Time-course RNA-Seq profiling reveals isoform-level gene expression dynamics of the cGAS-STING pathway

Jing Sun, Lu Li, Jiameng Hu, Yan Gao, Jinyi Song, Xiang Zhang, Haiyang Hu

State Key Laboratory of Natural Medicines, School of Life Science and Technology, China Pharmaceutical University, 639 Longmian Avenue, Nanjing, China

Abstract

The cGAS-STING pathway, orchestrating complicated transcriptome-wide immune responses, is essential for host antiviral defense but can also drive immunopathology in severe COVID-19. Here, we performed time-course RNA-Seq experiments to dissect the transcriptome expression dynamics at the gene-isoform level after cGAS-STING pathway activation. The in-depth time-course transcriptome after cGAS-STING pathway activation within 12 h enabled quantification of 48,685 gene isoforms. By employing regression models, we obtained 13,232 gene isoforms with expression patterns significantly associated with the process of cGAS-STING pathway activation, which were named activation-associated isoforms. The combination of hierarchical and k-means clustering algorithms revealed four major expression patterns of activation-associated isoforms, including two clusters with increased expression patterns enriched in cell cycle, autophagy, antiviral innate-immune functions, and COVID-19 coronavirus disease pathway, and two clusters showing decreased expression pattern that mainly involved in ncRNA metabolism, translation process, and mRNA processing. Importantly, by merging four clusters of activation-associated isoforms, we identified three types of genes that underwent isoform usage alteration during the cGAS-STING pathway activation. We further found that genes exhibiting protein-coding and non-protein-coding gene isoform usage alteration were strongly enriched for the factors involved in innate immunity and RNA splicing. Notably, overexpression of an enriched splicing factor, EFTUD2, shifted transcriptome towards the cGAS-STING pathway activated status and promoted protein-coding isoform abundance of several key regulators of the cGAS-STING pathway. Taken together, our results revealed the isoform-level gene expression dynamics of the cGAS-STING pathway and uncovered novel roles of splicing factors in regulating cGAS-STING pathway mediated immune responses.

1. Introduction

The DNA molecule derived from pathogens has immensely immunogenic characteristics and therefore represents one of the most critical signals for the host’s innate immunity. The cGAS-STING pathway, which is conserved among mammalian species, has been well acknowledged as the most prominent innate-immune signaling pathway for DNA molecule sensing and host defense [1,2]. The DNA-sensing enzyme Cyclic GMP-AMP Synthase (cGAS) and the signaling adaptor protein Stimulator of Interferon Genes (STING) are two key components of the cGAS-STING pathway [3,4]. Upon double-stranded DNA binding, cGAS is activated and catalyzes the production of the critical secondary messenger called cyclic GMP-AMP (cGAMP) that binds to and activates STING. STING will then provoke the downstream signaling cascade leading to the whole transcriptome changes, including but not limited to induction of Type-I interferons, pro-inflammatory cytokines, and immune mediators, all of which are orchestrated together for creating an efficient host antiviral defense environment [1–4]. While the cGAS-STING pathway is required to achieve adequate anti-pathogen responses, over-activation or inappropriate self-DNA sensing could induce untoward damage, resulting in severe autoimmune and inflammatory diseases [5–8]. Notably, several recent studies demonstrated that the cGAS-STING pathway is a major driver of the detrimental immunopathology in severe coronavirus disease 2019 (COVID-19) [9,10]. The accumulation of aberrant cytosolic DNA molecules caused by the infection of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus, either from damaged mitochondria and engulfed dying cells or through cell fusion, all activate the cGAS-STING pathway, which

https://doi.org/10.1016/j.csbj.2022.11.044
2001-0370/© 2022 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Finally results in transcriptome-wide excessive immune responses and severe host tissue damages [10–12]. Therefore, dissecting the transcriptome changes led by the cGAS-STING pathway activation is of importance to fully elucidate the physiological and pathological roles of the cGAS-STING signaling.

The alternative splicing process has recently emerged as a critical mechanism by which innate immunity is modulated [13,14]. It is established that the alternative transcript isoform usage is widespread, with particular prevalence in the immune system [14,15]. Alternative splicing enables the production of multiple distinct proteins with altered, antagonistic, or unrelated functions from a single gene, which is essential to the rapid and flexible control of gene expression during innate immune responses [14]. For instance, mitochondrial antiviral-signaling protein (MAVS), a crucial adaptor in the RIG-I pathway, has multiple N-terminal truncated gene isoforms that are essential in preventing full-length MAVS from spontaneous aggregation to avoid accidental the misfiring of innate immune response for potential detrimental inflammation [16]. Similarly, alternative transcript isoform usage has been characterized for crucial signaling transduction components, including MyD88 and TBK1, which encode shorter RNA splice variants to control the extent of the pro-inflammatory responses [17,18]. In addition, alternative splicing can lead to gene isoform usage switch or mRNA turnover of inflammatory signaling molecules to fine-tune the immune response [19–21]. For the cGAS-STING pathway, the adaptor protein STING is the most well-characterized gene in the context of transcript isoform usages. Several gene isoforms of STING have been uncovered as the critical factors in modulating the magnitude of the cGAS-STING pathway, indicating that alternative transcript isoform usage plays a crucial role in shaping the transient nature of the immune response mediated by the cGAS-STING pathway [22–25]. However, except for several prominent examples of STING, our knowledge of the alternative transcript isoform usage involved in the regulation of the cGAS-STING pathway is quite limited, even for the repertoire of gene isoforms associated with this pathway is lacking.

