Computational and Structural Biotechnology Journal 20 (2022) 1785–1797

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

m^{6}A-mediated modulation coupled with transcriptional regulation shapes long noncoding RNA repertoire of the cGAS-STING signaling

Jinyi Song [a,1], Lele Zhang [b,1,*], Chenhui Li [a,1], Munire Maimaiti [a], Jing Sun [a], Jiameng Hu [a], Lu Li [a], Xiang Zhang [a], Chen Wang [a,*], Haiyang Hu [a,*]

[a] State Key Laboratory of Natural Medicines, School of Life Science and Technology, China Pharmaceutical University, 639 Longmian Avenue, Nanjing, China
[b] Shanghai Pulmonary Hospital, Tongji University School of Medicine, Shanghai, China

ARTICLE INFO

Article history:
Received 15 February 2022
Received in revised form 2 April 2022
Accepted 2 April 2022
Available online 09 April 2022

Keywords:
lncRNA
cGAS-STING signaling
N^6-methyladenosine (m^{6}A)
Transcriptional regulation
Transcription factor

ABSTRACT

The cGAS-STING signaling plays pivotal roles not only in host antiviral defense but also in various non-infectious contexts. Compared with protein-coding genes, much less was known about long noncoding RNAs involved in this pathway. Here, we performed an integrative study to elucidate the lncRNA repertoire and the mechanisms modulating lncRNA's expression following cGAS-STING signaling activation. We uncovered a reliable set of 672 lncRNAs closely linked to cGAS-STING signaling activation (cs-lncRNA), which might be associated with type-I interferon response and infection-related phenotypes. The ChIP-seq analysis demonstrated that cs-lncRNA was strongly regulated at the transcriptional level. We further found *N*6-methyladenosine (m^{6}A) regulatory machinery was indispensable for establishing cs-lncRNA repertoire via modulating m^{6}A modification on cs-lncRNA transcripts and promoting the expression of signaling transduction key components, including IFNAR1. Loss of IFNAR1 led to the dysregulation of cs-lncRNAs resembled that of loss of an essential subunit of m^{6}A writer METTL14. We also found m^{6}A system affected transcriptional machinery to modulate cs-lncRNAs by targeting multiple crucial transcription factors. Inhibiting an m^{6}A modification regulated transcription factor, EZH2, markedly enhanced the expression pattern of cs-lncRNAs. Taken together, our results uncovered the composition of the cs-lncRNAs and revealed m^{6}A-mediated modulation coupled with transcriptional regulation significantly shaped cs-lncRNA repertoire.

1. Introduction

The life of almost every organism is assaulted by pathogens. The pathogen-derived DNA molecule is one of the most crucial signals for innate immunity. Because of the immensely immunogenic feature of DNA, organisms, especially mammalian species, have evolved powerful signaling pathways for sensing intracellular foreign DNA molecules. One of the most prominent innate-immune signaling pathways for DNA sensing is the cGAS-STING signaling pathway, which has been acknowledged as the major signaling for DNA molecule detection and host defense [1]. Two pivotal components of the cGAS-STING signaling are the DNA-sensing enzyme cGAS (Cyclic GMP-AMP Synthase) and the signaling adaptor protein STING (Stimulator of Interferon Genes) [2–4]. Upon binding to double-stranded DNA, cGAS is activated and forms ladder-like networks in a DNA-length-dependent manner, which converts ATP and GTP into a cyclic dinucleotide, cGAMP (cyclic GMP-AMP). cGAMP is a key secondary messenger that binds and activates STING located in the endoplasmic reticulum (ER). STING will then dimerize and translocate from the ER to Golgi apparatus and phosphorylates TBK1 (TANK binding kinase 1), which further binds and phosphorylates IRF3 (interferon regulatory factor 3), provoking downstream signaling cascade leading to the induction of Type-I and Type-III interferons, and a series of other inflammatory and immune mediators, all of which are orchestrated together for host defense. In addition to foreign DNA, host DNA could also be recognized by the cGAS-STING signaling and might cause severe inflammation and potential auto-inflammation or autoimmune diseases [5]. More recently, the roles of the cGAS-STING signaling are expanding to many pathological conditions beyond the infectious context [6–7]. For instance, the dysregulation of the cGAS-STING signaling has been linked to cancer and metastasis [8–10], Huntington disease [11], age-dependent macular degeneration

* Corresponding authors.
E-mail addresses: murongdule@163.com (L. Zhang), cwang1971@cpu.edu.cn (C. Wang), haiyanghu@cpu.edu.cn (H. Hu).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.csbj.2022.04.002

(AMD) [12], amyotrophic lateral sclerosis (ALS) [13–14], Parkinson's disease [15], nonalcoholic steatohepatitis (NASH) [16] and liver fibrosis [17].

Because of the importance of the cGAS-STING signaling, substantial efforts have been paid to dissect this pathway. In addition to the critical signaling component proteins, a great number of regulators have been uncovered, which orchestrate the cGAS-STING signaling precisely at transcriptional, post-transcriptional, and post-translational levels to achieve adequate anti-pathogen responses while avoiding over-activation or the inappropriate self-DNA sensing induced untoward damages [18–19]. More recently, N^6-methyladenosine (m^6A) modification has also been reported to regulate cGAS-STING signaling. Being the most prevalent post-transcriptional modification on RNA, m^6A modification controls innate immune response to infection by regulating Type-I interferons and other crucial signaling proteins of innate immune pathways, although the exact regulatory roles seem to be controversial [20–21]. Intriguingly, the newly identified nuclear DNA sensor, hnRNPA2B1, is also an m6A reader that can partially promote DNA-induced innate immunity by facilitating m6A modification of cGAS, IFI16, and STING mRNAs. The abolishment of hnRNPA2B1 in vitro or in vivo impaired DNA virus- but not RNA virus-induced IFN-I production [22], indicating a close link between m^6A modification and the cGAS-STING signaling activation. One hallmark of innate immunity activation is vast and complex transcriptomic changes, including remarkable expression changes of hundreds of interferon-stimulated genes (ISGs), transcription factors, epigenetic modulators, and noncoding RNAs. Being a major fraction of the noncoding transcriptome, long noncoding RNAs (lncRNAs) have emerged as a critical regulatory component in controlling host innate immune responses in a sequence-dependent or sequence-independent manner [23–24]. The lncRNAs are defined as the transcripts that are longer than 200 nt in length and lack protein-coding ability. Most of the annotated lncRNAs are transcribed by RNA polymerase II (Pol II) and have 5′cap structure and polyA tails, which are similar in structure to traditional mRNAs [25]. With the advances of algorithms for estimating the transcript's protein-coding potential, lncRNAs can be identified by combining RNAseq-based transcriptome assembly and systemic estimation of the transcript's protein-coding potential [26–29]. Recent studies uncovered several critical lncRNAs that are essential to control the activity of cGAS-STING signaling [30–32]. For instance, being an HSV infection-induced lncRNA, NEAT1 is indispensable for activating the cGAS-STING pathway. Loss of NEAT1 results in dramatic impairment of the induction of IFNα, IFNβ, and downstream ISG gene MXA. Mechanistically, NEAT1 is the core component of the HDP-RNP complex by binding to HEXIM1, which is not only required for the DNA virus mediated innate immune response but also involved in the interplay between DNA damage repair and the inflammatory response. Nevertheless, except for several prominent lncRNA examples, our knowledge of the lncRNAs involved in the cGAS-STING signaling is quite limited, even for the repertoire of lncRNA associated with this pathway is lacking.

To fully elucidate the roles of lncRNAs in physiological and pathological processes governed by the cGAS-STING signaling, it is fundamental to build the catalog of lncRNA repertoire and determine how these lncRNAs are regulated following the cGAS-STING signaling activation. In this study, we conducted a systematic analysis to investigate the composition of lncRNA repertoire and explored the mechanisms governing lncRNA expression following the cGAS-STING signaling activation. We performed RNA-seq experiments to measure the transcriptomic profiling of innate immune response after treating human HFF-1 cells with three well-established cGAS-STING signaling stimulators (cGAMP, G3-YSD, and HT-DNA) and one DNA virus (HSV-1). By fully integrating the results of multiple omics data, we uncovered a reliable set of the cGAS-STING pathway activation associated lncRNAs (cs-lncRNA) that were consistently regulated across various well-established cGAS-STING signaling stimulators and DNA viruses. We found that the Pol II-mediated transcriptional machinery participated in regulating the expression of the cs-lncRNA. Moreover, we provided evidence that m^6A regulatory machinery was required for establishing cs-lncRNA repertoire. Following the cGAS-STING signaling activation, m^6A-mediated regulation modulated cs-lncRNA via not only controlling the m^6A modification on cs-lncRNA transcripts but also promoting the expression of cGAS, hnRNPA2B1, STAT1, and IFNAR1, all of which were the key components of the cGAS-STING signaling. Finally, we provided clues about the interplay between m^6A regulatory machinery and transcriptional regulation in controlling cs-lncRNA via targeting crucial transcription factors. Collectively, our results suggested m^6A-mediated modulation coupled with transcriptional regulation significantly contributed to establishing long noncoding RNA repertoire of the cGAS-STING signaling.

2. Materials and Methods

2.1. Cell culture and reagents

Human HFF-1 cells (ATCC Cat#SCRC-1041) were cultured in Dulbecco's modified Eagle's medium (DMEM) plus 10% fetal bovine serum (FBS) (Gibco), supplemented with 1% penicillin–streptomycin (Invitrogen). cGAMP (Cat#tlrl-nacga23-02), G3-YSD (Cat#tlrl-ydna) were purchased from Invivogen. Herring testis (HT) DNA (Cat#D6898) was purchased from Sigma. HSV-1 (F strain) was kindly provided by Dr. Wentao Qiao (Nankai University) [33].

2.2. G3-YSD and HT-DNA transfection

The transfection of G3-YSD and HT-DNA was performed with Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Briefly, 2×10^5 HFF-1 cells in 12 well-plate were transfected with G3-YSD (3 μg per well) or HT-DNA (4 μg per well) through Lipofectamine 2000 for 9 h before processing for RNA extraction and RNA-seq library construction.

2.3. cGAMP stimulation

For cGAMP stimulation, 2×10^5 HFF-1 cells in 12 well-plate were incubated for 30 min at 37 °C with 1 μg cGAMP in 500 μl of permeabilization buffer (50 mM HEPES, pH 7.0, 100 mM KCl, 3 mM MgCl2, 0.1 mM DTT, 85 mM sucrose, 0.2% BSA, 1 mM ATP and 0.1 mM GTP) with 10 μg/ml digitonin (Sigma). Permeabilization buffer was then removed and replaced with DMEM or RPMI 1640 medium plus 10% FBS for 8.5 h before processing for RNA extraction and RNA-seq library construction.

2.4. HSV-1 infection

For HSV-1 infection, 2×10^5 HFF-1 in 12 well-plate were infected with HSV-1 (MOI = 1) for 2 h in serum-free DMEM or RPMI 1640 medium, and then the medium was removed and replaced with DMEM or RPMI 1640 medium plus 10% FBS for 7 h before processing for RNA extraction and RNA-seq library construction.

