6698–6716 *Nucleic Acids Research, 2017, Vol. 45, No. 11* *Published online 17 March 2017*
doi: 10.1093/nar/gkx187

CDK12 regulates alternative last exon mRNA splicing and promotes breast cancer cell invasion

Jerry F. Tien[1,†], Alborz Mazloomian[2,3,†], S.-W. Grace Cheng[1], Christopher S. Hughes[1], Christalle C.T. Chow[1], Leanna T. Canapi[1], Arusha Oloumi[3], Genny Trigo-Gonzalez[1], Ali Bashashati[3], James Xu[4], Vicky C.-D. Chang[1], Sohrab P. Shah[1,3,4], Samuel Aparicio[1,3,4] and Gregg B. Morin[1,5,*]

[1]Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver V5Z 1L3, Canada, [2]Graduate Bioinformatics Training Program, University of British Columbia, Vancouver V5Z 4S6, Canada, [3]Department of Molecular Oncology, BC Cancer Agency, Vancouver V5Z 1L3, Canada, [4]Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver V6T 2B5, Canada and [5]Department of Medical Genetics, University of British Columbia, Vancouver V6H 3N1, Canada

Received June 17, 2016; Revised March 07, 2017; Editorial Decision March 07, 2017; Accepted March 09, 2017

ABSTRACT

CDK12 (cyclin-dependent kinase 12) is a regulatory kinase with evolutionarily conserved roles in modulating transcription elongation. Recent tumor genome studies of breast and ovarian cancers highlighted recurrent *CDK12* mutations, which have been shown to disrupt DNA repair in cell-based assays. In breast cancers, *CDK12* is also frequently co-amplified with the *HER2* (*ERBB2*) oncogene. The mechanisms underlying functions of CDK12 in general and in cancer remain poorly defined. Based on global analysis of mRNA transcripts in normal and breast cancer cell lines with and without *CDK12* amplification, we demonstrate that CDK12 primarily regulates alternative last exon (ALE) splicing, a specialized subtype of alternative mRNA splicing, that is both gene- and cell type-specific. These are unusual properties for spliceosome regulatory factors, which typically regulate multiple forms of alternative splicing in a global manner. In breast cancer cells, regulation by CDK12 modulates ALE splicing of the DNA damage response activator ATM and a DNAJB6 isoform that influences cell invasion and tumorigenesis in xenografts. We found that there is a direct correlation between CDK12 levels, DNAJB6 isoform levels and the migration capacity and invasiveness of breast tumor cells. This suggests that *CDK12* gene amplification can contribute to the pathogenesis of the cancer.

INTRODUCTION

Cyclin-dependent kinases (CDKs) and their activating cyclin partners integrate numerous signal transduction pathways to regulate a variety of critical cellular processes (1,2). CDK12 (CRK7, CrkRS) is one of several CDKs that regulate transcription through the differential phosphorylation of the C-terminal domain (CTD) of RNA Polymerase II (3). Specifically, CDK12 pairs with Cyclin K (CCNK) and phosphorylates the CTD to maintain processive elongation (4–9). CDK13 (CDC2L5, CHED), a paralog of CDK12, also pairs with Cyclin K and phosphorylates the CTD (4–6,10), How CDK12 and CDK13 regulate elongation remains poorly understood and their distinct contributions to transcription are unclear. Human CDK12 and CDK13 (∼164 kDa) are much larger than other CDKs (typically 33–56 kDa); in addition to their kinase domains, each has an arginine/serine (RS) domain and two proline-rich domains (3,11–12). RS domains are commonly found in proteins that regulate pre-mRNA splicing (13) and proline-rich domains frequently function in signal transduction proteins (14). These features led to the proposal that CDK12 and CDK13 may integrate signal transduction processes to coordinately regulate pre-mRNA transcription, splicing and alternative splicing (AS) (11,12).

Splicing of pre-mRNA is performed by the spliceosome, a large and dynamic complex composed of snRNPs (small nuclear ribonucleoproteins) and many accessory proteins (15). Multicellular eukaryotes also carry out AS, a highly regulated mechanism for generating a diverse set of proteins from pre-mRNA precursors. It is estimated that 88-100% of human genes are alternatively spliced (16) and 15% of genetic diseases may stem from aberrant splicing (17). AS is also increasingly recognized as a major contributor to

*To whom correspondence should be addressed. Tel: +1 604 675 8150; Fax: +1 604 675 8178; Email: gmorin@bcgsc.ca
†These authors contributed equally to the paper as first authors.

cancer progression (18,19). AS is regulated through splicing factors that bind to *cis*-acting sequences on the pre-mRNA and influence splice-site choice (20). These factors include SR (serine/arginine-rich) proteins which contain RS domains, hnRNPs (heterogeneous nuclear ribonucleoproteins) and members of the RBM (RNA binding motif) family of proteins, all of which generally contain RNA recognition motifs. The expression of these splicing factors is often tissue-specific and the genes encoding them are commonly misregulated or mutated in cancer (19). Furthermore, global transcriptome studies have found that depletion or inhibition of many of these splicing factors have broad effects on AS and often affects multiple types of AS events (21–24). There are reports of CDK12 and CDK13 regulating AS, mostly with model splicing substrates. Rat Cdk12 altered the splice site utilization of E1a model splicing substrates (25) and CDK13 affects constitutive and AS of TNF-β and E1a model splicing substrates, respectively (12). In *Drosophila*, Cdk12 appears to regulate AS of Neurexin IV pre-mRNA during development (26). There is also a report that depletion of CDK12 affects the splicing of *SRSF1* in cultured human colorectal cancer cells (27). A detailed understanding of how CDK12 and CDK13 globally affect AS is not known.

Several recent studies have implicated CDK12 in cancer pathology. The Cancer Genome Atlas (TCGA) project identified recurrent somatic alterations in *CDK12* (bi-allelic deletions, genomic amplifications and mutations) in 13% of breast cancers and 5% of ovarian cancers (28–31). *CDK12* mutations are commonly nonsense mutations or impair CDK12 kinase activity (32) and are frequently coupled with loss of heterozygosity (28,33). Recent studies show that CDK12 functions in maintaining genome stability. In cell-based assays and xenograft models, depletion or inhibition of CDK12 is associated with defects in DNA damage response (DDR) and decreases expression of genes involved in the homology-directed repair (HDR) pathway (5,32,34–37). A direct effect of CDK12 on the expression of HDR genes is currently under debate (38). Although the best characterized alterations in CDK12 are mutations that likely disrupt its activity, the most prominent alterations in breast cancers are amplifications. *CDK12* is located on chromosome 17, 165–267 kb proximal to *HER2* (*ERBB2*), an oncogene that is frequently amplified in breast cancers. *CDK12* is co-amplified with *HER2* in 27–92% of breast tumors or tumor cell lines (39–47). Similar to HER2, overexpression of CDK12 also correlates with high proliferative index and grade 3 tumor status based on tissue microarrays of invasive breast carcinomas (48). It is unknown if CDK12 over-expression contributes to the pathogenesis of the tumor, or if it is predominantly a passenger within the *HER2* amplicon. It is also noteworthy that in about 13% of $HER2^+$ (*HER2*-amplified) breast tumors, the amplification breakpoint resides in the *CDK12* allele and likely results in the functional loss of one *CDK12* allele (35). Recurrent *CDK12-HER2* gene fusions in gastric cancers also result in impaired CDK12 protein levels (49). In synthetic lethality studies with *BRCA*-deficient triple-negative breast cancer cells having acquired resistance to poly (ADP-ribose) polymerase (PARP) inhibition, treatment with dinaciclib, a pan-CDK inhibitor used in clinical trials, acts through CDK12 inhibition to re-sensitize these cells to PARP inhibitors (37). However, it is currently unclear how alterations in *CDK12* contribute to the myriad of changes seen in breast tumors.

To address the cellular functions of CDK12, we performed comprehensive and systematic genomic and proteomic analyses of CDK12 function in normal and cancer breast cell lines that include cancer cells with and without genomic amplification of *CDK12*. We sought to determine if a role of CDK12 in tumorigenesis and DDR could involve its hypothesized ability to regulate splicing or AS in addition to its role in transcription. Instead of having a general effect on transcription (27) or splicing, we found that CDK12 regulated the expression and AS of a distinct set of mRNAs in a cell type-specific manner. Furthermore, CDK12 predominantly regulated only the alternative last exon (ALE) subtype of AS. Functionally, events regulated by CDK12 potentiated tumorigenic processes such as cell invasion, suggesting that aberrant CDK12 expression may have oncogenic properties.

MATERIALS AND METHODS

Mammalian cell culture

The SK-BR-3 (ATCC, HTB-30) and MDA-MB-231 (ATCC, HTB-26) cells were provided as a generous gift from Dr M. Bally (British Columbia Cancer Agency) and were independently verified by Short Tandem Repeat (STR) profiling (The Centre for Applied Genomics, SickKids Hospital, Toronto, ON, Canada). The 184-hTERT cell line (L9 clone) was isolated and characterized as previously described (50). SK-BR-3, MDA-MB-231 and 184-hTERT cells were cultured in McCoy's 5A media supplemented with 10% fetal bovine serum (FBS) (Invitrogen), Dulbecco's modified eagle's medium:F12 supplemented with 5% FBS and Mammary Epithelial Cell Growth Media (MEGM) (Lonza, CC-3150), respectively.

Antibodies

The polyclonal rabbit anti-CDK12 antibody was generated using a commercial service (ImmunoPrecise Antibodies Ltd) against a glutathione S-transferase (GST) fusion of the CTD of CDK12 as previously described (11). Other antibodies used were: Anti-β-actin (AB20272, Abcam), anti-ATM (Ataxia Telangiectasia Mutated) (AB2618, Abcam) and anti-DNAJB6 (H00010049-M01, Cedarlane Laboratories).

Quantification of RNA transcripts by qRT-PCR

Cells were harvested in Trizol (Invitrogen, 10296010) and total RNA was isolated using the RNeasy kit (Qiagen, 74106) as per the manufacturer's protocol. Two-step quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR) was performed using the SuperScript VILO cDNA Synthesis (Invitrogen, 11755) and the SYBR Select Master Mix (Applied Biosystems, 4472908) kits, run in 384-well format on the 7900HT Real-time PCR system (Applied Biosystems), according to the manufacturer's protocol. Primers used for qRT-PCR are listed in Supplementary Table S6. Data analyses were performed using the RQ

manager software (version 1.2.2, Applied Biosystems). All samples were normalized to *ACTB* and levels of *TUBA1B* served as a second internal control.

Transient expression of CDK12

Plasmids expressing 3×FLAG-CDK12 (6) or an empty vector control were transfected with polyethyleneimine (PEI, Polysciences, 24765) into SK-BR-3 cells. In brief, cells were grown to ∼80% confluence and transfected with plasmid DNA and PEI at a 1:3 ratio (wt/wt). Cells were harvested 48–72 h post-transfection and used for downstream analyses, including immunoprecipitation-mass spectrometry, western blot analysis and qRT-PCR. For scratch wound assays, transient expression of recombinant plasmids in MDA-MB-231 cells was performed using Lipofectamine LTX (Invitrogen, 15338100) as per the manufacturer's protocol.

Immunoprecipitation

SK-BR-3 cells (2–5 × 10^7 cells per replicate) were transfected with 3×FLAG-CDK12 or an empty vector control. At 72 h post-transfection, cells were harvested and lysed in 5 ml lysis buffer (Tris-buffered saline, pH 7.5, 1 mM ethylenediaminetetraacetic acid (EDTA), 0.1% NP-40, 0.05% wt/vol deoxycholate, 10 mM β-glycerophosphate, 2 mM Na_3VO_4 and Roche cOmplete EDTA-free protease inhibitors) for 30 min at 4°C. Lysed cells were passed through a 21G syringe, clarified by centrifugation at 12 000 x *g* for 10 min at 4°C and mixed with anti-FLAG M2 magnetic beads (Sigma-Aldrich, M8823) overnight at 4°C. Beads were washed three times with lysis buffer. For experiments with the benzonase endonuclease, beads were washed once in lysis buffer and twice with benzonase buffer (50 mM Tris pH 8.0, 20 mM NaCl and 2 mM $MgCl_2$). Beads were resuspended in 100 μl benzonase buffer and 25 units of benzonase (Novagen, 70664) and incubated for 15 min at 25°C. Beads treated with the same buffer and incubation conditions but not exposed to benzonase served as negative controls. To isolate immunoprecipitated proteins, beads were boiled twice, sequentially, in 40 μl elution buffer (50 mM HEPES pH 8.5, 4% sodium dodecyl sulphate (SDS) and 5 mM dithiothreitol (DTT)) for 5 min. Eluted proteins were incubated for 30 min at 45°C followed by alkylation with 1 μl of 400 mM iodoacetamide for 30 min at 25°C. Reactions were quenched by adding 2 μl 200 mM DTT. Immunoprecipitated proteins were identified and quantified by tandem mass tag (TMT) labeling (ThermoFisher Scientific, 90406) and mass spectrometry, as described in Supplementary Methods.

Depletion of genes with siRNA

Sequences for all siRNA constructs used are presented in Supplementary Table S5. CDK12 siRNA-1, CDK12 siRNA-3, CCNK siRNA, CDK9 siRNA and the scrambled control siRNA were previously described (6). CDK12 siRNA-2 (Dharmacon, M-004031-03-0020) is composed of four unique siRNA constructs. CDK13 Stealth siRNA (Invitrogen) was designed against the 3′ untranslated region (UTR) of the gene. Different methods were used for transfecting siRNA into the different cell types to achieve sufficient depletion (>70%) at the protein level. SK-BR-3 cells were transfected sequentially three times with CDK12 siRNA using Lipofectamine 2000 (Invitrogen, 11668019) as per the manufacturer's protocol over the course of 11 days to achieve a sufficient decrease in CDK12 protein expression. The scrambled control siRNA was likewise transfected. MDA-MB-231 cells were reverse transfected sequentially two times with CDK12 siRNA over the course of 7 days using Lipofectamine RNAiMax (Invitrogen) according to the manufacturer's protocol. 184-hTERT cells were transfected once with CDK12 siRNA using Lipofectamine 2000. Quantification of CDK12 protein depletion was determined by western blot (Supplementary Figure S2).

