Integrative multi-omics analysis reveals the landscape of Cyclin-Dependent Kinase (CDK) family genes in pan-cancer

Peng Wang
Affiliated Hospital of Nantong University

Mingjie Xie
Affiliated Hospital of Nantong University

Dongmei Yang
Affiliated Hospital of Nantong University

Feiran Wang
Affiliated Hospital of Nantong University

Erlin Chen (calnjmu@163.com)
Affiliated Hospital of Nantong University

Research Article

Keywords: Cyclin-dependent kinase, multi-omics, pan-cancer, bioinformatics

Posted Date: August 3rd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-24985/v2

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

Objective

Cyclin-Dependent Kinases (CDKs) get widely involved in cancer development. However, a wealth of conflicting data raise the question of whether CDKs serve as oncogenes or cancer suppressors. Direct evidence from a same-batch cohort with matched multi-omics sequencing data is still lacking.

Methods

Here, we integrated multi-omics analysis to explore CDKs across multiple cancer types using data from The Cancer Genome Atlas (TCGA) database. First, we evaluated the expression levels of CDKs in pan-cancer. Second, we conducted copy number variants (CNV) and somatic mutation analysis of CDKs in pan-cancer. Third, the biological functions of CDKs were obtained through pathway analysis. Finally, in order to explore effective drugs for tumors with obvious effects of CDKs, drug sensitivity analysis is also explored.

Results

To our surprise, CDKs are overexpressed in a wide range of cancers (39 upregulated, 12 downregulated cases), among which CDK1 is the top overexpressed gene (7 of 10 cancers). We next observed that CDK gene regions are heterozygously amplified in cancer genomes. Further pathway analysis revealed that CDK genes cancer hallmarks such as apoptosis and cell cycle. CDK genes associate with 58 drug sensitivity including Trametinib and I-BET-762.

Conclusions

Taken together, our study identified a panel of clinically relevant CDK genes, defined the genomic/epigenomic/transcriptomic landscape of CDKs at a system level, and opened potential therapeutic opportunities for cancer patients.

1. Introduction

Cyclin-dependent kinases (CDKs), a family of serine/threonine kinases, play an important role in the regulation of cell cycle transition (1), consisting of 20 CDKs (CDK1-CDK20) (2). CDK dysregulation is the hallmark in a broad spectrum of cancers, which may result in uncontrolled proliferation of cancer cells (3). For example, CDK8 was believed to be an oncogene, where its expression tightly associates with aggressive phenotypes in breast cancer patients (4). Another example was CDK12, whose expression links with HER2 status (5). Besides breast cancer, CDK5 overexpression may induce tumor cell motility in
hepatocellular carcinoma and head and neck squamous cell carcinoma (6, 7). Those data raise the critical need to define CDK expression landscape across common cancers.

Cancer cells can exhibit high frequency of genomic changes (8). We propose that DNA containing CDK genes may be amplified or deleted which contributes to transcriptional changes. Moreover, the initiation and progression of cancers can be viewed as the accumulation of many dysfunctional genes caused by gene mutation and epigenetic modification. DNA methylation as a well-studied epigenetic hallmark that has been determined to cause the abnormal gene expression in human cancers (9), which was associated with genomic instability, such as mutations, may cause chromosomal instability in human cancers (10). Although the linkage between methylation change and chromosomal instability has been widely reported, the direct correlation between CDK gene expression and differential methylation and the frequency of somatic mutation has not been directly estimated in across human cancers.

In this study, we comprehensively integrated multi-omics data to analyze the genetic and methylated alteration of CDK family genes, and explore the regulatory pathways of these genes and the effects of gene-related drugs across various human cancers. Our findings will provide new insights into the molecular regulatory mechanisms of CDK family members in cancer occurrence and progression.