To fully elucidate the physiological and pathological roles of the cGAS-STING signaling, it is fundamental to build the catalog of gene isoform repertoire and determine the dynamic expression patterns at the gene-isoform level following the cGAS-STING signaling activation. In this study, we conducted time-course RNA-Seq experiments to explore the transcriptome expression dynamics at the gene-isoform level after cGAS-STING pathway activation. Based on an in-depth time-course transcriptome of 1.58 billion reads after cGAS-STING pathway activation within 12 h, we built a catalog of gene isoform repertoire with 48,685 gene isoforms and obtained 13,232 activation-associated isoforms displaying expression patterns significantly associated with the process of cGAS-STING pathway activation by regression analysis. We further uncovered four major expression patterns of activation-associated isoforms, including two clusters showing increased expression patterns enriched in cell cycle, autophagy, innate-immune functions, and COVID-19 coronavirus disease pathway, and two clusters showing decreased expression patterns that are mainly involved in ncRNA metabolism, translation process, and mRNA processing. Importantly, we identified three types of genes that underwent gene-isoform usage alteration during the cGAS-STING pathway activation and found genes showing protein-coding and non-protein-coding gene isoform usage switch were strongly enriched in innate immune response and RNA splicing factors. Of note, we showed that overexpression of an enriched splicing factor, EFTUD2, significantly increased protein-coding isoform expression levels of HSP90C1 and SAMHD1, both are known regulators of the cGAS-STING pathway. Collectively, our results dissect transcriptome dynamics of the cGAS-STING pathway at the gene-isoform level and highlight a novel role of splicing factors in modulating the cGAS-STING pathway. The catalog of activation-associated isoforms also provides a rich resource for investigating the cGAS-STING pathway and potentially enables the identification of more precise targets at the gene-isoform level for therapeutic intervention of the cGAS-STING pathway associated diseases.

2. Materials and Methods

2.1. Cell culture

MEF cells (ATCC Cat#CRSC-1008) were maintained in Dulbecco’s modified Eagle’s medium (DMEM) supplemented with 10 % fetal bovine serum (FBS) (Gibco) and 1 % penicillin–streptomycin (Invitrogen).

2.2. The transfection experiments of G3-YSD and G3-YSD control

The transfection of G3-YSD (InvivoGen), a cGAS-specific agonist [26], was performed using Lipofectamine 2000 (Invitrogen) according to the manufacturer’s instructions. Briefly, MEF cells were transfected with G3-YSD (3 μg per well) with Lipofectamine 2000. We conducted the experiments twice and collected total RNAs at 2, 4, 6, 8, 10, and 12 h after transfection in replicates. In parallel, the same experiments were also performed by transfection of the G3-YSD control (InvivoGen) that differs from the G3-YSD in the hairpin-flanking nucleoside trimers. While G3-YSD flanking with guanosine trimers (G3) confers its cGAS agonist activity, G3-YSD control flanking with cytidine trimers (C3) abrogates cGAS activation [26].

2.3. The RNA-Seq experiment

The total RNA was extracted with the Trizol reagent (Invitrogen), subjected to polyA plus RNA enrichment using TruSeq RNA Library Preparation Kit v2 (Illumina) and further prepared into the cDNA library according to the Illumina strand-specific RNA-Seq instruction. Before library construction, the RNA integrity was estimated according to the RIN (RNA integrity number) value using Agilent 2100, and samples with the RIN value of no less than nine were retained for library construction. The constructed cDNA libraries were sequenced using Illumina HiSeq4000 in the 2 × 150 nt paired-end layout.

2.4. The RNA-Seq quality control

The raw RNA-Seq data were processed using Trimmomatic [27] to remove low-quality reads and potential adaptor contaminations with the default parameter. FastQC and MultiQC was used to visualize the quality of processed RNA-Seq data [28].

2.5. The gene isoform expression quantification

We used two popular RNA-Seq data quantification methods, RSEM [29] and Salmon [30], for gene isoform expression abundance estimation. While RSEM is an alignment-based gene isoform quantification method, Salmon is an alignment-free gene isoform quantification method. Both methods have high performance in gene isoform quantification estimated previously [31,32]. We used these two methods to estimate the potential quantification biases caused by RNA-Seq reads alignment algorithm and specific gene isoform quantification algorithm used. The mouse genome (GRCm38), gene loci annotation file, and gene cDNA annotation file were downloaded from Ensembl [33]. For RSEM quantitation, we used default parameters except for specifying strand-specific reads mapping with Bowtie2. For Salmon quantitation, we used default
parameters except for specifying strand-specific reads and using a decoy-aware transcriptome file by combining the mouse gene cDNA annotation and the mouse genome sequences. The Pearson correlation was calculated for the gene isoform quantification results between RSEM and Salmon to estimate the quantification consistency. For each isoform, we used the average expression of RSEM and Salmon to estimate its expression abundance in each sample.

2.6. Identification of gene isoforms significantly associated with the cGAS-STING signaling activation process

We tested the effect of the cGAS-STING signaling activation process (the time-course intervals of cGAS-STING pathway activation) on isoform-level gene expression abundance using polynomial regression models, following the method we developed previously [34,35]. Based on G3-YSD stimulated RNA-Seq data, we tested all possible linear to cubic regression models for each gene and chose the best regression model with the cGAS-STING signaling activation process as the predictor and gene isoform expression abundance as the response based on the “adjusted r²” criterion. The significance of the chosen regression model was estimated using the F-test. The gene isoforms with an F-test p-value <0.05 were considered as cGAS-STING signaling activation associated gene isoforms (activation-associated isoforms). The same method was applied to G3-YSD control treated RNA-Seq data to obtain the background number of activation-associated gene-isoforms that was expected by chance. The empirical FDR was estimated by the ratio between the background number of activation-associated gene-isoforms based on G3-YSD control treated RNA-Seq data and the number of activation-associated gene-isoforms based on G3-YSD treated RNA-Seq data.

