Systematic Determination of Human Cyclin Dependent Kinase (CDK)-9 Interactome Identifies Novel Functions in RNA Splicing Mediated by the DEAD Box (DDX)-5/17 RNA Helicases*[S]

Jun Yang‡§¶, Yingxin Zhao‡§¶, Mridul Kalita§¶, Xueling Li¶‖, Mohammad Jamaluddin‡¶, Bing Tian‡§¶, Chukwudi B. Edeh‡, John E. Wiktorowicz§¶‖, Andrzej Kudlicki§¶‖, and Allan R. Brasier‡§¶**

Inducible transcriptional elongation is a rapid, stereotypic mechanism for activating immediate early immune defense genes by the epithelium in response to viral pathogens. Here, the recruitment of a multifunctional complex containing the cyclin dependent kinase 9 (CDK9) triggers the process of transcriptional elongation activating resting RNA polymerase engaged with innate immune response (IIR) genes. To identify additional functional activity of the CDK9 complex, we conducted immunoprecipitation (IP) enrichment-stable isotope labeling LC-MS/MS of the CDK9 complex in unstimulated cells and from cells activated by a synthetic dsRNA, polyinosinic/polycytidylic acid [poly (I:C)]. 245 CDK9 interacting proteins were identified with high confidence in the basal state and 20 proteins in four functional classes were validated by IP-SRM-MS. These data identified that CDK9 interacts with DDX 5/17, a family of ATP-dependent RNA helicases, important in alternative RNA splicing of NFAT5, and mH2A1 mRNA two proteins controlling redox signaling. A direct comparison of the basal *versus* activated state was performed using stable isotope labeling and validated by IP-SRM-MS. Recruited into the CDK9 interactome in response to poly(I:C) stimulation are HSPB1, DNA dependent kinases, and cytoskeletal myosin proteins that exchange with 60S ribosomal structural proteins. An integrated human CDK9 interactome map was developed containing all known human CDK9- interacting proteins. These data were used to develop a probabilistic global map of CDK9-dependent target genes that predicted two functional states controlling distinct cellular functions, one important in immune and stress responses. The CDK9-DDX5/17 complex was shown to be functionally important by shRNA-mediated knockdown, where differential accumulation of alternatively spliced NFAT5 and mH2A1 transcripts and alterations in downstream redox signaling were seen. The requirement of CDK9 for DDX5 recruitment to NFAT5 and mH2A1 chromatin target was further demonstrated using chromatin immunoprecipitation (ChIP). These data indicate that CDK9 is a dynamic multifunctional enzyme complex mediating not only transcriptional elongation, but also alternative RNA splicing and potentially translational control. ***Molecular & Cellular Proteomics 14: 10.1074/mcp.M115.049221, 2701–2721, 2015.***

The mammalian respiratory tract is a large, contiguous epithelial surface that is continuously exposed to environmental agents, antigens and respiratory pathogens (1, 2). As a result, the lining epithelium plays a critical role in the early detection of pathogens and initiation of the protective innate immune response (IIR)[1]. Pathogen detection is mediated by germline-encoded pattern recognition receptors that bind to cognate

From the ‡Department of Internal Medicine; §Sealy Center for Molecular Medicine; ¶Institute for Translational Sciences; ‖Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, Galveston, Texas

Received February 21, 2015, and in revised form, June 23, 2015
Published, July 21, 2015, MCP Papers in Press, DOI 10.1074/mcp.M115.049221
Author contributions: J.Y., Y.Z., M.K., X.L., J.E.W., A.K., and A.R.B. designed research; J.Y., Y.Z., M.K., X.L., M.J., B.T., C.B.E., A.K., and A.R.B. performed research; Y.Z., M.K., X.L., M.J., B.T., C.B.E., J.E.W., A.K., and A.R.B. contributed new reagents or analytic tools; J.Y., Y.Z., M.K., X.L., M.J., B.T., C.B.E., J.E.W., A.K., and A.R.B. analyzed data; J.Y., Y.Z., M.K., X.L., M.J., B.T., J.E.W., A.K., and A.R.B. wrote the paper.

[1] The abbreviations used are: IIR, innate immune response; BRD4, bromodomain containing 4; CAN508, 4-(3,5-Diamino-1H-pyrazol-4-ylazo)-phenol; CDK9, cyclin dependent kinase 9; DDX5/17, DEAD-box RNA helicase 5/17; FDR, false discovery rate; GO, gene ontology; HEXIM, hexamethylene bis-acetamide inducible; HOA1, hydroxyacid oxidase 1; IP, immunoprecipitation;LC-MS/MS, Liquid chromatography - tandem mass spectrometry; NF-κB, nuclear factor kappa-light-chain-enhancer of activated B cells; Poly (I:C), polyinosinic/polycytidylic acid; P-TEFb, the positive transcription elongation factor complex; PPI, protein–protein interactions; Q-gPCR, quantitative genomic real time PCR; Q-RT-PCR, quantitative reverse transcription real time PCR; RFESD, Rieske (Fe-S) domain containing; ROS, reactive oxygen species; RP, ribosomal protein; SAEC, small airway epithelial cells; shRNA, small or short hairpin RNA; SOD3, extracellular superoxide dismutase 3; SRM, selected reaction monitoring; STAT1, signal transducer and activator of transcription 1; XChIP, two-step cross-linking chromatin immunoprecipitation.

molecular patterns, such as dsRNA, 5′phosphorylated RNA, and lipopolysaccharide, to trigger intracellular signaling cascades converging on nuclear factor (NF)κB and interferon regulatory factor (IRF)-3 (3, 4). NFκB- and IRF3 are key transcription factors whose inducible nuclear translocation trigger the expression of inducible "immediate-early" genes important in cellular anti-viral response and inflammation to limit pathogen spread and activate adaptive immunity (4, 5). Recent work has shown that these key immediate-early genes are induced by a biochemical process of transcriptional elongation (6–8).

Transcriptional elongation enables rapid genomic reprogramming of the epithelial cell in response to cellular stress. Work has shown that immediate early genes induced by the IIR are maintained in an "open" chromatin configuration whose promoters are engaged by a hypo-phosphorylated "paused" RNA polymerase. Hypo-phosphorylated RNA Pol II cycles nonproductively on the 5′ upstream of IIR genes, producing short noncoding (~30–50 nt) RNA transcripts. Upon innate pathway activation, activated NFκB and IRF3 bind to cognate cis regulatory regions in the proximal promoters of inducible genes to recruit CDK9, the major effector of the positive transcription elongation factor (P-TEFb) complex (9). CDK9 catalyzes phosphorylation at Ser 2 of the heptad repeats of the carboxyl terminus of the large subunit of RNA Pol II, RPB1, as well as the negative elongation factor (NELF). Phospho-Ser 2 RNA Pol II then enters a processive mode, producing full-length, fully spliced mRNAs (7, 8, 10). Functional evidence for the role of CDK9 has been generated in experiments by either its chemical inhibition or siRNA-mediated silencing. Here, CDK9 inhibition blocks inducible gene expression in both the NFκB and IRF3 signaling arms, demonstrating its essential, nonredundant role in mediating the innate response (8, 10, 11). In this way, CDK9 plays a central role in triggering productive mRNA elongation, enabling a rapid anti-viral response, limiting pathogen spread and activation of protective immunity.

CDK9 is found in a heterogeneous protein complex that exists in several states- one an inactive state bound with the abundant small nuclear RNA, 7SK snRNA, the hexamethylene bisacetamide-inducible proteins (HEXIM1/2) and the methyl phosphate capping enzyme (BCDIN3) (12, 13), and the other, an active state associated with bromodomain containing 4 (BRD4). BRD4 is a chromatin reader protein that recognizes and binds acetylated histones H4 and H3 associated with open chromatin (14). Our previous work suggests that CDK9 is recruited to NFκB-dependent targets through direct association with the NFκB/RelA transcriptional activator subunit (11, 15), but binds to IRF3-dependent genes independent of direct IRF3 association (8). In this way, CDK9's association with distinct classes of proteins modifies its activity and chromatin targeting.

We sought to identify functional activities associated with activated CDK9 in biologically relevant sentinel epithelial cells using a functional proteomics approach. Taking advantage of our previous discovery that activation of the IIR triggers a shift of CDK9 from the inactive-7SK snRNA complex into the activated state with BRD4 (8), we conducted antibody enrichment assays of basal and activation-enriched CDK9 complexes isolated from poly(I:C)-stimulated cells. High confidence interactions were identified in duplicate immunoprecipitates using stable isotope labeling by trypsin-catalyzed $H_2{}^{16}O/H_2{}^{18}O$ exchange. Key interactions were validated using immunoprecipitation (IP) enrichment-selected reaction monitoring (SRM) assays, suggesting the basal CDK9 complex participates in RNA splicing and ribosomal function/translation. Stimulus-dependent changes in the CDK9 interactome was directly demonstrated independently after CDK9 enrichment from control *versus* poly(I:C) stimulated cells and $^{16}O/^{18}O$ exchange. Functional experiments were developed to explore the role of the ATP-dependent RNA helicases, DDX5/17, in alternative mRNA splicing. shRNA depletion demonstrated that CDK9 and DDX5/17 were important in mH2A1/H2AFY and NFAT5 pre-mRNA splicing. These effects were functionally significant because of the effect of CDK9 and DDX5/17 silencing inhibited expression of downstream redox target genes. Finally, we observed that CDK9 mediates DDX5 recruitment to native NFAT5 and mH2A1 promoter. These data indicate that CDK9 interaction with DDX5/17 couples transcriptional elongation with RNA splicing functions in a single molecular machine. Our studies extend the functional activities of CDK9 complex to implicate its multifaceted role in RNA expression and post-transcriptional processing.

EXPERIMENTAL PROCEDURES

Cell Culture and Treatment—Human A549 type II pulmonary epithelial cells (American Type Culture Collection, Manassas, VA) were grown in F-12K medium (Invitrogen, Carlsbad, CA) with 10% fetal bovine serum, penicillin (100 U/ml), and streptomycin (100 μg/ml) at 37 °C in a 5% CO_2 incubator (27). Poly(I:C) was obtained from Sigma (St. Louis, MO), and the CDK9 inhibitor II (CAN508) was purchased from Calbiochem-EMD Millipore (Billerica, MA).

For poly(I:C) electroporation, A549 cells were trypsinized, washed in phosphate-buffered saline (PBS), and pelleted by centrifugation. The cell pellet was then resuspended in 100 μl Nucleofector solution T (Amaxa; Lonza, Basel, Switzerland) with poly(I:C) into an electroporation cuvette. The cell suspension was electroporated using a Nucleofector II device (Amaxa; Lonza). The Nucleofector Program X-001 was used.

Subcellular Fractionation and Western Blot Analyses—Nuclear and cytoplasmic proteins were fractionated as previously described (16, 17). For Western blotting, equal amounts of nuclear protein were fractionated by SDS-PAGE and transferred to polyvinylidene difluoride (PVDF) membranes. The membranes were incubated with indicated affinity-purified rabbit polyclonal antibodies (Abs). Abs were against DDX5, 17 CDK9 (Santa Cruz Biotechnology, Santa Cruz, CA) and phospho-Thr186 CDK9 (Cell Signaling, Danvers, MA). Washed membranes were then incubated with IRDye 800-labeled anti-rabbit IgG Abs (Rockland Immunochemicals, Gilbertsville, PA), and immune complexes were quantified using the Odyssey infrared imaging system (Li-Cor Biosciences, Lincoln, NE).

Quantitative Real-time Reverse Transcription-PCR (Q-RT-PCR)—Total RNA was extracted using acid guanidium phenol extraction (Tri reagent; Sigma). For gene expression analyses, 1 μg of RNA was reverse transcribed using SuperScript III in a 10-μl reaction mixture (27). 0.5 μl of cDNA product was amplified in a 10-μl reaction mixture containing 5 μl of SYBR green Supermix (Bio-Rad, Hercules, CA) and 0.4 μM (each) forward and reverse gene-specific primers (Supplemental Table S1). The reaction mixtures were aliquoted into a Bio-Rad 96-well clear PCR plate, and the plate was sealed with Bio-Rad Microseal B film. The plates were denatured for 90 s at 95 °C and then subjected to 40 cycles of 15 s at 94 °C, 60 s at 60 °C, and 1 min at 72 °C in a CFX96™ Real-Time PCR Detection System (Bio-Rad). PCR products were subjected to melting curve analysis to ensure that a single amplification product was produced. Quantification of relative changes in gene expression was done using the threshold cycle (ΔΔCT) method. In brief, the ΔCT value was calculated (normalized to glyceraldehyde-3-phosphate dehydrogenase [GAPDH]) for each sample by using the equation ΔCT = CT (target gene) - CT (GAPDH). Next, the ΔΔCT was calculated by using the equation ΔΔCT = ΔCT (experimental sample) - ΔCT (control sample). Finally, the fold differences between the experimental and control samples were calculated using the formula $2^{-\Delta\Delta CT}$.

Quantification of CDK9-Associated 7SK snRNA—A549 cells were electroporated with 10 μg poly(I:C) for 0 and 3h, and whole cell extracts were prepared in RIPA buffer containing 10 U/ml SUPERase In™ RNase inhibitor (Ambion, Life Technologies, San Jose, CA) and complete protease inhibitor mixture (Sigma-Aldrich) by sonication for 10 s with a pulse setting of 4 in a Branson Sonifier 150 (Branson Ultrasonics Corporation, Dansbury, CT). Sonicated cellular extract was clarified by centrifugation (12,000 rpm, 10 min at 4 °C) and immunoprecipitated (IPed) for 4 h at 4 °C with 4 μg of CDK9 antibody (Santa Cruz) in RIPA buffer with complete protease inhibitor mixture and SUPERase. IPs were collected with 40 μl protein A magnetic beads (Dynal Inc.) for 1 h at 4 °C, captured on a magnetic stand, and washed 4 times with PBS. Afterward, the magnetic beads were dissolved in 1 ml Tri reagent (Sigma-Aldrich) for total RNA extraction and processed according to the manufacturer's recommendation. The extracted RNA was measured by Q-RT-PCR using 7SK snRNA qPCR primers as described previously (8). The amount of 7SK snRNA in each sample was quantified relative to human 5s rRNA (Q-RT-PCR primers are shown in supplemental Table S1).

Identification of the CDK9 Interactome—To broadly identify members of the CDK9 interactome, control or poly(I:C)-stimulated A549 cells (4×10^6 to 6×10^6 per 100-mm dish) were washed twice with PBS and nuclei isolated as above (16, 17). Initial discovery experiments were conducted after crosslinking the nuclear complexes to stabilize transient protein-protein interactions. Nuclei were suspended in 1 ml of sucrose wash buffer (0.25 M sucrose, 10 mM HEPES(pH 7.5), 1 mM $MgCl_2$, 100 mM KCl, 1 mM PMSF) and treated with a final concentration of 2 mM disuccinimidyl glutarate (DSG) 45 min at 22 °C. The cross-linked nuclei were suspended in radioimmunoprecipitation assay buffer (RIPA; 150 mM NaCl, 1 mM Na_2EDTA,1% IGEPAL CA630, 1% sodium deoxycholate, 20 mM Tris-HCl (pH 7.5)) with complete protease inhibitor mixture (Sigma-Aldrich) and sonicated 4 times and centrifuged at 12,000 rpm for 10 min. The supernatants were collected and were quantified for protein concentrations. Equal amounts of nuclear lysates were IPed overnight at 4 °C with 4 μg of IgG or CDK9 Ab in ChIP dilution buffer (11). IPs were collected with 40 μl protein A magnetic beads (Dynal Inc.). The beads were washed with PBS for three times and then resuspended in 30 μl of 50 mM ammonium hydrogen carbonate (pH 7.8). The samples were then subjected to on-beads tryptic digestion, with control samples being subsequently labeled with ^{18}O using trypsin-mediated isotopic exchange (Fig. 1). The experiment was conducted in duplicate, and enrichments validated by $^{16}O/^{18}O$ label "swap" experiments, described below.