Global proteome profiling

Biological triplicates of SK-BR-3 and MDA-MB-231 cells treated with CDK12 siRNA-1 or scrambled siRNA (∼2 × 10^6 cells per replicate) were harvested and lysed in 100 μl SDS buffer (200 mM HEPES pH 8.5, 1% SDS, Roche cOmplete EDTA-free protease inhibitors) for 5 min at 95°C. Twenty-five units of benzonase (Novagen) were added and the reaction was incubated for 37°C for 30 min. Reduction (5 μl of 200 mM DTT, 45°C for 30 min) and alkylation (10 μl of 400 mM iodoacetamide, 25°C for 30 min) of proteins was subsequently carried out. Reactions were quenched by adding 5 μl 200 mM DTT. Samples were prepared for trypsin digestion using the SP3 protein cleanup protocol as previously described (51) and labeled with TMT 10-plex kits (ThermoFisher Scientific, 90406). Analyses of TMT-labeled peptides were performed on an Orbitrap Fusion Tribrid Mass Spectrometer (Thermo Scientific). Mass spectrometry and data analyses are further described in Supplementary Methods.

RNA-seq

Library construction was performed on 4–5 μg of total RNA (RIN ≥ 9.0 from Agilent 6000 Nano analysis, Agilent Technologies) using the ssRNA-seq pipeline for Poly(A)-purified mRNA libraries at the Michael Smith Genome Sciences Centre (52). Biological triplicates of the mRNA libraries were sequenced on a Hi-Seq 2500 (Illumina) using 75 base paired-end sequencing. Analyses for differential gene expression and AS were performed with DESeq2 (53) and MISO (54) as described in Supplementary Methods.

RESULTS

CDK12 interacts with the splicing machinery

To investigate functional properties of CDK12, we identified proteins that it interacts with by performing immunoprecipitation and mass spectrometry on SK-BR-3 cells transfected with FLAG-tagged CDK12 (Figure 1A). SK-BR-3 cells are a $HER2^+$ epithelial breast cancer cell line where the *CDK12* gene is co-amplified with *HER2* and the CDK12 protein is over-expressed (35). The CDK12-interacting proteins were highly enriched for RNA splicing function (Figure 1B) (55) and could be generally classified into core spliceosome components (pre-catalytic complexes A and B, and the associated Prp19 complex) and

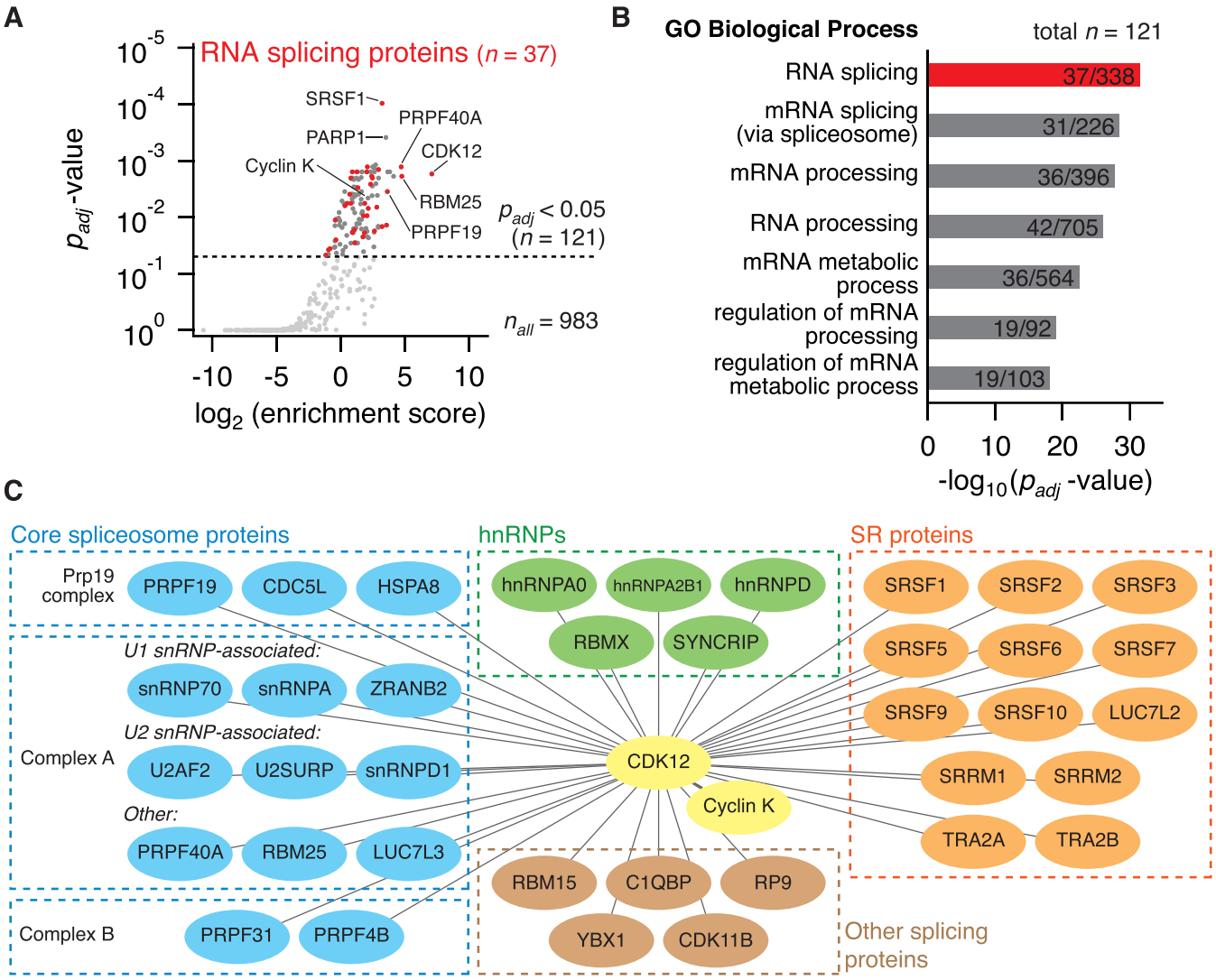

Figure 1. CDK12 interacts with the RNA splicing machinery. (**A**) Immunoprecipitation of FLAG-CDK12 and mass spectrometry was used to identify 121 CDK12-interacting proteins in SK-BR-3 cells (enrichment score > 0, p_{adj} < 0.05, Supplementary Table S1). (**B**) Interacting proteins were highly enriched for RNA splicing functions as determined by gene ontology analysis (55). (**C**) CDK12-interacting splicing proteins can be generally divided into core spliceosome proteins (blue) and regulatory splicing factors (green, orange and brown).

regulators of constitutive and AS (SR proteins, RBM proteins and hnRNPs) (Figure 1C and Supplementary Table S1) (20). The interactions between CDK12 and hnRNPs were sensitive to nuclease treatment (Supplementary Figure S1) and were therefore likely dependent on RNA intermediates, such as the pre-mRNA upon which hnRNPs are assembled. By contrast, interactions between CDK12 and core spliceosome and SR proteins were largely unaffected by nuclease treatment. The universality of interactions between CDK12 and core spliceosome components was further supported by immunoprecipitation experiments in HEK-293T cells (27,56), Jurkat T-cells (57) and HeLa cells (58,59); however, many of the regulatory splicing components differ across cell types. This could be a product of cell type-specific regulation or differences in experimental methodology. Together, these results suggest that CDK12 is a *bona fide* component of the splicing machinery.

CDK12 regulates alternative last exon splicing of genes with long transcripts and many exons

To explore the function of CDK12 in normal splicing regulation and in splicing misregulation in breast cancer, we performed mRNA sequencing (RNA-seq) on three breast cell lines: a HER2$^+$ cancer cell line with *CDK12* amplification (SK-BR-3), a triple-negative breast cancer cell line (MDA-MB-231) and an immortalized normal mammary epithelial cell line (184-hTERT). Cells were treated with a scrambled siRNA control or siRNA directed to CDK12 (CDK12 siRNA-1, Supplementary Figure S2). The RNA-seq was performed on three independent pairs of CDK12 siRNA:scrambled siRNA samples for each cell line (Supplementary Figure S3), with 103 ± 11 million reads per sample, to enable the identification of low level AS events. To identify changes in RNA splicing events, we used the MISO package (54), which applies a statistical framework to dis-

tinguish eight different types of annotated AS events in pairwise RNA-seq comparisons. We identified 102 AS events common to all SK-BR-3 samples, 724 AS events common to all MDA-MB-231 samples and 86 AS events common to all 184-hTERT samples (Figure 2A). The regulation of specific AS events by CDK12 was cell type-specific and only 22 AS events were common to all three cell lines (Figure 2B). However, the mechanism of regulation appears conserved: 86, 61 and 79% of AS events observed in CDK12-depleted SK-BR-3, MDA-MB-231 and 184-hTERT cells, respectively, were ALE splicing. Furthermore, 92% of AS events common to two or more cell lines and 100% of AS events common to all three cell lines were ALE events (Figure 2B). ALE events regulated by CDK12 had an average MISO $|\Delta\Psi|$ value of 0.27 ± 0.13 (range 0.10–0.72; Figure 2C) and the regulated genes were highly expressed with average FPKM (fragments per kb of exon per million fragments mapped) values of 20, 27 and 24 in SK-BR-3, MDA-MB-231 and 184-hTERT cells, respectively (Supplementary Figure S4A). The cell type-specific AS effects we observed were likely not an indirect result of low gene expression, as genes with CDK12-regulated AS events in one cell type, but not in the other two cell types, had similar overall expression across all three cell types (Supplementary Figure S4B). On a technical note, we observed that biological replicates for the RNA-seq analysis greatly increased the confidence of identified AS events associated with CDK12 depletion (Supplementary Figure S3). For example, in SK-BR-3 cells, ALE events represented 41, 81 and 86% of all AS events ($n = 819$, 202 and 102) after one, two and three replicates respectively. To further explore the universality of ALE regulation by CDK12, we performed MISO analysis on published RNA-seq data of HCT-116 (colorectal cancer) cells treated with CDK12 shRNAs (27). The experiments in HCT-116 were performed with two different shRNA constructs in duplicates. Consistent with our findings in breast cell lines, ALE events accounted for 33 and 41% of all AS types in HCT-116 cells for each of the two shRNAs, respectively (Supplementary Figure S5A). All eight AS events common to the four cell lines were all ALEs (Supplementary Figure S5B; CDK12 depleted by siRNA-1 (SK-BR-3, MDA-MB-231 and 184-hTERT cells) and either of the two shRNAs (HCT-116)).

The regulation of AS by CDK12 is largely cell type-specific, but the preponderance of ALE events suggests the regulated genes may possess common features. In 82% of all identified ALE events, CDK12 depletion resulted in the enrichment of mRNA isoforms utilizing the proximal ALE (Figure 3A). These results were independently validated by performing qRT-PCR on a select number of ALE events ($n = 19$) in cells depleted of CDK12; there was high correlation of $\Delta\Psi$ values between the MISO and qRT-PCR data (Supplementary Figure S6A). These observations were also not due to off-target effects; we obtained similar results with a different CDK12 siRNA construct (CDK12 siRNA-2; Supplementary Figure S6B), but not with siRNA constructs targeting CDK9 or CDK13 (Supplementary Figure S6C).

It was previously reported that genes transcriptionally regulated by CDK12 generally had longer transcripts (5). In our analysis, we found that the pre-mRNA transcripts

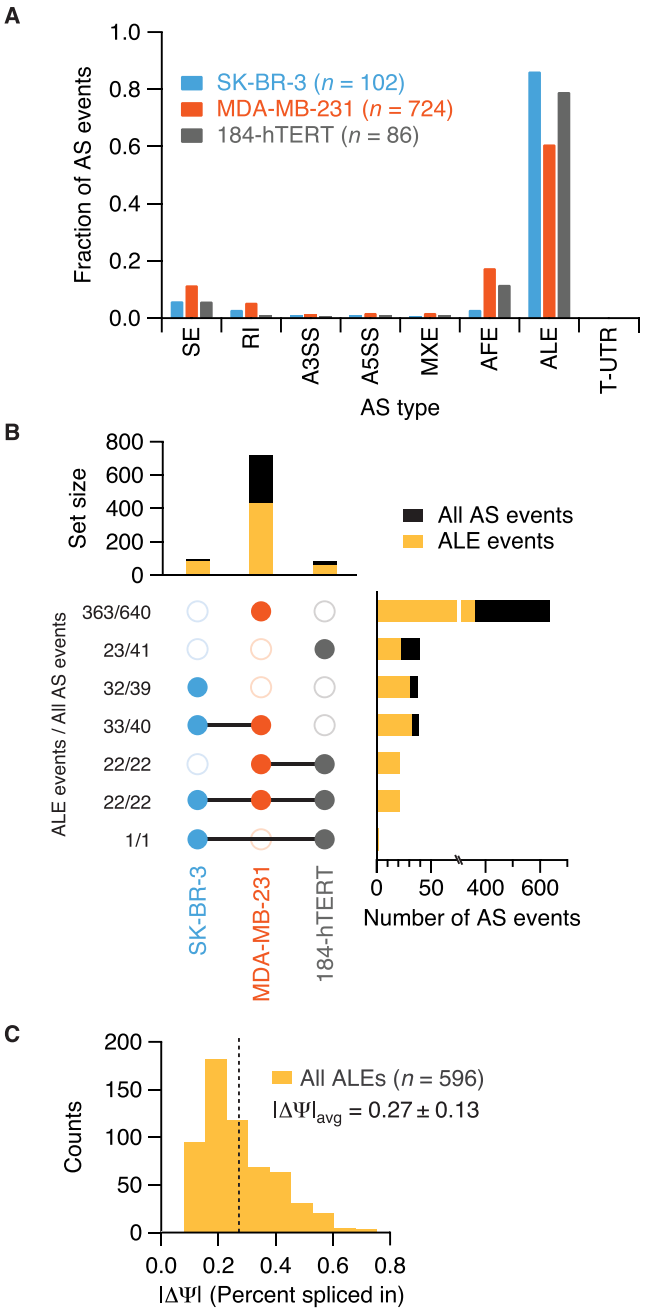

Figure 2. CDK12 regulates ALE splicing. (**A**) MISO analysis identified AS events that resulted from depletion of CDK12 in SK-BR-3, MDA-MB-231 and 184-hTERT cells (Bayes Factor ≥ 20, $|\Delta\Psi| \geq 0.1$, present in all three RNA-seq replicates). SE, skipped exons; RI, retained introns; A3SS, alternative 3′ splice sites; A5SS, alternative 5′ splice sites; MXE, mutually exclusive exons; AFE, alternative first exons; ALE, alternative last exons; T-UTR (untranslated region), tandem 3′ UTR. Three biological RNA-seq replicates assisted in the identification of ALEs as the predominate AS event in all three cell types (Supplementary Figure S3). (**B**) The majority of AS events (black bars) are cell type-specific and events common to SK-BR-3, MDA-MB-231 and 184-hTERT cells are all ALEs (orange bars). (**C**) Distribution of $|\Delta\Psi|$ values for ALE events (total $n = 596$) regulated by CDK12 in SK-BR-3 ($n = 88$), MDA-MB-231 ($n = 440$) and 184-hTERT ($n = 68$) cells. The mean $|\Delta\Psi|$, as denoted by the dotted vertical line, is 0.27 ± 0.13 s.d.

