LINE-1 expression in cancer correlates with p53 mutation, copy number alteration, and S phase checkpoint

Wilson McKerrow[a,b], Xuya Wang[a,b], Carlos Mendez-Dorantes[c,d], Paolo Mita[a,b], Song Cao[e,f], Mark Grivainis[a,b], Li Ding[e,f], John LaCava[g,h], Kathleen H. Burns[c,d], Jef D. Boeke[a,b,i,1], and David Fenyö[a,b,1]

[a]Institute for Systems Genetics, New York University Grossman School of Medicine, New York, NY 10016; [b]Department of Biochemistry and Molecular Pharmacology, New York University Grossman School of Medicine, New York, NY 10016; [c]Department of Oncologic Pathology, Dana-Farber Cancer Institute, Boston, MA 02215; [d]Department of Pathology, Harvard Medical School, Boston, MA 02115; [e]Department of Medicine and Genetics, Siteman Cancer Center, Washington University in St. Louis, St. Louis, MO 63110; [f]McDonnell Genome Institute, Washington University in St. Louis, St. Louis, MO 63108; [g]Laboratory of Cellular and Structural Biology, The Rockefeller University, New York, NY 10065; [h]European Research Institute for the Biology of Ageing, University Medical Center Groningen, 9713 GZ Groningen, The Netherlands; and [i]Department of Biomedical Engineering, Tandon School of Engineering, Brooklyn, NY11201

Contributed by Jef D. Boeke; received August 31, 2021; accepted January 14, 2022; reviewed by Molly Gale Hammell and Michael MacCoss

Retrotransposons are genomic DNA sequences that copy themselves to new genomic locations via RNA intermediates; LINE-1 is the only active and autonomous retrotransposon in the human genome. The mobility of LINE-1 is largely repressed in somatic tissues but is derepressed in many cancers, where LINE-1 retrotransposition is correlated with p53 mutation and copy number alteration (CNA). In cell lines, inducing LINE-1 expression can cause double-strand breaks (DSBs) and replication stress. Reanalyzing multiomic data from breast, ovarian, endometrial, and colon cancers, we confirmed correlations between LINE-1 expression, p53 mutation status, and CNA. We observed a consistent correlation between LINE-1 expression and the abundance of DNA replication complex components, indicating that LINE-1 may also induce replication stress in human tumors. In endometrial cancer, high-quality phosphoproteomic data allowed us to identify the DSB-induced ATM-MRN-SMC S phase checkpoint pathway as the primary DNA damage response (DDR) pathway associated with LINE-1 expression. Induction of LINE-1 expression in an in vitro model led to increased phosphorylation of MRN complex member RAD50, suggesting that LINE-1 directly activates this pathway.

retrotransposon | LINE-1 | cancer | DNA damage response | copy number alteration

LINE-1 (Long Interspersed Element 1) is a family of autonomous retrotransposons that remains active in the human genome. As such, they encode proteins (ORF1p and ORF2p) necessary for the spread of LINE-1 to new genomic loci via an RNA intermediate (a phenomenon referred to as retrotransposition). ORF1p is a RNA binding protein that can form trimers and is thought to bind LINE-1 RNA (1, 2). ORF2p is an enzyme with endonuclease (3) and reverse transcriptase (4) activities that binds to the LINE-1 RNA, most likely at or near its poly(A) tail (5). In dividing cells, this LINE-1 RNA/ORF1p/ORF2p ribonucleoprotein (RNP) complex can enter the nucleus in M phase and retrotranspose via a process called target primed reverse transcription (TPRT) (6, 7) during S phase (8). LINE-1 elements are expressed during early development (9–11), but in somatic tissues, many host factors contribute to LINE-1 silencing, via DNA methylation, histone H3K9 methylation, and RNA silencing (12). This somatic silencing is not always complete, and some LINE-1 RNA can "leak through" (13, 14). In particular, silencing of transposable elements, including LINE-1, appears to decrease with age (15–18).