A second experiment was conducted to directly compare the effect of poly(I:C) on the CDK9 interactome. In this experiment, nuclei from control- or poly(I:C)-stimulated A549 cells were prepared, DSG cross-linked, and sonicated as above. Equal amounts of control or poly(I:C) stimulated nuclei were IPed with 4 μg of anti-CDK9 Ab in ChIP dilution buffer (11), captured on magnetic beads, washed and subjected to on-beads tryptic digestion followed by trypsin-mediated isotopic exchange (Fig. 4). The experiment was conducted in duplicate, and enrichments validated by $^{16}O/^{18}O$ label "swap" experiments, described below.

On-beads Trypsin Digestion—The proteins on the beads were reduced with dithiothreitol (10 mM, 30 min at room temperature), alkylated with iodoacetamide (30 mM, 2 h at 37 °C) and digested with trypsin (2 μg 24 h at 37 °C) as described (18). After digestion, the supernatant was collected. The beads were washed with 50 μl of 50% acetonitrile (ACN) three times and the supernatant was pooled, and dried. The tryptic digests were then reconstituted in 30 μl of 5% formic acid-0.01% TFA. An aliquot of 10 μl of diluted stable isotope labeled signature peptides were added to each tryptic digest. These samples were desalted with ZipTip C18. The peptides were eluted with 80% ACN and dried with a SpeedVac system.

Trypsin-catalyzed $^{16}O/^{18}O$ Labeling—Peptide C-terminal $^{16}O/^{18}O$ labeling was performed as described previously (19). For ^{18}O labeling, the tryptic peptides from on-beads digestion of IgG pull down were redissolved in 20 μl of acetonitrile and diluted with 80 μl of ^{18}O-enriched water (97%, Sigma-Aldrich). Sequencing grade modified trypsin (Promega, Madison, WI) dissolved in ^{18}O-water was added to the samples at a ratio of 50:1 (w/w, protein-to-trypsin), and the mixture was incubated at 37 °C for overnight. The reactions were quenched by boiling the samples for 10 min in a water bath and then cooling down to room temperature. Before the heavy and light labeled peptides were mixed, a small fraction of the ^{18}O-labeled sample was tested for ^{18}O-labeling efficiency with LC-MS/MS. 2,764 peptides were identified with 1% FDR. Among them, 2,752 (99.6%) were peptides with double ^{18}O incorporated. We further inspected the spectra manually. As shown in supplemental Fig. S5, the ^{18}O labeling efficiency is higher than 95% based on the abundance of 2 ^{18}O-labeled peptides and their ^{16}O-labeled counterparts. The ^{16}O-H_2O and ^{18}O- H_2O labeled samples in each experiment were then mixed together and desalted with SepPak C18 cartridge (Waters, Milford, CT). The desalted peptides were stored at −80 °C for LC-MS/MS analysis.

In the label swap experiments, the identical procedure was conducted, only control samples were labeled with $H_2{}^{16}O$ and the peptides from the CDK9 pull-down were labeled with ^{18}O- H_2O. All the IP experiments were done in duplicate.

LC-MS/MS Analysis and Data Processing—Dried peptide samples were redissolved in 2 μl of acetonitrile and diluted with 40 μl of 0.1% formic acid. LC-MS/MS analysis was performed with a Q Exactive Orbitrap mass spectrometer (Thermo Scientific, San Jose, CA) equipped with a nanospray source with an on-line Easy-nLC 1000 nano-HPLC system (Thermo Scientific). Ten microliters of each peptide solution were injected and separated on a reversed phase nano-HPLC C18 column (75 μm × 150 cm) with a linear gradient of 0–35% mobile phase B (0.1% formic acid-90% acetonitrile) in mobile phase A (0.1% formic acid) over 120 min at 300 nL/min. The mass spectrometer was operated in the data-dependent acquisition mode with a resolution of 70,000 at full scan mode and 17,500 at MS/MS mode. The ten most intense ions in each MS survey scan were automatically selected for MS/MS. The acquired MS/MS spectra were analyzed by

MaxQuant 1.4 (20) using default parameters (supplemental Table S12) with a SWISSPROT protein databases (downloaded on February 2013, 20,247 protein entries) a mass tolerance of ± 20 ppm for precursor and product ions; a static mass modification on cysteinyl residues that corresponded to alkylation with iodoacetamide; differential modifications were defined to be ^{18}O-labeled C-terminal and oxidized methionine; maximum two missed cleavage. Single peptide spectra are in supplemental Table S13. Peptide identifications are in supplemental Tables S14 and S15. Protein identification data (Accession numbers, peptides observed, sequence coverage) are in supplemental Tables S16 and S17. The FDR cutoff for peptide and protein identification is 0.01. The relative protein abundance changes were quantified by MaxQuant using default parameters. A protein ratio is calculated as the median of all H/L ratios of the peptides mapped to a protein. The normalized proteins ratios were calculated using the median of the ratio. CDK9 interacting proteins were identified by low values of heavy (H)/light (L) O ratios observed in duplicate experiments that showed an absolute value of log 2 H/L ratio of >1.5, corresponding to an absolute difference of greater ± 2.83 fold.

Stable Isotope Dilution (SID)-Selected Reaction Monitoring (SRM)-MS—The SID-SRM-MS assays of CDK9 interactors were developed as described previously (21). For each targeted proteins, two or three peptides were initially selected and then the sensitivity and selectivity of these were experimentally evaluated as described previously (18). The peptide with best sensitivity and selectivity was selected as the surrogate for that protein. For each peptide, 3–5 SRM transitions were monitored. The signature peptides and SRM parameters are listed in supplemental Table S5. The peptides were chemically synthesized incorporating isotopically labeled [$^{13}C_6{}^{15}N_4$] arginine or [$^{13}C_6{}^{15}N_2$] lysine to a 99% isotopic enrichment (Thermo Scientific). The amount of SIS peptides were determined by amino acid analysis. The proteins immunoprecipitated with anti-CDK9 antibody were captured by protein A magnetic beads (Dynal Inc.). The proteins were trypsin digested on the beads as described above. The tryptic digests were then reconstituted in 30 μl of 5% formic acid-0.01% trifluoroacetic acid (TFA). An aliquot of 10 μl of 50 fmol/μl diluted stable isotope-labeled standard (SIS) peptides was added to each tryptic digest. These samples were desalted with a ZipTip C18 cartridge. The peptides were eluted with 80% ACN and dried. The peptides were reconstituted in 30 μl of 5% formic acid-0.01% TFA and were directly analyzed by liquid chromatography (LC)-SRM-MS. LC-SRM-MS analysis was performed with a TSQ Vantage triple quadrupole mass spectrometer equipped with nanospray source (Thermo Scientific, San Jose, CA). 8–10 targeted proteins were analyzed in a single LC-SRM run. The online chromatography were performed using an Eksigent NanoLC-2D HPLC system (AB SCIEX, Dublin, CA). An aliquot of 10 μl of each of the tryptic digests was injected on a C18 reverse-phase nano-HPLC column (PicoFrit™, 75 μm x 10 cm; tip ID 15 μm) at a flow rate of 500 nL/min with a 20-min 98% A, followed by a 15-min linear gradient from 2–30% mobile phase B (0.1% formic acid-90% acetonitrile) in mobile phase A (0.1% formic acid). The TSQ Vantage was operated in high-resolution SRM mode with Q1 and Q3 set to 0.2 and 0.7-Da Full Width Half Maximum (FWHM). All acquisition methods used the following parameters: 2100 V ion spray voltage, a 275 °C ion transferring tube temperature, a collision-activated dissociation pressure at 1.5 mTorr, and the S-lens voltage used the values in S-lens table generated during MS calibration.

All SRM data were manually inspected to ensure peak detection and accurate integration. The chromatographic retention time and the relative product ion intensities of the analyte peptides were compared with those of the stable isotope labeled standard (SIS) peptides. The variation of the retention time between the analyte peptides and their SIS counterparts should be within 0.05 min, and the difference in the relative product ion intensities of the analyte peptides and SIS peptides were below 20%. The peak areas in the extract ion chromatography of the native and SIS version of each signature peptide were integrated using Xcalibur® 2.1. The default values for noise percentage and base-line subtraction window were used. The ratio between the peak area of native and SIS version of each peptide were calculated.

Integrated database of CDK9 interacting proteins—IP-LC-MS/MS experiments led us to identify two data sets of CDK9 interacting proteins: basal (inactive CDK9 state, control cells) and active (active CDK9 state, poly(I:C) stimulated cells). In addition to these, three other major sources were used to gather published CDK9 interactions to create a comprehensive database of human CDK9 protein - protein interactions (PPIs): PubMed (ftp://ftp.ncbi.nih.gov/gene/GeneRIF/interactions.gz), Protein Interaction Network Analysis platform (PINA) (22), and a recent systematic study on protein interaction network of the basal transcriptional machinery (13). PINA is a nonredundant database based on integration of data from six public and curated PPI databases: IntAct, MINT, BioGRID, DIP, HPRD, and MIPS MPact. Integrating basal, activated, and published data sets helped us to create a comprehensive "integrated" data set of human CDK9 PPIs. All interaction entries in our database were manually annotated and reviewed by these public resources. Details are given in Supplemental Table S2.

Functional Analysis—We used the following approaches to identify functional annotation of CDK9 interacting proteins from the above three data sets: basal, activated and integrated data sets. These were: (1). molecular functions based Gene Ontology (GO) classification (23). The GO analysis was performed using BiNGO (24), a Cytoscape plugin (25) to determine the significantly overrepresented GO (Functional) terms in these data sets at Benjamini and Hochberg's false discovery rate (FDR) cutoff at 0.05 using human genome as a reference data set; (2). Protein class enrichment using Panther (www.pantherdb.org). (3). Interactive network analysis using the Search Tool for the Retrieval of Interacting Genes (STRING) database (26); this PPI network analysis was performed using Homo sapiens as the reference organism. Benjamini and Hochberg's FDR correction cutoff at 0.05 was implemented in the analysis; and, (4). complexes and pathway analysis - we used MCODE (27) for detecting clusters with networks in which proteins in a cluster can represent to interact with each other to form complexes or pathways. We investigated the possible protein complexes and pathways to be formed by these interacting proteins along with CDK9 protein also included. The parameters chosen were Human modules (MCODE level 3) with a FDR cutoff of 0.05 and a reference set as "Proteome."

shRNA Knockdown of DDX5/DDX17—TRIPZ inducible lentiviral shRNAs of human CDK9 were purchased from Thermo Scientific (V2THS_112919 and V2THS_112920). Both CDK9 shRNAs target 3′-UTR of the human CDK9 mRNA (NM_001261) with the mature antisense sequences of AGGATTGTGGGTGGGTGAG and TCTAACGGACCAAACTGTG, respectively. TRIPZ inducible lentiviral nonsilencing shRNA (RHS4743) was used here as a negative control. To construct a lentiviral plasmid encoding a specific shRNA targeting both human DDX5 and DDX17 mRNA, we first used an online siRNA prediction tool DSIR (Designer of Small Interfering RNAs, http://biodev.extra.cea.fr/DSIR/DSIR.html) to screen the potential shRNA target sequences from each transcript (28). Sequences identified by DSIR were then cross-checked against the human transcript database to exclude sequences that showed high similarity or exact seed matches to off-target genes. A total of three putative seed sequences (21-mer) were identified as the common targets of both DDX5 and DDX17 (#1, GGCTAGATGTGGAAGATGT; #2, CCATGG(T/A)GACAAGAGTCAA; #3, GCTTGATATGGG(G/C)TTTGAA). Among them, the

single sequence #1 targeting a conserved region in human DDX5 and DDX17 mRNAs without any mismatch was assembled into the shRNA template. The corresponding complementary linkers (linker A and Linker B, 110-mer each) were generated, annealed, and cloned into the pTRIPZ vector at the XhoI and EcoRI sites (supplemental Table S3).

To produce lentivirus carrying the specific shRNA, HEK293FT cells were cotransfected with shRNA plasmid together with the lentiviral packaging plasmids (Invitrogen). Forty-eight hours later the virus-containing supernatant was harvested and then target A549 cells were infected. Stably expressed A549 cells were selected with 8 μg/ml puromycin 48 h after infection. The stable transfectants from a mixed population were used in the experiments.

Two-step Chromatin IP (XChIP)—A549 cells at a density of 4–6 $\times$ 10^6 per 100-mm dish were washed twice with phosphate-buffered saline. Protein-protein cross-linking was first performed with DSG (2 mM, 45 min at 22 °C) followed by protein-DNA cross-linking with formaldehyde as previously described (29). Equal amounts of sheared chromatin were iIPed overnight at 4 °C with 4 μg indicated Ab in ChIP dilution buffer. IPs were collected with 40 μl protein-A magnetic beads (Dynal Inc), washed and eluted in 250 μl elution buffer for 15 min at room temperature. Samples were treated with RNase A before proteinase K digestion to reduce background.

Quantitative Genomic PCR (Q-gPCR)—Gene enrichment in XChIP was determined by Q-gPCR as previously described (30, 31) using region-specific PCR primers (supplemental Table S4). The fold change of DNA in each IP was determined by normalizing the absolute amount to input DNA reference and calculating the fold change relative to that amount in unstimulated cells.

Genome-wide prediction of the CDK9 modulatory network—Possible biological effects of CDK9 transcriptional complexes were characterized by a probabilistic method recently developed by us for inferring genome-wide modulator-transcription factor (CDK9)-target gene sets, or "triplets" (32, 33). In this analysis, all the potential significant triplets were calculated based on a compendium of transcriptional expression profiles of 534 CDK9 interactors identified, CDK9 itself, and predicted CDK9 regulated genes. We used a compendium of 2158 microarray expression profiles from expO (expression for Oncology) for these calculations (32). Five hundred and thirty-four CDK9 binding proteins mapped into 516 unique genes were used as modulators and 13,162 genes in the Affymetrix HG-U133 plus 2 array were used as target gene candidates in the prediction.

The 2158 conditions in the expO data set were rank-ordered according to the transcript concentrations (at the probeset level) and discretized into three equal bins to identify the high and low expression conditions for each triplet (32). The three-gene interactions pointing to coregulation, or modulation of CDK9 activity were derived from the analysis of the estimated conditional probabilities of high (or low) expression of the affected target (TG) given a concentration of CDK9 (F) and the gene coding for the interacting protein (M), see equations 1–5 below and (32, 33).

$$\hat{\alpha}_f = \mathrm{P}(\text{high TG}|\text{low M, high F}) - \mathrm{P}(\text{high TG}|\text{low M, low F}) \quad \text{(Eq. 1)}$$

$$\hat{\beta}_f = \mathrm{P}(\text{high TG}|\text{high M, high F}) - \mathrm{P}(\text{high TG}|\text{high M, low F}) \quad \text{(Eq. 2)}$$

$$\hat{\alpha}_m = \mathrm{P}(\text{high TG}|\text{high M, low F}) - \mathrm{P}(\text{high TG}|\text{low M, low F}) \quad \text{(Eq. 3)}$$

$$\hat{\beta}_m = \mathrm{P}(\text{high TG}|\text{high M, high F}) - \mathrm{P}(\text{high TG}|\text{low M, high F}) \quad \text{(Eq. 4)}$$

$$\hat{\gamma} = \hat{\beta}_m - \hat{\alpha}_m = \hat{\beta}_f - \hat{\alpha}_f$$
$$= \mathrm{P}(\text{high TG}|\mathrm{P}\ (\text{high TG}|\text{high M, high F}))$$
$$- \mathrm{P}(\text{high TG}|\text{low M, high F}) - \mathrm{P}(\text{high TG}|\text{high M, low F})$$
$$+ \mathrm{P}(\text{high TG}|\text{low M, low F}) \quad \text{(Eq. 5)}$$

At a FDR of 1%, 4298 significant modulatory triplet interactions (supplemental Table S10) were predicted with defined action modes, composing of 391 modulators and 2771 target genes. Note that we set the *p* value thresholds of the parameters to 0.1 for significant action modes definition of the significant triplets. This will include as many significant triplets with defined action modes as possible.