2.5. Western blot

Cell pellets were collected and resuspended in Radio-Immunoprecipitation-Assay buffer (50 mM Tris-HCl, pH 7.4,

1 mM EDTA, 0.5% NP40, 150 mM NaCl, 1 mM Na3VO4, 0.25% Na-deoxycholate, 0.1% SDS, 0.1 mM PMSF, Roche complete protease inhibitor set). The resuspended cell pellet was vortexed for 20 s and then incubated on ice for 20 min, followed by centrifugation at 1200 g for 15 min. Afterward, supernatants were collected for subsequent Western blot analysis. The following antibodies were used: anti-STING (9664S, Cell Signaling), anti-TBK1 (ab40676, Abcam), anti-IRF3 (D83B9, Cell Signaling), anti-GAPDH (sc-32233, Santa Cruz Biotechnology), anti-Phaspho-STING (85735, Cell Signaling), anti-Phaspho-TBK1 (D52C2, Cell Signaling), anti-Phaspho-IRF3 (4D4G, Cell Signaling).

2.6. The RNA-seq experiment

For each sample, total RNA was extracted using the Trizol reagent (Invitrogen), and the RNA integrity was estimated according to the RIN (RNA integrity number) value using Agilent 2100. The RNA sample was subject to polyA + RNA enrichment using TruSeq RNA Library Preparation Kit v2 (Illumina) and prepared into the cDNA library according to the standard Illumina RNA-seq instruction. The generated cDNA library was sequenced in 2 × 150 nt paired-end layout using Illumina HiSeq4000.

2.7. Quality control of RNA-seq data

The raw RNA-seq data were preprocessed using Trimmomatic (v0.36) to remove low-quality reads and potential adaptor contamination [34]. For the obtained reads longer than 75 nt, FastQC (v0.11.7) was used to inspect overall read quality in terms of read sequencing base quality, read G + C content, and adaptor contamination. MultiQC (1.8.0) [35] was used to visualize read G + C content and the read sequencing quality in the format of the average Phred score of each base. The resulting high-quality reads were aligned to the human genome (GRCh38) using HISAT2 (v2.1.0) [36] with default parameters except for adding '-dta' option for downstream transcriptome assembly. The human genome was downloaded from the FTP of Ensembl database [37] (ftp://ftp.ensembl.org/pub). The gene body coverage of mapped reads was estimated based on the script 'geneBody_coverage.py' of the RSeQC package (v4.0.0) [38], using the transcripts of all annotated protein-coding genes as the template.

2.8. Gene expression quantification and differential expression analysis

For the protein-coding gene and lncRNA expression quantification, only uniquely mapped reads were retained to estimate gene expression abundance at count level using featureCounts (v1.6.3) [39]. The human reference gene annotation in GTF format was downloaded from the FTP of Ensembl database [37] (ftp://ftp.ensembl.org/pub). We merged human reference gene annotation and novel lncRNA annotation as the quantification template. The significantly differentially expressed (DE) genes (FDR < 0.05) and log2-transformed gene expression fold-changes (LFC) were calculated using edgeR (v3.20.9) [40] after TMM (Trimmed Mean of M−values) normalization. The Interferon-stimulated gene list of human was downloaded from Wu et al. [41]. The same analysis protocol was applied for quantification and DE gene analysis for the following datasets: VACV-70mer (dsDNA) stimulation RNAseq dataset (SRP141411), IFNAR1 KO RNAseq dataset (SRP157933), EZH2 inhibition RNAseq dataset (GSE99839).

2.9. Transcriptome assembly

The transcriptome assembly was conducted using StringTie2 (v2.0.3) [42]. Specifically, for each sample, the mapped reads in sam format resulting from running HISAT2 with '-dta' option was transformed into sorted bam format using SAMtools (v1.8) [43], and then the sorted bam file was used as the input for StringTie2 to assemble the transcriptome using the human reference annotation to guide the assembly process. The assembly transcripts were required to be longer than 300nt by using the option (-m 300). The resulting transcript assembly files of each sample were further merged into the final transcriptome assembly result in gtf format, using the merge function of StringTie2.

2.10. Novel lncRNA identification

Based on transcriptome assembly results, the novel lncRNAs were predicted using the combination of three transcript coding potential estimation tools: CPAT (v2.0.0) [26], CNCI (v2.0) [28], and CPC (v0.1) [44]. Specifically, all novel transcripts were obtained by excluding the one overlapping with known annotations with at least 1 bp. For CPAT, the novel noncoding transcripts were selected using the recommended cutoff '0.364′ according to the CPAT website. For both CNCI and CPC, the novel noncoding transcripts were obtained based on the 'noncoding' flag in the output of CNCI and CPC coding potential prediction results, respectively. The novel transcripts that fulfill the noncoding criteria of all three methods were considered as novel lncRNAs.

2.11. ChIP-seq experiment

HFF-1 cells were transfected with HT-DNA (2 μg/mL) through Lipofectamine 2000 for 9 h, and then cells were harvested with Chromatin IP kit (CST) according to the manufacturer's recommendations with some modifications. Briefly, the culture medium was removed, and cells were washed with PBS three times, then 1% formaldehyde (Sigma) was used to crosslink proteins to DNA for 10 min at room temperature. Subsequently, 10 × glycine was added to stop the reaction for 5 min. After removing the mixed medium, cells were washed with precooled PBS twice and harvested. The chromatin was further sheared with the Bioruptor Twin instrument (Diagenode). The ChIP antibodies against Pol II was purchased from Active Motif (Cat#61085, RNA Pol II CTD phosphor Ser5 antibody). ChIP was performed with chromatin from 3 million HFF-1 cells and 5 μg of antibody each time. Three replicates were performed in total. The ChIP-seq libraries were prepared and sequenced on HiSeq 2000 according to Illumina standard protocols.

2.12. ChIP-seq data analysis

The raw ChIPseq data was processed using fastp [45] with default parameters to remove adaptor sequences and reads in low quality. The resultant data were further mapped to the human genome (GRCh38) using Bowtie by allowing three mismatches. The duplicated reads were further excluded using SAMtools [43]. For the samples before and after HT-DNA stimulation, MACS2 (Model-based Analysis of ChIP-seq) [46] was used to identify Pol II binding peaks with the FDR cutoff 0.05.

2.13. ChIP-qPCR experiment

Based on two selected cs-lncRNAs, we extracted the promoter region of corresponding lncRNAs based on transcription start site annotation and Pol II ChIPseq peak regions. The extracted regions of human genome (GRCh38) were as follows: ENSG00000204261 (chr6:32843963–32844112); ENSG00000285967 (chr5:36876677–36876826). For the quantitative PCR experiment, the following primers were used to amplify ChIP DNA: For ENSG00000204261, sense primer (5′-GGGCTCCGCATACATCTAGT-3′) and antisense primer (5′-CTTGTTCCCTCCCCTTCGAT-3′); For ENSG00000285967,

sense primer (5′-TGGTTGTTAGTGTTTGGCGC-3′) and antisense primer (5′- TGTGTCTCTCTCTCGTTCCG-3′).

2.14. m6A-seq data analysis

The raw data of m6A-seq before and after VACV-70mer stimulation in human NHDF cells were downloaded from SRA (SRP141411). After removing potential adaptor sequence and reads in low quality using fastp, reads in high quality were mapped to the human genome (GRCh38) using HISAT2. For each sample, MACS2 was used to identify m^{6}A peaks by taking the corresponding input sample as the control with the default parameters at the FDR cutoff 0.05, except for '-nomodel' and '-keepdup all' to turn off fragment size estimation step and to keep all mapped reads. The high confidential m^{6}A peaks were identified by requiring support from at least two replicate samples.

2.15. The analysis related to transcription factor

To obtain a comprehensive list of transcription factors (TFs) of human, we downloaded annotated TFs from nrg2538 [47], TRANSFAC [48], JASPAR [49], GO [50], DBD [51], AnimalTFDB [52], TFCat [53], and further merged them into a final list of TF dataset. The cs-TFs were obtained by using the same approach for cs-lncRNA identification based on the compiled TF dataset. The large-scale ChIP-seq dataset of human was downloaded from ENCODE [54], which includes 338 TFs and transcription cofactors. To identify the TFs that displayed enriched binding sites in the promoter region of cs-lncRNAs, we obtained the binding sites of each TF in the promoter region (2 kb upstream and downstream centered on TSS) of cs-lncRNAs based on ChIP-seq dataset, and further used fisher's exact test to get the enriched TFs by using that of all TFs as the background (FDR < 0.05).

2.16. Permutation analysis

We used 1000 times random sampling based permutation test to calculate whether increased cs-lncRNA harbored significantly more gain of Pol II peaks in their promoters. Specifically, we randomly selected the same number of lncRNA according to the number of increased cs-lncRNA from all expressed lncRNAs 1000 times to the estimated expected frequency of the number of lncRNA harboring specific gain of Pol II peaks in the promoter region, and then compared the observed and expected frequency to estimate the statistical significance. This protocol was used for estimating the statistical significance for [1] increased cs-lncRNA harboring gain or loss of Pol II peaks; (2) decreased cs-lncRNA harboring gain or loss of Pol II peaks; (3) increased cs-lncRNA harboring gain or loss of m^{6}A peaks; (4) decreased cs-lncRNA harboring gain or loss of m^{6}A peaks; (5) increased cs-TF harboring gain or loss of m^{6}A peaks; (6) decreased cs-TF harboring gain or loss of m^{6}A peaks.

2.17. Real-time PCR

Total cellular RNA was isolated using TRIzol. The quantification of selected lncRNAs 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:

ENSG00000249669, Forward: GTTGTTCTGCAGCCATCAGC.
Reverse: GAGACCGCTTGGGTCTCAAT.
MSTRG.7460, Forward: GTTGGCCAAACAAGCTCCAC.
Reverse: GTTGGCCAAACAAGCTCCAC.
ENSG00000197536, Forward: CACCGCGTCTGTGTCTGTAT.
Reverse: CCACCCAGAAGTAACCAGGC.
MSTRG.2419, Forward: AGCTGCTGTCTTGCAGCTAA.
Reverse: CGTGCTCCGGAATCCGAATA.

2.18. Functional enrichment analysis

For each RNA-seq dataset of innate-immune stimulator and pathogen transfection, the gene ontology (GO) enrichment analyses of DE genes were performed using David Bioinformatics (https://david.ncifcrf.gov/) [55]. The significantly enriched GO items of biological process (BP) were obtained based on the FDR cutoff<5%. All expressed genes were used as the background. To infer the putative function of cs-lncRNA, we utilized GREAT [56] by taking the protein-coding genes within the 1 Mb region of the cs-lncRNA as the input. The significantly enriched items of biological processes and MGI phenotypes based on single gene KO experiments were obtained using the binomial FDR cutoff<5%.

2.19. Data records

All original sequencing data were deposited in GEO database with accession number: GSE142735.