In contrast, pervasive LINE-1 derepression occurs in ~50% of human cancers (19, 20). However, we still do not fully understand the role that LINE-1 plays in cancer. In rare cases, derepressed LINE-1 elements drive cancer progression by inserting into and disrupting key tumor suppressor genes, most notably *APC* in colon cancer (21–23). New evidence indicates that LINE-1 may also have additional effects on genome stability. Building on previous experiments indicating that LINE-1 can induce widespread double-strand breaks (DSBs) (24) and trigger p53-mediated apoptosis (25), two recent papers proposed a model in which LINE-1 functionally interacts with the replication fork and factors involved in replication-coupled DNA repair, promoting replication stress and negatively impacting the fitness of replicating cells (26, 27). The fact that LINE-1 overexpression can cause DNA damage and/or replication stress suggests that LINE-1 derepression may aid tumor development and plasticity by promoting genome instability beyond the rare LINE-1 insertion into a protein coding or regulatory

Significance

In addition to canonical genes, our genomes encode repetitive copies of the LINE-1 retrotransposon. These elements duplicate themselves by cutting a single-strand break in genomic DNA and then reverse transcribing a new LINE-1 DNA copy into that breakpoint. In most contexts, LINE-1 elements are epigenetically repressed, but they are dramatically derepressed in many cancers, where they have the potential to impact genome integrity. We probed publicly available multiomic data from the CPTAC to identify the DNA damage response correlates of LINE-1 expression and validated the potential for LINE-1 overexpression to trigger RAD50 phosphorylation—a key step in an S phase double-strand break response pathway that we found to be highly correlated with LINE-1 expression in endometrial cancer.

Author contributions: W.M., X.W., P.M., L.D., J.L., K.H.B., J.D.B., and D.F. designed research; W.M., X.W., C.M.-D., S.C., and M.G. performed research; X.W., C.M.-D., and S.C. contributed new reagents/analytic tools; W.M., X.W., and S.C. analyzed data; and W.M., P.M., J.L., J.D.B., and D.F. wrote the paper.

Reviewers: M.G.H., Cold Spring Harbor Laboratory; and M.M., University of Washington.

Competing interest statement: D.F. is a founder and president of The Informatics Factory and serves or served on the scientific advisory board (SAB) or consults for Spectragen Informatics, Protein Metrics, and Preverna. J.D.B. is a founder and director of CDI Labs, Inc.; a founder of and consultant to Neochromosome, Inc.; a founder SAB member of and consultant to ReOpen Diagnostics, LLC; and serves or served on the SAB of the following: Sangamo, Inc., Modern Meadow, Inc., Rome Therapeutics, Inc., Sample6, Inc., Tessera Therapeutics, Inc., and the Wyss Institute. K.H.B. contributed a review with one of the reviewers of this manuscript within the last three years.

[1]To whom correspondence may be addressed. Email: jef.boeke@nyulangone.org or david@fenyolab.org.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2115999119/-/DCSupplemental.

Published February 15, 2022.

sequence. However, evidence for this hypothesis in actual human tumors is lacking.

Large-scale multiomic studies, such as those provided by the Clinical Proteomic Tumor Analysis Consortium (CPTAC), that combine DNA and RNA sequencing with shotgun proteomics and phosphoproteomics provide an opportunity to reveal the mutation, expression, and signaling correlates of LINE-1 RNA and protein expression. In particular, identifying the gene expression and phosphorylation correlates of LINE-1 expression will reveal whether there is evidence for LINE-1–induced replication stress or DNA damage in human tumors. For this study, we looked at five cancer types from the CPTAC project: breast, ovarian, colon, endometrial, and kidney. However, we excluded the kidney results from most analyses due to low LINE-1 expression. In agreement with a recent study that identified positive correlations between LINE-1 retrotransposition and several types of structural genomic alteration in many cancer types (28), we found positive correlations between LINE-1 ORF1p protein expression and copy number alteration (CNA) burden in breast, ovarian, and endometrial cancer. This correlation was strongest in endometrial tumors, particularly those with p53 mutation, where we found positive correlations between LINE-1 ORF1p expression and phosphosites indicative of S phase checkpoint activation via ATM phosphorylation of the MRN complex and cohesin (29). We were further able to show that the induction of LINE-1 expression in RPE-1 cells with p53 knockdown leads to modestly increased RAD50-S635 phosphorylation, a key step in this pathway. Together, these results suggest that LINE-1 can contribute to DNA damage and/or replication stress in human tumors and that this may contribute to genome instability in excess of what can be directly attributed to LINE-1 insertions.