We integrated all the significant triplets into a whole modulatory network and characterized the potential functional involvement of the CDK9 binding proteins. We validated the set of predicted target genes by enrichment in the CDK9 regulated genes, where the CDK9 regulated genes were obtained by two sample *t* test based on the control and CDK9 inhibitor only treated GSE48258 data set from GEO. For clustering, modulators and target genes involved in fewer than three triplets were discarded, which results in 257 modulator genes and 344 target genes. We validated the set of predicted target genes by enrichment in the CDK9 regulated genes, where the CDK9 regulated genes were obtained by two sample *t* test based on the control and CDK9 inhibitor only treated GSE48258 data set from GEO.

The modulatory network was then represented by a matrix, where each element is the γ (in Eq. 5) of the triplet. Module composition of the modulatory network was then characterized by hierarchical bi-clustering of the modulators and target genes using Matlab built-in clustergram function, with default parameters. We clustered the matrix of βm (in Eqn.4) as well (supplemental Fig. S4), which yields similar results if CT3 and CT4 are merged to a unique cluster of target genes, which agrees with functional similarity of the two clusters.

RESULTS

IP-LC-MS/MS Identification of Human CDK9 Interacting Proteins—In an attempt to comprehensively profile the functional activities of CDK9 in a model airway epithelial cell, we applied differential quantitative proteomics in control and activated human type II alveolar cells (A549) using IP enrichment coupled with LC-MS/MS (schematically diagrammed in Fig. 1*A*). In resting cells, CDK9 association with 7SK snRNA sequesters the kinase into an inactive complex, resulting in a mixed population of activated and inactive CDK9 complexes in a ~1:1 ratio (34). Previously we made the observation that poly(I:C) mediated activation of the IIR reduces the abundance of 7SK snRNA and increased BRD4 association with CDK9, shifting the equilibrium of CDK9 to that of the activated complex (8). In Fig. 1*B*, we IPed CDK9-associated 7SK snRNA from epithelial cells using IgG or anti-CDK9 Abs. 7SK snRNA was extracted from the immune complex and its expression was measured by Q-RT-PCR. CDK9-associated 7SK snRNA was detected in nonstimulated cells, whereas its abundance was largely reduced in response to poly(I:C) treatment revealed by Q-RT-PCR (Fig. 1*B*). These data confirm

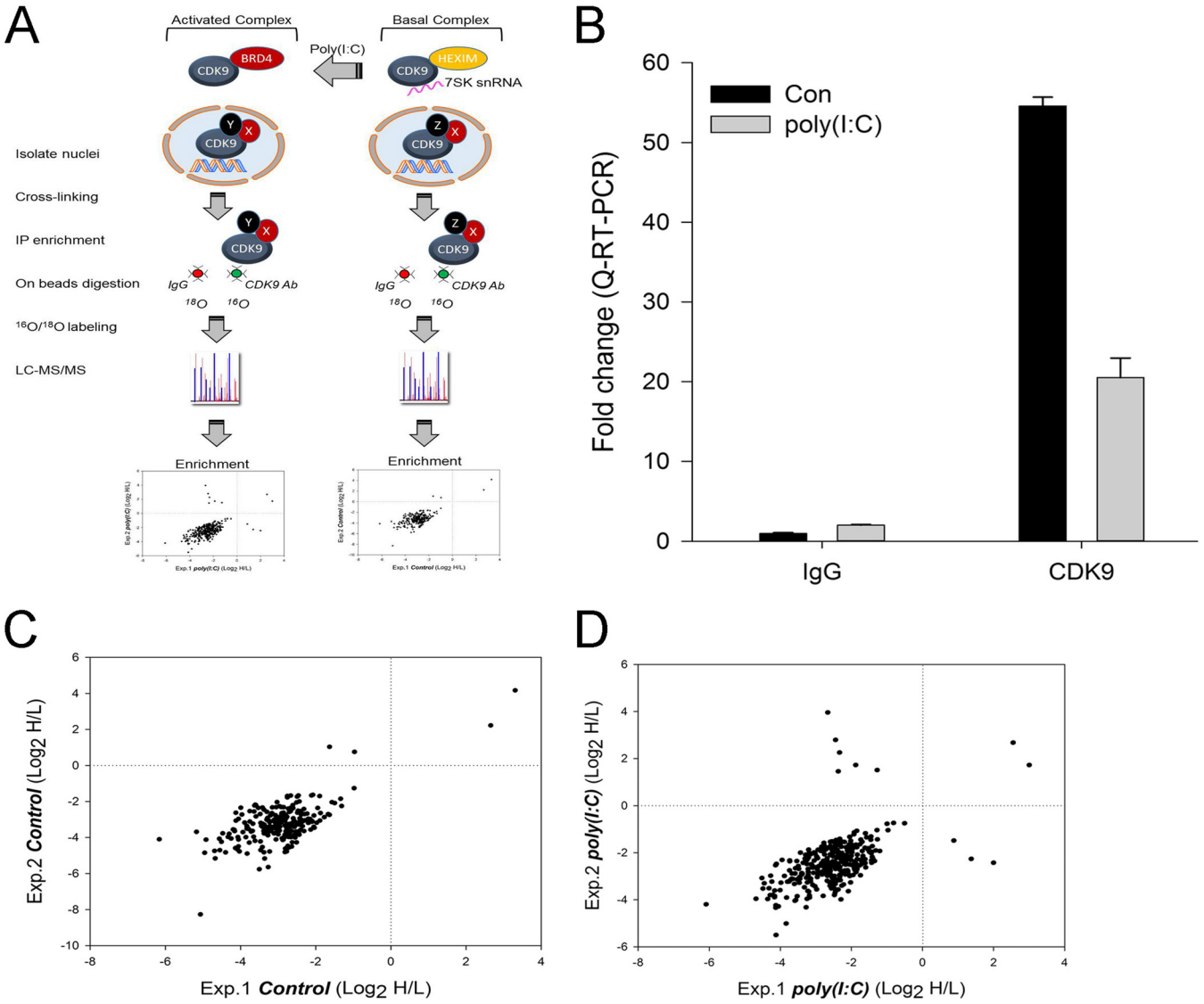

FIG. 1. **Experimental workflow and data acquisition.** *A*, Overall work flow of the proteomics strategy for the identification of functional CDK9 protein complexes from control- or poly(I:C)- stimulated A549 nuclear extracts (in duplicates). *B*, Quantification of CDK9-associated 7SK small nuclear RNA in A549 cells. A549 cells were electroporated with 10 μg poly(I:C) for 0 and 3 h. CDK9 complexes were enriched by IP and associated 7SK snRNA was quantified relative to 5s rRNA signal by Q-RT-PCR. ** $p < 0.01$ (*t* test). *C–D*, Reproducibility of IP-MS/MS measurements for the CDK9 interacting proteins, evaluated by the comparison between the duplicate experiments. The ratios of the heavy and light labeled versions of the control (*C*) or poly(I:C) (*D*) samples are plotted against each other in two independent duplicate measurements, respectively (Pearson correlation coefficients are 0.662 with control and 0.535 with poly(I:C) data sets, *p* values are < 0.001). The proteins on the left-bottom quarter are enriched in CDK9 IP compared with IgG IP.

that poly(I:C) stimulation disrupts CDK9 association with inhibitory 7SK snRNA, thereby inducing the activated CDK9 complex, resulting in a robust system to identify protein–protein interactions of activated CDK9.

To enhance detection of transiently-associated nuclear proteins, we conducted the initial discovery analysis using DSG-mediated cross-linking of nuclear fractions. An IP enrichment using anti-CDK9 Ab was then used to enrich CDK9 binding partners (Fig. 1*A*). Proteins captured by anti-CDK9 IP were labeled with light water (^{16}O), whereas nonspecific binding proteins captured by control IgG were labeled with heavy water (^{18}O). In our hands, the ^{18}O labeling efficiency is higher than 95% (Experimental Procedures). Both pull-down fractions (heavy and light) were mixed and analyzed together. Specific interacting proteins were identified as those enriched by low heavy/light ratios in the duplicate experiments by mass spectrometry analysis.

The CDK9 Interactome in its Basal State—The "basal" CDK9 interactome was determined in two replicate IP-LC-MS/MS experiments (Experimental Procedures). From this analysis, 245 proteins were observed with consistent enrichment (or depletion) in duplicate experiments with an absolute

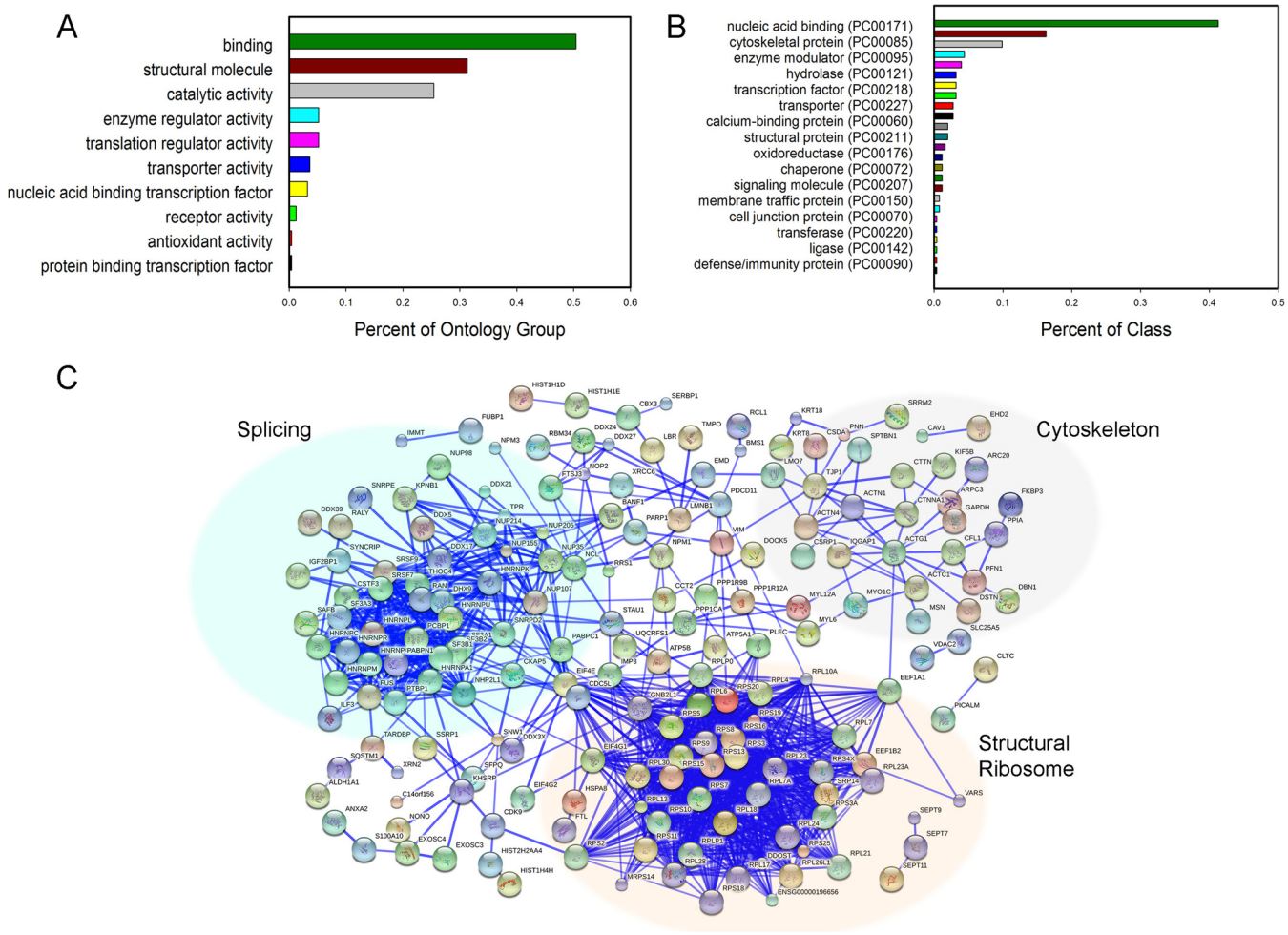

FIG. 2. **The human CDK9 protein-protein interaction network in the basal state.** *A*, Genome Ontology classification of 245 high confidence proteins in the basal CDK9 interactome. *B*, Protein class analysis of the CDK9 basal interactome. For each protein class, the percent of the class within the CDK9 interactome is shown. *C*, STRING network prediction - interaction network of 245 CDK9 interacting proteins in basal state. Each protein is represented by a node, and interactions are shown as blue colored edges. Interactions are shown at a high confidence score cutoff of 0.7. Stronger associations are represented by thicker edges. Core networks of the splicing factors, structural components of the ribosome and cytoskeleton are highlighted. All the significant modules are listed in supplemental Table S7.

value of log 2 H/L ratio of $>$ 1.5, corresponding to an absolute difference of greater $\pm$ 2.83 fold. The experimental concordance of high confidence interactors (after filtering, described in Experimental Procedures) are shown in Figs. 1*C* and 1*D*. The concordance of all proteins in the experiments are shown in supplemental Fig. S6.

The 245 high confidence protein interactions were subjected to genome ontology and protein class analysis. The top 3 GO classifications of the basal interactome included binding activity, structural molecule activity and catalytic activity (Fig. 2*A*). Twenty-seven molecular functions were enriched notably those of RNA synthesis, including pyrophosphatase activity, exonuclease activity, ATPase activity, RNA helicase, and mRNA binding (supplemental Tables S6, S9). A protein class analysis indicated significant enrichment of nucleic acid binding, cytoskeletal protein and enzyme modulators (Fig. 2*B*).

To further infer functional activities of the CDK9 interacting proteins, we used features of the STRING database to cluster the networks and conduct enrichment analysis. The STRING database predicts functional interactions integrating experimental observations, computational prediction methods, and published interactions. This network topology shows two primary dense subnetworks, enriched in RNA-binding (poly(A), rRNA) and structural constituents of ribosomes, with a smaller network enriched in actin binding proteins (Fig. 2*C*). In particular, we noted the RNA-binding enriched module consists of predicted interactions between CDK9, DDX5, DDX17, suggesting the presence of RNA splicing factors within the basal CDK9 complex (highlighted in teal, Fig. 2*C*). The network containing structural component of ribosomes (highlighted in orange, Fig. 2*C*) suggested potential roles of the CDK9 complex in translation. The actin binding module consisted of 8 proteins including actin adapters (ACTN1), perhaps linking the

CDK9 complex with the cytoskeleton (a detailed list is shown in (supplemental Fig. S3; a detailed list is shown in supplemental Table S7. A KEGG pathway analysis is in supplemental Table S8).

Validation of Basal CDK9 Interacting Proteins by IP-SRM-MS—To independently confirm interaction of the proteins identified in IP-LC-MS/MS of DSG cross-linked with CDK9, we conducted separate confirmation studies using IP-SRM measurements. To ensure we observed high stringency protein interactions, these studies were conducted in the absence of cross-linking. SRM assays were developed to 20 focus proteins involved in ribosomal assembly, nonmuscle actin complexes, histones and RNA-helicases (Supplemental Table S5). In these experiments, CDK9 complexes from unstimulated A549 cells were enriched by IP using anti-CDK9 antibody and the abundance of candidate interacting proteins in the immune complex was determined by SRM. In parallel, the same amount of extract was IPed using IgG as nonspecific binding controls. We observed a significant, 106-fold enrichment of CDK9 in the immune complexes using the CDK9 Abs *versus* that of IgG, validating IP-SRM assays for candidate interactors (Fig. 3*A*).

