane complex:JMC; plasma membrane:PM; endoplasmic reticulum:ER.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent to publish

Not applicable.

Availability of data and materials

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

Competing Interests

The authors declare no conflicts of interest.

Authors' contributions

CJX and YK conceived and designed the experiments; CJX, LWN and DJB performed the experiments and wrote the paper; CJX, WPC and WJT analyzed the data. All authors have read and approved the final manuscript.

Acknowledgments

The authors thank members of their laboratory and their collaborators for their research work.

References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin*. 2015;65:87–108.
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68:394–424.
3. Bu Z, Ji J. Controversies in the diagnosis and management of early gastric cancer. *Chin J Cancer Res*. 2013;25:263–6.
4. Karimi P, Islami F, Anandasabapathy S, Freedman ND, Kamangar F. Gastric cancer: descriptive epidemiology, risk factors, screening, and prevention. *Cancer Epidemiol Biomarkers Prev*. 2014;23:700–13.
5. Van Cutsem E, Dico M, Geva R, Arber N, Bang Y, Benson A, Cervantes A, Diaz-Rubio E, Ducreux M, Glynne-Jones R, et al: **The diagnosis and management of gastric cancer. expert discussion and recommendations from the 12th ESMO/World Congress on Gastrointestinal Cancer, Barcelona, 2010.** *Ann Oncol* 2011, **22 Suppl 5**:v1-9.
6. Shiratsu K, Higuchi K, Nakayama J. Loss of gastric gland mucin-specific O-glycan is associated with progression of differentiated-type adenocarcinoma of the stomach. *Cancer Sci*. 2014;105:126–33.
7. Ye W, Held M, Lagergren J, Engstrand L, Blot WJ, McLaughlin JK, Nyren O. Helicobacter pylori infection and gastric atrophy: risk of adenocarcinoma and squamous-cell carcinoma of the esophagus and adenocarcinoma of the gastric cardia. *J Natl Cancer Inst*. 2004;96:388–96.
8. Woo HG, Choi JH, Yoon S, Jee BA, Cho EJ, Lee JH, Yu SJ, Yoon JH, Yi NJ, Lee KW, et al. Integrative analysis of genomic and epigenomic regulation of the transcriptome in liver cancer. *Nat Commun*. 2017;8:839.
9. Shi J, Zhou W, Zhu B, Hyland PL, Bennett H, Xiao Y, Zhang X, Burke LS, Song L, Hsu CH, et al. Rare Germline Copy Number Variations and Disease Susceptibility in Familial Melanoma. *J Invest Dermatol*. 2016;136:2436–43.

10. Peng Y, Wu Q, Wang L, Wang H, Yin F. A DNA methylation signature to improve survival prediction of gastric cancer. *Clin Epigenetics*. 2020;12:15.
11. Song P, Wu L, Guan W. Genome-Wide Identification and Characterization of DNA Methylation and Long Non-Coding RNA Expression in Gastric Cancer. *Front Genet*. 2020;11:91.
12. Cai L, Li L, Ren D, Song X, Mao B, Han B, Zhang H. Prognostic impact of gene copy number instability and tumor mutation burden in patients with resectable gastric cancer. *Cancer Commun (Lond)*. 2020;40:63–6.
13. Feng Y, Shi C, Wang D, Wang X, Chen Z. **Integrated Analysis of DNA Copy Number Changes and Gene Expression Identifies Key Genes in Gastric Cancer.** *J Comput Biol* 2019.
14. Ebrahimi V, Soleimani A, Ebrahimi T, Azargun R, Yazdani P, Eyvazi S, Tarhriz V. Epigenetic modifications in gastric cancer: Focus on DNA methylation. *Gene*. 2020;742:144577.
15. Choi SJ, Jung SW, Huh S, Chung YS, Cho H, Kang H. Alteration of DNA Methylation in Gastric Cancer with Chemotherapy. *J Microbiol Biotechnol*. 2017;27:1367–78.
16. Balgkouranidou I, Matthaos D, Karayiannakis A, Bolanaki H, Michailidis P, Xenidis N, Amarantidis K, Chelis L, Trypsianis G, Chatzaki E, et al. Prognostic role of APC and RASSF1A promoter methylation status in cell free circulating DNA of operable gastric cancer patients. *Mutat Res*. 2015;778:46–51.
17. Xu X, Chang X, Xu Y, Deng P, Wang J, Zhang C, Zhu X, Chen S, Dai D. **SAMD14 promoter methylation is strongly associated with gene expression and poor prognosis in gastric cancer.** *Int J Clin Oncol* 2020.
18. Han Q, Zhou H, Xie W, Sun T, Wei R, Nie C, Hong J, Zhu L, Tian W. **Association between the methylation of the STAT1 and SOCS3 in peripheral blood and gastric cancer.** *J Gastroenterol Hepatol* 2020.
19. Zhang K, Gao J, Ni Y. Screening of candidate key genes associated with human osteosarcoma using bioinformatics analysis. *Oncol Lett*. 2017;14:2887–93.
20. Zeng Z, Vo AH, Mao C, Clare SE, Khan SA, Luo Y. Cancer classification and pathway discovery using non-negative matrix factorization. *J Biomed Inform*. 2019;96:103247.
21. Shen R, Mo Q, Schultz N, Seshan VE, Olshen AB, Huse J, Ladanyi M, Sander C. Integrative subtype discovery in glioblastoma using iCluster. *PLoS One*. 2012;7:e35236.
22. Arakawa N, Sugai T, Habano W, Eizuka M, Sugimoto R, Akasaka R, Toya Y, Yamamoto E, Koeda K, Sasaki A, et al. Genome-wide analysis of DNA copy number alterations in early and advanced gastric cancers. *Mol Carcinog*. 2017;56:527–37.
23. Li B, Severson E, Pignon JC, Zhao H, Li T, Novak J, Jiang P, Shen H, Aster JC, Rodig S, et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol*. 2016;17:174.
24. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.