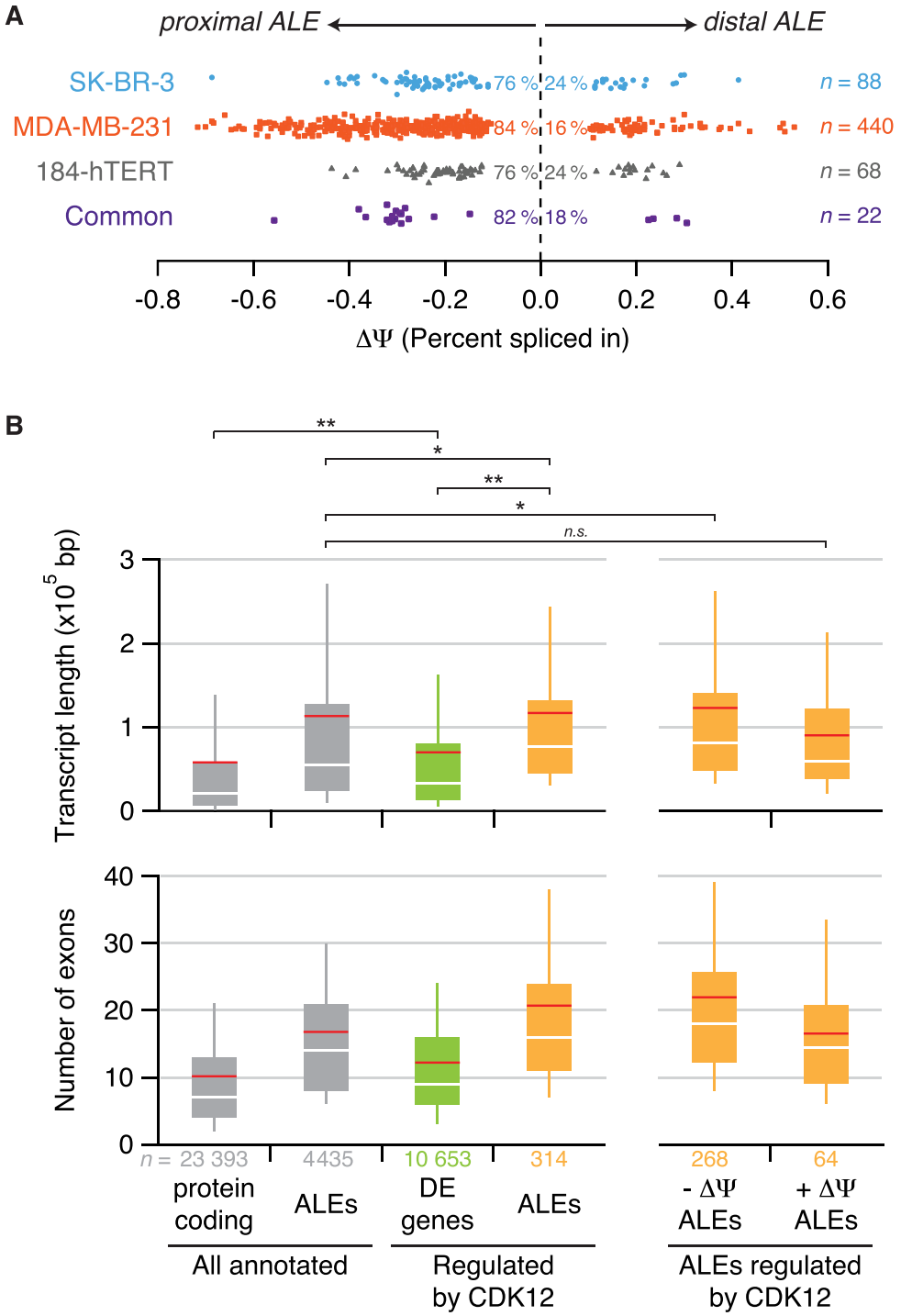

Figure 3. CDK12 regulates ALE splicing of genes with long transcripts and a large number of exons. (**A**) Depletion of CDK12 generally results in the utilization of proximal ALEs ($-\Delta\Psi$ values). (**B**) Distributions of gene pre-mRNA transcript length and number of exons. All protein coding genes ($n = 23\,393$) and genes with annotated ALE events ($n = 4435$) are compared to genes regulated by CDK12 via differential expression (DE, $n = 10\,653$) or ALE splicing ($n = 314$) in SK-BR-3, MDA-MB-231 or 184-hTERT cells. Genes regulated by ALE splicing can be further subdivided into those that utilize the proximal ALE ($-\Delta\Psi$, $n = 268$) or distal ALE ($+\Delta\Psi$, $n = 64$) after CDK12 depletion. Box plots denote the 10th, 25th, 50th, 75th and 90th percentiles. Red lines represent the means. Pairwise statistical comparisons performed using the Kolmogorov–Smirnov test (*$P < 1 \times 10^{-6}$, **$P \approx 0$, *n.s.* not significant), and apply to both plots describing transcript length and number of exons, respectively.

of genes with ALE events regulated by CDK12 were significantly longer and had more exons than those transcriptionally regulated by CDK12 (Figure 3B). Genes with ALE events regulated by CDK12 also had longer transcripts and more exons than those from the total set of annotated ALEs (Figure 3B). This trend, however, was only observed for genes with greater utilization of the proximal ALE after CDK12 depletion (negative $\Delta\Psi$ values) and not for genes with positive $\Delta\Psi$ values after CDK12 depletion. Therefore, there is a correlation between pre-mRNA transcript length and a requirement of CDK12 to form longer transcripts by ALE splicing. Notably, only a small proportion of genes with long transcripts were regulated by CDK12. When considering all genes with annotated ALEs, only ~3.5% with transcripts longer than the average were regulated by CDK12 (2.5, 6.9 and 1.3% in SK-BR-3, MDA-MB-231 and 184-hTERT cells; Supplementary Figure S7). In other words, only a small subset of genes with long transcripts was regulated by CDK12, suggesting additional gene-specific factors direct AS by CDK12. Taken together, our results suggest that CDK12 associates with core spliceosome components and regulates ALE splicing of long transcripts in multiple cell types. Furthermore, native expression of CDK12 likely promotes the usage of distal ALEs, which largely correspond to longer mRNA transcripts.

Proximal ALEs regulated by CDK12 contain higher densities of polyadenylation motifs

While the regulation of ALE usage by CDK12 can be achieved through its association with regulatory splicing factors, it could also be influenced by transcription termination processes (such as alternative polyadenylation) initiated by termination signals in the 3′ UTRs (60). To address this possibility, we searched for polyadenylation motifs in the 3′ UTRs of proximal and distal ALEs that were regulated by CDK12 (Figure 4). We observed that the density of polyadenylation motifs was slightly increased in the 3′ UTR of proximal ALEs regulated by CDK12, as compared to control ALEs not regulated by CDK12. If polyadenylation signals were the sole factor directing ALE splicing by CDK12, polyadenylation motifs should be enriched in the proximal ALE 3′ UTR of all genes with negative $\Delta\Psi$ values and not in any gene with a positive $\Delta\Psi$ value. However, the distributions of polyadenylation motif density were broad and there was also a slight increase in the density of polyadenylation motifs in the proximal ALE 3′ UTR of genes with positive $\Delta\Psi$ values. This observation was made with less statistical confidence, likely due to the smaller number of ALEs with positive $\Delta\Psi$ values, as compared to negative $\Delta\Psi$ values. While gene-specific recruitment of polyadenylation factors may be involved, the regulation of ALEs by CDK12 likely involves additional mechanisms.

Tumors defective in CDK12 function exhibit misregulation of ALE splicing

Alterations in *CDK12* have been described in numerous tumor types, including breast, ovarian, uterine, prostate, gastric and bladder cancers (29,30,35,47,49,61). The TCGA consortium has performed large-scale analyses on collections of tumor samples, including RNA-seq for 311 cases

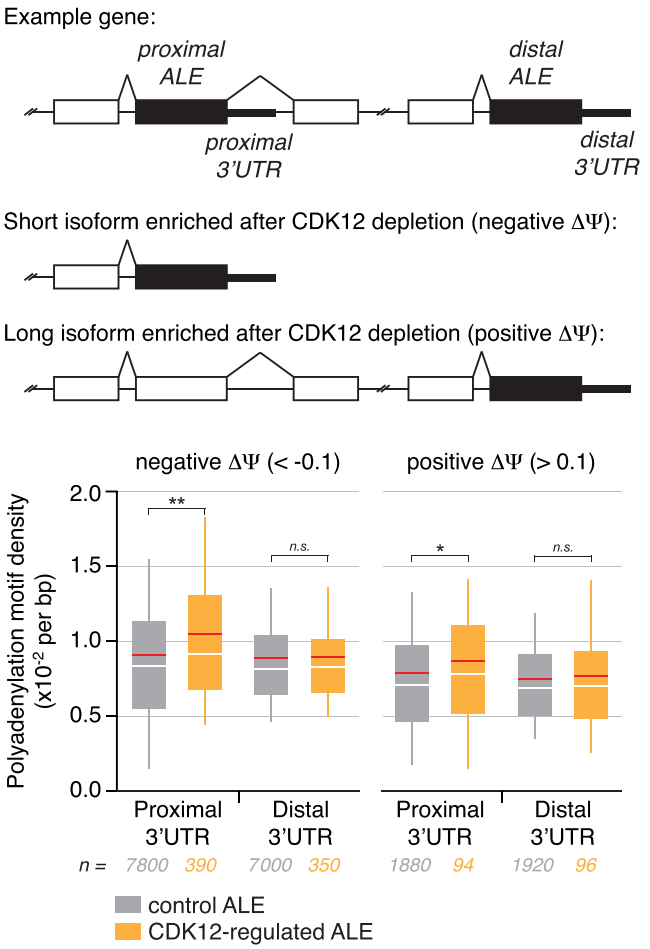

Figure 4. The 3′ UTRs of proximal ALEs regulated by CDK12 contain higher densities of polyadenylation motifs. Distributions of the densities of polyadenylation motifs in the 3′ UTRs of proximal and distal ALEs regulated by CDK12, as compared to control ALEs not regulated by CDK12 (detailed in Supplementary Methods). ALE events are divided into those that result in greater usage of the proximal ALE ($\Delta\Psi < -0.1$) and those that favor the distal ALE ($\Delta\Psi > 0.1$) after CDK12 depletion. Box plots denote the 10th, 25th, 50th, 75th and 90th percentile. Red lines denote the means. $^{**}P = 7 \times 10^{-7}$, $^{*}P = 0.03$, *n.s.* not significant (Mann–Whitney *U* test, Benjamini–Hochberg corrected).

of ovarian serous cystadenocarcinoma (29). *CDK12* is recurrently altered in 6% of these cases (Figure 5A and Supplementary Table S2). Tumors containing the *CDK12* mutations are notably not amplified for *HER2* and previous studies demonstrated that these ovarian cancer mutations impair the kinase activity of CDK12 *in vitro* (32,36). Therefore, these samples are well suited to explore the changes in AS as a consequence of modulating CDK12 function in tumors.