2. Materials And Methods

2.1 Datasets

We downloaded the datasets containing transcriptomic data, CNV, and DNA methylation data of various cancer types from The Cancer Genome Atlas (TCGA) database, such as Thyroid Carcinoma (THCA), Kidney Renal Papillary Cell Carcinoma (KIRP), Liver Hepatocellular Carcinoma (LIHC), Stomach Adenocarcinoma (STAD), Breast Invasive Carcinoma (BRCA), Colon Adenocarcinoma (COAD), Uterine Corpus Endometrial Carcinoma (UCEC), Bladder Urothelial Carcinoma (BLCA), Kidney Renal Clear Cell Carcinoma (KIRC), Kidney Chromophobe (KICH), and Prostate Adenocarcinoma (PRAD).

2.2 DNA methylation analysis

We obtained the DNA methylation data from TCGA public methylation profiles produced by the Illumina Human Methylation 450K BeadChip and calculated the Pearson coefficient to achieve a correlation analysis. The FDR was calculated by an unpaired t-test.

2.3 Copy number variants (CNV) and somatic mutation analysis

Copy number variations (CNVs) containing heterozygous CNV and homozygous CNV were shown in our TCGA data, and the GISTIC2 method (11) was used to identify CNV segments in various cancers. Pearson coefficient of paired mRNA expression and CNV data was calculated by using the R package. After downloading the somatic mutation data from the TCGA database, we obtained somatic mutations in
various cancer types. Then, we analyzed the mutation frequency across various cancer types for estimating the percentage of a specific mutation.

2.4 Pathway analysis

We obtained the pan-cancer gene expression profile and used clusterProfiler R package to analyze the enriched pathways, including apoptosis, cell cycle, DNA damage response, EMT, PI3K/AKT, RAS/MAPK, RTK, and TSC/mTOR pathways. The threshold $P$-value < 0.05 was set as significant enrichment.

2.5 Statistical analysis

$P$-value less than 0.05 was considered as statistical significance. As for correlation analysis, $|R| > 0.3$ and $P$-value < 0.05 was considered as statistical significance.

3. Results

3.1 CDK genes are globally upregulated in cancers and associate with patient survival

RNA sequencing data from the TCGA database were used to analyze the expression levels of CDK family genes in a wide range of cancers. The results showed that most of the CDK family genes, such as CDK1, CDK5, CDK4, CDK2, CDK16, CDK7, CDK6, CDK12, CDK8, CDK17, and CDK13, were up-regulated across various cancer types (Fig. 1A). In contrast, we also observed that CDK14, CDK15, and CDK20 were down-regulated in some cancer types, and CDK9, CDK11A, and CDK11B remained unchanged (Fig. 1B). In general, CDK genes show upregulation in most cases (39 upregulated cases, and 12 downregulated cases). Among with, CDK1, CDK2, CDK4, CDK5 are upregulated in more than 50% of tumor types, indicating those molecules show high frequency of upregulation in human cancer.

Next, we asked whether CDK genes may associate with patient prognosis in common cancers. As a result, CDK2 and CDK1 survival worse than other genes in KIRC, KIRP, and KICH (Figure S1A). In BLCA, BRCA, KIRC, and STAD, mRNA expression levels of CDK family genes were also shown in Figure S1B. Additionally, gene expression analysis of the CDK family genes in normal tissues showed that CDK9 has the highest expression value in most normal tissues, such as fallopian tube, nerve, ovary, pituitary, spleen, and uterus, followed by CDK4 and CDK16 (Figure S1C).

To further understand the cause of the core CDK family genes is transcriptionally dysregulated, we questioned the copy number variation status of CDKs at the genomic level. Importantly, heterozygous amplification and heterozygous deletion both exist across cancers. However, more CDK genes show heterozygous amplification. Those results are consistent with the transcriptional level of CDK genes. In detail, CDK family genes had the highest heterozygous amplification in Kidney Renal Papillary Cell Carcinoma (KIRP) (Fig. 1C and Figure S2A). However, the frequencies of homozygous amplification and deletion were very low (Fig. 1D and Figure S2A). We further explored the analysis of the Pearson correlation between CNV and mRNA expression and found that the expression levels of almost all CDK
family genes were significantly associated with CNV in BRCA. Among them, CDK12 was the most relevant (Figure S2B). Taken together, these above findings suggested that heterozygous amplification may contribute to the transcriptional upregulation of CDK family genes.