25. Cristescu R, Lee J, Nebozhyn M, Kim KM, Ting JC, Wong SS, Liu J, Yue YG, Wang J, Yu K, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat Med*. 2015;21:449–56.
26. Oh SC, Sohn BH, Cheong JH, Kim SB, Lee JE, Park KC, Lee SH, Park JL, Park YY, Lee HS, et al. Clinical and genomic landscape of gastric cancer with a mesenchymal phenotype. *Nat Commun*. 2018;9:1777.
27. Seo AN, Kang BW, Bae HI, Kwon OK, Park KB, Lee SS, Chung HY, Yu W, Jeon SW, Kang H, Kim JG. Exon 9 Mutation of PIK3CA Associated With Poor Survival in Patients With Epstein-Barr Virus-associated Gastric Cancer. *Anticancer Res*. 2019;39:2145–54.
28. Wang K, Kan J, Yuen ST, Shi ST, Chu KM, Law S, Chan TL, Kan Z, Chan AS, Tsui WY, et al. Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. *Nat Genet*. 2011;43:1219–23.
29. Ashizawa M, Saito M, Min AKT, Ujiie D, Saito K, Sato T, Kikuchi T, Okayama H, Fujita S, Endo H, et al. Prognostic role of ARID1A negative expression in gastric cancer. *Sci Rep*. 2019;9:6769.
30. Shi J, Li W, Ding X. Assessment of the association between ZBTB20 rs9841504 polymorphism and gastric and esophageal cancer susceptibility: a meta-analysis. *Int J Biol Markers*. 2017;32:e96–101.
31. Xiong W, Deng Z, Tang Y, Deng Z, Li M. Downregulation of KMT2D suppresses proliferation and induces apoptosis of gastric cancer. *Biochem Biophys Res Commun*. 2018;504:129–36.
32. Chen LH, Wang LP, Ma XQ. Circ_SPECC1 enhances the inhibition of miR-526b on downstream KDM4A/YAP1 pathway to regulate the growth and invasion of gastric cancer cells. *Biochem Biophys Res Commun*. 2019;517:253–9.
33. Hu Q, Zhang B, Chen R, Fu C, Fu AJ, Li X, Fu J, Zhang L, Dong Z. JT: ZFH3 is indispensable for ERbeta to inhibit cell proliferation via MYC downregulation in prostate cancer cells. *Oncogenesis*. 2019;8:28.
34. Walker CJ, Miranda MA, O'Hern MJ, McElroy JP, Coombes KR, Bundschuh R, Cohn DE, Mutch DG, Goodfellow PJ. **Patterns of CTCF and ZFH3 Mutation and Associated Outcomes in Endometrial Cancer.** *J Natl Cancer Inst* 2015, 107.
35. Packham S, Warsito D, Lin Y, Sadi S, Karlsson R, Sehat B, Larsson O. Nuclear translocation of IGF-1R via p150(Glued) and an importin-beta/RanBP2-dependent pathway in cancer cells. *Oncogene*. 2015;34:2227–38.
36. Wang HY, Li JY, Liu X, Yan XY, Wang W, Wu F, Liang TY, Yang F, Hu HM, Mao HX, et al. A three ion channel genes-based signature predicts prognosis of primary glioblastoma patients and reveals a chemotherapy sensitive subtype. *Oncotarget*. 2016;7:74895–903.
37. Wang HY, Wang W, Liu YW, Li MY, Liang TY, Li JY, Hu HM, Lu Y, Yao C, Ye YY, et al. Role of KCNB1 in the prognosis of gliomas and autophagy modulation. *Sci Rep*. 2017;7:14.
38. Hu X, Kuang Y, Li L, Tang H, Shi Q, Shu X, Zhang Y, Chan FK, Tao Q, He C. Epigenomic and Functional Characterization of Junctophilin 3 (JPH3) as a Novel Tumor Suppressor Being Frequently Inactivated by Promoter CpG Methylation in Digestive Cancers. *Theranostics*. 2017;7:2150–63.