Results

Quantifications of LINE-1 Expression, ORF1p Phosphorylation, and Retrotransposition Were Highly Correlated. LINE-1 activity can be measured at several points in its life cycle, including: LINE-1 RNA expression, ORF1 protein (ORF1p) expression, and somatic insertion. We also identified and quantified phosphorylation sites of ORF1p. We first wanted to know how these measurements relate to each other to determine the extent that one measurement can stand as a proxy for overall LINE-1 "activity." We reanalyzed data from the CPTAC project to quantify LINE-1 mRNA, ORF1p (L1RE1 in uniprot), and ORF1p phosphorylation in five cancer types: breast (n = 94) (30), ovarian (n = 97) (31), colon (n = 93) (32), clear cell kidney (n = 106) (33), and endometrial (n = 88) (34). LINE-1 RNA was quantified using L1EM (35), a tool that uses the expectation maximization to remove LINE-1 reads that are incorporated into other transcripts and only quantifies LINE-1 RNA expression derived from the LINE-1 5′ untranslated region (UTR)/promoter. Relative LINE-1 ORF1p quantifications were derived from 20 proteotypic peptides that we identified. We used the mobile element locator tool (MELT) (36) to quantify somatic LINE-1 insertions in kidney and endometrial tumors, where matched tumor and whole blood whole genome sequencing (WGS) was available. WGS was not available for the breast, ovarian, or colon tumors, but other studies have shown evidence of LINE-1 retrotransposition in these cancer types (28). Raw LINE-1 quantifications and individual somatic insertions from endometrial cancer from each of these data types can be found in Dataset S1. We did attempt to identify ORF2p peptides in these data, but the putative matches tended to have high expectation (e) values, suggesting that they may be false positives, and their peptide abundance (measured by comparing to an internal standard using tandem mass tag [TMT]) did not correlate with ORF1p. We previously observed that endogenous ORF2p is not readily detectable by standard shotgun proteomics (37). Because LINE-1 is subject to regulation throughout its life cycle and different quantifications measure distinct life cycle stages, we postulated that they may be uncorrelated or weakly correlated in certain contexts. Fig. 1 shows how these quantifications measure different phases of the LINE-1 life cycle: LINE-1 mRNAs may be present as "standalone" transcripts (Fig. 1*B*) or as RNPs coassembled with ORF1p and ORF2p (Fig. 1 *C* and *D*); ORF1p, which may be present as part of a LINE-1 RNP or perhaps separate from the LINE-1 RNA (not shown in this drawing) is primarily cytoplasmic (Fig. 1*C*), but can also be found in the nucleus (Fig. 1*D*) in the G1 phase of the cell cycle (prior to retrotransposition, which occurs in S phase) (8); somatic insertions indicate completed cycles of retrotransposition (Fig. 1*E*).

Overall, we found strong correlation among all three measures of LINE-1 activity: LINE-1 RNA, ORF1p, and somatic insertion. We also found strong correlations between ORF1p expression and phosphorylation (see next paragraph). However, we were not able to ascertain correlation with somatic insertion rate in breast, ovarian, and colon tumors as WGS was lacking in these cancer types. LINE-1 RNA and ORF1 protein correlate very well with each other (Fig. 2 *A*–*E*) except in kidney cancer (Fig. 2*D*), a lack of correlation attributable to low LINE-1 expression. Indeed, no more than one high-confidence somatic insertion was identified in any kidney sample analyzed here. A similar lack of retrotransposition was observed in The Cancer Genome Atlas sequenced clear cell kidney tumors (38). Given this lack of evidence for LINE-1 activity, we excluded kidney cancer from most of the subsequent analyses. For the other four cancer types, we found LINE-1 RNA/protein Spearman correlations (ρ values) ranging from 0.55 to 0.79. Despite the challenges associated with quantifying the expression of repetitive elements in both RNA and protein, LINE-1 RNA/protein correlations exceeded median cellular RNA/protein correlations, which were generally around ρ = 0.45 (*SI Appendix*, Fig. S1). We also found a significant correlation between LINE-1 ORF1 protein expression and somatic insertions in endometrial cancer after we removed low (<40%) purity tumors (ρ = 0.38, P = 0.004; Fig. 2*F*). Mobile element identification methods can have diminished sensitivity in low-purity tumors (39).

We consistently identified phosphorylation at three known sites in ORF1p: S18, S27, and T203 (40), as well as another site: S33. Except in two cases in which a particular phosphosite was detected never or rarely in a particular cancer type (T203 in colon cancer and S33 in endometrial cancer), these phosphosites correlated strongly with ORF1p abundance, with Spearman correlations mostly falling in the 0.5 to 0.7 range (*SI Appendix*, Fig. S2).