2.7. The expression pattern clustering analysis

We used hierarchical and k-means clustering algorithms to cluster activation-associated gene-isoforms with similar expression profiles into groups. To cluster the expression pattern instead of the expression abundance, the expression abundances of each gene across the cGAS-STING activation process were standardized with the heuristic nature of k-means clustering, we performed k-means clustering 10,000 times to obtain the most stable clustering constellation.

2.8. Permutation analysis

To obtain the expected number of isoforms that were associated with the cGAS-STING pathway activation only at the isoform level in each cluster, we performed 1,000 times permutation test. Specifically, we identified in total 13,232 activation-associated isoforms, including 5,408 activation-associated isoforms derived from 3297 genes that were associated with the cGAS-STING pathway activation at the isoform level but not at the gene level. To obtain the expected number of isoforms that were associated with the cGAS-STING pathway activation only at the isoform level in each cluster, we randomly selected 5,408 activation-associated isoforms from 13,232 activation-associated isoforms, and then counted the number of activation-associated isoforms that were associated with the cGAS-STING pathway activation only at the isoform level in each cluster. The distribution of the expected number of activation-associated isoforms in each cluster were estimated based on 1,000 times permutations.

2.9. Enrichment analysis

The gene ontology (GO) and KEGG enrichment analysis was performed using David Bioinformatics (https://david.ncifcrf.gov/). The significantly enriched GO items were obtained based on the FDR cutoff <10 %. All expressed genes were used as the background.

2.10. Biological pathway prediction using large-scale transcriptome data

We predicted the biological pathways in which the splicing factors from Type-I genes might be involved based on >8,000 RNA-Seq data of 53 human tissues from the GTEx database [36]. Specifically, for each splicing factor, we first calculated its expression correlation with all expressed genes to obtain the genes that were significantly correlated with its expression after multiple testing corrections (Pearson correlation FDR < 0.01 & Spearman correlation FDR < 0.01) and then calculated the enriched biological pathways of the obtained genes based on KEGG pathway annotation using Fisher's exact test at 5 % FDR cutoff. The absolute value of the FDR value in the log10-scale was used as the association score. All expressed genes of 53 human tissues were used as the background. To obtain the biological pathways that were commonly enriched for the splicing factors from Type-I genes, we first applied the above procedure to get the biological pathways for each gene and then used 1,000 times permutation test and Fisher's exact test to estimate the enrichment significance.

2.11. The analysis of the EFTUD2 overexpression RNA-Seq data

To examine the effects of EFTUD2 on the cGAS-STING pathway, we defined the induced and repressed genes caused by the GAS-STING pathway activation using negative binomial test (FDR < 0.05, absolute LFC > 0.25) based on G3-YSD transfection RNA-Seq data of GSE142735 that we generated previously [37], we then estimated the effects of EFTUD2 overexpression by comparing the expression changes of the obtained induced and repressed genes to that of all expressed genes using Wilcoxon rank sum test based on EFTUD2 overexpression RNA-Seq data of GSE154438 that published previously [38].

2.12. Real-time PCR

The quantification of selected gene isoforms was conducted by real-time PCR using the SYBR Green PCR mix (Applied Biosystems). The obtained values were normalized to the level of GAPDH mRNA. The primers used are listed as follows:

- ENSMUST00000057725, Forward: TCGGAGAGACTGTCACCAT.
- Reverse: TGTCACATGTCACCAT.
- ENST000000646673, Forward: TGAGCTAGATCACGCCATT.
- Reverse: GAATGTACCTGCGAGCTT.
- ENSMUST000000020238, Forward: CTCACTAGAGCTTGGATGA.
- Reverse: CTCTGTGGCTTCCGAGCCT.
- ENST00000681861, Forward: TTGTACCCACATCTGCTCCA.
- Reverse: TGTAAGCAGTTGCGAGTGTG.

2.13. Data records

All original sequencing data were deposited in GEO database with accession number: GSE207458.
3. Results

3.1. The overview of the time-course RNA-Seq profiling after the cGAS-STING pathway activation

To explore the gene expression dynamics of the cGAS-STING pathway, we conducted time-course RNA-Seq experiments after activating the cGAS-STING pathway using a well-established cGAS-specific agonist, G3-YSD [26]. The transcriptome data were collected every two hours after the cGAS-STING pathway activation within 12 h. In parallel, the same experiments were also performed by transfection of the G3-YSD control that differs from the G3-YSD in the hairpin-flanking nucleoside trimers and abrogates cGAS activation (Fig. 1A, Materials and Methods). To measure the innate-immune responses more precisely, the experiments were conducted twice, resulting in an in-depth time-course transcriptome profiling of 1.58 billion reads (Fig. 1A, Table S1). The quality of sequencing reads per base calculated using the Phred score was mostly close to 40, corresponding to an estimated sequencing error of 0.1%. Moreover, the G + C content was distributed very similarly among samples (Figure S1). These results demonstrated the high quality of RNA-Seq data. We then used an alignment-based gene isoform quantification method RSEM and an alignment-free gene isoform quantification method Salmon to quantify gene expression at the isoform level. For both methods, the isoform expression abundance was perfectly correlated between replicates at each time point (Pearson correlation > 0.99, Figures S2, S3). Furthermore, the gene-isoform abundance estimated using RSEM was highly correlated with that calculated using Salmon at each time point (Pearson correlation > 0.9, Fig. 1B, Figure S4). In addition, by grouping the gene isoforms into ten percentiles based on their expression abundance, we found the high expression correlation of isoforms between RSEM and Salmon was largely independent of isoform expression abundance (Fig. 1C), demonstrating the good consistency of isoform expression abundance estimated between RSEM and Salmon, even though these two methods utilized different quantification strategy. We then used the average expression values of RSEM and Salmon to estimate the isoform expression abundance at each time point and quantified 48,685 gene isoforms in total (Table S2). The principal component analysis (PCA) of 48,685 gene isoforms indicated gradual expression changes of gene isoforms across the time course after the cGAS-STING pathway activation within 12 h (Fig. 1D). Notably, the first two principal components explained >79% of total expression variance, suggesting that a large number of gene isoforms might undergo markedly expression changes induced by the cGAS-STING pathway activation (Fig. 1D).