IP-SRM assays were performed for the 60S ribosomal structural proteins, ribosomal protein (RP)L-13, -18 and -24, and the alpha subunit of the elongation factor-1 complex (EEF1A2). In these experiments, abundance of the target protein was expressed relative to the stable isotope dilution standard in each IP (referred to as the "Aqua peptide"). Here, we observed that RPL-24, 18 and -13 were significantly enriched in the CDK9 IP relative to that observed in IgG (Fig. 3*B*). We also observed significant enrichment of EEF1A2, a protein responsible for the recruitment of aminoacyl tRNAs to the ribosome for protein elongation (Fig. 3*B*). Together these validation experiments indicate that the CDK9 interacting proteins include those of the structural ribosome and those controlling translational initiation.

To validate the association of CDK9 with the actin cytoskeleton, IP- SRM-MS assays showed enrichment of α-actinin (ACTN)-1, -4, heat shock protein (HSP) B1/HSP27, actin β (ACTB), β-tubulin (TUBB) and vimentin (VIM; Fig. 3*C*). Although various levels of nonspecific binding were observed in the IgG control IPs, the signals were significantly greater in the anti-CDK9 IPs for each. These are cytoskeletal components of actin microfilaments (35, 36), suggesting that the CDK9 transcriptional elongation complex interacts with nuclear nonmuscle actin (see Discussion).

We next confirmed that CDK9 interacts with nuclear histones H3A, H4A, peroxiredoxin 1 (PRDX1), and the catalytic subunit of the DNA-dependent protein kinase (PRKDC), confirming that CDK9 interacts with nucleosomal and chromatin modifying enzymes (Fig. 3*D*).

Finally, we validated our observation that the basal CDK9 interactome includes multiple RNA helicases of the DEAD-box families (DDXs), including DDX -5, -17, -21, -24, and 3X (Fig. 3*E*), suggesting that the basal complex is involved in pre-mRNA processing. Finally, comparisons of the IP-SRM measurement and the $^{16}O/^{18}O$ isotopic enrichment were conducted for the 20 proteins validated (Fig. 3*F*). A high degree of concordance was observed.

Functional Analysis of the Activated CDK9 Interactome—We tentatively identified 162 proteins whose enrichment was greater in the CDK9 complex isolated from poly(I:C)-stimulated cells than that from unstimulated cells, suggesting that poly(I:C) activation changes the composition of the CDK9 complex. To better understand the effects of poly(I:C) on the CDK9 interactome, we conducted a separate quantitative proteomics study directly comparing the activated *versus* basal state. For this purpose, nuclei from control or poly(I:C)-stimulated cells were subjected to CDK9 IP, and the enrichment of proteins determined by trypsin-mediated $^{16}O/^{18}O$ stable isotopic labeling (Fig. 4*A*). Under these conditions, poly(I:C) stimulation has no detectable effect on steady state CDK9 activity, nor the abundance of phospho-Thr 186 CDK9, a T-loop phosphorylation shown to be essential for CDK9 kinase activity (37) (Fig. 4*B*). Seventy-seven proteins were observed in at least two observations with an absolute value of the log2 H/L ratio of $>$ 1.5 (Table I), indicating that poly(I:C) significantly restructures composition of the complex. The enriched proteins were subjected to genome ontology and protein classification analysis. The genome ontology showed increased representation of cellular process, metabolic process and cellular component organization (Fig. 4*C*) and the protein class enrichment showed cytoskeletal protein and nucleic acid binding were highly represented in the enriched protein set (Fig. 4*D*). A STRING network topology map showed that the poly(I:C) induced proteins are composed of a cluster of cytoskeletal myosin and distinct group of ribosomal subunit proteins (Fig. 4*E*).

Activation of the IIR Induces Dynamic Changes in the CDK9 Interactome—To more precisely examine the exchange of ribosomal subunit proteins and RNA helicases, IP-SRM experiments were conducted in using CDK9 Ab IP of unstimulated and poly(I:C)-stimulated cell extracts; enrichment is expressed as fold change relative to that signal produced by the basal complex. Equivalent amounts of CDK9 was detected in both IPs (Fig. 5*A*), further confirming that poly(I:C) treatment does not significantly affect steady state CDK9 abundance (c.f. Fig. 4*B*). Here we observed that poly(I:C) reduced association of RPLs and increased the association with EEF1A2 (Fig. 5*B*). ACTN-1 and -4, TUBB and VIM was reduced, whereas HSPB1 association was increased (Fig. 5C). For the chromatin modifying complexes, poly(I:C) increased H3F3A and PRKDC association but slightly decreased histone 1H4A and PRDX1 (Fig. 5*D*).

Finally, poly(I:C) stimulation significantly reduced the association of all the DDX -5, -17, -21, -24, and 3X isoforms (Fig. 5*E*). These data were confirmed in replicate $^{16}O/^{18}O$ quantification experiments (Fig. 5*F*), and separately vali-

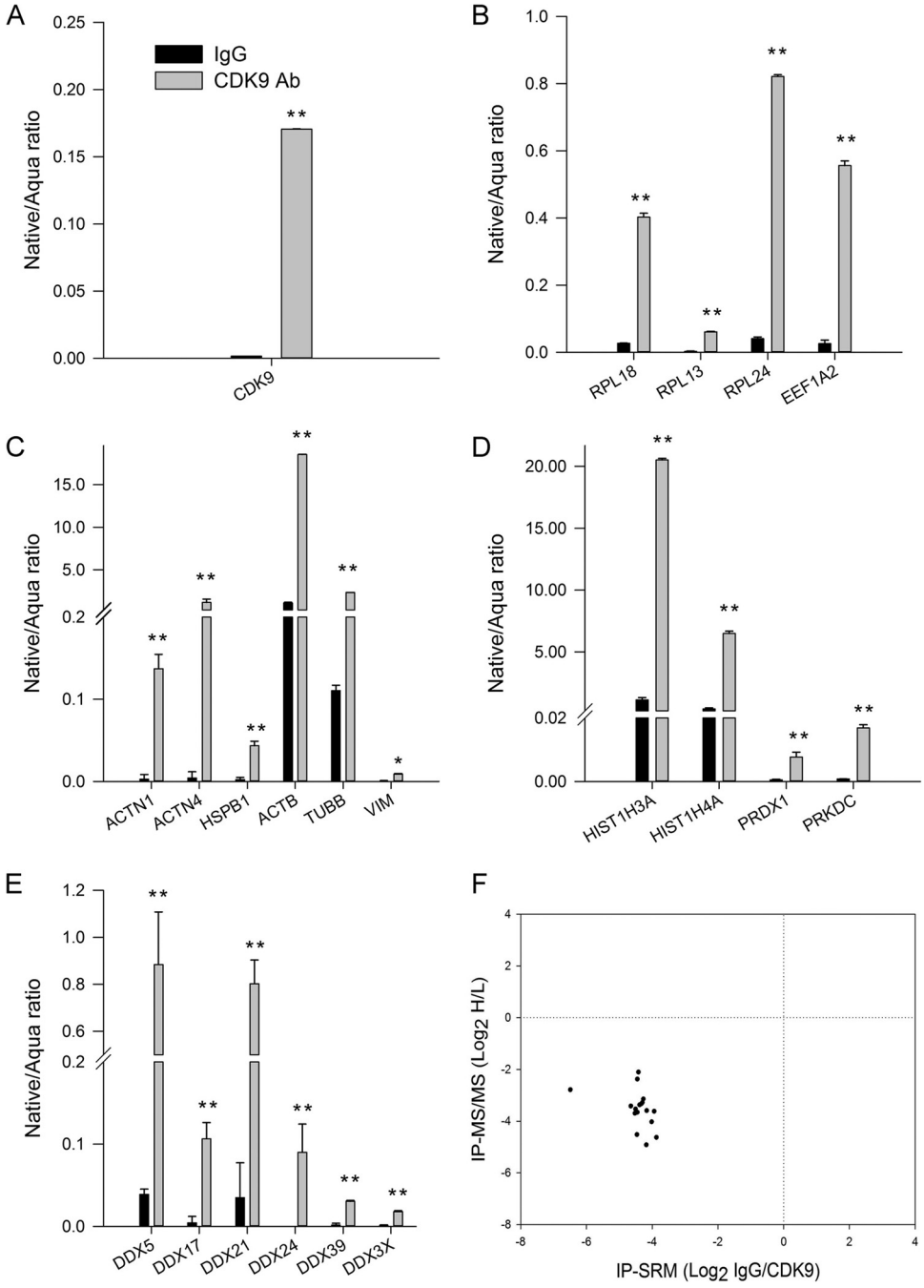

FIG. 3. **Validation of the basal CDK9 interacting proteins by IP-SRM-MS.** *A*, IP-SRM-MS shows CDK9 signal in control or anti-CDK9 IPs. *B–E*, IP-SRM-MS assays for selected target proteins in control (IgG) or anti-CDK9 IPs. Data are expressed as the ratio of measured protein relative to stable isotope standard (native/Aqua). Black, IgG; gray, anti-CDK9 IP. Error bars represent technical replicates. CDK9 interacts with proteins involved in ribosomal function/translational elongation (*B*), cytoskeletal assembly (*C*), chromosome structure (*D*) and DEAD-box RNA helicases (*E*). *F*, concordance between 16O/18O enrichment and IP-SRM. For each target validated by IP-SRM, the relationship of IP-SRM enrichment is plotted *versus* the H/L ratio. * $p < 0.05$; ** $p < 0.01$ (*t* test). Abbreviations: ACTN, actinin; DDX, DEAD box RNA helicase; HIST1H4A, histone cluster 1 H4A; HSP, heat shock protein; H3F3 histone H3 family A; RP, ribosomal protein; PRKDC, DNA-dependent protein kinase catalytic subunit; PRDX, peroxiredoxin; TUBB, tubulin; VIM, vimentin.

dated by IP-Western immunoblot. In this experiment, control or poly(I:C) stimulated nuclei were IPed with IgG or anti-CDK9 Abs, and the abundance of DDX5 and -17 measured by Western immunoblot (Fig. 5*G*). A reverse IP was performed using anti-DDX5 to detect CDK9 (Fig. 5*H*). Together, these results are consistent that the basal association of CDK9 with the DDX RNA helicases is reduced upon poly(I:C) activation.

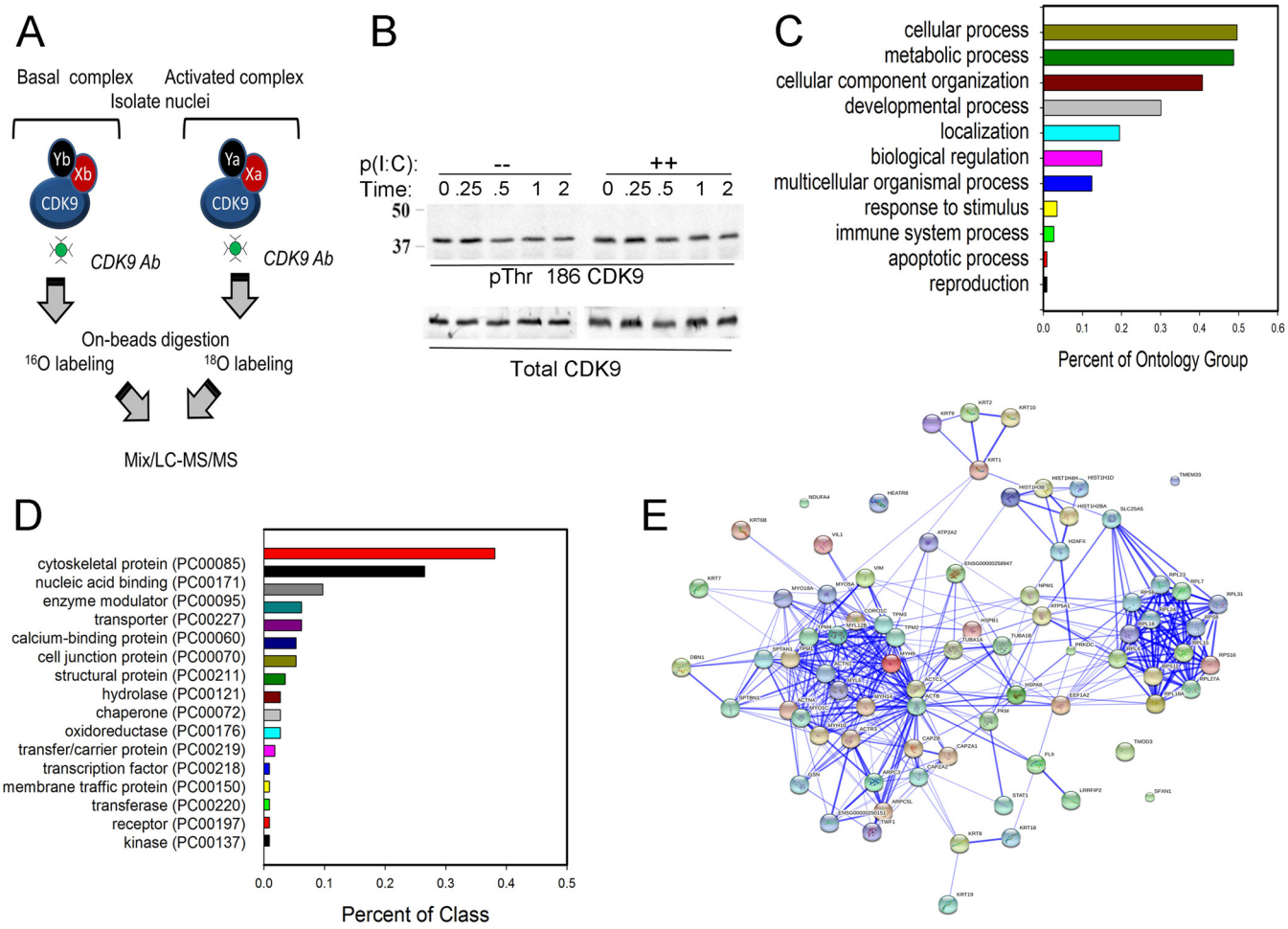

FIG. 4. **Direct comparison of the CDK9 interactome in basal *versus* activated state.** *A*, Schematic diagram of proteomics strategy for the identification of functional CDK9 protein complexes from control- or poly(I:C)- stimulated A549 nuclear extracts (in duplicates). Experiments were replicated in duplicate, with each labeling subjected to label swap. *B*, Effect of poly(I:C) on abundance of CDK9 and phospho-Thr CDK9. A time course of A549 cells treated in the absence (-) or presence of poly(I:C) (+) were lysed in SDS-sample loading buffer and assayed for CDK9 and phospho-Thr CDK9 by Western blot. Molecular weight sizes are shown at left. *C*, Genome ontology of proteins enriched in activated state. *D*, Protein class analysis of the CDK9 basal interactome. *E*, STRING network prediction of the 77 enriched proteins by poly(I:C) stimulation (from Table I).

Integration of Human CDK9 Interactions—A total of 407 proteins were identified through our IgG/CDK9 IP-LC-MS/MS experiments. To compare the CDK9 interactome identified in our IP-LC-MS/MS analysis with those previously reported, we identified 184 nonredundant human CDK9-interacting proteins from peer-reviewed sources (PubMed literature and public databases; see Experimental Procedures). The number of human CDK9 interacting proteins identified from each source is given in supplemental Table S2. The collection of interacting proteins is displayed as a Venn diagram (supplemental Fig. S2). Thirty proteins were found common between our IP-LC-MS/MS data and all other published nonredundant data sets. Therefore, 377 proteins identified in this study by IP-LC-MS/MS represent novel interacting proteins. The union of the basal, active and published data sets reveal an integrated data set of 561 proteins as the human CDK9 interactome (supplemental Table S9).

The integrated human CKD9 interactome was subjected to GO-functional analysis (Fig. 6). We identified 98 clusters, of which 11 clusters are involved in functions at a significance of p value $\leq$ 1E-10. Among these are RNA binding, structural constituent of ribosome, and RNA polymerase II transcription factor activity.