Overall the agreement between these LINE-1 quantifications indicates that we were able to accurately measure LINE-1 readouts in this dataset. The particularly high correlation between LINE-1 RNA and ORF1p supports the hypothesis that ORF1p binds *in cis* and protects its own RNA (*Discussion*). Given this concordance, we chose to primarily focus on the LINE-1 ORF1p quantification as a measure of LINE-1 expression. While LINE-1 RNAs are potentially transcribed from hundreds of different loci, they code for a highly concordant ORF1 protein sequence. Thus, we expect that the ORF1p quantification may be a simpler and more robust measure of LINE-1 expression than RNA, and somatic insertion, both of which present complex measurement challenges.

Analysis of LINE-1 RNA Showed the Highest Expression in Ovarian Cancer and High Expression of the "Hot" LINE-1 Located at 22q12.1 in the TTC28 Gene. Our ORF1p quantifications are relative to an internal standard that varies between cancer types and prevents

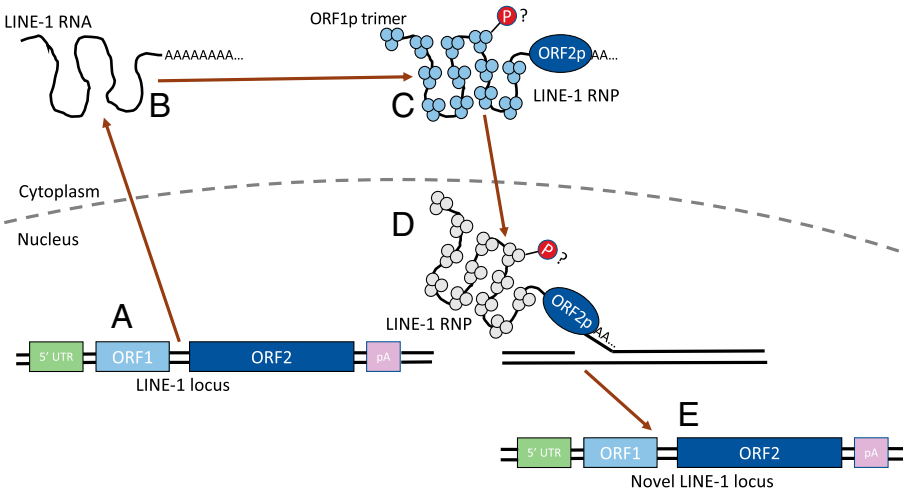

Fig. 1. Schematic showing the LINE-1 life cycle and the points at which LINE-1 products are quantified. (*A–E*) The path of retrotransposition from existing insertion (*A*) to new insertion (*E*). LINE-1 RNA may be loose (*B*), in a cytoplasmic RNP (*C*), or in a nuclear RNP (*D*). LINE-1 ORF1p may be in a LINE-1 RNP (*C* and *D*) or isolated (not shown in this drawing). ORF1p in the nuclear RNP (*D*) is shown in gray as it is present when the LINE-1 RNP enters the nucleus, but is removed before S phase. The timing of ORF1p phosphorylation is unknown, but likely occurs at the LINE-1 RNP stage (*C* and *D*) as it seems to play a role in retrotransposition. Novel insertions (*E*) indicate completed LINE-1 retrotransposition events.

straightforward comparison of different proteins in the same sample. Thus, we did not directly compare ORF1p expression across cancers. However, because read counts can be compared across genes and samples if properly normalized, it is more straightforward to compare RNA quantifications. We found that LINE-1 RNA is most abundant in ovarian cancer and least