3.2. 13,232 gene-isoforms were significantly associated with the cGAS-STING pathway activation process and exhibited four major expression patterns

We next investigated the expression dynamics of these 48,685 gene isoforms. Based on the time-course RNA-Seq profiling, we performed the regression analysis using the method we developed previously [34] and identified 13,232 gene isoforms exhibiting expression curves significantly associated with the time-course of cGAS-STING pathway activation, which were named activation-associated isoforms (Regression test, p < 0.05, Table S3, Materials and Methods). By contrast, only 764 gene isoforms were calculated to be significantly associated with the cGAS-STING pathway activation based on G3-YSD control transfection RNA-Seq data, indicating the empirical false discovery rate of the identified activation-associated isoforms was <6%. The interferon-stimulated genes (ISGs) are known to be induced by the cGAS-STING pathway [2]. Based on the regression analysis, the longest isoform of several well-known ISGs, such as IFIT1, IFITM3, ISG15 and MX1, were classified as activation-associated isoforms (Regression test, p < 0.005, Figure S5), thus demonstrating the validity of the method used. We next examined the expression patterns of 13,232 activation-associated isoforms. The hierarchical clustering analysis classified activation-associated isoforms into four major expression clusters (Fig. 2A, Table S4). To confirm this result, we clustered activation-associated isoforms into four groups using the K-means algorithm (Fig. 2B, Table S4) and found that 12,179 (92%) out of 13,232 activation-associated isoforms were consistently grouped into the same cluster between hierarchical clustering and K-means algorithm. The isoforms within the same cluster were correlated better based on K-means algorithm than that of hierarchical clustering, therefore we used the clustering results of K-means algorithm for more detail analysis. Overall, there were more activation-associated isoforms exhibiting decreased expression patterns (8236, 62%) than those exhibiting increased expression patterns (4996, 38%) (Fig. 2B). The C1 and C2 clusters showed increased expression pattern, and the C1 cluster was induced earlier than the C2 cluster after the cGAS-STING activation. The C3 and C4 clusters were in decreased expression patterns, and the C4 cluster decreased more dramatically while the C3 cluster showed gradually decreased expression patterns (Fig. 2B). We then investigated biological pathways and putative functions of these four clusters (Table S5). The C1 cluster was more significantly enriched in cell cycle, cell division, and autophagy processes and also involved in innate immune response (Fig. 2C). The C2 cluster was mainly enriched in the biological processes highly linked to anti-viral immune processes, including defense response to virus, cellular response to interferon-beta, negative regulation of viral genome replication, and positive regulation of T cell mediated cytotoxicity (Fig. 2D). Consistent with the differences in the enriched biological processes, the C1 cluster was preferentially expressed in pre-autophagosomal structure, autophagosome, and mitotic spindle, while C2 cluster was enriched in the cellular components including cytoplasm, cytosol, endoplasmic reticulum exit site, and MHC class I peptide loading complex (Figure S6 and Table S6). For the clusters in decreased expression pattern, the C3 cluster was enriched in the processes including chromatin organization, mRNA processing, RNA splicing, and regulation of translation (Fig. 2E), while the C4 cluster was more preferentially enriched in the processes related to the non-coding RNA biogenesis and processing, including RNA processing, maturation of SSU-rRNA, tRNA processing, and ribosome biogenesis (Fig. 2F). Both decreased clusters were preferentially expressed in the nucleus and nucleoplasm (Figure S6 and Table S6). More recently, the cGAS-STING pathway has been identified as a major driver of the detrimental immunopathology in COVID-19 [9–11]. Notably, the C2 cluster was strongly enriched in the Coronavirus disease - COVID-19 based on KEGG pathway enrichment analysis (FDR < 0.002, Figure S7).

3.3. 3,297 genes were associated with the cGAS-STING pathway activation process at the isoform expression level but not the gene expression level

We next examined the potential advantage of analyzing the expression dynamics of the cGAS-STING pathway at the isoform expression level than that at the gene expression level. The 13,232 activation-associated isoforms were derived from 7,512 genes. By applying the same regression analysis method, 4,636 genes were identified to be significantly associated with the time-course of cGAS-STING pathway activation at the gene expression level (Regression test, p < 0.05, FDR < 7%). Notably, 4,215 out of 4,636 genes (91%) contained at least one activation-associated
isoform exhibiting remarkable expression correlation to the corresponding gene (Pearson correlation coefficient > 0.8). Thus, 3,297 out of 7,512 genes were associated with the cGAS-STING pathway activation process at the isoform expression level only but not at the gene expression level (Fig. 3A). Furthermore, the activation-associated isoforms of 3,297 genes had similar isoform type constitutions to that of 4,215 genes (Fig. 3B). Based on 1,000 times permutation test, we found the activation-associated isoforms of 3,297 genes had significantly more proportion derived from C1 and C2 clusters compared with random background (permutation test, p < 0.001, Fig. 3C), suggesting they might be involved in the regulation of cell cycle, autophagy, and anti-viral innate-immune functions. Collectively, the above results demonstrated the necessity of analyzing the expression dynamics of the cGAS-STING pathway at the isoform expression level, and 3,297 genes containing activation-associated isoforms would otherwise be missed when analyzing at the gene expression level.