3. Results

3.1. The overview of RNAseq profiles following the cGAS-STING signaling activation

To obtain transcriptomic changes mediated by the cGAS-STING signaling activation, we performed RNA-seq experiments to measure the transcriptomic profiling of innate immune response in human HFF-1 cells with the treatments of three well-established cGAS-STING signaling stimulators (cGAMP, G3-YSD, and HT-DNA) and one DNA virus (HSV-1) (Fig. 1A). The Human HFF-1 cell has been widely used as a model cell for studying the mechanism of innate immunity [57–63]. The DNA molecules of HSV-1 (Herpes simplex virus 1) and HT-DNA (herring testis DNA) bind and activate cGAS in a sequence-independent manner [64–65]. G3-YSD is a 26mer DNA sequence derived from the HIV-1 RNA genome, which can lead to sequence-specific activation of cGAS specifically [66]. cGAMP functioning as a specific secondary messenger that binds and activates STING [67]. All these agonists are widely used and well recognized for the cGAS-STING signaling activation. The phosphorylations of STING, TBK1, and IRF3 are hallmarks of the activation of the cGAS-STING signaling pathway [68]. Notably, treating cells with all these four agonists strongly activated the cGAS-STING signaling, as demonstrated by the enhanced phosphorylation of STING, TBK1, and IRF3 compared with the control treatment. By contrast, transfecting CpG DNA, a TLR9 specific agonist, failed to activate the cGAS-STING signaling, demonstrating the validity of our experimental design for activating the cGAS-STING signaling specifically (Fig. 1B). To measure the innate-immune responses more precisely, the experiments were conducted in two replicates for each treatment (Fig. 1A, Table S1 in Supporting Information). All RNA samples were in high integrity, as reflected by high RNA integrity number (RIN) (Table S1 in Supporting Information). On average, 43 million (SD = 2.4) reads of high quality were obtained, corresponding to 92.2% of raw reads. The quality of sequencing per base measured using Phred score [69] was mostly close to 40, corresponding to an estimated sequencing error of $\sim$ 0.1%. The reads displayed very similar distributions of G + C content (Mean = 50.3%, SD = 0.4%) and were largely distributed uniformly without 3′ bias across transcript body of annotated genes for all samples (Figure S1 in Supporting Information). On average, approximately 96% (Mean = 96.2%, SD = 1.6%) of reads were mapped to the human genome (Fig. 1C). >13,000 protein-coding genes were reliably quantified (expression abundance > 0.5 TPM)

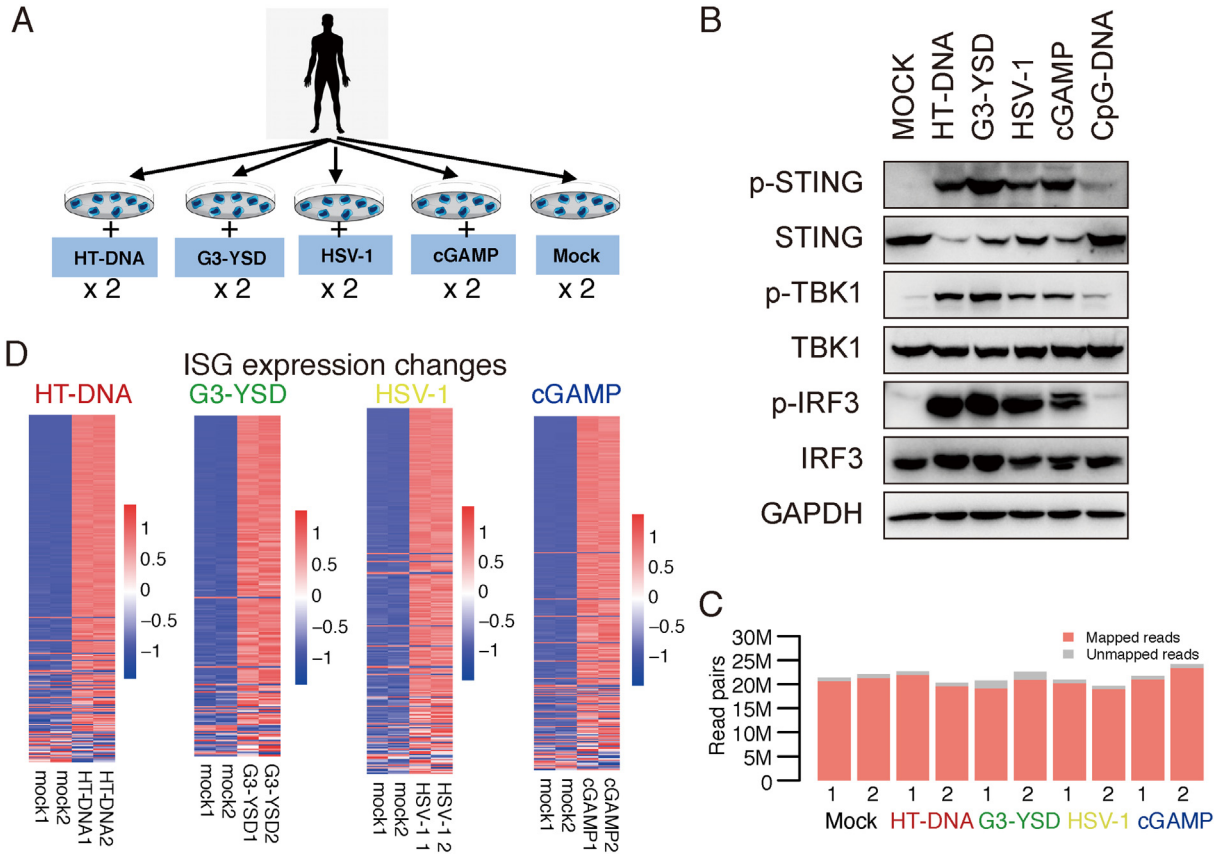

Fig. 1. The overview of RNA-seq experiments following the cGAS-STING signaling activation. **(A)** The schematic diagram of study design and sample collection for RNA-seq. **(B)** HFF cells were stimulated with HT-DNA, G3-YSD, HSV-1, cGAMP, and CpG-DNA for 9 h, respectively. Then, cell lysates were collected for western blot analysis of STING, TBK1, IRF3, GAPDH, and the phosphorylation of STING, TBK1, IRF3. **(C)** The barplot of the number of mapped paired-end reads per RNA-seq sample. **(D)** The heatmap of the expression pattern of known ISGs following the treatment of three cGAS-STING pathway stimulators and one DNA virus.

in each sample (Figure S1 in Supporting Information). Notably, the expression correlation between replicate samples was 0.99 (Figure S2 in Supporting Information). Furthermore, most of the known ISGs were remarkably increased compared to mock controls for each treatment (Fig. 1D). The differentially expressed genes were also strongly enriched in the biological processes closely associated with innate immune activation, including innate immune response, type I interferon signaling pathway, defense response to virus, and cellular response to interferon-beta (FDR < 0.01, Figure S3 in Supporting Information). Collectively, these results demonstrated the validity of experiment performance and high-quality of RNA-seq data, which provided a foundation for investigating lncRNAs.

3.2. The repertoire of the cGAS-STING signaling activation associated lncRNAs (cs-lncRNA)

To obtain a more comprehensive list of lncRNA associated with the cGAS-STING signaling, we firstly performed transcriptome assembly and novel lncRNA identification. Based on gene annotation of GENCODE (v36), we obtained 498 novel transcripts in multiple-exon structure and contained<20% repeat elements in their sequences. We then used three widely-used transcript coding potential estimation tools to evaluate their coding potential (Materials and Methods). CNCI, CPC2, and CPAT classified 96.2%, 88.7%, and 85.9% of the identified novel transcript as non-coding transcripts, respectively (Fig. 2A-C). By intersecting the results of three tools, 417 novel lncRNAs were obtained (Fig. 2D, Table S2 in Supporting Information). These novel lncRNAs mainly transcribed from the sense and antisense region of promoter as well as gene antisense region, representing almost 65% of all novel lncRNAs, while intergenic lncRNAs represented 75% of the remaining novel lncRNAs (Fig. 2E). The length of the novel lncRNA transcript was similar to that of known lnRNAs (Fig. 2F). At the expression level, novel lncRNAs were significantly more abundant than known lncRNAs following the cGAS-STING signaling activation (Fig. 2G; Wilcoxon rank sum test, p < 2.2e-16), suggesting a closer association between novel lncRNAs and the cGAS-STING pathway than all annotated lncRNAs.

We then merged known and novel lncRNAs together to screen the cGAS-STING signaling activation associated lncRNAs (cs-lncRNA), which were required to be consistently induced or repressed after activating the cGAS-STING pathway under different stimulations. Following the treatments of four cGAS-STING signaling stimulators, we found 7.4% to 30.1% of expressed known lncRNAs to be differentially expressed (DE) (Fig. 3A, B; Negative binomial test, FDR < 0.05). The proportion of DE novel lncRNAs ranged from 18.3% to 65.6% of all expressed novel lncRNAs (Fig. 3C, D), which was much higher than that of known lncRNA. To exclude the potential artifacts of DE lncRNAs caused by factors other than the cGAS-STING activation while maximizing the cs-lncRNA identification sensitivity, we merged DE lncRNAs of all treatments and obtained 672 cs-lncRNAs that were differentially expressed in at least two kinds of stimulations (Negative binomial test, FDR < 0.05) and displayed consistent expression changes across all stimulations (Fig. 3E; Table S3 in Supporting Information; Fisher's combined probability test, p < 0.005). The expression pattern of the obtained cs-lncRNAs was highly correlated between different