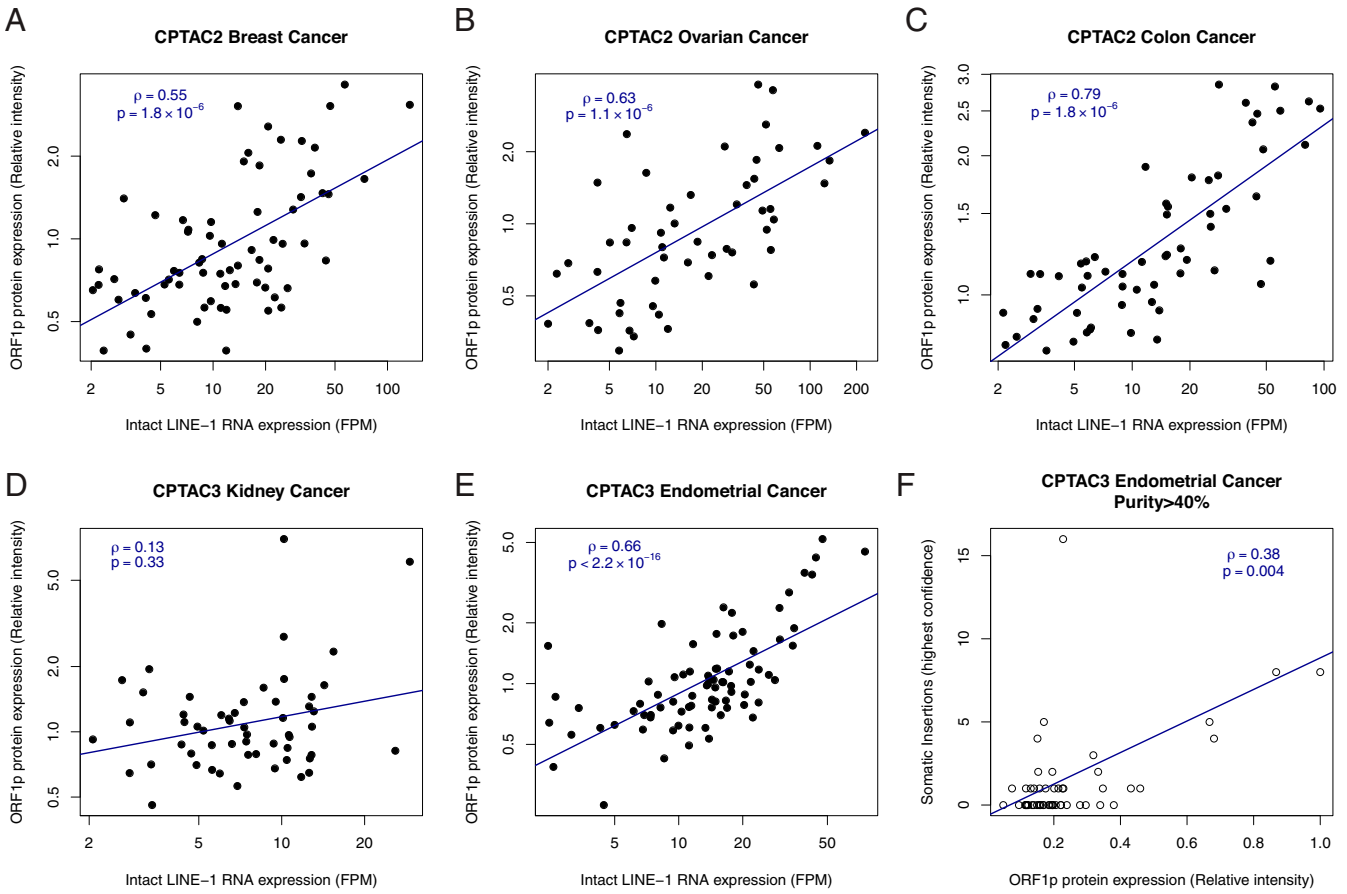

Fig. 2. Correlation between LINE-1 RNA, ORF1p, and insertion. (*A–E*) LINE-1 RNA/ORF1p correlation in breast, ovarian, colon, kidney, and endometrial cancers, respectively. (*F*) Correlation between LINE-1 ORF1p and high-confidence somatic insertions in endometrial tumors with purity >40% (sensitivity may suffer in low-purity tumors.) These high-confidence insertions may represent only a fraction of the total somatic insertions.