Prediction of the Global CDK9 Modulatory Network—CDK9 functions as a transcriptional regulator for multiple target genes. Proteins physically interacting with transcription factors can alter their mode of action, often on a target-specific manner. To further characterize the potential functional involvement of the CDK9 binding proteins as modulators in gene expression and regulation, we inferred the targets genes of CDK9 with different interacting proteins and their respective modes of regulation using our probabilistic approach (32). For this purpose, the inference considers the effects of the CDK9 interacting proteins as CDK9 "modulators," identifies

TABLE I
Differentially regulated proteins poly(I:C) vs control

Entry	Protein names	H/L (average)	# obs
A6NL28	Tropomyosin alpha-3 chain-like protein	3.11	4
O00159	Unconventional myosin-Ic	4.28	4
O14950	Myosin regulatory light chain 12B	2.98	4
O15145	Actin-related protein 2/3 complex subunit 3	2.97	4
O43707	Alpha-actinin-4	3.79	4
P05141	ADP/ATP translocase 2	2.75	4
P05783	Keratin, type I cytoskeletal 18	2.68	4
P05787	Keratin, type II cytoskeletal 8	2.72	4
P06396	Gelsolin	3.45	4
P06753	Tropomyosin alpha-3 chain	2.88	4
P08670	Vimentin	3.25	4
P08729	Keratin, type II cytoskeletal 7	2.83	4
P09327	Villin-1	1.88	4
P09493	Tropomyosin alpha-1 chain	1.71	4
P13645	Keratin, type I cytoskeletal 10	1.75	4
P16104	Histone H2AX	1.54	4
P35579	Myosin-9	2.12	4
P35580	Myosin-10	1.61	4
P35908	Keratin, type II cytoskeletal 2 epidermal	2.07	4
P46776	60S ribosomal protein L27a	1.93	4
P47755	F-actin-capping protein subunit alpha-2	2.74	4
P47756	F-actin-capping protein subunit beta	2.10	4
P52907	F-actin-capping protein subunit alpha-1	2.51	4
P59998	Actin-related protein 2/3 complex subunit 4	2.84	4
P60660	Myosin light polypeptide 6	2.19	4
P60709	Actin, cytoplasmic 1	1.73	4
P61313	60S ribosomal protein L15	1.82	4
P62829	60S ribosomal protein L23	2.42	4
P67936	Tropomyosin alpha-4 chain	0.17	4
P68363	Tubulin alpha-1B chain	2.89	4
P78527	DNA-dependent protein kinase catalytic subunit	3.49	4
Q02543	60S ribosomal protein L18a	3.34	4
Q07020	60S ribosomal protein L18	1.75	4
Q16643	Drebrin	1.98	4
Q7Z406	Myosin-14	2.27	4
Q92614	Unconventional myosin-XVIIIa	2.28	4
Q9NYL9	Tropomodulin-3	2.36	4
Q9ULV4	Coronin-1C	5.47	4
P04792	Heat shock protein beta-1	5.16	3
P14618	Pyruvate kinase PKM	2.71	3
P16615	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	2.29	3
P18124	60S ribosomal protein L7	1.54	3
P25705	ATP synthase subunit alpha, mitochondrial	2.18	3
P46781	40S ribosomal protein S9	3.25	3
P61158	Actin-related protein 3	2.64	3
P62241	40S ribosomal protein S8	2.38	3
Q9BPX5	Actin-related protein 2/3 complex subunit 5-like protein	3.30	3
Q9Y4I1	Unconventional myosin-Va	3.26	3
Q9Y608	Leucine-rich repeat flightless-interacting protein 2	2.48	3
O00483	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 4	1.59	2
P06748	Nucleophosmin	2.71	2
P07951	Tropomyosin beta chain	4.92	2
P08727	Keratin, type I cytoskeletal 19	2.26	2
P11142	Heat shock cognate 71 kDa protein	2.57	2
P12814	Alpha-actinin-1	2.62	2
P36578	60S ribosomal protein L4	2.13	2
P42224	Signal transducer and activator of transcription 1-alpha/beta	4.69	2
P42330	Aldo-keto reductase family 1 member C3	5.81	2
P57088	Transmembrane protein 33	3.63	2
P62249	40S ribosomal protein S16	2.67	2

TABLE I—*continued*

Entry	Protein names	H/L (average)	# obs
P62280	40S ribosomal protein S11	1.70	2
P62753	40S ribosomal protein S6	1.83	2
P62899	60S ribosomal protein L31	1.81	2
P83731	60S ribosomal protein L24	2.31	2
Q01082	Spectrin beta chain, non-erythrocytic 1	1.82	2
Q05639	Elongation factor 1-alpha 2	2.93	2
Q12792	Twinfilin-1	2.50	2
Q13045	Protein flightless-1 homolog	4.73	2
Q13509	Tubulin beta-3 chain	0.50	2
Q13813	Spectrin alpha chain, non-erythrocytic 1	1.84	2
Q68CQ1	Maestro heat-like repeat-containing protein family member 7	2.87	2
Q71U36	Tubulin alpha-1A chain	3.75	2
Q96A08	Histone H2B type 1-A	2.10	2
Q9H9B4	Sideroflexin-1	2.24	2

CDK9 regulated genes and characterizes the effect of the complex on the expression of the CDK9 regulated genes, based on three-gene correlations learned from a compendium of transcriptional expression profiles (see supplemental Table S11). The resulting CDK9 modulatory network was represented with a matrix of 257 rows (each row a modulator) and 344 columns (each column a predicted target gene, Fig. 7). The color for each element represents the secondary effect of a modulator on target gene expression, as activation or attenuation.

The CDK9 modulators clustered into two major clusters (Fig. 7). With all the clustered 257 modulators as background, cluster 1 modulators (CM1) are enriched in translation elongation, regulation of transcription, gene expression, regulation of DNA replication, chromatin modification, regulation of chromosome organization, telomere organization and telomere maintenance. Cluster 2 modulators (CM2) are enriched in cytoskeleton organization, cellular component movement, rRNA processing, actin filament-based process, anatomical structure morphogenesis, ribonucleoprotein complex biogenesis, ribosome biogenesis. For enriched GO categories with all genome as background, both clusters of the modulators are enriched in mRNA processing, RNA splicing, cell cycle. GO categories uniquely enriched in CM1 include telomere maintenance, replicative cell aging, cytokinesis, and regulation DNA replication. Biological processes uniquely enriched in CM2 include DNA repair, response to DNA damage stimulus, cellular response to stress, DNA repair, cytoskeleton organization, DNA packaging, regulation of mRNA statbility, and others (For details, see supplemental Table S11). We note that DDX5 and DDX17 are in cluster 1 (CM1) and cluster 2 (CM2) of modulators, respectively, which suggests that while DDX5 and DDX17 have similar splicing function, their effects on CDK9-dependent target gene expression may be divergent (Supplemental Table S11). Interestingly to us, CM2 modulators are significantly enriched in the differentially expressed proteins identified by our direct comparison analysis (Fig. 4 and Table I). Here 20 of the poly(I:C) enriched genes are found in the 193 modulators in CM2, whereas only 2 of the poly(I:C) enriched genes are found in CM1. This enrichment is highly statistically significant ($p < 5.96$ e-4).

The predicted CDK9 target genes (genes whose transcription depends on CDK9) were clustered into four clusters. Cluster 1 target genes (CT1), which has positive correlation with the CM2 modulator cluster, are enriched in protein ubiquitination, positive regulation of circadian rhythm, catabolic process, cell cycle, and vasculature development. We note the functional consistency of the CM2 modulator cluster and that of its positively correlated CT1 target genes cluster. Similarly, cluster 2 target genes (CT2), which also positively correlated with CM2 modulator cluster are more enriched in response to oxygen level, vasculature development, response to stress, defense response, wound healing involved in inflammatory response, regulation of production of molecular mediator of immune response; regulation of innate immune response, inflammatory response, response to nutrient levels, regulation of leukocyte mediated immunity. Target gene cluster CT4 have only one significant enriched GO term in biological process, that being asymmetric cell division (supplemental Table S11). Target gene cluster CT4 has no significant enriched GO terms after Benjamini-Hochberg correction for multiple testing. These results provide a global modulatory network for CDK9 and suggest functional consistency of the CDK9 interacting proteins and target genes. Our interpretation is that CDK9 binding proteins may be classified into two clusters, each controlling distinct biological pathways.

DDX5/17 RNA Helicases Function as Regulators of Alternative Splicing—Together these data confirmed that the basal CDK9 interacts with DDX RNA helicases and suggests that CDK9 may have novel functions on RNA splicing in unstimulated cells. DDX5/17 are highly conserved paralogs whose alternative splicing effects on histone macroH2A1 controls cellular redox signaling. To confirm the role of DDX5/17 in alternative mRNA splicing, we conducted experiments to knock down their expression by small hairpin (sh)-RNA mediated silencing. Because DDX-5 and -17 are highly conserved

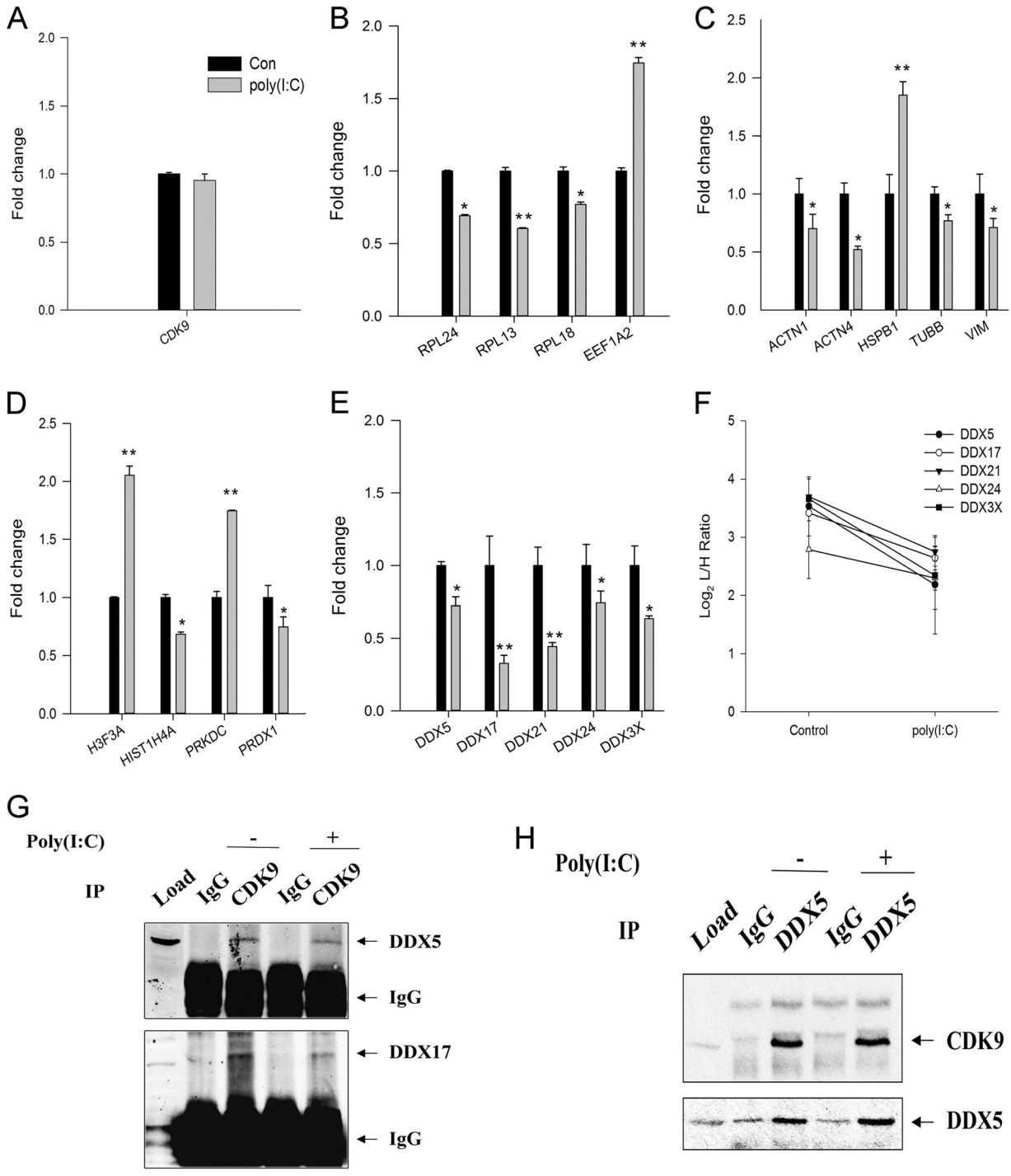

FIG. 5. **Validation of the differential CDK9 interacting proteins in basal *versus* activated states.** IP-SRM assays were conducted using anti-CDK9 pull-down of unstimulated (control) or poly(I:C) stimulated cells. Data are expressed as fold change relative to unstimulated native/aqua measurements. CDK9 measurements were similar between control and poly(I:C) stimulated cells (*A*), poly(I:C) affects CDK9 protein binding with proteins involved in ribosomal function/translational elongation (*B*), cytoskeletal assembly (*C*), chromosome structure (*D*) and DEAD-box RNA helicases (*E*). * $p < 0.05$; ** $p < 0.01$ (*t* test). Abbreviations are in Fig. 3. *F*, Combined annotation of DEAD-box RNA helicases in IP-MS/MS analysis, including DDX17, DDX21, DDX24, DDX3X, and DDX5. *G–H*, IP-Western blot assays of unstimulated or poly(I:C) stimulated cells IPed with IgG, anti-CDK9 (*G*) or anti-DDX5 (*H*) as indicated. Western blot using antibodies of anti-DDX5 (*G*, *top*), anti-DDX17 (*G*, *bottom*), anti-CDK9 (*H*, *top*) or anti-DDX5 (*H*, *bottom*). Specific complexes are marked.

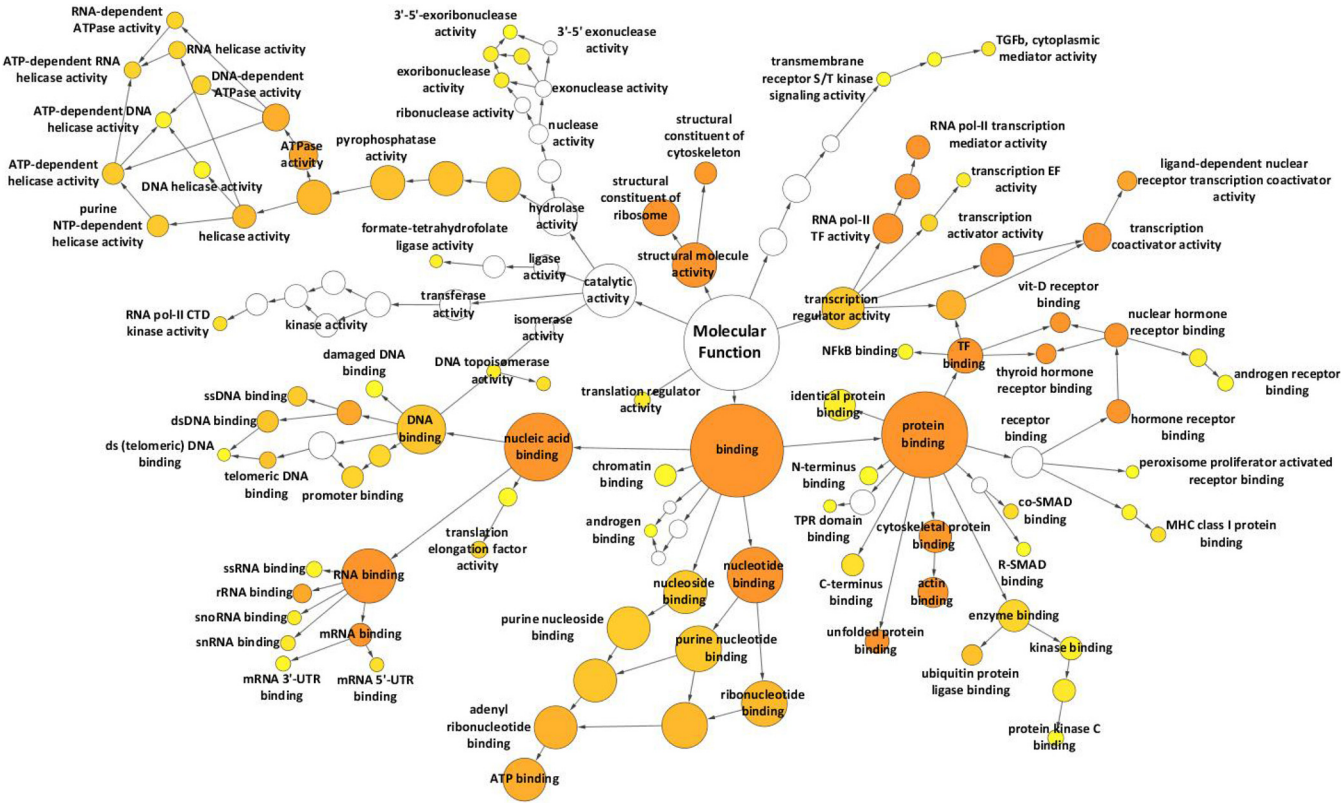

FIG. 6. **Functional gene ontology annotation of literature-completed human CDK9 interactome.** Gene ontology annotation - clustering analysis of 561 human CDK9 interacting proteins (184 published + 407 newly identified). Yellow nodes represent GO categories that are overrepresented at the significance level of p value < 0.05. Higher the significant p values are represented by darker node colors, increasing toward orange. Uncolored nodes are those that were not over-represented. For part of the lineage of a given GO category, their labels were removed to increase the clarity of significant categories.