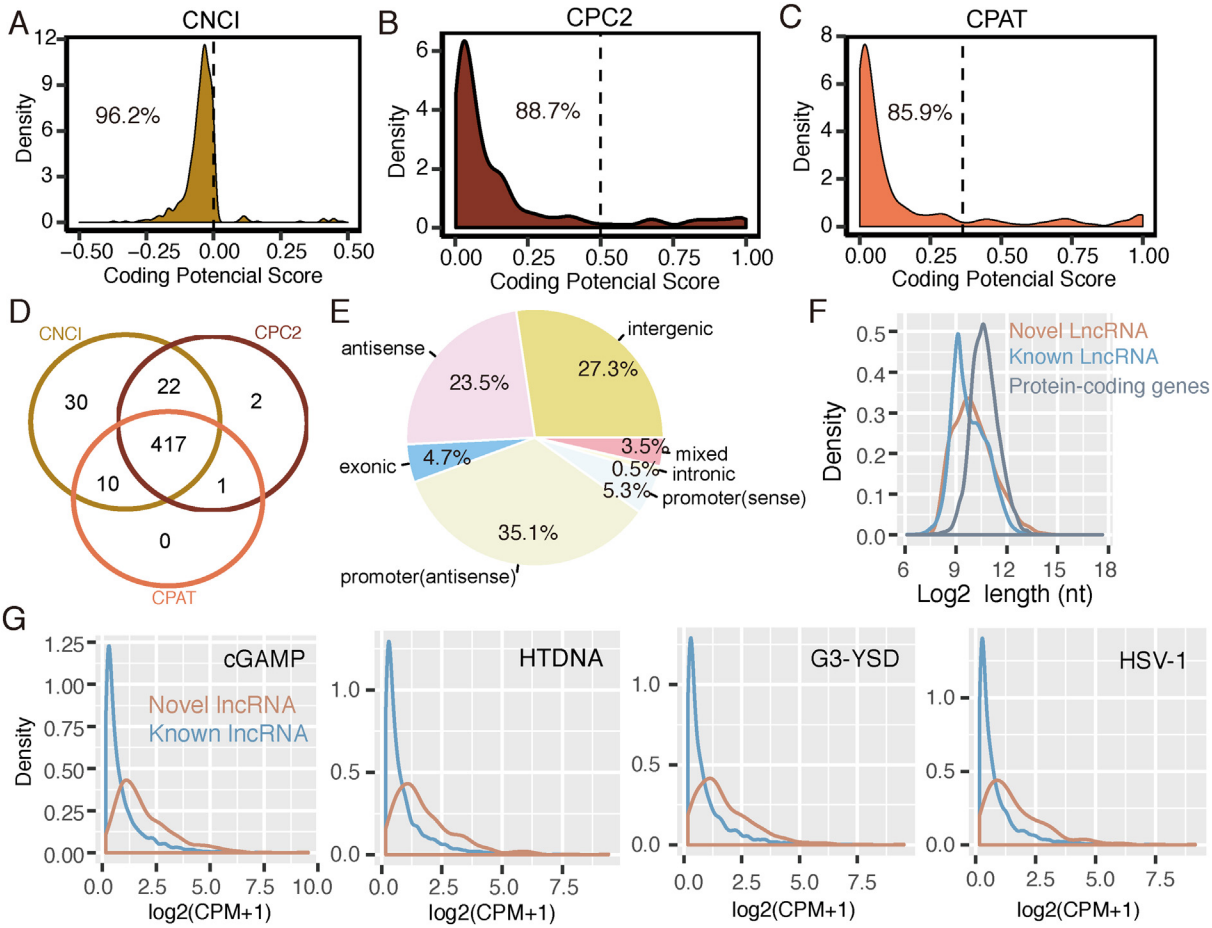

Fig. 2. The identification of novel lncRNAs. **(A, B, C)** The coding potential score distribution of novel transcripts using three coding potential estimation methods (CNCI, CPC2, and CPAT). **(D)** The overlaps of novel lncRNAs identified by CNCI, CPC2, and CPAT. **(E)** The genomic context of the identified novel lncRNAs. **(F)** The length distributions of novel lncRNAs, known lncRNAs, and protein-coding genes. **(G)** The expression abundance comparison between novel lncRNAs and known lncRNAs, after the treatments of cGAMP, HT-DNA, G3-YSD, and HSV-1.

cGAS-STING signaling stimulations (Fig. 3F, average Pearson correlation coefficient: 0.91). Notably, novel lncRNAs contributed to 28.1% of upregulated cs-lnRNAs, demonstrating the necessity for conducting novel lncRNA identification. To validate the expression pattern of cs-lncRNA, we measured expression changes of four selected cs-lncRNAs using qPCR. The result showed consistent expression changes of all four selected cs-lnRNAs between RNAseq and qPCR (Figure S4 in Supporting Information). To further check the validity of cs-lncRNAs, we examined an independent RNAseq dataset that measured transcriptome alterations following VACV-70mer (dsDNA) stimulations at 6 h and 12 h in human NHDF cells. Overall, 92.5% and 93.3% of cs-lncRNAs displayed consistent expression changes after dsDNA stimulations at 6 h and 12 h, respectively (Fig. 3G). Moreover, the expression changes of cs-lncRNA were markedly correlated between four cGAS-STING signaling stimulations in HFF-1 cells and dsDNA stimulation in NHDF cells at both 6 h and 12 h (Fig. 3H, Pearson correlation coefficient > 0.85). Collectively, these results indicated validity and reliability of the identified cs-lncRNAs, which were strongly associated with the activation of the cGAS-STING pathway, and exhibited highly consistent expression changes between multiple cGAS-STING signaling stimulations.

We next explored the putative functions of cs-lncRNA. By applying the same approach for cs-lncRNA identification, we identified 3,406 protein-coding genes associated with the cGAS-STING pathway activation (cs-PCG) (Table S4 in Supporting Information). The cs-lncRNA and cs-PCG exhibited similar genomic distributions on most chromosomes, suggesting cs-lncRNA may play roles in *cis*-regulation of nearby cs-PCG (Figure S5 in Supporting Information). We then utilized GREAT [56] to infer the putative functions of cs-lncRNA by taking the protein-coding genes within the 1 Mb region of the cs-lncRNA as the input. The results showed that cs-lncRNAs were significantly enriched in the biological processes, including type-I interferon signaling pathway, response to type-I interferon, defense response, and cytokine-mediated signaling pathway, all of which were strongly related to innate immunity (Fig. 3I; Table S5 in Supporting Information). In addition, based on mouse genes with phenotype annotation, cs-lncRNA might be associated with several infection-related phenotypes, such as abnormal response to infection, altered susceptibility to viral infection, and altered susceptibility to infection induced morbidity/mortality (Table S6 in Supporting Information). Intriguingly, cs-lncRNA might also be related to the phenotype of the decreased T cell number, which is an IRF3 independent novel function of the cGAS-STING pathway uncovered recently [70].

3.3. The expression of cs-lncRNA was regulated at the transcriptional level

We then investigated how cs-lncRNAs were regulated. Functional enrichment analysis suggested the biological processes such as DNA templated regulation of transcription and Pol II-mediated

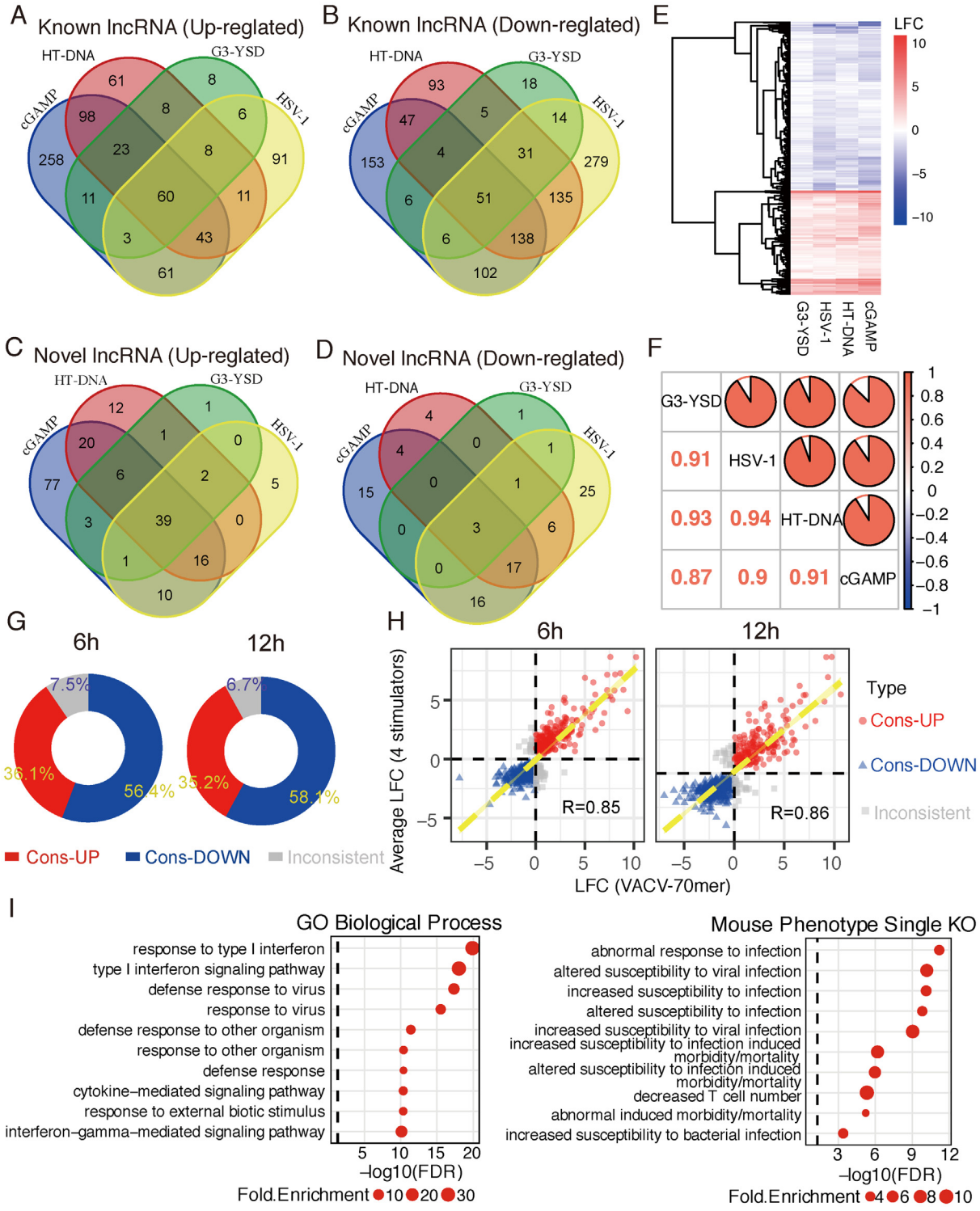

Fig. 3. The catalog of the cGAS-STING signaling associated lncRNAs (cs-lncRNA). **(A)** and **(B)** The overlaps of upregulated **(A)** and downregulated **(B)** differentially expressed (DE) known lncRNAs after the treatments of cGAMP, HT-DNA, G3-YSD, and HSV-1. **(C)** and **(D)** The overlaps of up-regulated **(C)** and down-regulated **(D)** DE novel lncRNAs after the treatments of cGAMP, HT-DNA, G3-YSD, and HSV-1. **(E)** The heatmap of the cGAS-STING signaling associated lncRNAs (cs-lncRNA) after the treatments of cGAMP, HT-DNA, G3-YSD, and HSV-1 in HFF-1 cells. **(F)** The pairwise expression correlation of cs-lncRNAs between the treatments of cGAMP, HT-DNA, G3-YSD, and HSV-1 in HFF-1 cells. **(G)** The proportions of consistently upregulated (Cons-UP), consistently downregulated (Cons-DOWN), and the rest (Inconsistent) cs-lncRNA after VACV-70mer (dsDNA) stimulation at 6 h (left) and 12 h (right) in NHDF cells, respectively. **(H)** The correlation of expression changes of cs-lncRNAs between four cGAS-STING signaling stimulations in HFF-1 cells and dsDNA stimulation in NHDF cells at 6 h and 12 h, respectively. **(I)** The top10 potential biological processes and associated phenotypes of cs-lncRNA that predicted using the protein-coding genes within the 1 Mb region of cs-lncRNAs by GREAT.