39. Coleman EA, Lee JY, Erickson SW, Goodwin JA, Sanathkumar N, Raj VR, Zhou D, McKelvey KD, Apewokin S, Stephens O, et al. GWAS of 972 autologous stem cell recipients with multiple myeloma identifies 11 genetic variants associated with chemotherapy-induced oral mucositis. Support Care Cancer. 2015;23:841–9.

Figures

Figure 1

Recognition of DNA methylation-correlated (METcor) and DNA copy-number-correlated (CNVcor) genes of STAD. (A) Correlation coefficient distribution of mRNA expression with DNA MET or DNA copy numbers among different specimens are presented, respectively. (B) The Venn diagram displays the METcor and CNVcor gene numbers. Overlapped gene number between METcor and CNVcor genes are indicated. (C) Proportions of CNVcor genes in the total number of genes in every chromosome arm. (D) Proportions of METcor genes in the total number of genes in every chromosome arm. (E-F) Genomic locations of DNA MET probes have been classified according to the relationships with genes (left) and CpG islands (right) in positions, separately.

Figure 2

Identification of STAD molecular subtypes using METcor and CNVcor genes. (A-B) Plots display the cluster results of non-negative factorization (NMF) for CNVcor in CNV data (A) as well as for METcor in MET data (B), separately. (C-D) OS Kaplan–Meier curves for subtypes stratified based on NMF clustering of CNVcor (C) as well as METcor (D) genes, separately. (E-F) Subtypes identified based on CNVcor show marked overlaps with those identified by METcor.

Figure 3

STAD molecular subtypes identified through iCluster analysis. (A) OS Kaplan–Meier curves for subtypes stratified based on iCluster (iC1, iC2 and iC3). (B) OS Kaplan–Meier curves for iC1 and iC2. (C-D) Subtypes identified based on iCluster analysis showed evident overlaps with those identified based on CNVcor (C) as well as METcor (D) genes

Figure 4

Coordinated DNA MET abnormalities and DNA CNVs in STAD. (A and F) DNA CNVs or DNA MET abnormalities were identified at the threshold fold change of >0.2 relative to the means in normal tissues, respectively. Directional DNA CNV-gain (CNVgain) and CNV-loss (CNVloss), together with DNA hypomethylation (METHypo) and hypermethylation (METHyper) in all samples are presented, respectively. (B-E) Plots display the pairwise occurrence rates of CNVloss, CNVgain, METHypo and METHyper genes in each sample.

Figure 5

Key STAD immune features. Sample immune scores among the 3 subtypes are computed and analyzed by the tumor immune estimation resource (TIMER) method. Scores of 6 immune cells in all samples are determined (B) and analyzed compared with those in samples of other subtypes (A).

Figure 6

Key STAD molecular features. (A) The KCNB1 EXP or DNA CNV or DNA MET levels for iC1 and iC2 subtypes are presented in the left. The gene expression of KCNB1 is divided into low (L1), medium (L2) or high (L3) group, and the Kaplan–Meier curves for L1, L2 and L3 groups are displayed in the right. (B) The LCXD3 EXP or DNA CNV or DNA MET levels for iC1 and iC3 subtypes are presented in the left, and the Kaplan–Meier curves for L1, L2 and L3 groups are displayed in the right.

Figure 7

Differentially mutated genes across various STAD subtypes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FS1.tif](#)
- [FS2.tif](#)

- [FS3.tif](#)
- [FS4.tif](#)
- [FS5.tif](#)
- [FS6.tif](#)
- [S1Table.xlsx](#)
- [S2Table.xlsx](#)
- [S3Table.xlsx](#)
- [S4Table.xlsx](#)
- [S5Table.xlsx](#)
- [S6Table.xlsx](#)
- [S7Table.xlsx](#)
- [S8Table.xlsx](#)
- [S9Tabel.xlsx](#)
- [S10Table.xlsx](#)
- [S11Table.xlsx](#)
- [S12Table.xlsx](#)