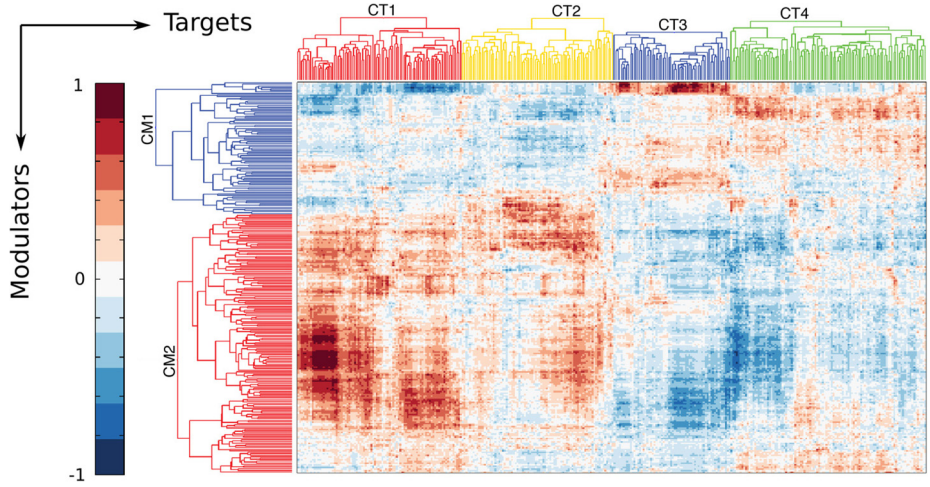

FIG. 7. **Global CDK9 modulatory network.** The CDK9 modulatory network was constructed from all known CDK9 binding proteins as candidate modulators and all genes with target gene candidates. Only modulators and candidates are involved in at least three triplets were retained. Shown is a biclustering heatmap based on the γ parameter. Each row represents a modulator probeset, each column represents a target gene probeset and each element represents the γ parameter. The modulators are clustered into two large clusters: cluster CM1 (red) and cluster CM2 (blue). The target genes are also clustered into four large clusters, cluster CT1 (red), cluster CT2 (yellow), cluster 3 (blue), and cluster CT4 (green).

at the gene level, we designed a inducible shRNA that targets both isoforms under control of a tetracycline inducible ("Tet-on") knockdown system. In this system, treatment with doxycycline (Dox) induces the expression of DDX5/7 shRNA. The resulting inducible knockdown cell lines were treated with Dox or left untreated and tested for cellular DDX5/17 levels by

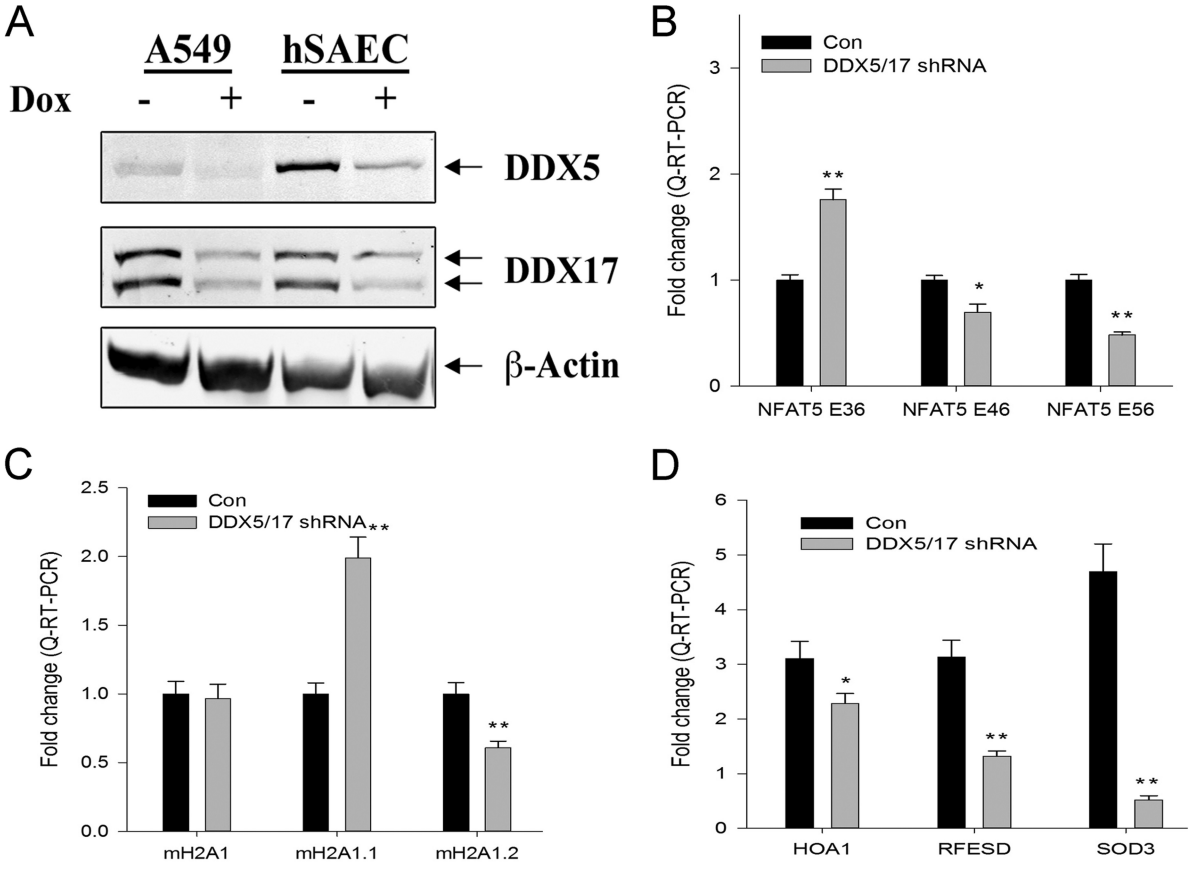

FIG. 8. **The DDX5/17 RNA helicases control alternative splicing of NFAT5 and mH2A1 transcripts.** *A*, Western blot analysis of DDX5 (68 kDa), DDX17 (72 and 82 kDa), and β-actin as loading control, using total protein exacts from A549 and human small airway epithelial cells (hSAEC) stably transfected with inducible lentivial DDX5/17 shRNA in the absence or presence of Dox (2 μg/ml) for 72 h. *B*, Q-RT-PCR analysis of NFAT5 splicing variants containing exon 5 (E56), none of exon 4 and exon 5 (E36), or exon 4 alone (E46) after 72 h of Dox treatment (2 μg/ml) in DDX5/17 shRNA stably transfected A549 cells. Data are shown as fold change normalized to unstimulated cells. *C*, Q-RT-PCR analysis of total mH2A1 and mH2A1.1 and mH2A1.2 splicing variants in the same experimental conditions as in panel *B*. *D*, Regulation of mH2A1 splicing isoforms by DDX5 and DDX17 affects genes involved in redox metabolism. Q-RT-PCR analysis of HOA1, RFESD, and SOD3 mRNA level in the same experimental conditions as in panel *B*. * $p < 0.05$; ** $p < 0.01$ (*t* test).

Western blot. As shown in Fig. 8*A*, both DDX-5 and -17 isoforms were depleted (70–80%) in a Dox-dependent manner in A549 cells as well as in primary human small airway epithelial cells (hSAECs).

It has been shown that DDX5 and DDX17 regulate the alternative splicing of the promigratory transcription factor nuclear factor of activated T-cells 5 (NFAT5) and the histone macroH2A1 (38). The human NFAT5 gene has known alternatively included exons 4 and 5. Q-RT-PCR using primers binding exons 3 and 6, flanking exons 4 and 5, revealed that DDX5/17 depletion in A549 cells increased the level of an NFAT5 variant that did not contain exon 4 and 5 (NFAT5 E36, Fig. 8*B*), and decreased the level of the NFAT5 splicing variants lacking exon 4 but containing exon 5 (NFAT5 E56) as well as those containing exon 4, but lacking exon 5 (NFAT5 E46). The human mH2A1 gene generates two splicing isoforms (mH2A1.1 and mH2A1.2) around exon 6 through the use of two mutually exclusive exons (39). The depletion of DDX5/17 in A549 cells had no effect on the global mH2A1 gene expression level but increased that of the mH2A1.1 isoform and decreased the level of the mH2A1.2 isoform (Fig. 8*C*). These data are in agreement with previous reports and confirmed that DDX5/17 depletion induced mH2A1 splicing switch from mH2A1.2 isoform to mH2A1.1 isoform.

DDX5 and DDX17 mediated regulation of mH2A1 alternatively spliced isoforms influences expression of mH2A1-dependent downstream genes involved in redox metabolism (38). We therefore measured the effect of DDX5/17 depletion on the expression profiles of hydroxyacid oxidase 1 (HAO1), Rieske (Fe-S) domain containing (RFESD), and extracellular superoxide dismutase 3 (SOD3). As shown in Fig. 8*D*, DDX5/17 knockdown decreased HAO1, RFESD, and SOD3 mRNA expression. Altogether, our results indicate that DDX5 and DDX17 regulate the alternative splicing of transcription factor NFAT5 and histone mH2A1 variant, controlling redox metabolism in a functionally significant manner.

CDK9 Regulates mH2A1 and NFAT5 Alternative Splicing—The observation that CDK9 interacts with DDX5/17 suggests

that CDK9 plays a potential role on alternative splicing. To test this hypothesis, we first evaluated the effects of the specific CDK9 kinase small molecule inhibitor, CAN508, on alternative splicing of NFAT5 and mH2A1 transcripts. CAN508 is a highly CDK9-selective inhibitor with a 50% inhibitory concentration (IC_{50}) 10-fold lower than other CDKs from CDK1 to CDK7 (40). A549 cells were treated in the absence or presence of CAN508 and the effect measured on expression of NFAT5 alternative splice forms. We observed that CAN508 treatment increased the NFAT5 isoform lacking exon 4 and 5 (NFAT5 E36, Fig. 9*A*), decreased the isoform containing exon 4 but not exon 5 (NFAT5 E46, Fig. 9*A*), and decreased isoform lacking exon 4 but containing exon 5 (NFAT5 E56, Fig. 9*A*). Importantly, this pattern was very similar to that observed after DDX5/17 depletion (c.f. Fig. 8*B*).

CDK9 inhibition produced a similar effect on the expression of mH2A1 splicing patterns. Similarly to what has been shown in DDX5/17 depletion (Fig. 8*C*), the CDK9 inhibition resulted in decreased level of mH2A1.2 isoform, increased level of mH2A1.1 isoform, whereas having no effect on the level of global mH2A1 transcripts (Fig. 9*B*). A time series of CAN508 inhibition showed time dependent increase of mH2A1.1 isoform but not mH2A1.2 isoform during short term (6 h) of CAN508 treatment (supplemental Fig. S1). Collectively, these results suggested that the alternative splicing events of NFAT5 and mH2A1 genes are regulated by CDK9 kinase activity in the same manner as that by the DDX5/17 RNA helicases.

To further understand the effects of CDK9 on alternative splicing, we depleted endogenous CDK9 protein using shRNA expression in "Tet-on" lentiviruses. Two Doxycyline (Dox)-inducible lentivial CDK9 shRNAs (CDK9 shRNA1 and shRNA2) were tested for knockdown efficiency in A549 cells. As shown in Fig. 9*C*, CDK9 shRNA1 significantly reduced the CDK9 protein level (80–90%) for both the 42 KDa and 55 KDa CDK9 isoforms after Dox induction. Consistent with our findings in CAN508-treated A549 cells (Figs. 9*A* and 9*B*), CDK9 depletion resulted in NFAT5 and mH2A1 mRNA splicing patterns similar to that of DDX5/17-depletion (Figs. 9*D* and 9*E*). Moreover, the splicing switch of mH2A1 isoforms induced by CDK9 depletion largely down-regulated the transcription of mH2A1-dependent HAO1, RFESD, and SOD3 redox metabolism genes (Fig. 9*F*). Taken together, our data demonstrated that CDK9 complexes with DDX5 and DDX17 to regulate the alternative splicing of NFAT5 and mH2A1 in a functionally significant manner, controlling redox regulation.

CDK9 Regulates Alternative Splicing by recruiting DDX5 to Target Gene Promoters—The functional codependence between CDK9 and DDX5/17 on alternative splicing of NFAT5 and mH2A1 led us to explore whether CDK9 regulates promoter recruitment of DDX-5 or -17. In order to determine the effect of CDK9 knockdown on the promoter localization of DDX5 and DDX17, CDK9 depletion was induced by Dox treatment of shRNA1-expressing stable A549 cells, and then stimulated with or without poly(I:C). The *in vivo* occupancy of NFAT5 and mH2A1 promoter regions by CDK9, DDX5, and DDX17 were assessed using a highly quantitative XChIP assay (31). Here we observed that CDK9 knockdown significantly decreased CDK9 recruitment to the NFAT5 and mH2A1 promoters, as would be expected (Fig. 10*A* and 10*B*). Interestingly, at both regions tested, we observed that DDX5 genomic occupancy decreased upon CDK9 knockdown, suggesting CDK9 is required to recruit DDX5 to the promoters of NFAT5 and mH2A1. Strikingly, the occupancy of DDX17 was not altered, suggesting that DDX17 may be found within multiple chromatin remodeling complexes in addition to that with CDK9.