To analyze the regulation of ALE events by CDK12 in TCGA tumor samples, we used the MISO package to perform pairwise comparisons of tumor samples containing CDK12 alterations to tumor samples without CDK12 alterations (Figure 5B). For this analysis, we utilized data from four types of available TCGA RNA-seq samples (29): tumors with *CDK12* mutations ($n = 7$), tumors with biallelic *CDK12* deletions ($n = 3$), tumors with genomic am-

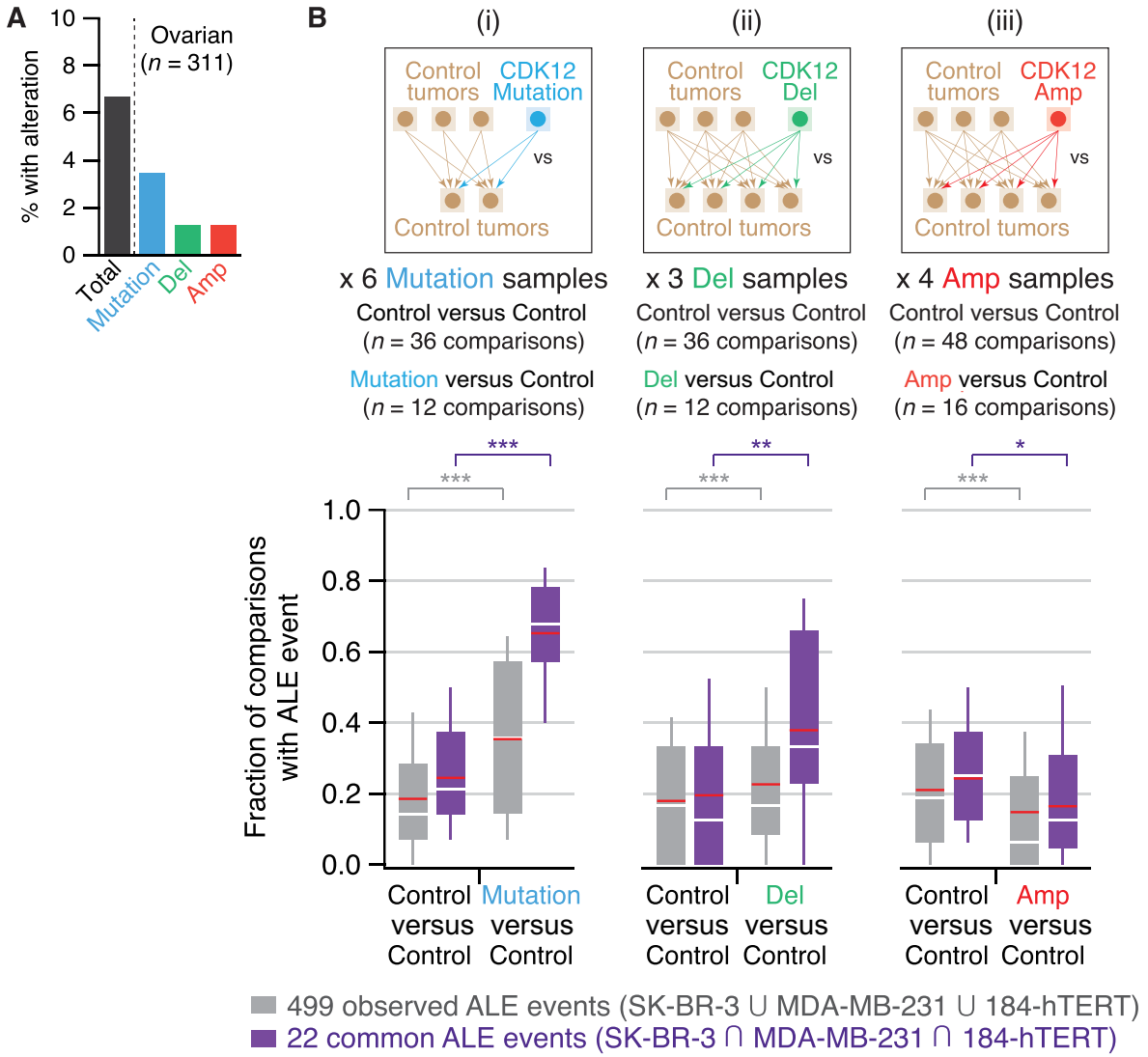

Figure 5. Alterations in *CDK12* correlate with misregulation of ALE splicing in ovarian tumor samples. (**A**) *CDK12* is recurrently altered in ovarian serous cystadenocarcinomas (29). From this dataset, RNA-seq data were available for tumors containing *CDK12* mutations (blue, $n = 7$; coupled to loss of heterozygosity, $n = 6$), bi-allelic deletions (green, $n = 3$) and amplifications (red, $n = 4$). (**B**) Using the MISO package, changes in AS (Bayes Factor ≥ 20) were determined based on the following comparisons: (i) *CDK12* mutation versus control, (ii) *CDK12* deletion versus control and (iii) *CDK12* amplification versus control. Changes in CDK12-regulated AS events were compared to AS events found in control versus control comparisons. To obtain a similar number of comparisons in each scenario, each mutation sample (i) was compared to two unique control samples ($n = 12$ comparisons), while each deletion (ii) and amplification sample (iii) was compared to four unique control samples ($n = 12$ and 16 comparisons, respectively). Control versus control comparisons were likewise paired and performed in triplicate ($n = 36$, 36 or 48 comparisons). A total of 499 ALE events were queried, representing the aggregate of events found in the SK-BR-3, MDA-MB-231 and 184-hTERT experiments (gray boxes, SK-BR-3 ∪ MDA-MB-231 ∪ 184-hTERT). We also queried 22 ALE events common to all three cell lines (purple boxes; SK-BR-3 ∩ MDA-MB-231 ∩ 184-hTERT). Box plots denote the 10th, 25th, 50th, 75th and 90th percentiles. Red lines represent the means. The significances of comparisons (SK-BR-3 ∪ MDA-MB-231 ∪ 184-hTERT, gray lines; SK-BR-3 ∩ MDA-MB-231 ∩ 184-hTERT, purple lines) were determined using the Mann–Whitney *U* test (*$P < 0.05$, **$P < 0.005$, ***$P < 1 \times 10^{-5}$).

plification of *CDK12* ($n = 4$) and tumors with no alterations in *CDK12* ($n = 56$ control samples). Six of the seven cases of *CDK12* mutations were coupled to loss of heterozygosity. We queried the mutation, deletion, amplification and control samples for the occurrence of the 499 aggregate ALE events that resulted from CDK12 depletion in SK-BR-3, MDA-MB-231 and 184-hTERT cells (Supplementary Table S3). Each ALE event in *CDK12*-mutated tumors (with loss of heterozygosity) was found in 35% of comparisons on average (mutation:control), as compared to 19% of control (control:control) comparisons ($P < 1 \times 10^{-5;}$ Figure 5B (i)). When considering only the 22 events common to all three cell lines, each ALE event was found in 65 and 25% of mutation and control comparisons on average, respectively ($P < 1 \times 10^{-5}$). Similar trends were obtained with tumors containing bi-allelic *CDK12* deletions (Figure 5B (ii)); each of the 22 ALE events common to all three cell lines were found in 38 and 20% of deletion and control comparisons on average, respectively ($P < 0.005$). The deletion:control comparison resulted in a smaller difference than the mutation:control possibly because of remaining *CDK12* copies after a bi-allelic deletion in a polyploid background. Nev-

ertheless, both sets of comparisons demonstrate that these ALE events occurred more frequently in tumors impaired in CDK12 function.

In breast cancers, *CDK12* is commonly co-amplified with *HER2*. Similarly, the four ovarian tumor samples with *CDK12* amplifications also contain *HER2* amplifications. Unlike cases containing *CDK12* mutations or deletions, the queried ALE events were found less frequently in tumors amplified for *CDK12* (15% of amplification:control and 21% of control:control comparisons, $P < 1 \times 10^{-5}$; Figure 5B (iii)). These observations with the *CDK12*-amplified tumor samples mirror our results in SK-BR-3 cells, where the ALE events were identified after depletion of CDK12 from an over-expressed state. Together, these results suggest that misregulation of ALE splicing occurs due to aberrations in *CDK12* and support a functional role of *CDK12* alterations in tumor development in ovarian tumors.

Regulation of gene expression by CDK12 is gene- and cell type-specific but modulates a core set of common pathways

The regulation of ALE splicing by CDK12 is both gene- and cell type-specific, and only a small subset of regulated genes are common to multiple cell types. We therefore evaluated the effects of CDK12 on global gene expression to determine if its regulation of transcription was also gene- and cell type-specific. We analyzed the triplicate CDK12 siRNA and control siRNA RNA-seq data from SK-BR-3, MDA-MB-231 and 184-hTERT cells using DESeq2 (53,62), a program that utilizes replicate data to establish high confidence identification of differential gene expression. The analysis found that depletion of CDK12 resulted in small to moderate changes in gene expression (Figure 6A), affecting 3163, 10 245 and 3940 genes ($p_{adj} < 0.01$) in SK-BR-3, MDA-MB-231 and 184-hTERT cells, respectively. These events were generally evenly divided into upregulated and downregulated genes in all three cell types. Of these events, only 386, 3699 and 671 exhibited more than a 2-fold change in gene expression in SK-BR-3, MDA-MB-231 and 184-hTERT cells, respectively (Figure 6A and B). A previous microarray study in HeLa cells found that after Cyclin K depletion, 3.9% of genes were downregulated and 2.6% were upregulated (5). Combined with our findings, these results contrast with a study in HCT-116 cells, which reported that 98% of differentially expressed genes were downregulated after CDK12 depletion (27). In general, we observed very little overlap between the genes regulated by each cell type, with only 23 differentially expressed genes common to all three cell types (Figure 6B). Taken together, our observations suggest that similar to the regulation of ALE splicing, regulation of gene expression by CDK12 is highly gene- and cell type-specific.

While the regulation of individual genes by CDK12 differed across the three cell lines, an examination of the affected cellular pathways offered additional insight. Using Gene Set Enrichment Analysis (GSEA) (63), we found that in all three cell lines, loss of CDK12 downregulated similar pathways (Figure 6C). These include pathways involved in the cell cycle, DNA replication and repair and RNA processing and splicing. In general, these pathways support previously reported functions of CDK12 (5,27,32,34–36,64,65). Since these processes were previously identified in different cell types, they appear to represent universal functions of CDK12. The pathway analysis also aided in determining cell type-specific properties of CDK12. For example, depletion of CDK12 in SK-BR-3 cells decreased expression of genes associated with mitochondrial function (Figure 6C). This trend was not observed in MDA-MB-231 or 184-hTERT cells. Instead, depletion of CDK12 in MDA-MB-231 cells upregulated pathways involved in translation. Depletion of CDK12 in 184-hTERT cells upregulated pathways associated with the plasma membrane, development and the extracellular matrix (Figure 6C). Taken together, these results demonstrate that while transcriptional regulation by CDK12 is largely gene- and cell type-specific, some cellular processes are commonly modulated by CDK12 activity in different cell types.

We next sought to determine how changes in gene expression due to CDK12 function manifest at the protein level to affect the expressed cellular phenotype. We applied a global proteomics approach to quantify alterations in protein expression after depletion of CDK12 in SK-BR-3 cells. Similar to the transcriptome data, only a small proportion of proteins were differentially expressed ($n = 444$, $p_{adj} < 0.01$) after depletion of CDK12 (Figure 7A). Differentially expressed proteins were both upregulated (61%, mean fold change = 1.3) and downregulated (39%, mean fold change = −1.3). When compared to the matching RNA-seq data, we found that the proteome data represented a smaller subset of the transcriptome data (Figure 7B). Of the 11 072 expressed genes in the RNA-seq data (defined as FPKM ≥ 1), 7031 (64%) were identified at the protein level by mass spectrometry (Figure 7B). These 7031 proteins represent almost all of the 7651 total identified proteins (92%) in the proteomic analysis. There was a high correlation ($r^2 = 0.88$) in the fold change values of the 197 genes that were differentially expressed in a statistically significant manner in both the transcriptome and proteome datasets (Figure 7C). We note that 242 genes were significantly changed at the protein level and not at the mRNA level, and that 1136 mRNAs were significantly changed at the transcriptome level and not at the protein level. Pathway analyses demonstrated that the core functions of CDK12 (e.g. RNA processing and DDR) were all observed in the proteomics experiment (Figure 7D). Functions specific to SK-BR-3 cells, such as the involvement of mitochondrial processes, were also found at the protein level. However, the regulation of proteins involved in the cell cycle, which was prominent in the transcriptome data, was not significantly enriched in the proteome data. These results could reflect additional layers of regulation at the protein level, including the modulation of translation, post-translational modifications and protein turnover/proteolysis. An additional factor to explain this observation could be a dominant effect of HER2 over-expression on many pathways (66). Consistent with this idea, loss of CDK12 significantly downregulates cell cycle and cell division proteins in MDA-MB-231 cells, which do not have *HER2* amplification (Supplementary Figure S8).

Combined with the immunoprecipitation interactome experiments and AS RNA-seq analyses, these results establish a function of CDK12 in regulating splicing and modulating