transcriptional regulation were strongly linked to the cGAS-STING pathway activation (Figure S3 in Supporting Information). We, therefore, explored whether the expression of cs-lncRNA was regulated transcriptionally. To check this, we performed Pol II ChIP-seq experiments and identified 36,606 and 35,578 Pol II peaks before and after HT-DNA stimulation (9 h), respectively. In line

with previous observations, the majority of Pol II peaks (>80%) were located in the promoter (2 kb upstream &downstream of TSS) and gene body region (Figure S6 in Supporting Information). Moreover, the known ISG genes, such as STAT1, TRIM56, MX1, MYD88, and IFI44, harbored clear Pol II peaks in their promoter region only after HT-DNA stimulation (Figure S7 in Supporting Information), demonstrating the validity of ChIP-seq experiment performance and data quality. By overlapping the identified Pol II peaks, we obtained 15,447 and 16,475 Pol II peaks that were specifically gained and lost after HT-DNA stimulation compared with unstimulated status, respectively (Fig. 4A). Notably, 31.2% of cs-lncRNAs with increased expression pattern (increased cs-lncRNA) harbored specific gain of Pol II peaks in their promoters while 24.5% of cs-lncRNAs with decreased expression pattern (decreased cs-lncRNA) harbored specific loss of Pol II peaks in their promoters, which were both significantly more than expected by chance (Fig. 4B, C; Table S7 in Supporting Information; permutation test, $p < 0.001$). By contrast, the proportion of increased cs-lncRNA harboring specific loss of Pol II peaks and the proportion of decreased cs-lncRNA harboring specific gain of Pol II peaks were similar to that of random simulations (Fig. 4B, C; permutation test, $p > 0.33$). These results suggested that the transcriptional regulation contributed to the induction and repression of cs-lncRNAs. To verify this result, we measured Pol II ChIP-seq signals quantitatively across the promoter and gene body region of cs-lncRNAs. Consistent with the result of gain and loss of Pol II peaks, we found the promoter region of increased cs-lncRNA displayed significantly more Pol II ChIPseq signals after HT-DNA stimulation (Fig. 4D, Wilcoxon rank sum test, $p < 1e\text{-}10$). Conversely, the opposite was true for decreased cs-lncRNA (Fig. 4E, Wilcoxon rank sum test, $p < 1e\text{-}10$). To further validate our finding, we conducted ChIP-qPCR to examine Pol II binding intensity on the promoter region of one increased cs-lncRNA (Fig. 4F) and one decreased cs-lncRNA (Fig. 4G). The ChIP-qPCR result fully supported the conclusion obtained based on ChIPseq (Fig. 4H, I; Wilcoxon rank sum test, $p < 0.001$). Taken together, the above results indicated that the expression of cs-lncRNA was regulated at the transcriptional level.

3.4. N^6-methyladenosine (m^6A) regulatory machinery was required for establishing cs-lncRNA repertoire

Recent studies reported that N^6-methyladenosine (m^6A) regulatory machinery regulates the innate immune response. We then investigated whether m^6A-mediated modulation was involved in regulating the expression of cs-lncRNA. METTL14 is a core component of the m^6A writer complex, depletion of which results in a substantial disturbance of the m^6A system. Notably, we found that loss of METTL14 led to dramatic dysregulation of cs-lncRNA after dsDNA stimulation. Compared with wildtype control counterparts treated with dsDNA, increased cs-lncRNA were mostly repressed while decreased cs-lncRNA were extensively upregulated in METTL14 abolished cells (Fig. 5A). The 1,000 times permutation test indicated that the downregulated proportion of increased cs-lncRNA and upregulated proportion of decreased lncRNA caused by METTL14 abolishment were both significantly more than expected (Fig. 5B, permutation test, $p < 0.001$). Moreover, the expression pattern of cs-lncRNA was markedly reversed in METTL14 knockout cells after dsDNA stimulation (Fig. 5C; Pearson correlation coefficient: −0.54). These results demonstrated that m^6A regulatory machinery might be involved in regulating cs-lncRNA expression. We then investigated the possible roles of the m^6A system in more detail. The m^6A-seq data analysis identified 15,648 m^6A peaks, including 3,818 and 1,904 ones specifically gained and lost after dsDNA stimulation compared with unstimulated status (Fig. 5D, Materials and Methods). Intriguingly, we found increased cs-lncRNAs harbored significantly more gain of m^6A peaks and significantly less loss of m^6A peaks while decreased cs-lncRNAs harbored significantly more loss of m^6A peaks (Fig. 5E, F; Table S8 in Supporting Information; permutation test, $p < 0.001$), suggesting m^6A system may play roles in controlling cs-lncRNA expression partially through affecting m^6A modification on their transcripts. We next explored whether the m^6A system might regulate cs-lncRNA via modulating signaling transduction of the cGAS-STING pathway. Interestingly, we found several signaling transduction key components, including two DNA sensors cGAS and hnRNPA2B1, an interferon receptor IFNAR1, and a signal transducer STAT1, were significantly diminished after METTL14 knockout (Fig. 5G, Negative binomial test, FDR < 0.05). Moreover, the expression of hnRNPA2B1 and IFNAR1 were consistently dampened before and after dsDNA stimulations in METTL14 knockout cells. Of note, we found that loss of IFNAR1 led to global dysregulation of cs-lncRNAs, which were significantly correlated with that of loss of METTL14 (Fig. 5H, Pearson correlation coefficient: 0.47). Taken together, these results indicated that m^6A regulatory machinery was indispensable for establishing cs-lncRNA repertoire, probably via modulating m^6A modification of cs-lncRNA transcripts and promoting the expression of signaling transduction key components.

3.5. m^6A-mediated regulation affected transcriptional machinery to modulate the cs-lncRNA expression

When checking the disrupted biological processes caused by METTL14 knockout, we found many items related to transcriptional regulation were strongly enriched, which displayed remarkable similarity to that associated with the cGAS-STING pathway activation (Fig. 6A). We, therefore, further explored the potential interplay between m^6A-mediated regulation and transcriptional machinery in modulating cs-lncRNA. To do this, we compiled a comprehensive list of transcription factors (TFs), and searched for transcription factors associated with the cGAS-STING pathway activation (cs-TFs) by using the same approach for cs-lncRNA identification (Materials and Methods). By examining their expression changes in METTL14 abolished condition, we obtained 91 cs-TFs that were significantly altered after dsDNA treatment in METTL14 knockout cells (Table S9 in Supporting Information). Strikingly, 95.6% of increased cs-TFs were repressed, while 73.9% of decreased cs-TFs were upregulated (Fig. 6B). The downregulated proportion of increased cs-TFs and upregulated proportion of decreased cs-TFs caused by METTL14 abolishment were both significantly more than expected (Fig. 6C, permutation test, $p < 0.008$). The expression pattern of cs-TFs was notably reversed in METTL14 knockout cells after dsDNA stimulation (Fig. 6D, Pearson correlation coefficient: −0.59). Furthermore, increased cs-TFs harbored significantly more gain of m^6A peaks while decreased cs-TFs harbored significantly more loss of m^6A peaks (Fig. 6E, F; Table S9 in Supporting Information; permutation test, $p < 0.001$). Collectively, these results indicated that cs-TFs were regulated by the m^6A system following the cGAS-STING pathway activation. We further checked whether cs-TFs were involved in controlling the transcription of cs-lncRNAs. By taking advantage of large-scale TF ChIP-seq data from ENCODE, we obtained 17 cs-TFs with genome-wide binding site information. Importantly, 10 out of 17 cs-TFs, including six cs-TFs with gain or loss of m^6A modifications, were also displayed enriched binding in the promoter region of cs-lncRNAs, which was significantly more than expected (Fig. 6G, hypergeometric test, $p < 0.001$). This result suggested that the expression of cs-lncRNAs might be controlled by cs-TFs. To verify this, we analyzed the expression changes of cs-lncRNAs after inhibiting a deceased cs-TF, EZH2, which has been reported to be regulated by m^6A modification [71–72]. Of note, we found that inhibition of EZH2 led to a globally enhanced expression pattern of cs-lncRNA. The increased

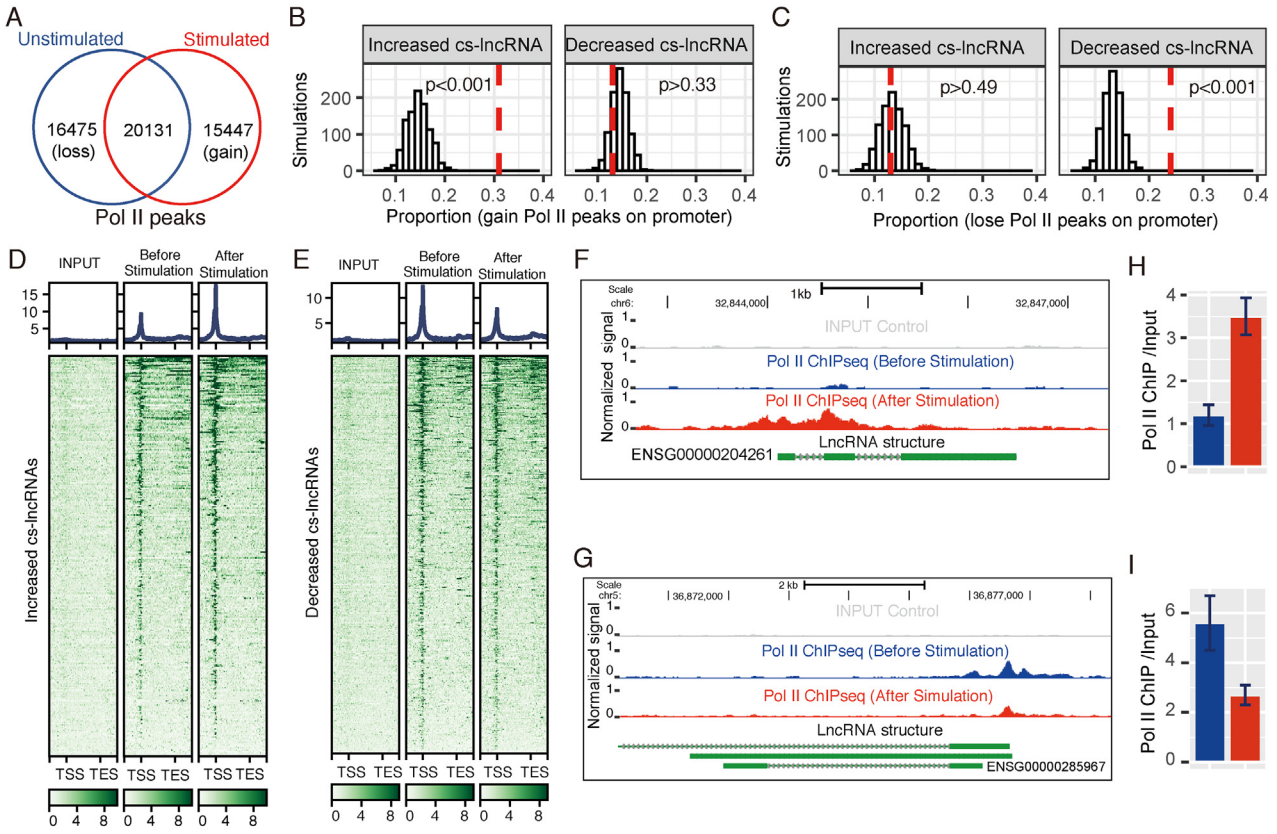

Fig. 4. The expression of cs-lncRNA was modulated transcriptionally. (A) The overlap of identified Pol II peaks before (unstimulated) and after (stimulated) HT-DNA treatment. In total, 15,447 Pol II peaks were specifically gained (gain) and 16,475 Pol II peaks were lost (loss) after HT-DNA stimulation compared with unstimulated status. **(B)** The observed (labeled as a vertical red line) and expected (represented as the histogram) proportion of increased cs-lncRNA (left) and decreased cs-lncRNA (right) harboring gain of Pol II peaks in their promoters. **(C)** The observed (labeled as a vertical red line) and expected (represented as the histogram) proportion of increased cs-lncRNA (left) and decreased cs-lncRNA (right) harboring loss of Pol II peaks in their promoters. For both **(B)** and **(C)**, the expected proportion was estimated using 1,000 permutations based on all expressed lncRNAs. **(D)** and **(E)** The heatmap of Pol II ChIP-seq signals across the promoter and gene body regions of increased cs-lncRNA **(D)** and decreased cs-lncRNA **(E)** before and after HT-DNA stimulation. **(F)** and **(G)** The Pol II ChIP-seq signals across the promoter and gene body regions of an increased cs-lncRNA example **(F)** and a decreased cs-lncRNA example **(G)**. All ChIP-seq signals were normalized and scaled to 1. **(H)** and **(I)** The ChIP-qPCR result of Pol II binding intensity on the promoter region of the increased cs-lncRNA **(H)** and decreased cs-lncRNA **(I)** corresponding **(F)** and **(G)**. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