DISCUSSION

CDK9 plays a central role in the rapid activation of epithelial innate immune response (IIR) genes. In the unstimulated state, IIR genes are found in an open chromatin configuration engaged with hypophosphorylated RNA Pol II. In response to IIR activation, the CDK9 complex shifts to an activated state, dissociating from the 7SK snRNA-HEXIM1/2 complex to bind BRD4. Upon transcription factor-dependent recruitment, the CDK9 complex phosphorylates Ser2 of the Pol II CTD, triggering elongation-competent RNA Pol II to make fully spliced inflammatory and antiviral genes that elicit protective host responses. In this manuscript, we seek to identify the spectrum of functions associated with the activated CDK9 complex by IP-LC-MS/MS and inference of the molecular functions of the interacting proteins. Using a stringent filter of enrichment on proteins to be identified in replicate experiments, we identify 407 high confidence unique CDK9 interacting proteins in the basal and activated states, extending previous work on CDK9 protein interactions (13, 41). Our study builds on previous stable isotope labeling methods to identify basal CDK9 interactors using ectopically expressed CDK9 in murine erythroleukemic cells (41). Of the proteins identified in this earlier study, 32 proteins were found common in our human basal CDK9 list, including cytoskeletal proteins, ribosomal proteins and DDX17. The differences between our studies may be explained by the differences in cell type (epithelial *versus* erythroleukemia), species (human *versus* mouse) or the earlier study using ectopic expression of tagged CDK9. These same confounding issues may explain why we have not identified 151 proteins published in the literature. More work will be required to understand these differences.

Control of Pol II-dependent gene expression in eukaryotic cells involves regulatory events at multiple transcriptional and post-transcriptional stages. Transcriptional regulatory complexes are multifunctional multicomponent complexes containing kinases, ubiquitin ligases, polymerases, and histone modifiers (acetylases, deacetylases, methylases). Gene expression is coordinated through a series of integrated steps including promoter initiation, elongation, processing, and ter-

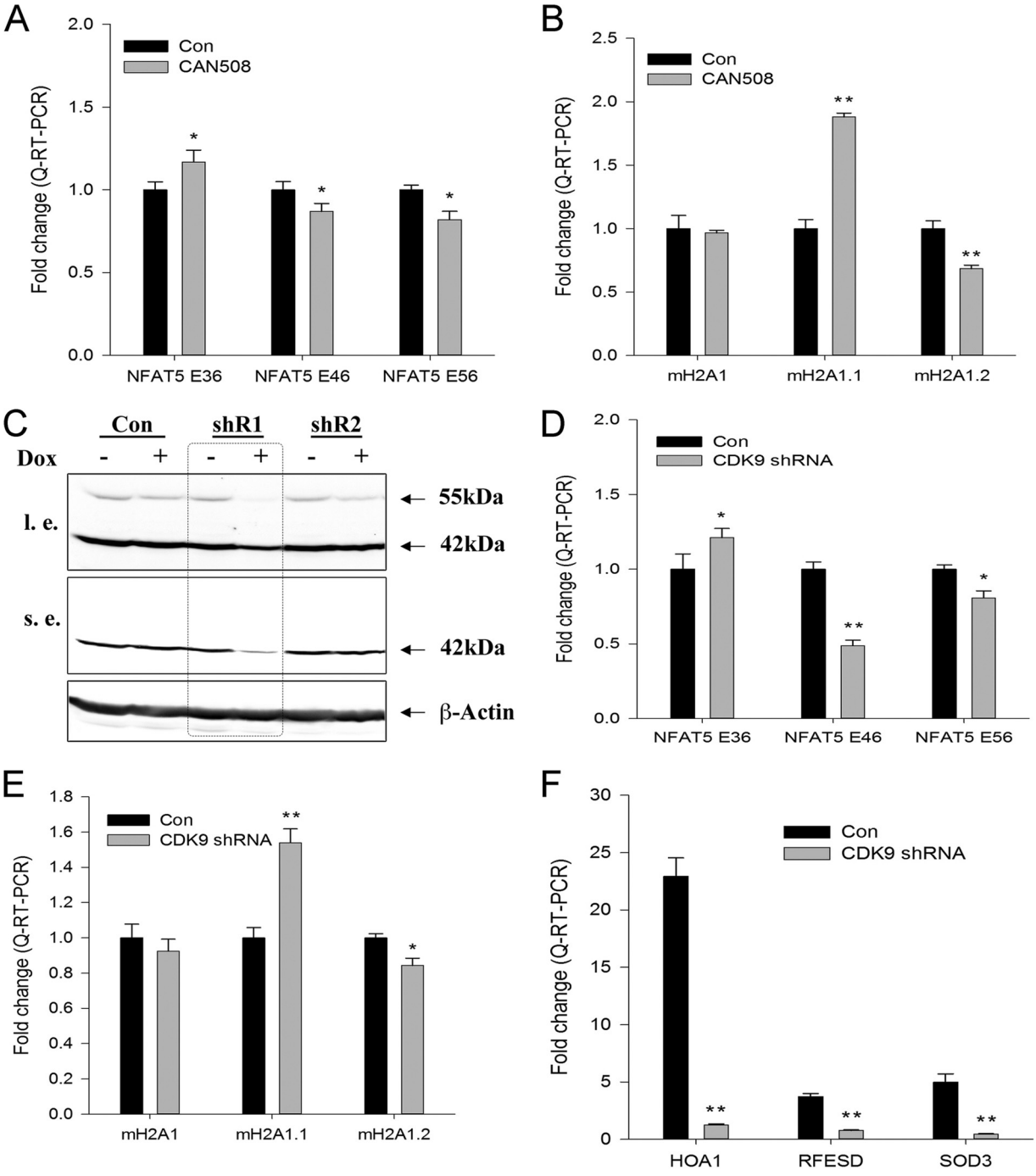

FIG. 9. **CDK9 controls alternative splicing of NFAT5 and mH2A1 isoforms.** *A*, Q-RT-PCR analysis of NFAT5 splicing variants containing exon 5 (E56), missing both exon 4 and exon 5 (E36), or exon 4 alone (E46) after CDK9 inhibitor CAN508 pretreatment (20 μM, 16 h) in A549 cells. *B*, Q-RT-PCR analysis of total mH2A1 as well as its splicing variants mH2A1.1 and mH2A1.2, in the same experimental conditions as in panel *A*. *C*, Western blot analysis of CDK9 (42 and 55 kDa) and β-actin as loading control, using total protein exacts from A549 cells stably transfected with inducible lentiviral control or two CDK9 shRNAs (shR1 or shR2) in the absence (- Dox) or presence (+ Dox) of 2 μg/ml doxycycline for 72 h. l.e., long exposure; s.e., short exposure. *D*, Q-RT-PCR analysis of NFAT5 splicing variants containing exon 5 (E56), missing both exon 4 and exon 5 (E36), or containing exon 4 alone (E46) with or without 72 h of Dox treatment (2 μg/ml) in CDK9 shRNA1-stably transfected A549 cells. *E*, Q-RT-PCR analysis of total mH2A1 as well as its splicing variants mH2A1.1 and mH2A1.2, in the same experimental conditions as in panel *D*. *F*, CDK9 mediated regulation of mH2A1 splicing isoforms by DDX5 and DDX17 affects genes involved in redox metabolism. Q-RT-PCR analysis of HOA1, RFESD, and SOD3 mRNA level in the same experimental conditions as in panel D. * $p < 0.05$; ** $p < 0.01$ (*t* test).

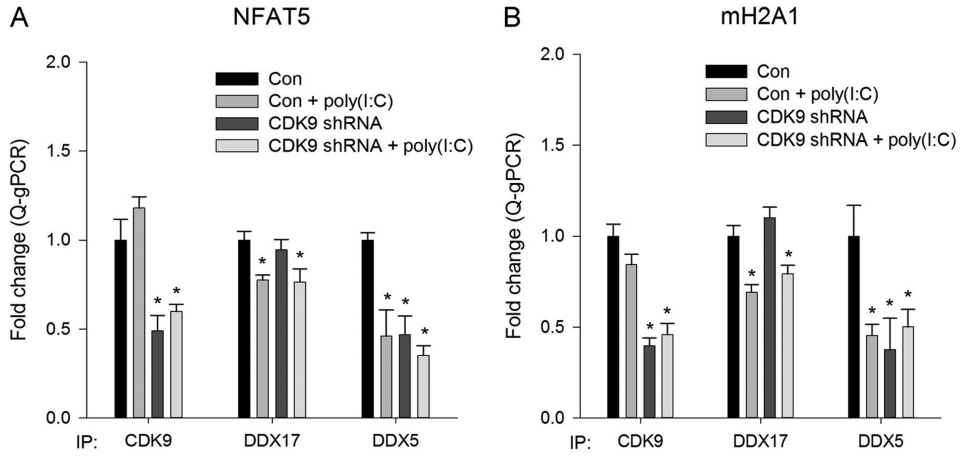

FIG. 10. **CDK9 regulates alternative splicing by recruitment of DDX5 to promoters of mH2A1 and NFAT5 genes.** A549 cells stably transfected with CDK9 shRNA1 were induced with or without Dox (2 μg/ml for 72 h), then electroporated with or without poly (I:C) for 3 h. Chromatin was immunoprecipitated with anti- CDK9, DDX5, or DDX17 Abs. Q-gPCR was performed using the specific primers spanning promoter regions of NFAT5 and mH2A1 genes on genomic DNA from ChIP experiment. Data are expressed as fold change relative to unstimulated cells. * $p < 0.05$ (*t* test).

mination. Although virtually all steps of transcriptional activation are regulated, recent work has shown that immediate early genes in the IIR are primarily controlled at the level of transcriptional elongation, coordinated by the phosphorylated CTD of RNA Pol II (8, 10, 42). Other studies have implicated distinct splicing complexes are associated with the RNA Pol II CTD. Deleted in breast cancer and ZNF326, proteins of the DBIRD complex, have been shown to be associated with A-T rich RNA-splicing activity (43). Our study provides the first direct evidence that the transcriptional elongation complex is associated with DDX family of RNA splicing factors, involved in G- rich RNA splicing. Here, the transcriptional elongation complex serves as a scaffold for promoting the process of mRNA capping, processing, splicing, polyadenylation and nuclear export, - an integrated cotranscriptional process coined, "cotranscriptionality" (44). How multicomponent and multifunctional protein complexes integrate the multiple steps in transcriptional elongation, 5′ capping, mRNA splicing and, 3′ processing is incompletely understood. Our work in the analysis of the CDK9 complex suggests that this complex is not only responsible for transcriptional elongation, but also mRNA splicing and cytoskeletal binding.

Our data discover and confirm that the basal CDK9 complex is enriched in DEAD-box RNA helicase proteins; this diverse family of RNA-dependent ATPases are involved in RNA processing and metabolism (45). DDXs share a central helicase core domain that function to unwind RNA, promote duplex formation, displace proteins from single stranded RNA, and serve as assembly platforms for larger ribonucleoprotein complexes. DDX5 is multi-functional enzyme that modulates multiple functions in gene expression: transcriptional initiation as a coactivator of p53, MyoD, SMADs and ERα transcription factors; pre-mRNA processing via splicing; and termination through transcript release from chromatin (46). DDX5 is a member of the Drosha complex and may play a role in miRNA processing (47). In future studies, it will be of interest to examine whether miRNA processing is a function of the CDK9-DDX5/17 complex.

Alternative splicing of mRNA precursors provides an important means of genetic control. Previous work has shown that DDX-5 and -17 cooperate with heterogeneous nuclear ribonucleoprotein (hnRNP) H/F splicing factors to define epithelial- and myoblast-specific splicing subprogram (48). Here genome-wide exon array-based profiling of cells depleted in DDX5/17 identified their roles in control of 233 skipped exons and 133 exon inclusion events (48). The DDX-5/17-hnRNP complex in alternative splicing is thought to be because of its affinity for 5′ splice sites, forming G-rich quadriplex structures where its RNA helicase activity produces different pre-mRNA conformations within the spliceosome. Indeed our data showed that hnRNPs, such as hnRNPH1, hnRNPF, and hnRNPU, also interact with activated CDK9. We show here that two targets, NFAT5 and mH2A1, are under DDX5/17 control in a CDK9-dependent manner. TOP2B, another novel protein identified in this project, may also have functional interaction with CDK9 during RNA transcription and processing.

Alternative mH2A1 splicing is involved in cell migration and tumor cell invasiveness (38). The expression of DDX5 is upregulated in colorectal, prostate, and breast cancers and has been shown to correlate with tumor progression and transformation. Our shRNA knockdown and ChIP studies here indicate that DDX5 binding to mH2A1 is CDK9-dependent. The DDX-5 and -17 proteins share greater than 90% sequence conservation and form heterodimeric complexes, explaining their significant functional redundancy. However our data suggest that DDX5 promoter recruitment to mH2A1 is CDK9-dependent, whereas DDX17 has separate modes of gene recruitment. These data suggest that DDX5/17 also form

independent protein complexes, allowing both CDK9-dependent and CDK9-independent mechanisms for targeting transcribed chromatin.

Our findings showed that the activated CDK9 complex is associated with ribosomal proteins. RPLs-13 and -24 are nuclear encoded cytosolic structural proteins associated with the large 60S ribosomal subunit. These proteins along with RPL7, 31, 27, and S8 are within a known interacting hub with EEF1A2 (Fig. 2*B*). Together we speculate that the CDK9 complex may be functionally involved in export, ribosomal engagement, and translation of IIR-activated genes. The role, if any, of CDK9 in RNA translational control will require additional exploration.

Our observations that the ACTN-1, -4, and HSPB1/HSP27 as validated components with the basal CDK9 complex indicate a role for nuclear cytoskeletal association in transcriptional elongation/mRNA splicing. Nuclear actin is an essential component of gene expression, being involved in RNA Pol I, II, and III transcription, chromatin remodeling, the formation of hnRPs, as well as in recruitment of histone modifiers to the actively expressed genes (49). These findings suggest that nuclear actin may play an important role in the formation, and perhaps transit, of the transcriptional elongation complex (50). It is known that actively transcribed genes are enriched in actin-containing nuclear matrix, and that chromatin remodeling SWI/SNF complex is associated with nuclear actin (51). Moreover, recent work has shown that ACTN4 is associated with the activated nuclear NFκB transcriptional factor as a gene-specific coactivator (52). More work will be required to understand the relationship between the cytoskeleton, CDK9, and mRNA processing in inducible gene expression.

Interestingly, although a family of DDX isoforms are associated with the basal CDK9 complex, our quantitative comparisons of DDX abundance between the resting and activated CDK9 complexes shows that the DDX association is reduced after poly(I:C) stimulation. This finding is consistent with our findings in XChIP where the amount of CDK9 dependent DDX-5 and -17 binding to NFAT5 and mH2A1 in native chromatin is reduced after poly(I:C) treatment despite having similar amounts of CDK9 binding to each promoter (Figs. 10*A* and 10*B*). NFAT5 and mH2A1 are constitutive genes, whose expression is not modified by poly(I:C) treatment. We interpret these findings to suggest that alternative RNA splicing by the CDK9-DDX complex is primarily seen in unstimulated cells and that this alternative splicing mechanism may be suppressed during activation of the innate immune response. In this setting, rapid recruitment of CDK9 without associated alternative splice factors mediates the rapid stereotypic expression of protective IFNs and ISG genes, necessary for organismal survival to an infectious organism. More work will need to be done to understand how signaling pathways induce dynamic changes in the CDK9 complex, and these complexes modulate mRNA expression, processing, and translation.

Our analysis of the inducible changes of the CDK9 interactome after poly(I:C) stimulation shows recruitment of cytoskeletal myosins, 60S ribosomal structural components, EEF1A2, HSPB1, histone H3F3A, and PRKDC into the complex. The presence of these proteins suggest CDK9 interactome controls steps in cytoskeletal rearrangement, translational regulation, and chromatin reorganization in the IIR. Although HSPB1 is associated with CDK9 in the basal complex, we observe an increase in HSPB1 after poly(I:C) stimulation. In response to stress, HSPB1 is phosphorylated by MAPKs and undergoes nuclear translocation. In the nucleus, HSPB1 is a component of SC35 nuclear speckles; interestingly, these macromolecular structures are involved in mRNA splicing (53). EEF1A2 is involved in bringing aminoacylated tRNA into the ribosome, the combined interaction of EEF1A2 and ribosomal structural proteins suggests to us CDK9 may be involved in promoting efficient translation of innate activated genes. Histone H3F3A is a Histone H3 isoform is incorporated into nucleosomes restructured by activated transcription (54); its association with CDK9 may play a role in chromatin restructuring induced by poly(I:C).