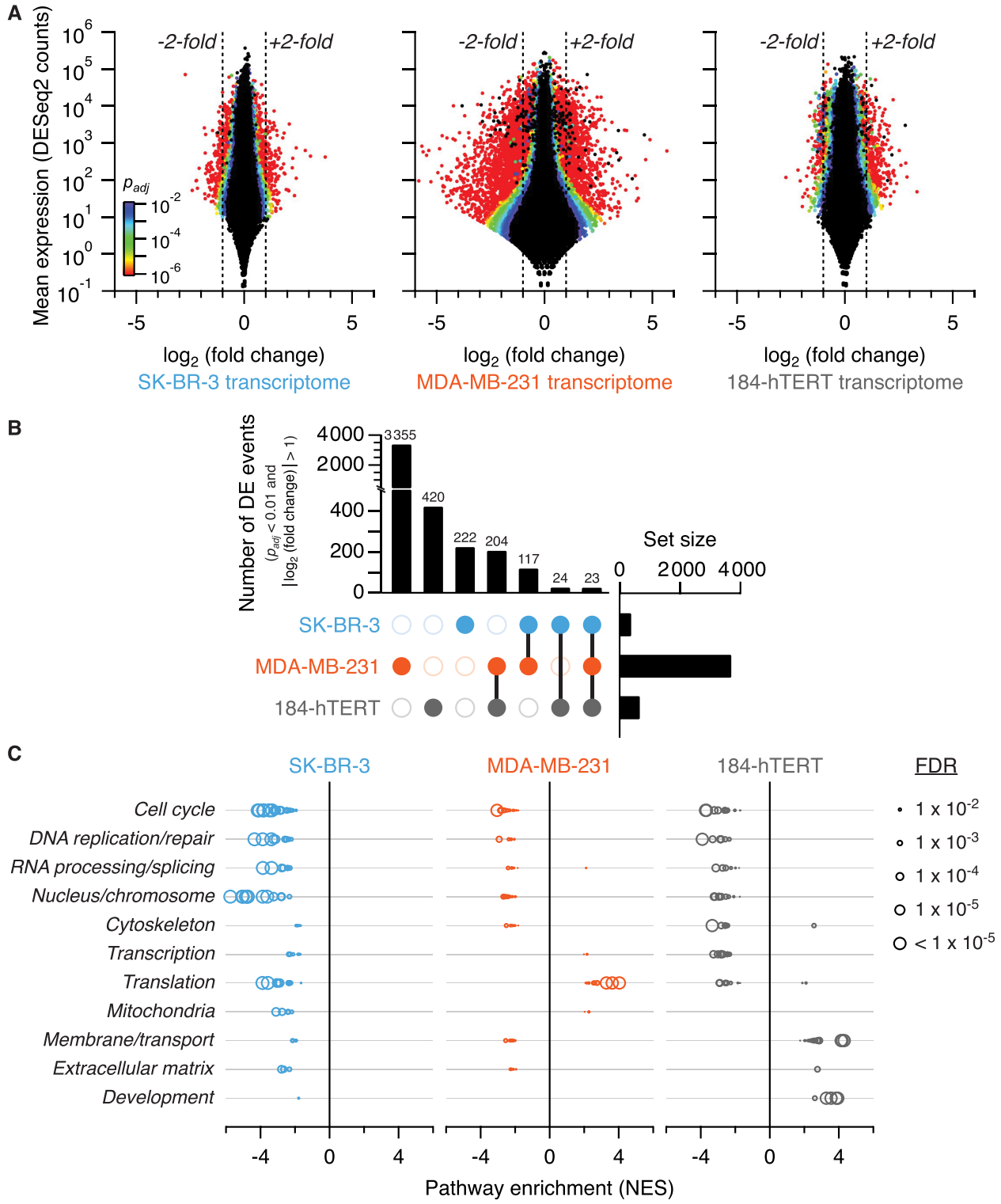

Figure 6. CDK12 differentially regulates gene expression in a cell type-specific manner, but affects a core set of genes and pathways. (**A**) Differential gene expression analysis by RNA-seq following CDK12 depletion in SK-BR-3 (left), MDA-MB-231 (middle) and 184-hTERT (right) cells. Mean expression (DESeq2 counts) is plotted against fold change (CDK12 siRNA-1 versus scrambled siRNA). Black dotted lines delineate events with $|\log_2$ (fold change)$| > 1$. Events with $p_{adj} < 0.01$ are colored. (**B**) Intersection set analysis showing that few differential gene expression events with $p_{adj} < 0.01$ and $|\log_2$ (fold change)$| > 1$ are common between the three cell lines. (**C**) Gene set enrichment analysis (GSEA) of differential gene expression (DE) resulting from CDK12 depletion in SK-BR-3, MDA-MB-231 and 184-hTERT cells (detailed in Supplementary Methods). For each pathway, a normalized enrichment score (NES) represents the extent of over-representation of genes of that pathway at the top or bottom of a ranked list. Positive and negative NES values represent up- and downregulated pathways after CDK12 depletion, respectively. Pathways are organized into 11 categories based on clustering of gene sets and general biological function (Supplementary Table S7). Sizes of markers represent the false discovery rate (FDR) values and only pathways where FDR < 0.1 are shown.

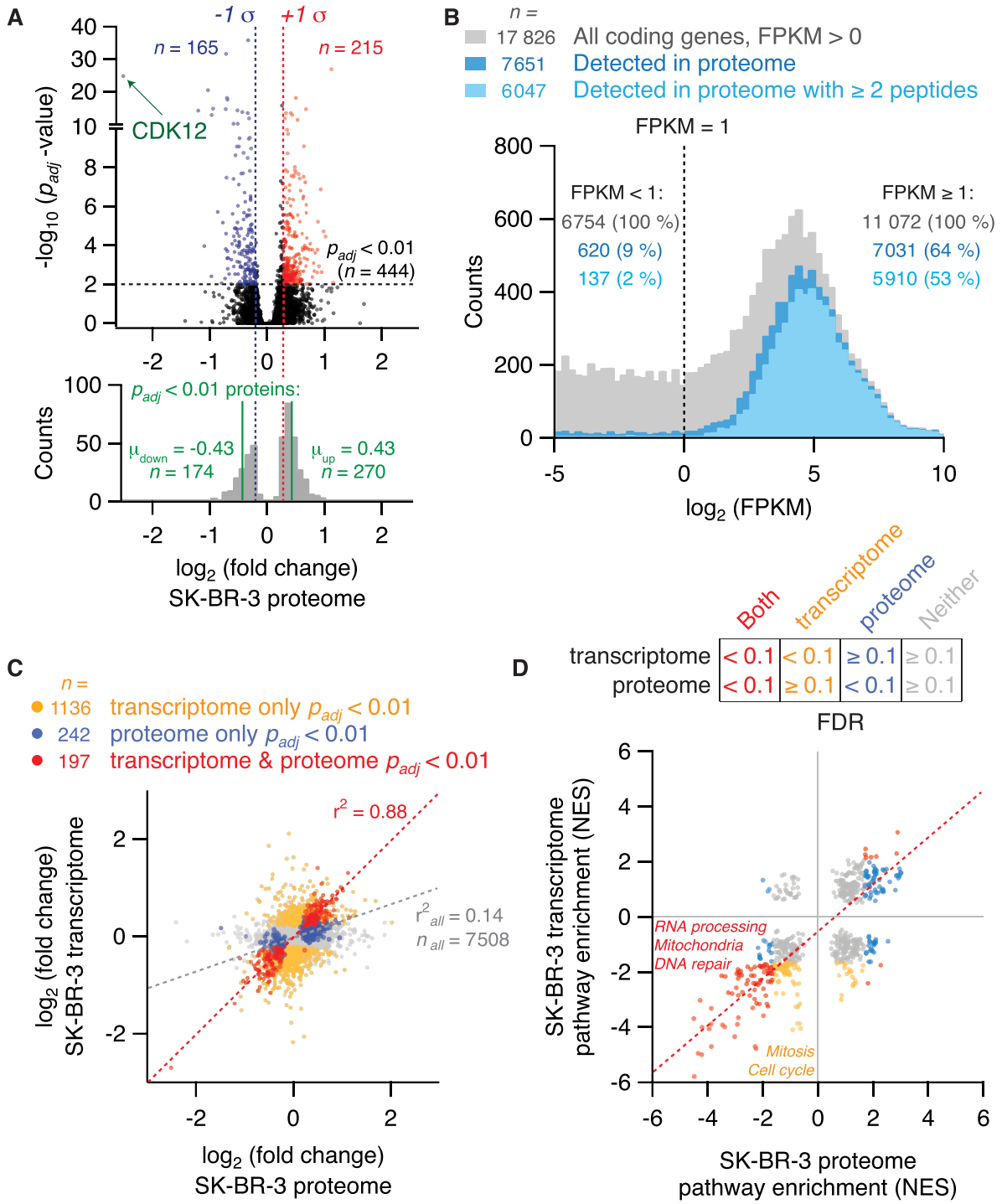

Figure 7. Differential protein expression due to CDK12 regulation represents a subset of differential gene expression events. (**A**) Top: volcano plot of the global proteome analysis in SK-BR-3 cells ($n_{all} = 6119$ proteins detected by ≥ 2 peptides). Dotted horizontal line denotes point at which $p_{adj} = 0.01$. Dotted vertical lines lineate events with |fold change| > 1 s.d. (σ) from the mean. Bottom: distribution of fold change values for all differential protein expression events with $p_{adj} < 0.01$. Green vertical lines denote mean $\log_2$ (fold change) values (μ) for up and downregulation. Dotted lines are the ± 1 σ lines extended from the top plot. (**B**) Histogram of RNA-seq expression values (fragments per kb of exon per million fragments mapped (FPKM)) for all coding genes and genes with corresponding proteins detected by mass spectrometry with ≥ 1 unique peptides (dark blue bars) or ≥ 2 unique peptides (light blue bars). 64% of proteins detected by global proteome analysis in SK-BR-3 cells had corresponding transcripts that were detected (as defined by FPKM >1) in the RNA-seq data. (**C**) Correlation of fold change values from global transcriptome and proteome analysis in SK-BR-3 cells ($r^2_{all} = 0.14$, $P < 10^{-5}$). Events with significant fold change values ($p_{adj} < 0.01$) in both datasets are shown in red ($r^2 = 0.88$, $P < 10^{-5}$). Events significant only in the transcriptome and proteome are colored yellow and blue, respectively. (**D**) GSEA pre-ranked analysis assigned a NES representing the extent of over-representation of genes of a pathway at the top or bottom of a ranked list. Positive and negative NES values represent up- and downregulated pathways, respectively. For each pathway, NES values in the SK-BR-3 transcriptome and proteome are shown. Red markers represent NES values significant in both datasets (FDR < 0.1). The dotted red line shows the general trend of these points. Blue and yellow markers represent NES values only significant in proteome and transcriptome, respectively.

core cellular processes such as the DDR. We further showed that CDK12 can affect cell type-specific pathways, but not all cellular processes identified as regulated at the mRNA level are translated into expressed phenotypes.

CDK12 can modulate the expression of DNA damage response genes through multiple mechanisms

One of the most consistently reported functions of CDK12 has been the regulation of the DDR. Differential expression of specific DDR genes was first identified by microarray analysis (5) and changes in DDR pathways were determined from transcriptome analysis (27). Furthermore, CDK12 depletion or inhibition was found to be synthetic lethal with PARP inhibition (32,34,35,37). This behavior is reminiscent of the sensitivity of *BRCA1/BRCA2*-deficient tumors to PARP inhibitors (67–69), suggesting that similar to BRCA1/BRCA2, CDK12 may be specifically involved in the HDR pathway. Indeed, ovarian tumors containing *CDK12* mutations exhibited downregulation of several HDR genes (36). In our analyses, there were 10 DNA repair proteins (MDC1, LIG1, MCM7, PARP1, SFN, HMGB2, XRCC6, TDP1, XAB2, HMGB1) that were significantly downregulated in both the SK-BR-3 and MDA-MB-231 proteome data, and many more that were regulated in a cell type-specific manner (Supplementary Table S4). Furthermore, our AS data suggest that ALE splicing may be a significant mechanism of regulation by CDK12, especially for genes with long transcripts and many exons. One such example we identified was the gene encoding the ATM protein, a key regulatory kinase that responds to DNA double-strand breaks and initiates the HDR pathway (70). The canonical ATM isoform is a 350 kDa protein translated from a 13 147-bp transcript containing 63 exons (Supplementary Figure S9A). Depletion of CDK12 in SK-BR-3 and MDA-MB-231 cells resulted in an increased usage of a proximal ALE, corresponding to the 32nd exon of the canonical isoform ($\Delta\Psi = -0.25$ and -0.49, respectively). Using a monoclonal antibody targeting ATM residues 980–1512 (exons 20–30), we found that full-length ATM protein was decreased 3-fold after CDK12 depletion in SK-BR-3 cells (Supplementary Figure S9B). While these data suggest that the expression of *ATM* could be regulated through AS, further experiments will be required to determine if this decrease occurs primarily by this mechanism. In 184-hTERT cells, however, ATM was not significantly regulated by ALE splicing; instead, treatment of 184-hTERT cells with CDK12 siRNA-1 resulted in a modest 1.5-fold transcriptional downregulation of *ATM* mRNA ($p_{adj} = 4 \times 10^{-5}$). While ALE splicing of *ATM* was cell type-specific, the regulation of *INTS6* ALE splicing by CDK12 was common to all cell lines we examined. INTS6 (DDX26A, DICE1) forms a complex containing INTS3 and localizes to DNA damage sites where it participates in HDR activation (71). While full-length INTS6 features 18 exons, depletion of CDK12 promoted the usage of exon 3 as an ALE ($\Delta\Psi = -0.32$, -0.23 and -0.41 in SK-BR-3, MDA-MB-231 and 184-hTERT cells, respectively, Supplementary Table S4). By compiling our data and previously published results (5,27) it is apparent that gene expression regulation and AS regulation of DDR genes by CDK12 is both cell type specific and gene specific.

CDK12 downregulates the long isoform of DNAJB6 and increases the invasiveness of breast cancer cells

Pathway analysis of differential gene and protein expression suggests that some CDK12 functions are conserved across cell types. In addition to cell type-specific regulation described above, we identified common ALE events that were regulated by CDK12 in multiple cell lines. From our experiments with SK-BR-3, MDA-MB-231 and 184-hTERT cells, and from the available datasets from HCT-116 cells (27), we found that depletion of CDK12 promotes distal ALE splicing of the *DNAJB6* (DnaJ homolog subfamily B member 6, MRJ) gene transcript (in SK-BR-3, MDA-MB-231 and 184-hTERT, $\Delta\Psi_{avg} = 0.24$). In an analysis of TCGA RNA-seq data for tumors containing *CDK12* mutations or bi-allelic *CDK12* deletions ($n = 18$ comparisons; mutation/deletion:control), we found the *DNAJB6* distal ALE event in 78% of comparisons on average, as compared to 46% of control ($n = 54$ comparisons; control:control) comparisons (Fisher's exact test $P = 0.03$). Unlike the long genes that were regulated in a cell type-specific manner, *DNAJB6* encodes two small protein isoforms (36 and 27 kDa) from transcripts containing 10 and 8 exons, respectively (Figure 8A). The short isoform of the DNAJB6 protein (DNAJB6-S) is a cytosolic HSP40 family chaperone with implicated roles in Huntington's disease (72,73). By contrast, ALE splicing introduces a nuclear localization signal into the long isoform of DNAJB6 (DNAJB6-L) and therefore it operates primarily in the nucleus. Increased nuclear localization of DNAJB6-L has been reported to mitigate tumorigenicity and metastasis of breast and esophageal cancer cells (74,75). We found that treatment of SK-BR-3 cells with CDK12 siRNA-1 increased expression of DNAJB6-L with a concomitant decrease of DNAJB6-S expression (Figure 8B and C). This suggests that the high native CDK12 levels in SK-BR-3 cells can reduce the expression of DNAJB6-L, consistent with overexpression of CDK12 functioning to promote tumorigenesis. We tested this hypothesis functionally in MDA-MB-231 cells, where DNAJB6-L had been previously shown to decrease cell migration potential (74). We first confirmed that treatment of MDA-MB-231 cells with CDK12 siRNA-1 increased gene and protein expression of DNAJB6-L (Figure 8B–D). To examine the cellular phenotype associated with CDK12 expression we used a scratch wound assay and live cell imaging of MDA-MB-231 cells as a functional test for cell migration (Figure 9). In separate assays we also coated the scratch wound with collagen-I to examine the ability of cells to invade into an extracellular matrix. Depletion of CDK12 by siRNA (Figure 9A) decreased the ability of MDA-MB-231 cells to migrate and invade into a matrix (Figure 9B and C, 'Dep'). In this experiment, cells were pre-treated with Mitomycin C to inhibit cell proliferation to ensure that the changes in migration and invasion rates were not due to impaired cell growth caused by the siRNA treatment (Supplementary Figure S10A). The same result was also observed using a different CDK12 siRNA construct (CDK12 siRNA-3), suggesting that these observa-