cs-lncRNAs and decreased cs-lncRNAs were significantly induced and repressed, respectively (Fig. 6H, KS test, p < 1e-7). Collectively, these results suggested that the m^{6}A system controlled cs-TF-mediated transcription machinery, which in turn modulated cs-lncRNA transcriptionally.

4. Discussion

Being a pivotal innate-immune pathway for sensing DNA molecules, the cGAS-STING signaling has been acknowledged as one of the essential components of the immune system for host defense in mammals. Decades of efforts significantly deepen our understanding of the core signaling components, critical regulators, and their interplays in regulating this pathway [1,18–19]. The function of the cGAS-STING signaling has also been expanding to many noninfectious settings of cellular stress and injury [6–7]. However, except for several prominent lncRNA examples, our knowledge of the lncRNAs involved in the cGAS-STING signaling is still quite limited, even for the repertoire of lncRNA associated with this pathway was lacking. In this study, we performed an integrative analysis to comprehensively build the lncRNA catalog associated with the cGAS-STING pathway activation. We designed the experiments to measure the innate immune responses across several well-acknowledged cGAS-STING signaling agonists, which largely excluded the potential artifacts caused by agonist-specific effects. Moreover, we also implemented an analysis protocol to obtain a consensus set of lncRNAs closely associated with the activation of the cGAS-STING pathway, named cs-lncRNA, which displayed highly consistent expression changes between multiple cGAS-STING signaling stimulations. Notably, the identified cs-lncRNAs were strongly supported by an independent RNAseq dataset that measured transcriptome alterations following VACV-70mer (dsDNA) stimulations at 6 h and 12 h in human NHDF cells. Along with verifying cs-lncRNA expression using the qPCR experiment, all these results demonstrated the reliability of cs-lncRNAs identified.

Determining how these cs-lncRNAs are regulated will be helpful to fully elucidating the roles of lncRNAs in physiological and pathological processes governed by the cGAS-STING signaling. Recently studies reported that the *N*6 methyladenosine (m^{6}A) regulatory machinery regulates innate immune responses [20–21]. Being the most abundant and prevalent post-translational modification throughout the whole transcriptome, the m^{6}A system also affects the metabolism of lncRNAs [73]. However, whether and how the m^{6}A system regulates cs-lncRNAs during the cGAS-STING pathway activation remain unclear. The validity and reliability of cs-lncRNA allowed us to explore the potential roles of the m^{6}A system in modulating cs-lncRNA by taking advantage of recently published high-quality omics datasets designed to explore the influence of

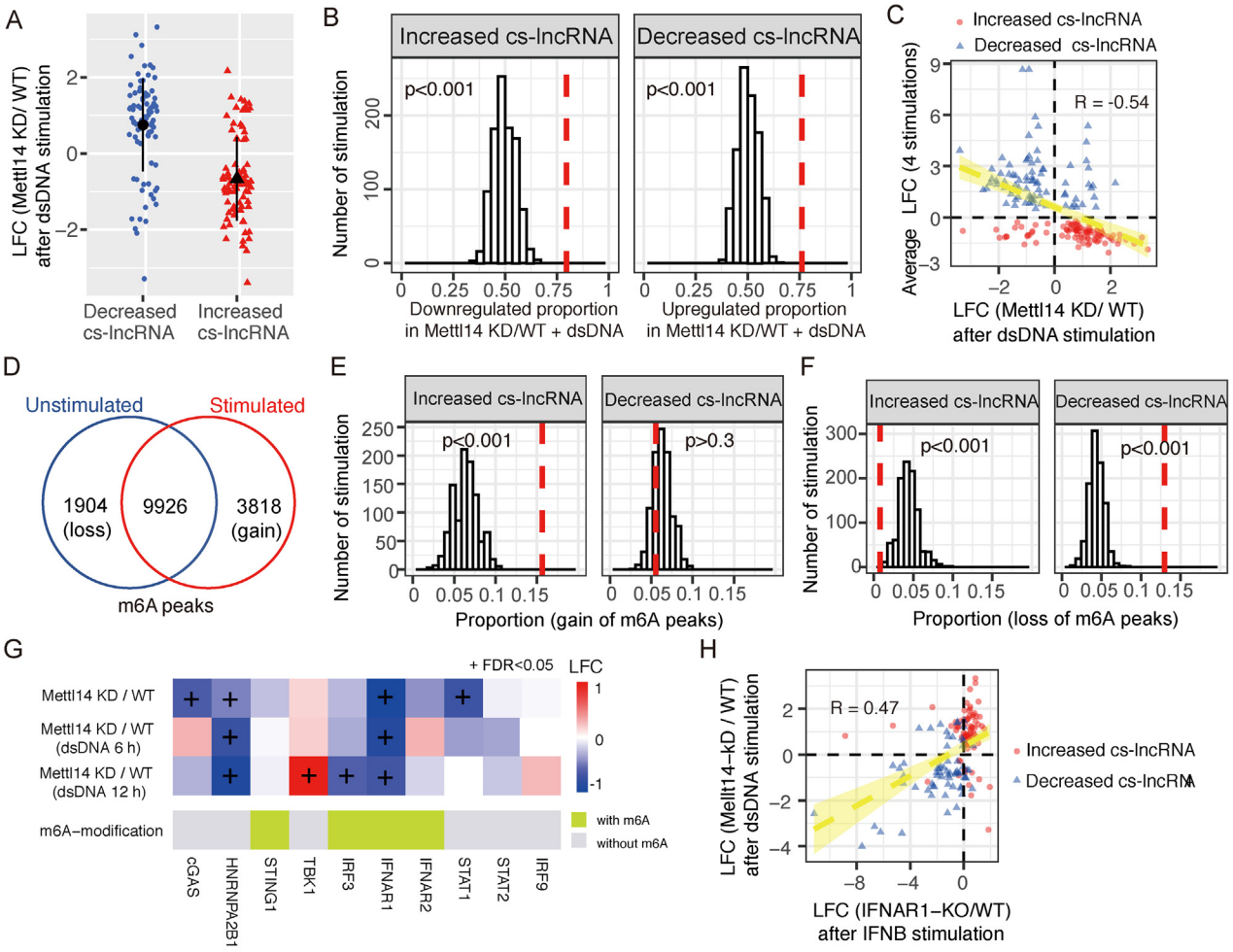

Fig. 5. N^6-methyladenosine (m^6A) regulatory machinery was indispensable for establishing cs-lncRNA repertoire. **(A)** The expression changes of decreased and increased cs-lncRNAs in METTL14 knockout NHDF cells compared with control after VACV-70mer (dsDNA) treatment. **(B)** The observed (labeled as a vertical red line) and expected (represented as the histogram) downregulated proportion of increased cs-lncRNA (left) and upregulated proportion of decreased cs-lncRNA (right) in METTL14 knockout NHDF cells compared with that of control after dsDNA treatment. The expected proportion was estimated using 1,000 permutations based on all expressed lncRNAs. **(C)** The correlation of expression changes of cs-lncRNAs between four cGAS-STING signaling stimulations in HFF-1 cells and dsDNA stimulation in METTL14 knockout NHDF cells. **(D)** The overlap of identified m^6A peaks before (unstimulated) and after (stimulated) dsDNA treatment. In total, 3,818 m^6A peaks were specifically gained (gain) and 1,904 Pol II peaks were lost (loss) after dsDNA stimulation compared with unstimulated status. **(E)** The observed (labeled as a vertical red line) and expected (represented as the histogram) proportion of increased cs-lncRNA (left) and decreased cs-lncRNA (right) harboring gain of m^6A peaks in their transcripts. **(F)** The observed (labeled as a vertical red line) and expected (represented as the histogram) proportion of increased cs-lncRNA (left) and decreased cs-lncRNA (right) harboring loss of m^6A peaks in their transcripts. For both **(E)** and **(F)**, the expected proportion was estimated using 1,000 permutations based on all expressed lncRNAs. **(G)** The heatmap of the expression changes of key signaling transduction components of the cGAS-STING pathway in METTL14 knockout NHDF cells compared with control with (6 h and 12 h) and without dsDNA treatment. **(H)** The correlation of expression changes of cs-lncRNAs between IFNβ treatment in IFNAR1-KO HFF-1 cells and dsDNA stimulation in METTL14 knockout NHDF cells. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the m^6A system on dsDNA elicited innate immunity. Our results demonstrated that an intact m^6A system was indispensable for cs-lncRNA expression. Mechanistically, we provided clues that the m^6A system might partially participate in controlling the expression of cs-lncRNA through manipulating m^6A modification on their transcripts. Intriguingly, we found increased cs-lncRNAs harbored significantly more gain of m^6A peaks and significantly less loss of m^6A peaks while decreased cs-lncRNAs harbored significantly more loss of m^6A peaks. Since the exact function of m^6A modification is governed by specific m^6A "reader" including m^6A-binding proteins YTHDF1/2/3, YTHDC1/2, IGF2BP1/2/3, and hnRNPA2B1 [74–75], it is appealing to speculate that the induction and repression of cs-lncRNA might be regulated by these m^6A "reader" proteins that recognized specific m^6A modification site and led to the stabilization or destabilization of corresponding cs-lncRNAs. However, the exact roles of m^6A modification required further exploration and validation. In addition to controlling m^6A modification on cs-lncRNA, we found m^6A system might regulate cs-lncRNA by promoting several signaling transduction key components of the cGAS-STING pathway, including two DNA sensors cGAS and hnRNPA2B1, an interferon receptor IFNAR1, and a signal transducer STAT1. Impaired expression of these key components will greatly dampen the whole cGAS-STING signaling, resulting in substantial disruption of the cs-lncRNA expression. Intriguingly, hnRNPA2B1 is also an m^6A reader that can facilitate m^6A modification of cGAS, IFI16, and STING mRNAs to promote DNA-induced innate immunity [22]. Notably, we found that Loss of IFNAR1 led to the dysregulation of cs-lncRNAs resembled that of loss of m^6A "writer" METTL14, which is in line with the observation of a recent study that degradation of another m^6A "writer" WTAP blocks Type-I interferon response by destabilizing IFNAR1 mRNA on which m^6A modification is reduced [76].