Our study has identified 407 novel, high-confidence proteins within the CDK9 interactome. Combined with 184 proteins previously published in the literature enables a significantly more comprehensive view of the role of CDK9 in transcription, pre-mRNA splicing, and potentially mRNA transport/translation through interactions with ribosomal structural proteins and actin/myosin cytoskeleton. Our computational, genome-wide inference of all possible triplets consisting of modulators-CDK9-predicted target genes of CDK9 genome-wide, provides insight into the innate pathway-induced changes in the CDK9 regulatory network (Fig. 7). Biclustering the network demonstrates how different CDK9 binding proteins as modulators might modulate the expression of distinct cellular functions. We note here that CDK9 modulator cluster 2 (CM2) is significantly enriched in proteins identified to be differentially recruited to the CDK9 complex in response to poly(I:C) stimulation. These modulators are associated with target genes involved in cytokine response, wound healing and inflammatory response, all essential functions of the epithelial innate immune response. These data suggest that this dynamic exchange of CM2 modulators of CDK9 mediates the basis of the IIR. Because the abundance and Ser- Thr 186 phosphorylation state of CDK9 are not affected by poly(I:C) stimulation, the post-translational modification of these proteins regulating their recruitment into the CDK9 complex will require further investigation.

In summary, IP-LC-MS/MS was used to identify PPIs of CDK9 and to understand the multiple functions of this essential component of the transcriptional elongation complex. Our work has demonstrated the association of DDX-5 and -17 in regulating alternative mRNA splicing of NFAT5 and mH2A1 and downstream redox genes. This work will be the foundation for additional exploration of the transcriptional elongation

complex in translational control and relationship with the nuclear cytoskeleton.

* This work was supported by the Keck center computational cancer biology training program of the gulf coast consortia (CPRIT Grant No. RP140113 to XL, AK and ARB), the NIAID Signaling in Airway inflammation, AI062885, NHLBI Proteomics Center in Airway Inflammation, HHSN272200800048C (ARB), UL1TR000071 UTMB CTSA (ARB), and NIEHS P30 ES006676 (ARB).

S This article contains supplemental Figs. S1 to S6, Tables S1 to S17.

** To whom correspondence should be addressed: Institute for Translational Sciences, University of Texas, MRB 8.128 301 University Blvd., Galveston, TX 77555-1060. Tel.: 409-772-2824; Fax: 409-772-8709; E-mail: arbrasie@utmb.edu.

REFERENCES

1. Knight, D. A., and Holgate, S. T. (2003) The airway epithelium: structural and functional properties in health and disease. *Respirology* **8,** 432–446
2. Lambrecht, B. N., and Hammad, H. (2012) The airway epithelium in asthma. *Nat. Med.* **18,** 684–692
3. Hopkins, P. A., and Sriskandan, S. (2005) Mammalian Toll-like receptors: to immunity and beyond. [Review] [176 refs]. *Clin. Exp. Immunol.* **140,** 395–407
4. Akira, S., Uematsu, S., and Takeuchi, O. (2006) Pathogen recognition and innate immunity. *Cell* **124,** 783–801
5. Brasier, A. R., Garcia-Sastre, A., and Lemon, S. M. (2008) *Cellular signaling and innate immune response to RNA virus infections*, ASM Press, Washington DC
6. Brasier, A. R. (2008) Expanding role of cyclin dependent kinases in cytokine inducible gene expression. *Cell Cycle* **7,** 2661–2666
7. Price, D. H. (2000) P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. *Mol. Cell. Biol.* **20,** 2629–2634
8. Tian, B., Zhao, Y., Kalita, M., Edeh, C. B., Paessler, S., Casola, A., Teng, M. N., Garofalo, R. P., and Brasier, A. R. (2013) CDK9-dependent transcriptional elongation in the innate interferon-stimulated gene response to respiratory syncytial virus infection in airway epithelial cells. *J. Virol.* **87,** 7075–7092
9. Marshall, N. F., and Price, D. H. (1995) Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *J. Biol. Chem.* **270,** 12335–12338
10. Brasier, A. R., Tian, B., Jamaluddin, M., Kalita, M. K., Garofalo, R. P., and Lu, M. (2011) RelA Ser276 phosphorylation-coupled Lys310 acetylation controls transcriptional elongation of inflammatory cytokines in respiratory syncytial virus infection. *J. Virol.* **85,** 11752–11769
11. Nowak, D. E., Tian, B., Jamaluddin, M., Boldogh, I., Vergara, L. A., Choudhary, S., and Brasier, A. R. (2008) RelA Ser276 phosphorylation is required for activation of a subset of NF-kappaB-dependent genes by recruiting cyclin-dependent kinase 9/cyclin T1 complexes. *Mol. Cell. Biol.* **28,** 3623–3638
12. Egloff, S., Van Herreweghe, E., and Kiss, T. (2006) Regulation of polymerase II transcription by 7SK snRNA: two distinct RNA elements direct P-TEFb and HEXIM1 binding. *Mol. Cell. Biol.* **26,** 630–642
13. Jeronimo, C., Forget, D., Bouchard, A., Li, Q., Chua, G., Poitras, C., Thérien, C., Bergeron, D., Bourassa, S., Greenblatt, J., Chabot, B., Poirier, G. G., Hughes, T. R., Blanchette, M., Price, D. H., and Coulombe, B. (2007) Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. *Mol. Cell* **27,** 262–274
14. Wu, S. Y., and Chiang, C. M. (2007) The double bromodomain-containing chromatin adaptor Brd4 and transcriptional regulation. *J. Biol. Chem.* **282,** 13141–13145
15. Barboric, M., Nissen, R. M., Kanazawa, S., Jabrane-Ferrat, N., and Peterlin, B. M. (2001) NF-kappaB binds P-TEFb to stimulate transcriptional elongation by RNA polymerase II. *Mol. Cell* **8,** 327–337
16. Forbus, J., Spratt, H., Wiktorowicz, J., Wu, Z., Boldogh, I., Denner, L., Kurosky, A., Brasier, R. C., Luxon, B., and Brasier, A. R. (2006) Functional analysis of the nuclear proteome of human A549 alveolar epithelial cells by HPLC-high resolution 2-D gel electrophoresis. *Proteomics* **6,** 2656–2672
17. Jamaluddin, M., Wiktorowicz, J. E., Soman, K. V., Boldogh, I., Forbus, J. D., Spratt, H., Garofalo, R. P., and Brasier, A. R. (2010) Role of peroxiredoxin 1 and peroxiredoxin 4 in protection of respiratory syncytial virus-induced cysteinyl oxidation of nuclear cytoskeletal proteins. *J. Virol.* **84,** 9533–9545
18. Zhao, Y., Widen, S. G., Jamaluddin, M., Tian, B., Wood, T. G., Edeh, C. B., and Brasier, A. R. (2011) Quantification of activated NF-kappaB/RelA complexes using ssDNA aptamer affinity-stable isotope dilution-selected reaction monitoring-mass spectrometry. *Mol. Cell. Proteomics* **10,** M111 008771
19. Starkey, J. M., Zhao, Y., Sadygov, R. G., Haidacher, S. J., Lejeune, W. S., Dey, N., Luxon, B. A., Kane, M. A., Napoli, J. L., Denner, L., and Tilton, R. G. (2010) Altered retinoic acid metabolism in diabetic mouse kidney identified by O isotopic labeling and 2D mass spectrometry. *PloS One* **5,** e11095
20. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26,** 1367–1372
21. Zhao, Y., and Brasier, A. R. (2013) Applications of selected reaction monitoring (SRM)-mass spectrometry (MS) for quantitative measurement of signaling pathways. *Methods* **61,** 313–322
22. Cowley, M. J., Pinese, M., Kassahn, K. S., Waddell, N., Pearson, J. V., Grimmond, S. M., Biankin, A. V., Hautaniemi, S., and Wu, J. (2012) PINA v2.0: mining interactome modules. *Nucleic Acids Res.* **40,** D862–D865
23. Gene Ontology, C. (2008) The Gene Ontology project in 2008. *Nucleic Acids Res.* **36,** D440–444
24. Maere, S., Heymans, K., and Kuiper, M. (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21,** 3448–3449
25. Saito, R., Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L., Lotia, S., Pico, A. R., Bader, G. D., and Ideker, T. (2012) A travel guide to Cytoscape plugins. *Nat. Methods* **9,** 1069–1076
26. Jensen, L. J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M., Bork, P., and von Mering, C. (2009) STRING 8–a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* **37,** D412–416
27. Bader, G. D., and Hogue, C. W. (2003) An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* **4,** 2
28. Vert, J. P., Foveau, N., Lajaunie, C., and Vandenbrouck, Y. (2006) An accurate and interpretable model for siRNA efficacy prediction. *BMC Bioinformatics* **7,** 520
29. Yang, J., Mitra, A., Dojer, N., Fu, S., Rowicka, M., and Brasier, A. R. (2013) A probabilistic approach to learn chromatin architecture and accurate inference of the NF-kappaB/RelA regulatory network using ChIP-Seq. *Nucleic Acids Res.* **41,** 7240–7259
30. Nowak, D. E., Tian, B., and Brasier, A. R. (2005) Two-Step Cross-linking method for Identification of NF- k B Gene Network by Chromatin Immunoprecipitation. *BioTechniques* **39,** 715–725
31. Tian, B., Yang, J., and Brasier, A. R. (2012) Two-step cross-linking for analysis of protein-chromatin interactions. *Methods Mol. Biol.* **809,** 105–120
32. Li, X., Zhao, Y., Tian, B., Jamaluddin, M., Mitra, A., Yang, J., Rowicka, M., Brasier, A. R., and Kudlicki, A. (2014) Modulation of gene expression regulated by the transcription factor NF-kappaB/RelA. *J. Biol. Chem.* **289,** 11927–11944
33. Li, X., Zhu, M., Brasier, A. R., and Kudlicki, A. S. (2015) Inferring Genome-Wide Functional Modulatory Network: A Case Study on NF-kappaB/RelA Transcription Factor. *J. Comput. Biol.* **22,** 300–312
34. Nguyen, V. T., Kiss, T., Michels, A. A., and Bensaude, O. (2001) 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. *Nature* **414,** 322–325
35. Doshi, B. M., Hightower, L. E., and Lee, J. (2010) HSPB1, actin filament dynamics, and aging cells. *Ann. N.Y. Acad. Sci.* **1197,** 76–84
36. Khurana, S., Chakraborty, S., Cheng, X., Su, Y. T., and Kao, H. Y. (2011) The actin-binding protein, actinin alpha 4 (ACTN4), is a nuclear receptor coactivator that promotes proliferation of MCF-7 breast cancer cells. *J. Biol. Chem.* **286,** 1850–1859
37. Li, Q., Price, J. P., Byers, S. A., Cheng, D., Peng, J., and Price, D. H. (2005) Analysis of the Large Inactive P-TEFb Complex Indicates That It Contains One 7SK Molecule, a Dimer of HEXIM1 or HEXIM2, and Two

P-TEFb Molecules Containing Cdk9 Phosphorylated at Threonine 186. *J. Biol. Chem.* **280,** 28819–28826
38. Germann, S., Gratadou, L., Zonta, E., Dardenne, E., Gaudineau, B., Fougère, M., Samaan, S., Dutertre, M., Jauliac, S., and Auboeuf, D. (2012) Dual role of the ddx5/ddx17 RNA helicases in the control of the pro-migratory NFAT5 transcription factor. *Oncogene* **31,** 4536–4549
39. Kustatscher, G., Hothorn, M., Pugieux, C., Scheffzek, K., and Ladurner, A. G. (2005) Splicing regulates NAD metabolite binding to histone macroH2A. *Nat. Struct. Mol. Biol.* **12,** 624–625
40. Krystof, V., Cankar, P., Frysová, I., Slouka, J., Kontopidis, G., Dzubák, P., Hajdúch, M., Srovnal, J., de Azevedo, W. F., Jr., Orság, M., Paprskárová, M., Rolcik, J., Látr, A., Fischer, P. M., and Strnad, M. (2006) 4-arylazo-3,5-diamino-1H-pyrazole CDK inhibitors: SAR study, crystal structure in complex with CDK2, selectivity, and cellular effects. *J. Med. Chem.* **49,** 6500–6509
41. Bezstarosti, K., Ghamari, A., Grosveld, F. G., and Demmers, J. A. (2010) Differential proteomics based on 18O labeling to determine the cyclin dependent kinase 9 interactome. *J. Proteome Res.* **9,** 4464–4475
42. Hargreaves, D. C., Horng, T., and Medzhitov, R. (2009) Control of inducible gene expression by signal-dependent transcriptional elongation. *Cell* **138,** 129–145
43. Close, P., East, P., Dirac-Svejstrup, A. B., Hartmann, H., Heron, M., Maslen, S., Chariot, A., Söding, J., Skehel, M., and Svejstrup, J. Q. (2012) DBIRD complex integrates alternative mRNA splicing with RNA polymerase II transcript elongation. *Nature* **484,** 386–389
44. Perales, R., and Bentley, D. (2009) "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. *Mol. Cell* **36,** 178–191
45. Putnam, A. A., and Jankowsky, E. (2013) DEAD-box helicases as integrators of RNA, nucleotide and protein binding. *Biochim. Biophys. Acta* **1829,** 884–893
46. Fuller-Pace, F. V. (2013) The DEAD box proteins DDX5 (p68) and DDX17 (p72): Multi-tasking transcriptional regulators. *Biochim. Biophys. Acta* **1829,** 756–763
47. Fukuda, T., Yamagata, K., Fujiyama, S., Matsumoto, T., Koshida, I., Yoshimura, K., Mihara, M., Naitou, M., Endoh, H., Nakamura, T., Akimoto, C., Yamamoto, Y., Katagiri, T., Foulds, C., Takezawa, S., Kitagawa, H., Takeyama, K., O'Malley, B. W., and Kato, S. (2007) DEAD-box RNA helicase subunits of the Drosha complex are required for processing of rRNA and a subset of microRNAs. *Nat. Cell Biol.* **9,** 604–611
48. Dardenne, E., Polay Espinoza, M., Fattet, L., Germann, S., Lambert, M. P., Neil, H., Zonta, E., Mortada, H., Gratadou, L., Deygas, M., Chakrama, F. Z., Samaan, S., Desmet, F. O., Tranchevent, L. C., Dutertre, M., Rimokh, R., Bourgeois, C. F., and Auboeuf, D. (2014) RNA helicases DDX5 and DDX17 dynamically orchestrate transcription, miRNA, and splicing programs in cell differentiation. *Cell Reports* **7,** 1900–1913
49. Zheng, B., Han, M., Bernier, M., and Wen, J. K. (2009) Nuclear actin and actin-binding proteins in the regulation of transcription and gene expression. *FEBS J.* **276,** 2669–2685
50. Percipalle, P. (2013) Co-transcriptional nuclear actin dynamics. *Nucleus* **4,** 43–52
51. Rando, O. J., Zhao, K., and Crabtree, G. R. (2000) Searching for a function for nuclear actin. *Trends Cell Biol.* **10,** 92–97
52. Aksenova, V., Turoverova, L., Khotin, M., Magnusson, K. E., Tulchinsky, E., Melino, G., Pinaev, G. P., Barlev, N., and Tentler, D. (2013) Actin-binding protein alpha-actinin 4 (ACTN4) is a transcriptional co-activator of RelA/p65 sub-unit of NF-kB. *Oncotarget* **4,** 362–372
53. Vos, M. J., Kanon, B., and Kampinga, H. H. (2009) HSPB7 is a SC35 speckle resident small heat shock protein. *Biochim. Biophys. Acta* **1793,** 1343–1353
54. Daury, L., Chailleux, C., Bonvallet, J., and Trouche, D. (2006) Histone H3.3 deposition at E2F-regulated genes is linked to transcription. *EMBO Rep.* **7,** 66–71