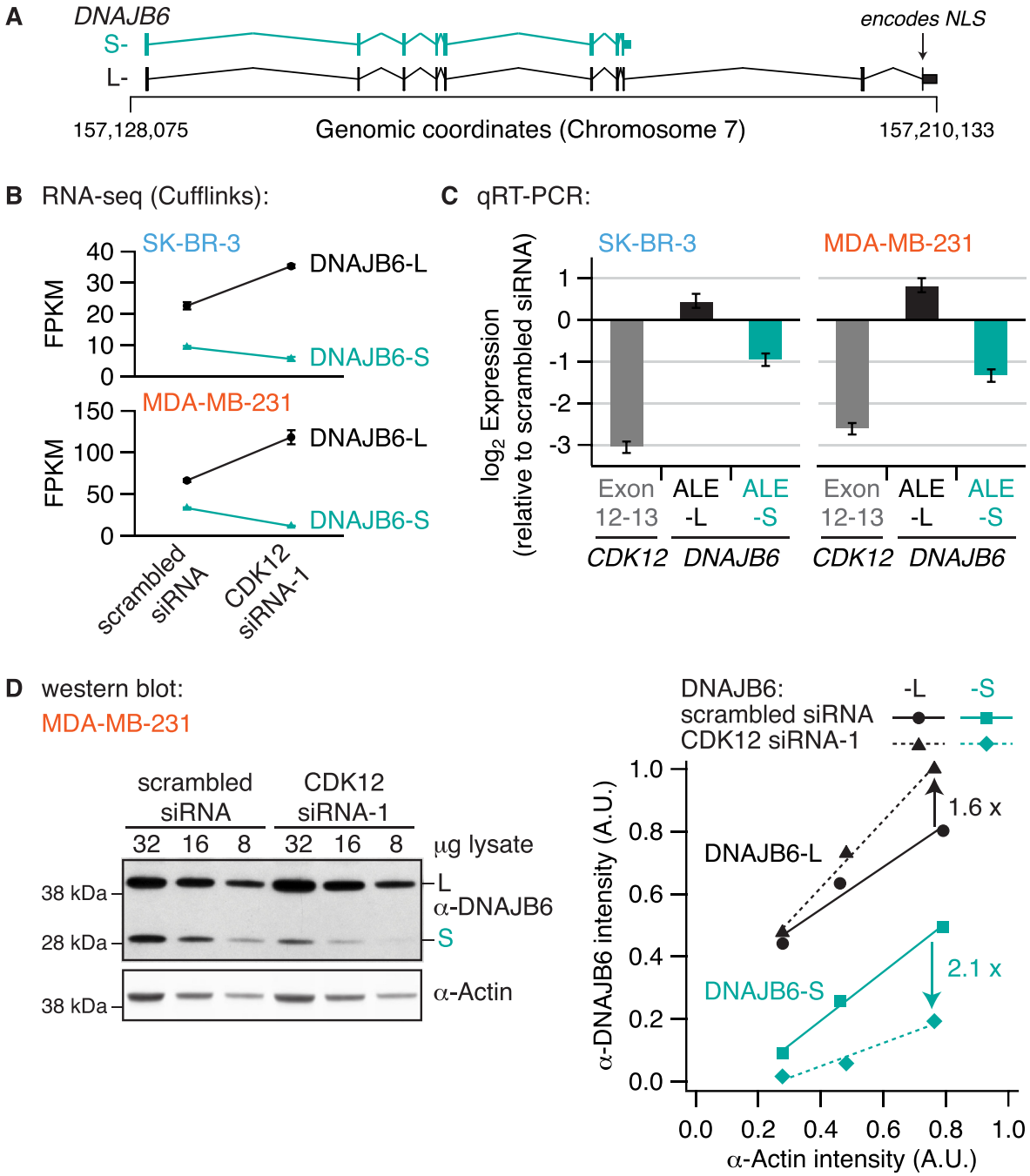

Figure 8. CDK12 downregulates the long isoform of DNAJB6 through ALE splicing. (**A**) Exon structure of the long (-L) and short (-S) isoforms of *DNAJB6*, corresponding to Ensembl transcripts ENST00000262177 and ENST00000429029, respectively. NLS, nuclear localization signal. (**B**) Quantification of *DNAJB6-L* and *DNAJB6-S* transcript levels (FPKM) after CDK12 depletion in SK-BR-3 and MDA-MB-231 cells by RNA-seq using Cufflinks. Error bars represent s.d. (**C**) Validation of changes in *DNAJB6-L* and *DNAJB6-S* transcript expression after CDK12 depletion in SK-BR-3 and MDA-MB-231 cells by quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR). Error bars denote the 99% confidence interval range. (**D**) Relative quantification of changes in DNAJB6-L and DNAJB6-S protein expression due to CDK12 depletion in MDA-MB-231 cells by western blot analysis (detailed in Supplementary Methods).

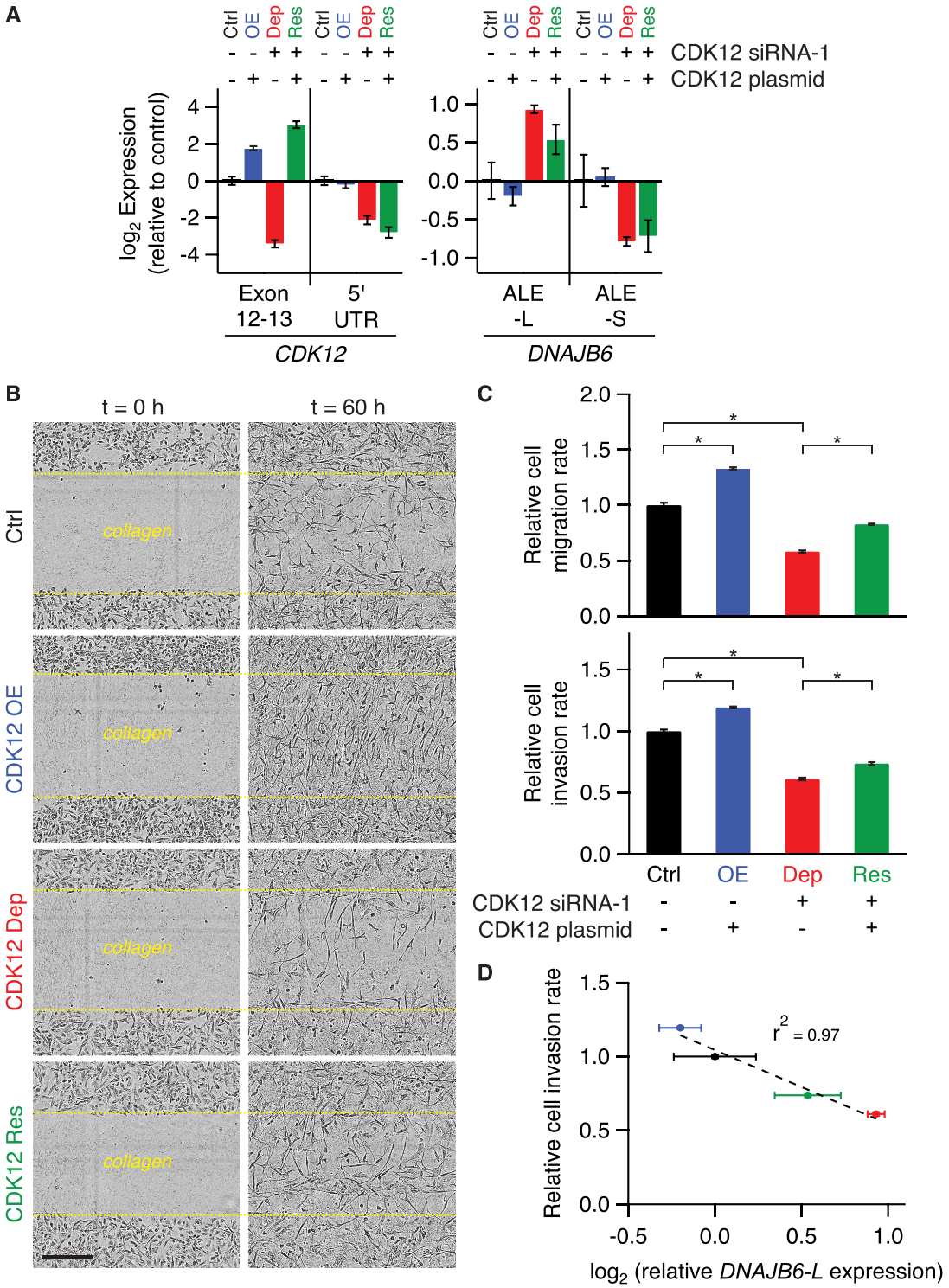

Figure 9. CDK12 promotes cell migration and invasion in MDA-MB-231 triple-negative breast cancer cells. (**A**) CDK12 siRNA-1 and a CDK12 cDNA plasmid were used to manipulate the expression levels of *CDK12* and for correlation to *DNAJB6-S/L* mRNA isoform expression, as measured by qRT-PCR. PCR primers targeting the exon 12–13 junction of *CDK12* amplify both the endogenous CDK12 mRNA and the CDK12 mRNA from the ectopic *CDK12* cDNA. PCR primers targeting the 5′ UTR of *CDK12* specifically amplify the endogenous mRNA only. Error bars denote the 99% confidence interval. (**B**) Representative images of MDA-MB-231 cell invasion into collagen matrix as assayed by scratch wound experiments. Images shown for control cells (Ctrl), cells over-expressing CDK12 (OE), cells depleted of CDK12 by siRNA (Dep) and cells depleted of CDK12 and transfected with CDK12 cDNA (Res). Scale bar represents 300 μm. (**C**) Rates of MDA-MB-231 cell migration and invasion. Error bars represent s.d. of curve fits to the migration and invasion data in Supplementary Figure S10 ($n = 4$ per condition, detailed in Supplementary Methods). The student's *t*-test was used for statistical comparisons (*$P < 0.0001$). (**D**) The rate of cell invasion is proportional to *DNAJB6-L* mRNA expression ($r^2 = 0.97$, $P = 0.03$). Horizontal error bars denote the 99% confidence interval range from qRT-PCR analysis. Vertical error bars represent the s.d. of curve fits to the data in Supplementary Figure S10.

tions were not due to off-target effects (Supplementary Figure S10B). Transfection of CDK12 siRNA-treated MDA-MB-231 cells with a *CDK12* cDNA to re-introduce CDK12 (Figure 9A) recovered the migratory and invasive properties (Figure 9B and C, 'Res'). Unlike SK-BR-3 cells, MDA-MB-231 cells do not over-express CDK12. Therefore, we also tested the effect of CDK12 over-expression on cell migration and invasion. Compared to a vector control, MDA-MB-231 cells transfected with a CDK12 cDNA decreased *DNAJB6-L* expression and were able to migrate and invade at a faster rate (Figure 9A–C, 'OE'). These experiments show that the ability of MDA-MB-231 cells to invade is correlated with CDK12 expression and inversely correlated with the expression level of *DNAJB6-L* (Figure 9D). Therefore, our results suggest that CDK12 can increase the invasiveness of a breast cancer cell line, likely through ALE splicing of the *DNAJB6* gene.

DISCUSSION

CDK12 regulates ALE splicing in a gene- and cell type-specific manner

Prior to this study, the global effects of CDK12 on AS were uncharacterized and opposing conclusions had been made regarding its role in gene expression. While several studies proposed that CDK12 specifically affects a small number of genes (5,76), another report suggested that CDK12 depletion causes a global downregulation of transcription (27). Here, we applied stringent criteria, combining RNA-seq datasets in biological triplicates from three different cell lines to identify AS and differential gene expression events with high confidence. In our global analysis of these three cell lines and from analysis of RNA-seq data from a previous study (27), we consistently identified ALE splicing as a novel mode of regulation by CDK12. The specific regulation of ALE events by CDK12 is striking, and contrasts to the broad effects across all classes of AS that occurs from inhibiting or depleting general splicing factors such as CLKs (21,77), hnRNPs (22), SR proteins (23) and exon junction complex proteins (24). Previous studies have implicated CDK12 in the 3′ end processing of *c-MYC* in HeLa cells (78) and *c-FOS* in HEK293-T cells (56). In these examples, it was proposed that through phosphorylation of the RNA Polymerase II CTD, CDK12 directs cleavage and polyadenylation of *c-MYC* and *c-FOS* transcripts. This is similar to the reported activity of the yeast homolog of CDK12, Ctk1 (79). However, Ctk1 lacks an RS domain and has no reported role in RNA splicing. It is therefore currently unknown how the Ctk1-like mechanism of regulation by CDK12 affects the total set of differentially expressed and alternatively spliced transcripts in humans.

It was recently proposed that following CDK9-dependent initiation of transcription, CDK12 is the predominant kinase responsible for regulating elongation, analogous to the roles of their respective homologs in yeast, Bur1 and Ctk1 (4). We found that the regulation of differential gene expression by CDK12 was limited to a small subset of genes and that the nature of this regulation was highly cell type-specific. These genes generally had long transcripts and a high numbers of exons, as previously reported for HeLa cells (5). Using RNA Polymerase II occupancy experiments, a recent study also demonstrated that the effect of CDK12 inhibition on elongation is not global (80). Chemical inhibition of CDK12 kinase activity resulted in reduced elongation processivity in a small number of target genes, resulting in decreased expression of those genes. Together, these observations suggest a parsimonious model where cell type-dependent factors regulate CDK12 at specific genes, whereupon CDK12 increases the processivity and/or rate of transcription elongation. Therefore, loss of CDK12 function manifests as the downregulation of targeted genes, especially those with long transcripts and are most reliant on productive elongation for expression. This model, however, does not explain how CDK12 depletion promotes the upregulation of genes, which accounts for almost half of all differential gene expression events in our data. Furthermore, it remains possible that CDK12 has a global effect, but CDK9 and CDK13 activity could largely compensate for the loss of CDK12 function after its depletion or inhibition.