3.2 Demethylation also links with the high expression of CDK family in cancers

DNA methylation is an important mechanism involved in aberrant gene expression and carcinogenesis (12). Therefore, we further performed the methylation analysis in various cancer types. As a result, CDK6 has the highest increase in methylation rate in BRCA, followed by CDK17 in UCEC and BLCA, and CDK9 in LIHC. The methylation levels of CDK16, CDK1, CDK11B, CDK10, and CDK5 were significantly decreased (Fig. 2A). Previous reports have revealed that the methylation levels may be associated with the regulation of gene expression in cancer tissues (13). Therefore, we applied the correlation analysis between expression and methylation levels of CDK family genes in various cancer types. The results showed that CDK6, CDK9, CDK4, and CDK18 genes with increased methylation levels, and CDK16 and CDK10 genes with decreased methylation levels were negatively correlated with gene expression in some cancer types (Fig. 2B). We also studied the patients’ overall survival difference between hypermethylation and hypomethylation, indicating that CDK2, CDK15, CDK10, CDK17, CDK12, CDK11B, and CDK11A survival worse than other CDK members in several cancer types (Figure S3). In a nut shell, those data indicate that demethylation of CDK gene promoters also associate with the overexpression of CDK genes.

3.3 Frequency analysis of CDK family genes mutation in cancers

Firstly, we detected the mutation profile in various cancer types and found that most of the CDK family genes were frequently mutated. CDK12 and CDK13 have the highest mutation frequency in UCEC, STAD, BLCA, and COAD (Fig. 3A). An integrated analysis of the mutation types of CDK family genes exhibited that SNPs were dominant (Figure S4A). Missense mutations occur most frequently (Figure S4B). In addition, the mutation has the highest frequency of C > T conversion (Figure S4C). The median variation of each tumor sample was 1 (Figure S4D). For each mutation type, missense mutation has the highest number of mutations per capita (Figure S4E). The high-frequency mutation CDK family genes were CDK12, CDK13, CDK11B, CDK14, CDK11A, CDK15, CDK19, CDK18, CDK16, and CDK17, respectively (Figure S4F). Moreover, among the variances in the CDK family genes, CDK12 had the highest mutation frequency (28%), followed by CDK13 (18%), and CDK11B (10%) (Fig. 3B).

3.4 Pathway analysis of CDK family genes

To further explore the biological functions of these CDK family genes, we performed the correlation pathway analysis. As expected, overexpression of CDK genes associates with activated cell cycle pathway, which is in accordance with the known functions of those genes (Fig. 4 and Figure S5). Besides, activation of CDK6, CDK4, CDK2, and CDK1 could activate apoptosis and cell cycle, and inhibit RAS/MAPK. CDK17, CDK15, and CDK14 activation could activate EMT and suppress cell cycle. On the
opposite, activation of CDK16 could activate the cell cycle, and CDK7 could inhibit RAS/MAPK, RTK, and TSC/mTOR pathways (Fig. 4). In summary, our results define the enriched pathways of CDK family genes in a systems level. CDK genes show conserved signaling pathways such as activating apoptosis.

3.5 Drug sensitivity analysis of CDK family genes

To further characterize the association between CDK genes and drug responses, we calculate the correlation between CDK genes and drug response score across ~1000 cancer cell lines. The results showed that BX-912, PIK-93, XMD13-2, and KIN001-236 were strongly negatively associated with CDK13 expression (Fig. 5). PLX4720 was negatively associated with CDK2. Moreover, NPK76-72-1 was negatively related to CDK11A. PIK-93 was negatively associated with CDK14. Furthermore, Trametinib was positively associated with CDK19 and CDK1. TGX221 was positively correlated with CDK8. Navitoclax was positively associated with CDK7. In addition, Methotrexate presented the most significant positive correlation with CDK16 expression and had a significant negative correlation with CDK9 and CDK11B (Fig. 5).