3.4. Identification of three types of genes displaying protein-coding isoform usage alteration of activation-associated gene-isoforms

The isoform usage alteration is an essential mechanism by which innate immunity is modulated, we therefore investigated the expression curves of 9,911 activation-associated isoforms from 3,339 genes containing multiple activation-associated isoforms. As expected, activation-associated isoforms within a single gene were mainly positively correlated (Fig. 4A). Intriguingly, we found 1,205 pairs of activation-associated isoforms that were strongly negatively correlated (Pearson correlation coefficient < -0.6, Fig. 4A, Table S7) and exhibited inverted expression patterns that were preferentially derived from the C2 and C4 clusters (Fisher’s exact test, p < 0.00001, Fig. 4B). Both down-regulated and up-regulated activation-associated isoforms mainly comprised four isoform types. The protein-coding transcript (PCT) isoform type represented ~50% of total isoform types, while the rest three types of non-coding transcript (NPCT) isoform, including retained intron, processed transcript, nonsense mediated decay, accounted for the other half part (Fig. 4C). We then merged three types of NPCT isoforms and identified three types of genes that underwent protein-coding isoform usage alteration, including two types of genes showing isoform switches of PCT and NPCT, and one type of genes underwent isoform switches of PCT only (Fig. 4C and D, Table S8). The Type-I genes contained 303 activation-associated isoform pairs, in which PCT isoforms were increased while NPCT isoforms were decreased after the cGAS-STING pathway activation. The Type-II genes contained 264 activation-associated isoform pairs, in which PCT and NPCT isoforms exhibited the opposite expression pattern compared with that of Type-I genes. The Type-III genes contained 394 activation-associated isoform pairs, in which both isoforms were protein-coding transcript isoforms. The three types of genes were largely not overlapped with each other, indicating the majority of them underwent one type of activation-associated isoform usage alteration event (Fig. 4E). Notably, the Type-I genes had significantly more overlaps with known innate immune genes based on 1000 times permutation test (Fig. 4F), suggesting a closer association of this type of genes with innate immunity.

3.5. The genes exhibiting protein-coding and non-protein-coding isoform usage alteration of activation-associated gene-isoforms were enriched with mRNA splicing factors

We next investigated Type-I and Type-II genes that underwent protein-coding and non-protein-coding usage alteration in more detail. In line with the previous permutation test result, functional enrichment analysis showed that Type-I genes were significantly enriched in the process of innate immune response (Fig. 5A). Intriguingly, both Type-I and Type-II genes were enriched for the factors of RNA splicing and mRNA processing, suggesting that the cGAS-STING pathway modulated these enriched splicing factors that worked coordinately to regulate protein-coding and non-protein-coding transcript usage alteration of Type-I and Type-II genes. To check this further, we analyzed the potential protein interactions/associations between the splicing factors from the
Type-I and Type-II genes. Based on the gene interaction information from the STRING database, the splicing factors from Type-I genes had significant more protein–protein interactions (PPI) between each other (PPI enrichment \( p < 1.92 \times 10^{-8} \)), the observed edges: 33, the expected edges: 4). Notably, several members of these splicing factors have been shown to be involved in the innate immune process (Fig. 5B). Moreover, tissue expression enrichment analysis indicated that those splicing factors were the most preferentially expressed in the lymphocyte (FDR < 2.54 \times 10^{-5}, Table S9).

Similarly, the splicing factors from Type-II genes also interacted more frequently than expected (PPI enrichment \( p < 1 \times 10^{-7} \)), the observed edges: 29, the expected edges: 3), which were preferentially involved in the process of nonsense mediated decay (Fig. 5C). Interestingly, significantly more interactions were also observed between the splicing factors between the Type-I and Type-II genes (PPI enrichment < 1.62 \times 10^{-8}), the observed edges: 48, the expected edges: 9, Fig. 5D), suggesting more intrinsic and complex splicing machinery was underling the regulation of protein-coding and non-protein-coding transcript usage alteration of activation-associated gene-isoforms.

3.6. Overexpression of an enriched splicing factor, EFTUD2, significantly promoted protein-coding isoform expression of known regulators of the cGAS-STING pathway