Similar to its regulation of transcription, CDK12 also regulates ALE splicing of genes with long transcripts and high number of exons. This trend was significantly more pronounced in ALE splicing events regulated by CDK12, compared to differential gene expression events regulated by CDK12. Furthermore, in a majority of events, native CDK12 promoted the splicing of the longer mRNA isoform. In the most parsimonious interpretation, the processivity model can be extended to the regulation of pre-mRNA splicing by CDK12, wherein CDK12 controls the processivity and/or rate of elongation to achieve successful splicing of one exon to the next exon. In the absence of CDK12, this splicing event is reduced and transcription defaults to termination and polyadenylation of what then becomes the last exon (the proximal ALE). While it is possible that this mechanism likely underlies ALE regulation by CDK12, this simple model alone cannot explain all our major observations. First, slow transcription elongation by RNA Polymerase II is typically associated with increased inclusion of alternative exon cassettes (81), which was not observed in our data. The specific regulation of ALE events in our dataset therefore suggests additional factors regulating this process. Second, the model does not explain how the proximal ALE is selected among all the exons within a long transcript, and how ALE splicing is distinguished from transcriptional regulation. Our analysis of RNA-binding motifs suggests that a subset of proximal ALEs regulated by CDK12 have a higher density of polyadenylation signals in the 3′ UTR. CDK12 could therefore be required to bypass the tendency of the transcript to terminate and successfully splice to the next exon. However, the enrichment of polyadenylation motifs was also observed in instances where CDK12 activity promoted the shorter mRNA isoform. Furthermore, the genes affected by CDK12 ALE regulation were mostly different across the three cell types. Therefore, polyadenylation motifs are likely not the sole factor influencing the regulation of AS by CDK12. Lastly, the processivity and elongation model is inconsistent with CDK12-dependent splicing events that promote utilization of the proximal ALE, as was observed with a minority of genes (∼20%; positive $\Delta\Psi$ values). Even though positive and negative $\Delta\Psi$ events result in opposite directions of

ALE splicing, re-analysis of a published dataset (80) show that loss of CDK12 reduced RNA Polymerase II processivity in both instances (Supplementary Figure S11). Furthermore, identifying genes with decreased RNA Polymerase II processivity in this dataset was not predictive of ALE splicing regulation by CDK12 based on our data (Supplementary Figure S12). As an important caveat, these comparisons were made between different cell types, though the same conclusions can be drawn when considering only ALE events common to all three cell lines in our study. These common ALE events were also enriched in TCGA ovarian tumors with *CDK12* alterations. Taken together, these observations suggest that processivity alone is not the sole differentiating mechanism for the specificity of ALE regulation.

Transcripts with positive $\Delta\Psi$ values after depletion of CDK12 are not significantly longer nor contain more exons that those with ALEs not regulated by CDK12. One such gene, *DNAJB6*, is regulated by CDK12 in multiple cell types and tumors, suggesting a gene-specific regulation that differs from the possible length-dependent regulation common to other ALE events. Therefore, it is probable that regulation of AS by CDK12 also requires additional splicing factors such as the SR proteins, hnRNPs and RNA processing factors identified in our immunoprecipitation experiments. The regulation of only a small subset of genes that differ depending on cell type is possibly accomplished by the various tissue-specific splicing regulatory factors that associate with CDK12 or by signal transduction processes that regulate the action of CDK12 and/or its interacting proteins. Future studies should be aimed at determining the precise role of these regulatory proteins in CDK12-dependent regulation of transcription and AS.

Evidence for oncogenic properties of CDK12 in MDA-MB-231 cells

In line with our findings, experiments exploring the effect of loss-of-function mutations in *CDK12* on the DDR suggest that *CDK12* is a tumor suppressor gene. However, several observations show that *CDK12* has properties that also resemble oncogenes. This is particularly pertinent in breast cancers, where *CDK12* is frequently co-amplified with the *HER2* oncogene. Over-expression of *CDK12* is correlated with aggressive tumor behavior and poor survival (28,31,48). Our RNA-seq experiments examining a breast cancer cell line over-expressing CDK12 (SK-BR-3 cells) identified AS splicing events that could promote tumor-like behavior. These events were also found in our analysis of TCGA RNA-seq data of ovarian tumors containing CDK12 amplifications. One notable AS event regulated by CDK12 and identified in multiple cell types and tumors was the ALE splicing of DNAJB6. Recent studies show that the long isoform of DNAJB6 (DNAJB6-L) suppresses cell migration and invasion in MDA-MB-231 cells (74). While the mechanism driving this activity was unclear, it was dependent on the ALE splicing and subsequent nuclear localization of DNAJB6-L. Using the same MDA-MB-231 cell line model, we showed that CDK12 expression is inversely correlated with ALE splicing of DNAJB6-L. The ability of cancer cells to migrate and invade is a fundamental mechanism underlying tumorigenesis and metastasis (82) MDA-MB-231 cells can seed tumors in mouse models, and increasing DNAJB6-L expression decreases tumor growth and metastasis in athymic mice (74). Therefore, the ability of CDK12 over-expression to downregulate DNAJB6-L through ALE splicing represents a specific cellular mechanism by which amplified *CDK12* can increase the aggressiveness of breast cancer cells. This could be a significant factor contributing to the progression of $HER2^+$ breast cancers, where *CDK12* is co-amplified in 27–92% of cases (39–47).

In this study, we applied a comprehensive genomic and proteomic approach to define the cellular functions of CDK12 and to investigate its effect on breast cancer cell lines. We showed that in multiple cell lines, CDK12 regulated a core set of cellular processes including RNA processing and DNA repair. We also found that CDK12 regulated ALE splicing, primarily of genes with long transcripts and a large number of exons. While this regulation mechanism appears conserved, the affected genes are highly cell type-specific. CDK12 regulated splicing of DNAJB6, whose nuclear localization attenuates tumor invasion. In MDA-MB-231 cells, CDK12 promoted migration and invasion in a dose-dependent manner. Together, these results show how loss of CDK12 can disrupt DNA repair and also suggest an AS-dependent mechanism by which CDK12 over-expression can increase the tumorigenicity of breast cancer cells.

ACCESSION NUMBERS

The RNA-seq datasets supporting the conclusions of this article are available in the Gene Expression Omnibus repository (http://www.ncbi.nlm.nih.gov/geo) with the accession number GSE80409. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://www.ebi.ac.uk/pride) with the dataset identifier PXD004184.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Tyler Funnell, Daniel Woodsworth and Scott Brown for valuable advice and Shane Colborne for technical assistance.

FUNDING

Canadian Breast Cancer Foundation Post-doctoral Fellowship (to J.F.T.); Michael Smith Foundation for Health Research Post-doctoral Fellowship (to J.F.T.); University of British Columbia Doctoral Fellowship (to A.M.); Canadian Breast Cancer Foundation Studentship (to C.C.T.C., L.T.C.); British Columbia Proteomics Network Studentship (to C.C.T.C., L.T.C.); Michael Smith Foundation for Health Research Scholarship (to S.P.S.); Canada Research Chairs (to S.P.S., S.A.); Natural Sciences and Engineering Research Council of Canada [RGPIN 326895-13 to G.B.M.]; Canadian Cancer Society Research Institute

[701584 to S.A.]; Terry Fox Research Institute [PPG1021 to S.A.]; Canadian Institutes of Health Research [MOP-126119 and FDN-148429 to S.A.]; Takeda Pharmaceutical Co. Ltd Shonan Incubation Laboratories (to G.B.M., S.A.); British Columbia Cancer Foundation. Funding for open access charge: Natural Sciences and Engineering Research Council of Canada.
Conflict of interest statement. None declared.

REFERENCES

1. Romano,G. and Giordano,A. (2008) Role of the cyclin-dependent kinase 9-related pathway in mammalian gene expression and human diseases. *Cell Cycle*, **7**, 3664–3668.
2. Loyer,P., Trembley,J.H., Katona,R., Kidd,V.J. and Lahti,J.M. (2005) Role of CDK/cyclin complexes in transcription and RNA splicing. *Cell Signal.*, **17**, 1033–1051.
3. Malumbres,M. (2014) Cyclin-dependent kinases. *Genome Biol.*, **15**, 122–131.
4. Bartkowiak,B., Liu,P., Phatnani,H.P., Fuda,N.J., Cooper,J.J., Price,D.H., Adelman,K., Lis,J.T. and Greenleaf,A.L. (2010) CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes Dev.*, **24**, 2303–2316.
5. Blazek,D., Kohoutek,J., Bartholomeeusen,K., Johansen,E., Hulinkova,P., Luo,Z., Cimermancic,P., Ule,J. and Peterlin,B.M. (2011) The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. *Genes Dev.*, **25**, 2158–2172.
6. Cheng,S.W., Kuzyk,M.A., Moradian,A., Ichu,T.A., Chang,V.C., Tien,J.F., Vollett,S.E., Griffith,M., Marra,M.A. and Morin,G.B. (2012) Interaction of cyclin-dependent kinase 12/CrkRS with cyclin K1 is required for the phosphorylation of the C-terminal domain of RNA polymerase II. *Mol. Cell. Biol.*, **32**, 4691–4704.
7. Bowman,E.A. and Kelly,W.G. (2014) RNA polymerase II transcription elongation and Pol II CTD Ser2 phosphorylation: a tail of two kinases. *Nucleus*, **5**, 224–236.
8. Bosken,C.A., Farnung,L., Hintermair,C., Merzel Schachter,M., Vogel-Bachmayr,K., Blazek,D., Anand,K., Fisher,R.P., Eick,D. and Geyer,M. (2014) The structure and substrate specificity of human Cdk12/Cyclin K. *Nat. Commun.*, **5**, 1–14.
9. Dixon-Clarke,S.E., Elkins,J.M., Cheng,S.W., Morin,G.B. and Bullock,A.N. (2015) Structures of the CDK12/CycK complex with AMP-PNP reveal a flexible C-terminal kinase extension important for ATP binding. *Scientific Rep.*, **5**, 1–13.
10. Greifenberg,A.K., Honig,D., Pilarova,K., Duster,R., Bartholomeeusen,K., Bosken,C.A., Anand,K., Blazek,D. and Geyer,M. (2016) Structural and functional analysis of the Cdk13/cyclin k complex. *Cell Rep.*, **14**, 320–331.
11. Ko,T.K., Kelly,E. and Pines,J. (2001) CrkRS: a novel conserved Cdc2-related protein kinase that colocalises with SC35 speckles. *J. Cell Sci.*, **114**, 2591–2603.
12. Even,Y., Durieux,S., Escande,M.L., Lozano,J.C., Peaucellier,G., Weil,D. and Geneviere,A.M. (2006) CDC2L5, a Cdk-like kinase with RS domain, interacts with the ASF/SF2-associated protein p32 and affects splicing in vivo. *J. Cell. Biochem.*, **99**, 890–904.
13. Fu,X.D. (1995) The superfamily of arginine/serine-rich splicing factors. *RNA*, **1**, 663–680.
14. Kay,B.K., Williamson,M.P. and Sudol,M. (2000) The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J.*, **14**, 231–241.
15. Will,C.L. and Luhrmann,R. (2011) Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.*, **3**, 1–23.
16. Lee,Y. and Rio,D.C. (2015) Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Anna. Rev. Biochem.*, **84**, 291–323.
17. Caceres,J.F. and Kornblihtt,A.R. (2002) Alternative splicing: multiple control mechanisms and involvement in human disease. *Trends Genet.*, **18**, 186–193.
18. Venables,J.P. (2004) Aberrant and alternative splicing in cancer. *Cancer Res.*, **64**, 7647–7654.
19. Sveen,A., Kilpinen,S., Ruusulehto,A., Lothe,R.A. and Skotheim,R.I. (2016) Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. *Oncogene*, **35**, 2413–2427.
20. Busch,A. and Hertel,K.J. (2012) Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley interdiscip. Rev. RNA*, **3**, 1–12.
21. Dominguez,D., Tsai,Y.H., Weatheritt,R., Wang,Y., Blencowe,B.J. and Wang,Z. (2016) An extensive program of periodic alternative splicing linked to cell cycle progression. *Elife*, **5**, 1–19.
22. Huelga,S.C., Vu,A.Q., Arnold,J.D., Liang,T.Y., Liu,P.P., Yan,B.Y., Donohue,J.P., Shiue,L., Hoon,S., Brenner,S. *et al.* (2012) Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell Rep.*, **1**, 167–178.
23. Li,H., Cheng,Y., Wu,W., Liu,Y., Wei,N., Feng,X., Xie,Z. and Feng,Y. (2014) SRSF10 regulates alternative splicing and is required for adipocyte differentiation. *Mol. Cell. Biol.*, **34**, 2198–2207.
24. Wang,Z., Murigneux,V. and Le Hir,H. (2014) Transcriptome-wide modulation of splicing by the exon junction complex. *Genome Biol.*, **15**, 1–18.
25. Chen,H.H., Wang,Y.C. and Fann,M.J. (2006) Identification and characterization of the CDK12/cyclin L1 complex involved in alternative splicing regulation. *Mol. Cell. Biol.*, **26**, 2736–2745.
26. Rodrigues,F., Thuma,L. and Klambt,C. (2012) The regulation of glial-specific splicing of Neurexin IV requires HOW and Cdk12 activity. *Development*, **139**, 1765–1776.
27. Liang,K., Gao,X., Gilmore,J.M., Florens,L., Washburn,M.P., Smith,E. and Shilatifard,A. (2015) Characterization of human cyclin-dependent kinase 12 (CDK12) and CDK13 complexes in C-terminal domain phosphorylation, gene transcription, and RNA processing. *Mol. Cell. Biol.*, **35**, 928–938.
28. Cerami,E., Gao,J., Dogrusoz,U., Gross,B.E., Sumer,S.O., Aksoy,B.A., Jacobsen,A., Byrne,C.J., Heuer,M.L., Larsson,E. *et al.* (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.*, **2**, 401–404.
29. Cancer Genome Atlas Research Network (2011) Integrated genomic analyses of ovarian carcinoma. *Nature*, **474**, 609–615.
30. Kandoth,C., McLellan,M.D., Vandin,F., Ye,K., Niu,B., Lu,C., Xie,M., Zhang,Q., McMichael,J.F., Wyczalkowski,M.A. *et al.* (2013) Mutational landscape and significance across 12 major cancer types. *Nature*, **502**, 333–339.
31. Cancer Genome Atlas Network (2012) Comprehensive molecular portraits of human breast tumours. *Nature*, **490**, 61–70.
32. Joshi,P.M., Sutor,S.L., Huntoon,C.J. and Karnitz,L.M. (2014) Ovarian cancer-associated mutations disable catalytic activity of CDK12, a kinase that promotes homologous recombination repair and resistance to cisplatin and poly(ADP-ribose) polymerase inhibitors. *J. Biol. Chem.*, **289**, 9247–9253.
33. Carter,S.L., Cibulskis,K., Helman,E., McKenna,A., Shen,H., Zack,T., Laird,P.W., Onofrio,R.C., Winckler,W., Weir,B.A. *et al.* (2012) Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.*, **30**, 413–421.
34. Bajrami,I., Frankum,J.R., Konde,A., Miller,R.E., Rehman,F.L., Brough,R., Campbell,J., Sims,D., Rafiq,R., Hooper,S. *et al.* (2014) Genome-wide profiling of genetic synthetic lethality identifies CDK12 as a novel determinant of PARP1/2 inhibitor sensitivity. *Cancer Res.*, **74**, 287–297.
35. Natrajan,R., Wilkerson,P.M., Marchio,C., Piscuoglio,S., Ng,C.K., Wai,P., Lambros,M.B., Samartzis,E.P., Dedes,K.J., Frankum,J. *et al.* (2014) Characterization of the genomic features and expressed fusion genes in micropapillary carcinomas of the breast. *J. Pathol.*, **232**, 553–565.
36. Ekumi,K.M., Paculova,H., Lenasi,T., Pospichalova,V., Bosken,C.A., Rybarikova,J., Bryja,V., Geyer,M., Blazek,D. and Barboric,M. (2015) Ovarian carcinoma CDK12 mutations misregulate expression of DNA repair genes via deficient formation and function of the Cdk12/CycK complex. *Nucleic Acids Res.*, **43**, 2575–2589.
37. Johnson,S.F., Cruz,C., Greifenberg,A.K., Dust,S., Stover,D.G., Chi,D., Primack,B., Cao,S., Bernhardy,A.J., Coulson,R. *et al.* (2016) CDK12 inhibition reverses de novo and acquired PARP inhibitor resistance in BRCA wild-type and mutated models of triple-negative breast cancer. *Cell Rep.*, **17**, 2367–2381.
38. Popova,T., Manie,E., Boeva,V., Battistella,A., Goundiam,O., Smith,N.K., Mueller,C.R., Raynal,V., Mariani,O., Sastre-Garau,X. *et al.* (2016) Ovarian cancers harboring inactivating mutations in CDK12 display a distinct genomic instability pattern characterized by large tandem duplications. *Cancer Res.*, **76**, 1882–1891.