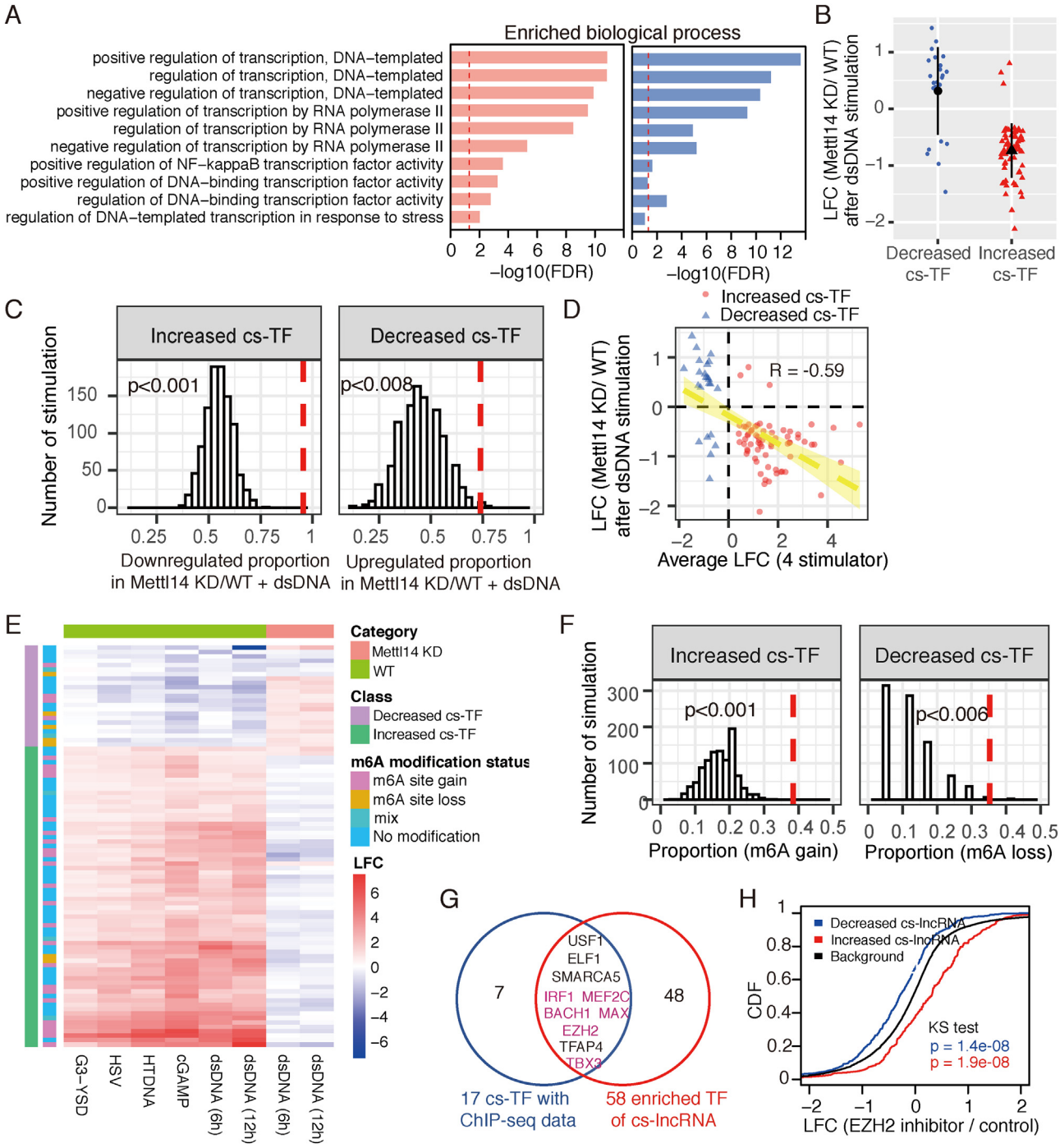

Fig. 6. m^{6}A-mediated regulation affected transcriptional machinery to modulate cs-lncRNA expression. **(A)** The enriched transcriptional regulation related biological processes that were significantly associated with the activation of the cGAS-STING pathway and disrupted by METTL14 knockout. **(B)** The expression changes of decreased and increased cs-TFs in METTL14 knockout NHDF cells compared with that of control after VACV-70mer (dsDNA) treatment. **(C)** The observed (labeled as a vertical red line) and expected (represented as the histogram) downregulated proportion of increased cs-TFs (left) and upregulated proportion of decreased cs-TFs (right) in METTL14 knockout NHDF cells compared with that of control after dsDNA treatment. The expected proportion was estimated using 1,000 permutations based on all expressed TFs. **(D)** The correlation of expression changes of cs-TFs between four cGAS-STING signaling stimulations in HFF-1 cells and dsDNA stimulation in METTL14 knockout NHDF cells. **(E)** The integrated heatmap depicted the m^{6}A status of increased and decreased cs-TFs, and reversed expression pattern of cs-TFs in METTL14 knockout NHDF cells after dsDNA treatment. **(F)** The observed (labeled as a vertical red line) and expected (represented as the histogram) proportion of increased cs-TFs (left) harboring gain of m^{6}A peaks and proportion of decreased cs-TFs (right) harboring loss of m^{6}A peaks. The expected proportion was estimated using 1,000 permutations based on all expressed TFs. **(G)** The overlap of 17 cs-TFs with ENCODE ChIPseq binding site information and 58 TFs with enriched binding sites on the promoters of cs-lncRNAs. **(H)** The cumulative distribution of gene expression changes (LFC) for increased cs-lncRNAs (red), decreased cs-lncRNAs (blue), and all expressed lncRNAs (black) after inhibiting a cs-TF, EZH2. The y-axis shows the cumulative distribution function (CDF) of LFC distribution. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In addition to the regulation on the post-transcriptional level, we also found evidence that Pol II-mediated transcriptional regulation significantly controlled the expression of cs-lncRNAs. Interestingly, our detailed analysis suggested a role of m^6A system mediated transcriptional machinery in modulating cs-lncRNA. At the global level, by leveraging large-scale ChIP-seq data of>300 transcription factors, we demonstrated that the m^6A system regulated cs-TFs displayed enriched binding on the promoter region of cs-lncRNA. By investigating the effect of inhibition of a specific m^6A modification regulated cs-TF, EZH2, we confirmed the role of m^6A modification regulated cs-TF in regulating cs-lncRNAs. Our results strongly indicated that the m^6A system played a role in controlling transcription machinery through regulating cs-TF, which in turn modulated cs-lncRNA transcriptionally.

Determining the function of individual cs-lncRNA remains a challenge. Nevertheless, chromosome distribution examination of cs-lncRNAs coupling with functional enrichment analysis using nearby protein-coding genes suggested a strong association of cs-lncRNA to many typical and critical biological processes of innate immunity, and infection-related phenotypes. Further investigation was required to verify the role of individual cs-lncRNA in the associated biological processes. In this study, we investigated lncRNAs of the cGAS-STING pathway after stimulating human cells with four different agonists/stimuli for 9 h. Further studies based on profiling the transcriptome after the cGAS-STING signaling activation in a time series manner would be helpful to obtain more comprehensive information about lncRNAs expression, especially for the potential dynamic expression pattern of cs-lncRNAs.

5. Conclusion

Overall, our study uncovered a reliable set of cs-lncRNA and revealed m^6A-mediated modulation coupled with transcriptional regulation strongly contributed to establishing cs-lncRNA repertoire. The cs-lncRNA repertoire will provide the foundation to investigate the roles of lncRNA in the cGAS-STING signaling. Moreover, the catalogs of cs-PCG and cs-TF also provided rich resources to fully elucidate the mechanisms underlying diverse physiological and pathological processes governed by the cGAS-STING signaling.

Author contributions

H.H. and C.W. conceptualized this study; L.Z., C.L., M.M., J.H., H. H., and X.Z. performed experiments; J.Y.S., J.S., and L.L. performed data analysis, data visualization, and data curation; H.H. L.Z. and C.W. acquired funding for this study. H.H., L.Z. and C.W. 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.

Acknowledgments

We thank Dr. Wentao Qiao (Nankai University) for sharing HSV-1 (F strain).

Fundings

Dr. Haiyang Hu was supported by the National Natural Science Foundation of China (31601068), the "Double First-Class" Project of China Pharmaceutical University (CPU2018GF10), Key R&D project of Jiangsu Province (BE2020725), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD). Dr. Chen Wang was supported by the National Key R&D Program of China (2016YFA0501800), the National Natural Science Foundation of China (31730018, 81672029), the Open Project of State Key Laboratory of Natural Medicines (SKLNMZZCX201802), the "Double First-Class" Project of China Pharmaceutical University (CPU2018GF10), the Jiangsu Innovative and Entrepreneurial Talents Program. Dr. Lele Zhang was supported by Shanghai Sailing Program (21YF1438300).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2022.04.002.