Given the observations that Type-I genes were enriched for splicing factors and had significant association with the innate immune process, we next investigated whether certain splicing factors from Type-I genes might play roles in regulating the cGAS-STING pathway mediated innate immune response. To test this, we developed a method to infer the biological pathways in which the splicing factors of Type-I genes might be involved based on large-scale transcriptome data of multiple human tissues (Materials and Methods). As a validation of the method developed, the top four predicted biological pathways of the splicing factors of Type-I genes re-discovered all three annotated biological pathways that were shared by at least two splicing factors (Fig. 6A, Tables S10 and S11). Of note, several biological pathways/processes that were identified to be enriched in four major clusters based on gene ontology analysis, such as cell cycle, autophagy, and ribosome biogenesis, were among the predicted biological pathways in which...
the splicing factors of Type-I genes involved (Fig. 6A). Intriguingly, seven out of 11 splicing factors from Type-I genes were predicted to be involved in the Cytosolic DNA-sensing pathway (Fisher’s exact test, FDR < 0.05, Fig. 6A), suggesting the potential roles of splicing factors in the regulation of the cGAS-STING pathway. To further test this hypothesis, we focused on a splicing factor, EFTUD2, that was predicted to be the most associated with the Cytosolic DNA-sensing pathway based on the association score (Table S10, Materials and Methods) and displayed high expression correlations with both cGAS and STING, the two pivotal compo-
ponents of the cGAS-STING pathway (Pearson correlation coefficient > 0.6, p < 1e-07, Figure S8 in the Supporting Information). We first examined the effects of manipulation of EFTUD2 based on transcriptome changes. To do this, we defined the induced and repressed genes after the cGAS-STING activation based on G3-YSD transfection RNA-Seq data (Negative binominal test, FDR < 0.05, absolute LFC > 0.25) and then checked the corresponding gene expression changes after EFTUD2 over-expression. Notably, the induced and repressed genes after the cGAS-STING activation were significantly increased and decreased after EFTUD2 over-expression, respectively (Wilcoxon rank sum test, p < 1e-08, Figure S9 in the Supporting Information), indicating a regulatory role of EFTUD2 in modulating the cGAS-STING pathway activation mediated transcriptome changes. We further explored the potential underlying mechanism. Through manually checking Type-I genes, we found two known regulators of the cGAS-STING pathway, HSP90β and SAMHD1, which regulate the sensing and magnitude of the cGAS-STING pathway activation [39,40]. We hypothesized that EFTUD2 might regulate the splicing process of these two genes, after the cGAS-STING activation mediated innate immune response. Based on RNA-Seq data, the activation-associated isoforms of HSP90β and SAMHD1 were upregulated after EFTUD2 overexpression. By performing q-PCR, we further confirmed the induction of the activation-associated isoforms of HSP90β and SAMHD1 in both human and mouse cells after EFTUD2 overexpression (Wilcoxon rank sum test, p < 0.05, Fig. 6B, C). Collectively, the above results indicated that the splicing factors of Type-I genes might play roles in modulating the cGAS-STING pathway at least partially through promoting the activation-associated coding isoform expression of known regulators of the cGAS-STING pathway.

4. Discussion

The cGAS-STING pathway is one of the essential innate-immune pathways for its pivotal role in host defense (1). Being an innate immune signaling deeply conserved across mammalian species, one hallmark of the activation of the cGAS-STING pathway is the profound and complex transcriptomic changes, which are important to create a flexible and efficient immune environment. Indeed, recent studies found STING undergoes intrinsic modulation at the isoform level, several gene isoforms of which have been uncovered as the critical factors in modulating the magnitude of the cGAS-STING pathway [22–25]. This observation strongly demonstrates that alternative transcript isoform usage is one of the critical mechanisms in shaping the transient nature of the innate immune responses. However, except for the prominent examples of STING, little is known about the gene expression dynamics at gene isoform levels after the cGAS-STING signaling activation. In this study, we conducted time-course RNA-Seq experiments after the cGAS-STING pathway activation to address this problem. By performing...
systemic computational analyses based on high quality in-depth RNA-Seq data of 1.58 billion reads, we uncovered the dynamic expression patterns at the gene-isoform level following the cGAS-STING signaling activation, which provided the basis to further elucidate the physiological and pathological roles of the cGAS-STING signaling.

The time-course RNA-Seq data generated in this study enabled the dissection of the cGAS-STING pathway more comprehensively and precisely. By performing the regression analysis, we built the first catalog of gene isoform repertoire of the cGAS-STING pathway, which encompassed >13,000 activation-associated isoforms. Clustering analysis further uncovered four major expression patterns of the activation-associated isoforms, including two clusters in Fig. 6.
increased expression patterns and two clusters in decreased expression patterns. The advantage of the time-course transcriptome allowed us to distinguish closely related sequential biological events after the cGAS-STING pathway activation. Of note, although the two upregulated clusters displayed similar expression patterns, they were enriched in distinct biological processes. The C1 cluster was induced earlier than the C2 cluster, which was strongly enriched in the autophagy-related processes. Consistently, recent studies have showed that the induction of autophagy is a primordial function of the cGAS-STING pathway, which is vital for the clearance of DNA and viruses in the cytosol [41]. Intriguingly, our results demonstrated that the biological processes of cell cycle and cell division were more significantly enriched than the autophagy process in the C1 cluster, suggesting a link between the cell cycle and cell division processes and the cGAS-STING pathway, which is largely not reported before. Several recent studies revealed an intrinsic relationship between genome stability and the cGAS-STING pathway. It was established that the precise control of cell cycle and cell division processes is essential to genome stability [42,43]. In addition, micronuclei, as a consequence of genome stability, represents an important source of immunostimulatory DNA, which is recognized by the cGAS-STING pathway as a cellular immune surveillance mechanism [44]. Our results further showed that the biological processes of the cell cycle and cell division were quickly induced by the cGAS-STING pathway activation. All these cues suggested a potentially closed association between the cell cycle and cell division processes and the cGAS-STING pathway. However, the exact mechanism and functions underlying this observation required further exploration and validation. The C2 cluster was strongly enriched for the processes related to antiviral innate immune response. Importantly, we found that the biological pathway of Coronavirus disease - COVID-19 was the top-ranked pathway of the C2 cluster based on KEGG pathway enrichment analysis, which was in line with recent observations that the cGAS-STING pathway is a major driver of the detrimental immunopathology in severe COVID-19 [9,11]. Therefore, the activation-associated isoforms of C2 clusters might be helpful for the identification of more precise targets at the gene-isoform level for therapeutic intervention of the cGAS-STING pathway mediated severe COVID-19. Intriguingly, the activation of the cGAS-STING pathway also has been reported to play roles in the protection of SARS-CoV-2 virus infection [45,46]. These observations further emphasized the importance of dissection of transcriptome the cGAS-STING pathway in resolving this seemingly paradoxical problem and the in-depth time-course transcriptome generated in this study would be helpful to explore the underlying mechanism.