39. Kauraniemi,P., Kuukasjarvi,T., Sauter,G. and Kallioniemi,A. (2003) Amplification of a 280-kilobase core region at the ERBB2 locus leads to activation of two hypothetical proteins in breast cancer. *Am. J. Pathol.*, **163**, 1979–1984.
40. Hyman,E., Kauraniemi,P., Hautaniemi,S., Wolf,M., Mousses,S., Rozenblum,E., Ringner,M., Sauter,G., Monni,O., Elkahloun,A. *et al.* (2002) Impact of DNA amplification on gene expression patterns in breast cancer. *Cancer Res.*, **62**, 6240–6245.
41. Kao,J. and Pollack,J.R. (2006) RNA interference-based functional dissection of the 17q12 amplicon in breast cancer reveals contribution of coamplified genes. *Genes Chromosomes Cancer*, **45**, 761–769.
42. Kauraniemi,P., Barlund,M., Monni,O. and Kallioniemi,A. (2001) New amplified and highly expressed genes discovered in the ERBB2 amplicon in breast cancer by cDNA microarrays. *Cancer Res.*, **61**, 8235–8240.
43. Neve,R.M., Chin,K., Fridlyand,J., Yeh,J., Baehner,F.L., Fevr,T., Clark,L., Bayani,N., Coppe,J.P., Tong,F. *et al.* (2006) A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell*, **10**, 515–527.
44. Pollack,J.R., Sorlie,T., Perou,C.M., Rees,C.A., Jeffrey,S.S., Lonning,P.E., Tibshirani,R., Botstein,D., Borresen-Dale,A.L. and Brown,P.O. (2002) Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 12963–12968.
45. Curtis,C., Shah,S.P., Chin,S.F., Turashvili,G., Rueda,O.M., Dunning,M.J., Speed,D., Lynch,A.G., Samarajiwa,S., Yuan,Y. *et al.* (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, **486**, 346–352.
46. Lawrence,R.T., Perez,E.M., Hernandez,D., Miller,C.P., Haas,K.M., Irie,H.Y., Lee,S.I., Blau,C.A. and Villen,J. (2015) The proteomic landscape of triple-negative breast cancer. *Cell Rep.*, **11**, 630–644.
47. Ciriello,G., Gatza,M.L., Beck,A.H., Wilkerson,M.D., Rhie,S.K., Pastore,A., Zhang,H., McLellan,M., Yau,C., Kandoth,C. *et al.* (2015) Comprehensive molecular portraits of invasive lobular breast cancer. *Cell*, **163**, 506–519.
48. Capra,M., Nuciforo,P.G., Confalonieri,S., Quarto,M., Bianchi,M., Nebuloni,M., Boldorini,R., Pallotti,F., Viale,G., Gishizky,M.L. *et al.* (2006) Frequent alterations in the expression of serine/threonine kinases in human cancers. *Cancer Res.*, **66**, 8147–8154.
49. Zang,Z.J., Ong,C.K., Cutcutache,I., Yu,W., Zhang,S.L., Huang,D., Ler,L.D., Dykema,K., Gan,A., Tao,J. *et al.* (2011) Genetic and structural variation in the gastric cancer kinome revealed through targeted deep sequencing. *Cancer Res.*, **71**, 29–39.
50. Burleigh,A., McKinney,S., Brimhall,J., Yap,D., Eirew,P., Poon,S., Ng,V., Wan,A., Prentice,L., Annab,L. *et al.* (2015) A co-culture genome-wide RNAi screen with mammary epithelial cells reveals transmembrane signals required for growth and differentiation. *Breast Cancer Res.*, **17**, 1–21.
51. Hughes,C.S., Foehr,S., Garfield,D.A., Furlong,E.E., Steinmetz,L.M. and Krijgsveld,J. (2014) Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol. Syst. Biol.*, **10**, 1–14.
52. Morrissy,A.S., Garzia,L., Shih,D.J., Zuyderduyn,S., Huang,X., Skowron,P., Remke,M., Cavalli,F.M., Ramaswamy,V., Lindsay,P.E. *et al.* (2016) Divergent clonal selection dominates medulloblastoma at recurrence. *Nature*, **529**, 351–357.
53. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 1–21.
54. Katz,Y., Wang,E.T., Airoldi,E.M. and Burge,C.B. (2010) Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods*, **7**, 1009–1015.
55. Mi,H., Lazareva-Ulitsky,B., Loo,R., Kejariwal,A., Vandergriff,J., Rabkin,S., Guo,N., Muruganujan,A., Doremieux,O., Campbell,M.J. *et al.* (2005) The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res.*, **33**, D284–D288.
56. Eifler,T.T., Shao,W., Bartholomeeusen,K., Fujinaga,K., Jager,S., Johnson,J.R., Luo,Z., Krogan,N.J. and Peterlin,B.M. (2015) Cyclin-dependent kinase 12 increases 3′ end processing of growth factor-induced c-FOS transcripts. *Mol. Cell. Biol.*, **35**, 468–478.
57. Ingham,R.J., Colwill,K., Howard,C., Dettwiler,S., Lim,C.S., Yu,J., Hersi,K., Raaijmakers,J., Gish,G., Mbamalu,G. *et al.* (2005) WW domains provide a platform for the assembly of multiprotein networks. *Mol. Cell. Biol.*, **25**, 7092–7106.
58. Bartkowiak,B. and Greenleaf,A.L. (2015) Expression, purification, and identification of associated proteins of the full-length hCDK12/CyclinK complex. *J. Biol. Chem.*, **290**, 1786–1795.
59. Jung,S.Y., Malovannaya,A., Wei,J., O'Malley,B.W. and Qin,J. (2005) Proteomic analysis of steady-state nuclear hormone receptor coactivator complexes. *Mol. Endocrinol.*, **19**, 2451–2465.
60. Elkon,R., Ugalde,A.P. and Agami,R. (2013) Alternative cleavage and polyadenylation: extent, regulation and function. *Nat. Rev. Genet.*, **14**, 496–506.
61. Grasso,C.S., Wu,Y.M., Robinson,D.R., Cao,X., Dhanasekaran,S.M., Khan,A.P., Quist,M.J., Jing,X., Lonigro,R.J., Brenner,J.C. *et al.* (2012) The mutational landscape of lethal castration-resistant prostate cancer. *Nature*, **487**, 239–243.
62. Anders,S. and Huber,W. (2010) Differential expression analysis for sequence count data. *Genome Biol.*, **11**, 1–12.
63. Subramanian,A., Tamayo,P., Mootha,V.K., Mukherjee,S., Ebert,B.L., Gillette,M.A., Paulovich,A., Pomeroy,S.L., Golub,T.R., Lander,E.S. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 15545–15550.
64. Juan,H.C., Lin,Y., Chen,H.R. and Fann,M.J. (2016) Cdk12 is essential for embryonic development and the maintenance of genomic stability. *Cell Death Differ.*, **23**, 1038–1048.
65. Iorns,E., Martens-de Kemp,S.R., Lord,C.J. and Ashworth,A. (2009) CRK7 modifies the MAPK pathway and influences the response to endocrine therapy. *Carcinogenesis*, **30**, 1696–1701.
66. Moasser,M.M. (2007) The oncogene HER2: its signaling and transforming functions and its role in human cancer pathogenesis. *Oncogene*, **26**, 6469–6487.
67. Bryant,H.E., Schultz,N., Thomas,H.D., Parker,K.M., Flower,D., Lopez,E., Kyle,S., Meuth,M., Curtin,N.J. and Helleday,T. (2005) Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature*, **434**, 913–917.
68. Farmer,H., McCabe,N., Lord,C.J., Tutt,A.N., Johnson,D.A., Richardson,T.B., Santarosa,M., Dillon,K.J., Hickson,I., Knights,C. *et al.* (2005) Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature*, **434**, 917–921.
69. Fong,P.C., Boss,D.S., Yap,T.A., Tutt,A., Wu,P., Mergui-Roelvink,M., Mortimer,P., Swaisland,H., Lau,A., O'Connor,M.J. *et al.* (2009) Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N. Engl. J. Med.*, **361**, 123–134.
70. Guleria,A. and Chandna,S. (2016) ATM kinase: much more than a DNA damage responsive protein. *DNA Repair*, **39**, 1–20.
71. Zhang,F., Ma,T. and Yu,X. (2013) A core hSSB1-INTS complex participates in the DNA damage response. *J. Cell Sci.*, **126**, 4850–4855.
72. Chuang,J.Z., Zhou,H., Zhu,M., Li,S.H., Li,X.J. and Sung,C.H. (2002) Characterization of a brain-enriched chaperone, MRJ, that inhibits huntingtin aggregation and toxicity independently. *J. Biol. Chem.*, **277**, 19831–19838.
73. Fayazi,Z., Ghosh,S., Marion,S., Bao,X., Shero,M. and Kazemi-Esfarjani,P. (2006) A Drosophila ortholog of the human MRJ modulates polyglutamine toxicity and aggregation. *Neurobiol. Dis.*, **24**, 226–244.
74. Mitra,A., Fillmore,R.A., Metge,B.J., Rajesh,M., Xi,Y., King,J., Ju,J., Pannell,L., Shevde,L.A. and Samant,R.S. (2008) Large isoform of MRJ (DNAJB6) reduces malignant activity of breast cancer. *Breast Cancer Res.*, **10**, 1–13.
75. Yu,V.Z., Wong,V.C., Dai,W., Ko,J.M., Lam,A.K., Chan,K.W., Samant,R.S., Lung,H.L., Shuen,W.H., Law,S. *et al.* (2015) Nuclear localization of DNAJB6 is associated with survival of patients with esophageal cancer and reduces AKT signaling and proliferation of cancer cells. *Gastroenterology*, **149**, 1825–1836.
76. Li,X., Chatterjee,N., Spirohn,K., Boutros,M. and Bohmann,D. (2016) Cdk12 is a gene-selective RNA polymerase II kinase that regulates a subset of the transcriptome, including Nrf2 target genes. *Sci. Rep.*, **6**, 1–13.
77. Funnell,T., Tasaki,S., Oloumi,A., Araki,S., Kong,E., Yap,D., Nakayama,Y., Hughes,C.S., Cheng,S.G., Tozaki,H. *et al.* (2017) CLK-dependent exon recognition and conjoined gene formation revealed with a novel small molecule inhibitor. *Nat. Commun.*, **8**, 1–15.
78. Davidson,L., Muniz,L. and West,S. (2014) 3′ end formation of pre-mRNA and phosphorylation of Ser2 on the RNA polymerase II

CTD are reciprocally coupled in human cells. *Genes Dev.*, **28**, 342–356.
79. Ahn,S.H., Kim,M. and Buratowski,S. (2004) Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3′ end processing. *Mol. Cell*, **13**, 67–76.
80. Zhang,T., Kwiatkowski,N., Olson,C.M., Dixon-Clarke,S.E., Abraham,B.J., Greifenberg,A.K., Ficarro,S.B., Elkins,J.M., Liang,Y., Hannett,N.M. *et al.* (2016) Covalent targeting of remote cysteine residues to develop CDK12 and CDK13 inhibitors. *Nat. Chem. Biol.*, **12**, 876–884.
81. Kornblihtt,A.R., Schor,I.E., Allo,M., Dujardin,G., Petrillo,E. and Munoz,M.J. (2013) Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.*, **14**, 153–165.
82. Price,J.T. and Thompson,E.W. (2002) Mechanisms of tumour invasion and metastasis: emerging targets for therapy. *Expert Opin. Ther. Targets*, **6**, 217–233.