References

[1] Hu MM, Shu HB. Innate immune response to cytoplasmic DNA: mechanisms and diseases. Annu Rev Immunol 2020;38:79–98.
[2] Ishikawa H, Barber GN. STING is an endoplasmic reticulum adaptor that facilitates innate immune signalling. Nature 2008;455:674–U674.
[3] Zhong B, Yang Y, Li S, Wang YY, Li Y, Diao F, et al. The adaptor protein MITA links virus-sensing receptors to IRF3 transcription factor activation. Immunity 2008;29:538–50.
[4] Kato K, Omura H, Ishitani R, Nureki O. Cyclic GMP-AMP as an endogenous second messenger in innate immune signaling by cytosolic DNA. Annu Rev Biochem 2017;86:541–66.
[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;35:831–44.
[6] Motwani M, Pesiridis S, Fitzgerald KA. DNA sensing by the cGAS-STING pathway in health and disease. Nat Rev Genet 2019;20:657–74.
[7] Decout A, Katz JD, Venkatraman S, Ablasser A. The cGAS-STING pathway as a therapeutic target in inflammatory diseases. Nat Rev Immunol 2021;21:548–69.
[8] Ahn J, Xia T, Konno H, Konno K, Ruiz P, Barber GN. Inflammation-driven carcinogenesis is mediated through STING. Nat Commun 2014;5:5166.
[9] Chen Q, Boire A, Jin X, Valiente M, Er EE, Lopez-Soto A, Jacob LS, Patwa R, Shah H, Xu K, Cross JR, Massague J. Carcinoma-astrocyte gap junctions promote brain metastasis by cGAMP transfer (vol 533, pg 493, 2016). Nature 2017;544. 124-124.
[10] Bakhoum SF, Ngo B, Laughney AM, Cavallo JA, Murphy CJ, Ly P, et al. Chromosomal instability drives metastasis through a cytosolic DNA response. Nature 2018;553:467-+.
[11] Sharma M, Rajendrarao S, Shahani N, Ramirez-Jarquin UN, Subramaniam S. Cyclic GMP-AMP synthase promotes the inflammatory and autophagy responses in Huntington disease. PNAS 2020;117:15989–99.
[12] Kerur N, Fukuda S, Banerjee D, Kim Y, Fu D, Apicella I, et al. cGAS drives noncanonical-inflammasome activation in age-related macular degeneration. Nat Med 2018;24:50–61.
[13] Yu CH, Davidson S, Harapas CR, Hilton JB, Mlodzianoski MJ, Laohamonthonkul P, et al. TDP-43 Triggers mitochondrial DNA release via mPTP to Activate cGAS/STING in ALS. Cell 2020;183(636–649):e618.
[14] McCauley ME, O'Rourke JG, Yanez A, Markman JL, Ho R, Wang X, et al. C9orf72 in myeloid cells suppresses STING-induced inflammation. Nature 2020;585:96–101.
[15] Sliter DA, Martinez J, Hao L, Chen X, Sun N, Fischer TD, et al. Parkin and PINK1 mitigate STING-induced inflammation. Nature 2018;561:258-+.
[16] 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;129:546–55.
[17] Iracheta-Vellve A, Petrasek J, Gyongyosi B, Satishchandran A, Lowe P, Kodys K, et al. Endoplasmic Reticulum Stress-induced Hepatocellular Death Pathways Mediate Liver Injury and Fibrosis via Stimulator of Interferon Genes. J Biol Chem 2016;291:26794–805.
[18] Hopfner KP, Hornung V. Molecular mechanisms and cellular functions of cGAS-STING signalling. Nat Rev Mol Cell Biol 2020;21:501–21.
[19] Galluzzi L, Vanpouille-Box C, Bakhoum SF, Demaria S. SnapShot: CGAS-STING Signaling. Cell 2018;173(276–276):e271.
[20] Rubio RM, Depledge DP, Bianco C, Thompson L, Mohr I. RNA m(6) A modification enzymes shape innate responses to DNA by regulating interferon beta. Genes Dev 2018;32:1472–84.
[21] Winkler R, Gillis E, Lasman L, Safra M, Geula S, Soyris C, et al. m(6)A modification controls the innate immune response to infection by targeting type I interferons. Nat Immunol 2019;20:173–82.
[22] Wang L, Wen M, Cao X. Nuclear hnRNPA2B1 initiates and amplifies the innate immune response to DNA viruses. Science 2019;365.
[23] Zhou Y, Li M, Xue Y, Li Z, Wen W, Liu X, et al. Interferon-inducible cytoplasmic lncLrrc55-AS promotes antiviral innate responses by strengthening IRF3 phosphorylation. Cell Res 2019;29:641–54.

[24] Xu H, Jiang Y, Xu X, Su X, Liu Y, Ma Y, et al. Inducible degradation of lncRNA Sros1 promotes IFN-gamma-mediated activation of innate immune responses by stabilizing Stat1 mRNA. Nat Immunol 2019;20:1621–30.
[25] Palazzo AF, Koonin EV. Functional Long Non-coding RNAs Evolve from Junk Transcripts. Cell 2020;183:1151–61.
[26] Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res 2013;41:e74.
[27] Hu HY, He L, Khaitovich P. Deep sequencing reveals a novel class of bidirectional promoters associated with neuronal genes. BMC Genomics 2014;15:457.
[28] Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, et al. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. Nucleic Acids Res 2013;41:e166.
[29] Li J, Zhang X, Liu C. The computational approaches of lncRNA identification based on coding potential: Status quo and challenges. Comput Struct Biotechnol J 2020;18:3666–77.
[30] Morchikh M, Cribier A, Raffel R, Amraoui S, Cau J, Severac D, et al. HEXIM1 and NEAT1 Long Non-coding RNA Form a Multi-subunit Complex that Regulates DNA-Mediated Innate Immune Response. Mol Cell 2017;67(387–399):e385.
[31] Chen JH, Feng DD, Chen YF, Yang CX, Juan CX, Cao Q, et al. Long non-coding RNA MALAT1 targeting STING transcription promotes bronchopulmonary dysplasia through regulation of CREB. J Cell Mol Med 2020;24:10478–92.
[32] Liu W, Wang Z, Liu L, Yang Z, Liu S, Ma Z, et al. LncRNA Malat1 inhibition of TDP43 cleavage suppresses IRF3-initiated antiviral innate immunity. Proc Natl Acad Sci U S A 2020;117:23695–706.
[33] Zhang L, Wei N, Cui Y, Hong Z, Liu X, Wang Q, et al. The deubiquitinase CYLD is a specific checkpoint of the STING antiviral signaling pathway. PLoS Pathog 2018;14:e1007435.
[34] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30:2114–20.
[35] Ewels P, Magnusson M, Lundin S, Kaller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 2016;32:3047–8.
[36] Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 2019;37:907–15.
[37] Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, et al. Ensembl 2020. Nucleic Acids Res 2020;48:D682–8.
[38] Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. Bioinformatics 2012;28:2184–5.
[39] Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 2014;30:923–30.
[40] Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;26:139–40.
[41] Wu X, Dao Thi VL, Huang Y, Billerbeck E, Saha D, Hoffmann HH, et al. Intrinsic Immunity Shapes Viral Resistance of Stem Cells. Cell 2018;172(423–438):e425.
[42] Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol 2019;20:278.
[43] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing, S. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078–9.
[44] Kang YJ, Yang DC, Kong L, Hou M, Meng YQ, Wei L, et al. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. Nucleic Acids Res 2017;45:W12–6.
[45] Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018;34:i884–90.
[46] Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol 2008;9:R137.
[47] Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. A census of human transcription factors: function, expression and evolution. Nat Rev Genet 2009;10:252–63.
[48] Wingender E, Dietze P, Karas H, Knuppel R. TRANSFAC: a database on transcription factors and their DNA binding sites. Nucleic Acids Res 1996;24:238–41.
[49] Fornes O, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA, et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic Acids Res 2020;48:D87–92.
[50] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25:25–9.
[51] Wilson D, Charoensawan V, Kummerfeld SK, Teichmann SA. DBD–taxonomically broad transcription factor predictions: new content and functionality. Nucleic Acids Res 2008;36:D88–92.
[52] Hu H, Miao YR, Jia LH, Yu QY, Zhang Q, Guo AY. AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. Nucleic Acids Res 2019;47:D33–8.
[53] Fulton DL, Sundararajan S, Badis G, Hughes TR, Wasserman WW, Roach JC, et al. TFCat: the curated catalog of mouse and human transcription factors. Genome Biol 2009;10:R29.
[54] Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, et al. The Encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Res 2018;46:D794–801.
[55] Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4:44–57.
[56] McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol 2010;28:495–501.
[57] Sun X, Liu T, Zhao J, Xia H, Xie J, Guo Y, et al. DNA-PK deficiency potentiates cGAS-mediated antiviral innate immunity. Nat Commun 2020;11:6182.
[58] Orzalli MH, Broekema NM, Diner BA, Hancks DC, Elde NC, Cristea IM, et al. cGAS-mediated stabilization of IFI16 promotes innate signaling during herpes simplex virus infection. PNAS 2015;112:E1773–1781.
[59] Hu MM, Yang Q, Xie XQ, Liao CY, Lin H, Liu TT, et al. Sumoylation Promotes the Stability of the DNA Sensor cGAS and the Adaptor STING to Regulate the Kinetics of Response to DNA Virus. Immunity 2016;45:555–69.
[60] Zhang J, Zhao J, Xu S, Li J, He S, Zeng Y, et al. Species-Specific Deamidation of cGAS by Herpes Simplex Virus UL37 Protein Facilitates Viral Replication. Cell Host Microbe 2018;24(234–248):e235.
[61] Liu ZS, Cai H, Xue W, Wang M, Xia T, Li WJ, et al. G3BP1 promotes DNA binding and activation of cGAS. Nat Immunol 2019;20:18–28.
[62] Zhao Q, Wei Y, Pandol SJ, Li L, Habtezion A. STING Signaling Promotes Inflammation in Experimental Acute Pancreatitis. Gastroenterology 2018;154(1822–1835):e1822.
[63] Dai J, Huang YJ, He XH, Zhao M, Wang XZ, Liu ZS, et al. Acetylation Blocks cGAS Activity and Inhibits Self-DNA-Induced Autoimmunity. Cell 2019;176:1447-+.
[64] Ma Z, Damania B. The cGAS-STING defense pathway and its counteraction by viruses. Cell Host Microbe 2016;19:150–8.
[65] Li XD, Wu JX, Gao DX, Wang H, Sun LJ, Chen ZJJ. Pivotal Roles of cGAS-cGAMP Signaling in Antiviral Defense and Immune Adjuvant Effects. Science 2013;341:1390–4.
[66] Herzner AM, Hagmann CA, Goldeck M, Wolter S, Kubler 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;16:1025-+.
[67] Wu JX, Sun LJ, Chen X, Du FH, Shi HP, Chen C, et al. Cyclic GMP-AMP is an endogenous second messenger in innate immune signaling by cytosolic DNA. Science 2013;339:826–30.
[68] Balka KR, De Nardo D. Molecular and spatial mechanisms governing STING signalling. FEBS J 2021;288:5504–29.
[69] Zhang S, Wang B, Wan L, Li LM. Estimating Phred scores of Illumina base calls by logistic regression and sparse modeling. BMC Bioinf 2017;18:335.
[70] Wu J, Chen YJ, Dobbs N, Sakai T, Liou J, Miner JJ, et al. STING-mediated disruption of calcium homeostasis chronically activates ER stress and primes T cell death. J Exp Med 2019;216:867–83.
[71] Li F, Chen S, Yu J, Gao Z, Sun Z, Yi Y, et al. Interplay of m(6) A and histone modifications contributes to temozolomide resistance in glioblastoma. Clin Transl Med 2021;11:e553.
[72] Meng QZ, Cong CH, Li XJ, Zhu F, Zhao X, Chen FW. METTL3 promotes the progression of nasopharyngeal carcinoma through mediating M6A modification of EZH2. Eur Rev Med Pharmacol Sci 2020;24:4328–36.
[73] Liu L, Song B, Ma J, Song Y, Zhang SY, Tang Y, et al. Bioinformatics approaches for deciphering the epitranscriptome: Recent progress and emerging topics. Comput Struct Biotechnol J 2020;18:1587–604.
[74] Yang Y, Hsu PJ, Chen YS, Yang YG. Dynamic transcriptomic m(6)A decoration: writers, erasers, readers and functions in RNA metabolism. Cell Res 2018;28:616–24.
[75] Jiang X, Liu B, Nie Z, Duan L, Xiong Q, Jin Z, et al. The role of m6A modification in the biological functions and diseases. Signal Transduct Target Ther 2021;6:74.
[76] Ge Y, Ling T, Wang Y, Jia X, Xie X, Chen R, et al. Degradation of WTAP blocks antiviral responses by reducing the m(6) A levels of IRF3 and IFNAR1 mRNA. EMBO Rep 2021;e52101.