The isoform usage alternation is a critical mechanism for modulating innate immunity. Based on the time-course transcriptome data, we greatly expanded the gene repertoire that underwent isoform usage alternation. The detailed analysis identified three types of genes showing protein-coding isoform usage alteration after the cGAS-STING pathway activation. Interestingly, Type-I and Type-II genes that underwent protein-coding and non-protein-coding usage alteration were strongly enriched for the factors of mRNA processing and splicing, suggesting that the cGAS-STING pathway activation modulated these enriched splicing factors that worked coordinately to regulate protein-coding and non-protein-coding transcript usage alteration of Type-I and Type-II genes. Indeed, the PPI between those splicing factors was significantly more than expected. We focused on the Type-I genes for more detailed analysis because Type-I genes were significantly enriched in the biological process of the innate immune response, and several splicing factors of Type-I genes have been found to play roles in regulating innate immune responses. Based on these observations, we hypothesized that the splicing factors of Type-I genes might be involved in the regulation of the cGAS-STING pathway. We further provided computational and experimental evidence to support our hypothesis. In the aspect of computational analysis, we developed a method to infer the biological pathways in which the splicing factors of Type-I genes might be involved based on large-scale transcriptome data of multiple human tissues. Notably, seven out of 11 splicing factors from Type-I genes were predicted to be involved in the Cytosolic DNA-sensing pathway, which was significantly more than expected. To further provide experimental evidence to support our hypothesis, we focused on an enriched splicing factor, EFTUD2, that was predicted to be the most associated with the Cytosolic DNA-sensing pathway based on the association score and displayed high expression correlations with both cGAS and STING. Strikingly, based on the transcriptome study, we found that the overexpression of EFTUD2 shifted the transcriptome towards the cGAS-STING pathway activated status. Mechanistically, we found EFTUD2 overexpression promoted coding isoform expression abundance of several known regulators of the cGAS-STING pathway. We further confirmed these results using q-PCR in both human and mouse cells. These results indicated that the splicing factors of Type-I genes might play roles in modulating the cGAS-STING pathway at least partially through promoting the activation-associated coding isoform expression of known regulators of the cGAS-STING pathway. Of note, EFTUD2 is also involved in the restriction of Hepatitis C virus infection through regulating the alternative splicing of RIG-I and MDA5 [47]. In addition, EFTUD2 also play a role in controlling the alternative mRNA splicing of MyD88, a critical signaling adaptor in multiple Toll-like receptor (TLR) signaling pathways [48]. These observations further highlight the potential roles of alternative splicing in the modulation of the innate-immune response.

5. Conclusion

In summary, our study uncovered the isoform-level gene expression dynamics of the cGAS-STING pathway and identified novel roles of splicing factors in regulating cGAS-STING pathway mediated immune responses. Moreover, the activation-associated isoforms and the time-course high quality transcriptome generated also provided rich resources to fully elucidate the mechanisms underlying diverse physiological and pathological processes governed by the cGAS-STING pathway.

Funding

Dr. Haiyang Hu was supported by the National Key R&D Program of China (2021YFF0702003), National Natural Science Foundation of China (31601068), Key R&D Project of Jiangsu Province (BE2020725), the “Double First-Class” Project of China Pharmaceutical University (CPU2022Q201), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

CRediT authorship contribution statement

H.H. conceptualized this study; L.L., J.H., X.Z., Y.G., and H.H. performed experiments; J.S., J.H., L.L., Y.G. and J.Y.S. performed data analysis, data visualization, and data curation; H.H. acquired funding for this study. H.H. wrote the original draft with contributions from all authors.

Declaration of 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.
Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2022.11.044.

References

[1] Hu MM, Shu HB. Innate immune response to cytoplasmic DNA: mechanisms and diseases. Annu Rev Immunol 2020;38:79–98.

[2] Bogler KO, Horung V. Molecular mechanisms and cellular functions of cGAS-STING signaling. Nat Rev Mol Cell Biol 2020;21(9):501–21.

[3] Zhong B, Yang Y, Li S, Wang YG, Li Y, Diao F, et al. The adaptor protein MITA links virus-sensing receptors to IRF3 transcription factor activation. Immunity 2008;29(4):538–50.

[4] Wu J, Sun L, Chen X, Du F, Shi H, Chen C, et al. Cyclic GMP-AMP is an endogenous second messenger in innate immune signaling by cytosolic DNA. Science 2013;339(6121):826–30.

[5] Mackenzie KJ, Carroll P, Lettice L, Tarnauskaite Z, Reddy K, Dix F, et al. Ribonuclease H2 mutations induce a cGAS/STING-dependent innate immune response. EMBO J 2016;35(8):831–44.

[6] Kerur N, Fukuda S, Banerjee D, Kim Y, Fu D, Apicella I, et al. cGAS drives non-canonical-inflammation activation in age-related macular degeneration. Nat Med 2018;24(1):50–61.

[7] Sitter DA, Martinez J, Hao L, Chen X, Sun N, Fischer TD, et al. Parkin and PINK1 mitigate STING-induced inflammation. Nature 2018;561(7722):258–62.

[8] Yu Y, Liu Y, An W, Song J, Zhang Y, Zhao X. STING-mediated inflammation in Kupffer cells contributes to progression of nonalcoholic steatohepatitis. J Clin Invest 2019;129(2):546–55.

[9] Andreakos E. STINGing type I IFN-mediated immunopathology in COVID-19. Nat Immunol 2022;23(4):478–80.

[10] Li H, Zhou F, Zhang L. STING, a critical contributor to SARS-CoV-2 immunopathology. Signal Transduct Target Ther 2022;7(1):106.

[11] Domizio JD, Guleen MF, Saidoune F, Thacker VV, Yatin A, Sharma KX, et al. The cGAS-STING pathway drives type I IFN immunopathology in COVID-19. Nature 2022;603(7899):145–51.