Interestingly, I-BET-762, which can block the epigenetic readers - bromodomain and extra-terminal (BET) proteins (14), was found to be negatively correlated with CDK13, CDK9, CDK19, and CDK11A. Those results further confirm the epigenetics role of CDK genes. These above candidate small molecule drugs could reverse the expression of CDK family genes, thus providing novel directions and molecular mechanisms for treating cancers.

4. Discussion

Cancer is a genetic disease, as occurrence, development, and metastasis are controlled by some genetic and epigenetic alterations in the genome (15). Pan-cancer analysis of multi-omics data, combined with bioinformatics methods, can provide a special platform to identify the common molecular characteristics and the molecular mechanisms of various cancer types (16). In this study, we analyzed the molecular alterations of CDK family genes in multiple aspects, such as the genome, transcriptome, and epigenome. Firstly, gene expression levels of CDK family members were extracted from TCGA Pan-cancer. Across various cancers, we observed that CDK1, CDK5, CDK4, CDK2, CDK16, CDK7, CDK6, CDK12, CDK8, CDK17, and CDK13 were up-regulated in a wide range of cancers, whereas, CDK14, CDK15, and CDK20 were down-regulated in some cancer types, and CDK9, CDK11A, and CDK11B remained unchanged. Moreover, aberrant methylation of many genes has been associated with transcriptional inactivation of genes in various cancers (17). To determine the correlation between methylation and expression, we screened tumors with different levels of methylation and analyzed the correlation between methylation and target gene expression in different types of cancers. We found that the methylation levels of CDK4, CDK6, and CDK18 genes were significantly associated with gene expression in some cancer types. Previous evidence demonstrated that CDK4 and CDK6 could be frequently considered together as promoters of G1 progression (18). As a new regulator of genome stability, CDK18 could prevent DNA damage accumulation and genome instability (19).
Due to the deregulation of cell cycle across a broad range of cancers, cancer cells can frequently show aberrant proliferation, genomic instability consisting of increased DNA mutations and chromosomal aberrations, as well as chromosomal instability (1). In the present study, we identified the genetic mutations in the frequently mutated genes of the CDK family in different cancers. The most mutation frequency was CDK12, followed by CDK13. In line with the above, CDK12 mutations have also been reported in some cancer types, such as lung cancer (20), lymphoma (21), and advanced carcinoma of unknown primary (22). Notably, it is well known that CDK12 and CDK13 have similar biological processes, with both regulating RNA splicing and alternative splicing, maintaining self-renewal in embryonic stem cells (23). Collectively, these findings indicated that multilevel data integration exhibited CDK members with epigenetic phenotypes and distinct mutations.