[12] Liu X, Wei L, Xu F, Zhao F, Huang Y, Fan Z, et al. SARS-CoV-2 spike protein-induced cell fusion activates the cGAS-STING pathway and the interferon response. Sci Signal 2022;15(729):eaeb8744.

[13] Martinez NM, Lynch KW. Control of alternative splicing in immune responses: many regulators, many predictions, much still to learn. Immun Rev 2013;253(1):216–36.

[14] Lynch KW. Consequences of regulated pre-mRNA splicing in the immune system. Nat Rev Immunol 2004;4(12):531–40.

[15] Wang ET, Sandberg R, Luo S, Kruehlichova I, Zhang L, Mayr C, et al. Alternative isoform regulation in human tissue transcriptions. Nature 2008;456(7221):470–6.

[16] Qi N, Shi Y, Zhang R, Zou W, Yuan B, Li X, et al. Multiple truncated isoforms of MAVS prevent its spontaneous aggregation in antiviral innate immune signalling. Nat Med 2018;14(7):1567–76.

[17] De Arras L, Alper S. Limiting of the innate immune response by SF3A-dependent control of Dyb88 alternative mRNA splicing. PLoS Genet 2013;9(10):e1003855.

[18] Deng W, Shi M, Han M, Zhong J, Li Z, Li W, et al. Negative regulation of virus-triggered IFN-beta signaling pathway by alternative splicing of TRIB1. J Biol Chem 2008;283(51):33590–7.

[19] Lovett DH, Mahimkar R, Kaffai BL, Cape I, Maklashina E, Cecchini G, et al. A novel intracellular isoform of matrix metalloproteinase-2 induced by oxidative stress activates innate immunity. PLoS One 2012;7(4):e34177.

[20] Raum T, Jiang YJ, Hedrich CM, Kis-Toth K, Tisokos GC. A novel isoform of the orphan receptor RORgammaT suppresses IL-17 production in human T cells. Genes Immun 2012;13(4):346–50.

[21] Han KJ, Yang Y, Xu LG, Shu HB. Analysis of a TIR-less splice variant of TRIF reveals an unexpected mechanism of TRLR3-mediated signaling. J Biol Chem 2010;285(17):12543–50.

[22] Li C, Feng L, Luo WW, Lei CQ, Li M, Shu HB. The RNA-binding protein LUC7L2 mediates MITA/STING intron retention to negatively regulate innate antiviral response. Cell Discov 2021;7(1):46.

[23] Pang WH, Fung SY, Gao WW, Deng JJ, Cheng Y, Chaudhary V, et al. A novel transcript isoform of STING that sequesters cGAMP and dominantly inhibits innate nucleic acid sensing. Nucleic Acids Res 2018;46(8):4045–71.

[24] Li X, Zou Y, Zhang X, An X, Weng M, Shi J, et al. An alternatively spliced STING isoform localizes in the cytoplasmic membrane and directly senses extracellular cGAMP. J Clin Invest 2022;132(3):1162–70.

[25] Herzer AM, Hagmann CA, Coldeek M, Wolter S, Kubiker K, Wittmann S, et al. Sequence-specific activation of the DNA sensor cGAS by Y-form DNA structures as found in primary HIV-1 cDNA. Nat Immunol 2015;16(10):1025–33.

[26] Bolger AM, Lohse M, Usadel B. Trimomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30(15):2114–20.

[27] Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 2016;32(19):3047–8.

[28] Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinf 2011;12:323.

[29] Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 2017;14(4):417–9.

[30] Sarantopoulou D, Brooks TG, Nayak S, Mrcela A, Lahens NF, Grant GR. Comparative evaluation of full-length isoform quantification from RNA-Seq. BMC Bioinf 2021;22(1):266.

[31] Zhang C, Zhang B, Lin LL, Zhao S. Evaluation and comparison of computational tools for RNA-seq isoform quantification. BMC Genomics 2017;18(1):583.

[32] Cunningham F, Achuthan P, Akanni W, Allen J, Amode MR, Ardean IM, et al. Ensembl 2019. Nucleic Acids Res 2019;47(1):D745–51.

[33] Somel M, Franz H, Lorenc A, Guo S, Giger T, et al. Transcriptional neoteny in the human brain. Proc Natl Acad Sci USA 2009;106(14):5743–8.

[34] Hu H, Liu JM, Hu Z, Jiang X, Yang X, Li J, et al. Recently evolved tumor suppressor transcript TPF73-A5 functions as sponge of human-specific miR-941. Mol Biol Evol 2018;35(5):1063–77.

[35] Consortium GT. The GTeX Consortium atlas of genetic regulatory effects across human tissues. Science 2020;369(6509):1318–30.

[36] Song J, Zhang L, Li C, Maimaiti M, Sun J, Hu J, et al. miR-6-mediated modulation coupled with transcriptional regulation shapes long noncoding RNA repertoire of the cGAS-STING signaling. Comput Struct Biotechnol J 2020;22:10785–97.

[37] Tu M, He L, You Y, Li J, Yao N, Qi C, et al. EFTUD2 maintains the survival of tumor cells and promotes hepatocellular carcinoma progression via the activation of STAT3. Cell Death Dis 2020;11(10):830.

[38] Maelfait J, Bridgeman A, Benlahrech A, Cursi C, Rehwinkel J. Restriction by Rauen T, Juang YT, Hedrich CM, Kis-Toth K, Tsokos GC. A novel isoform of the immune regulator restricting hepatitis C virus infection through the RIG-I/MDA5 pathway. J Virol 2015;89(13):6508–18.

[39] De Arras L, Laws R, Leach SM, Pontis K, Kasten J, Freedman JH, Schwartz DA, et al. Comparative genomics RNAi screen identifies Eftud2 as a novel regulator of innate immunity. Genetics 2014;197(2):485–96.