To understand the functional relevance of CDK family genes, we further assessed their involved signaling pathways. We showed that activation of CDK6, CDK4, CDK2, and CDK1 could activate apoptosis and cell cycle, and inhibit RAS/MAPK. CDK17, CDK15, and CDK14 activation could activate EMT and suppress cell cycle. Besides, activation of CDK16 could activate the cell cycle, and CDK7 could inhibit RAS/MAPK, RTK, and TSC/mTOR pathways. Since cell cycle abnormalities are common in different types of cancer, it has always been considered a potential therapeutic target (24). Thus, it is of great significance to elucidate the functional roles of these CDK genes as biomarkers for therapeutic intervention. In addition, due to their role in cell cycle control, CDKs are viewed as targets of genetic manipulations in various cancers, leading to accelerate the development of small molecule drugs against CDKs as an anticancer approach (25). Therefore, we screened the CDK family genes with small molecule drugs to find potential candidate drugs that could reverse abnormally expressed CDK genes in various cancer tissues. Our analysis revealed that BX-912, PIK-93, XMD13-2, and KIN001-236 were negatively associated with CDK13 expression. Z-LLNle-CHO was negatively correlated with CDK6. PLX4720 was negatively associated with CDK2. Besides, NPK76--72-1 was negatively related to CDK11A. PIK-93 was negatively associated with CDK14. Trametinib was positively associated with CDK19 and CDK1. TGX221 was positively correlated with CDK8. Navitoclax was positively associated with CDK7. Furthermore, Methotrexate presents the most significant positive correlation with CDK16 expression and has a significant negative correlation with CDK9 and CDK11B. Previous studies have demonstrated that P276-00, as an inhibitor of CDK1, CDK4, and CDK9, could make pancreatic cancer cells sensitive to gemcitabine-induced apoptosis, and inhibit tumor growth and angiogenesis (26). Flavopiridol could also act as a pan-CDK-inhibitor of CDK1, 2, 4, 6, 7 and 9 (27), and the application of this drug in the treatment of chronic lymphocytic leukemia has achieved satisfactory results (28). Likely, these drugs of our research results could act as chemotherapeutic agents and be widely used to improve the effects of cancer therapeutics.

5. Conclusion

In conclusion, this study has demonstrated the differential expression of CDK family genes and revealed the cause of abnormal gene expression. We also found that CDK genes associates with cancer hallmarks such as RAS/MAPK and apoptosis. Further drug resistance analysis showed that CDK globally impacts drug effectiveness including Trametinib and Methotrexate. These findings provide new insights into
carcinogenesis and unravel new mechanisms of CDK genes that may be further investigated in the future.

**Abbreviations**

CDKs
Cyclin-Dependent Kinases
TCGA
The Cancer Genome Atlas
CNV
copy number variants
THCA
Thyroid Carcinoma
KIRP
Kidney Renal Papillary Cell Carcinoma
LIHC
Liver Hepatocellular Carcinoma
STAD
Stomach Adenocarcinoma
BRCA
Breast Invasive Carcinoma
COAD
Colon Adenocarcinoma
UCEC
Uterine Corpus Endometrial Carcinoma
BLCA
Bladder Urothelial Carcinoma
KIRC
Kidney Renal Clear Cell Carcinoma
KICH
Kidney Chromophobe
PRAD
Prostate Adenocarcinoma
BET
bromodomain and extra-terminal.

**Declarations**

**Authors’ contributions**
P. W. and M. X. wrote the manuscript text and put forward the idea of the article. P. W., M. X. and D. Y. contributed to completing the picture modification. E. C. and F. W. completed the revision and review of the article. All authors reviewed the manuscript.

Funding

None.

Data availability

All data were from TCGA, which are publicly available.

Ethics approval and consent to participate

Not required.

Consent for publication

Not applicable.

Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References


Figures
Figure 1

CDK family genes are widely overexpressed across human cancers. **A** The workflow of this study. **B** CDK genes are transcriptionally up-regulated in a broad spectrum of cancers. Copy number variations of CDK genes including **C** Heterozygous amplification and **D** Homozygous Amplification are common.
Figure 2

CDK family is epigenetically de-methylated. **A** The differential methylation profile of CDK genes in human cancers. **B** Correlation between methylation and expression of CDK genes.

Figure 3
Mutation also contributes to the dysfunctional CDK family in cancers. **A** The Mutation frequency of CDK genes. **B** A summary plot of variants in each sample.

**Figure 4**

CDK genes widely associate with hallmark cancer pathways in pan-cancer. Heatmap of genes that have a function (inhibit or activate) in at least 5 cancer types. Pathway_A represents activation of this pathway, inhibition in a similar way showed as pathway_I.
**Figure 5**

CDK family genes widely impact drug sensitivity in cancers. The dot represents the correlation between the CDK gene expression and drug sensitivity. The positive correlation represents that the high gene expression is resistant to the drug.

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

- FigureS1.tif
- FigureS2.tif
- FigureS3.tif
- FigureS4.tif
- FigureS5.tif