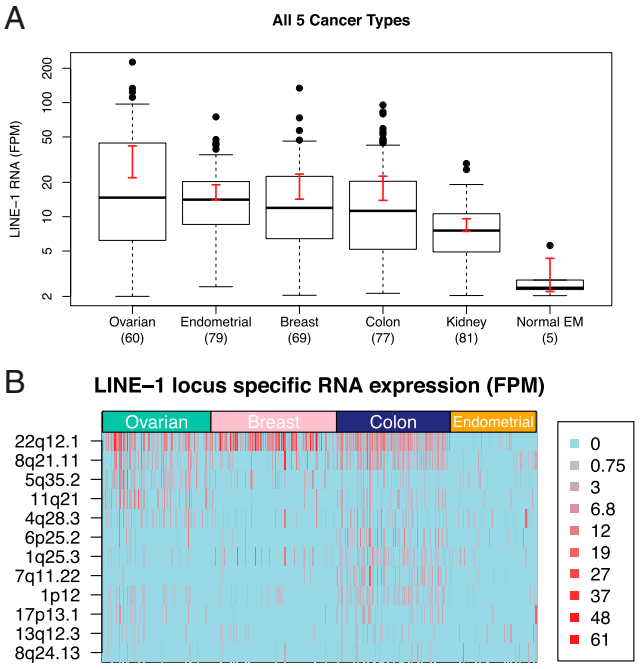

Fig. 3. Insights from analysis of LINE-1 RNA. (*A*) Comparison of LINE-1 RNA quantification across cancer types. Expression is highest in ovarian cancer. Normal EM, normal endometrium. (*B*) Highly expressed (hot) intact LINE-1 reference loci. Rows indicate an intact locus that is highly expressed (top five) in at least one cancer type. Columns are individual cases. Darker shading indicates greater expression. Across samples/cancers, the *TTC28* locus at 22q12.1 is most highly expressed, accounting for two-thirds of all intact LINE-1 expression in breast cancer.

abundant in clear cell kidney cancer (Fig. 3*A*). Some caution should be exercised as the RNA sequencing for endometrial and kidney cancer was done using shorter reads, but at greater depth. In the case of endometrial cancer, RNA from adjacent normal tissue was also available. We found intact LINE-1 RNA to be much less abundant in normal endometrial tissue than in any of the five cancer types considered here, including kidney (Fig. 3*A*).

Because the tool we used to quantify LINE-1 RNA, L1EM, assigns RNA levels to specific loci, we were able to compare locus-specific LINE-1 RNA expression between cancer types (Fig. 3*B*). However, it should first be noted that L1EM assigns all LINE-1 RNA expression to reference loci despite the presence of potentially active, polymorphic, nonreference LINE-1 loci that exist in a given tumor. Thus, some LINE-1 RNA reads assigned to a locus, x, may actually be derived from a nonreference locus (or loci) for which x is likely the immediate parent. In breast, ovarian, and colon, the most highly expressed intact (full-length and lacking nonsense mutations in ORF1 and ORF2) LINE-1 locus in the hg38 reference genome was located at 22q12.1, antisense to an intron of the *TTC28* gene. This locus has been previously pinpointed as a highly active (hot) LINE-1 in several cancer types (28, 39, 41). Among significantly expressed loci (two or more read pairs per million), this locus accounted for 27% of intact LINE-1 mRNA expression in ovarian cancer samples, 67% in breast cancer samples, and 30% in colon cancer. However, the 22q12.1 locus accounted for only about 1% of intact LINE-1 mRNA expression in endometrial cancer, where no single locus consistently dominates LINE-1 RNA expression. A locus at 4q28.3 (embedded in the lncRNA gene *RP11-775H9.2*) was the most highly expressed, yet it still accounted for only 4.5% of intact LINE-1 mRNA expression across the endometrial cancer samples.

LINE-1 ORF1 Protein Expression Is Positively Correlated with p53 Mutation, can, and DNA Replication Initiation Protein Complexes. Several studies have shown that LINE-1 expression and retrotransposition is higher in cancers with p53 mutations compared to those without (19, 28, 42). Furthermore, in several cancer types, LINE-1 retrotransposition (as measured by the number of detected somatic insertions) is correlated with structural genomic alteration and chromosomal instability (28). We first wanted to know whether LINE-1 expression (using ORF1p as a proxy) is also correlated with p53 mutation in these cohorts and whether the correlation between retrotransposition and chromosomal instability extends to expression as well. We found LINE-1 ORF1p expression to be about twice as high on average in p53 mutant endometrial cancers (Wilcoxon test $P = 0.0014$, Fig. 4*A*) and about 50% higher in p53 mutant breast cancers (Wilcoxon test $P = 0.011$, Fig. 4*B*), but we found no significant relationship between LINE-1 ORF1p expression and p53 mutation in colon cancer (Wilcoxon test $P = 0.2$, Fig. 4*C*). There was also no significant correlation in ovarian cancer, where p53 mutations are nearly universal. The correlation between LINE-1 ORF1p expression and CNA burden (average of the absolute value of GISTIC2 estimated CNA across the genome) was highest in endometrial cancer (Spearman $\rho = 0.44$, $P = 3.6 \times 10^{-5}$, Fig. 4*D*), but also significant in breast cancer ($\rho = 0.25$, $P = 0.015$, Fig. 4*E*) and ovarian cancer ($\rho = 0.22$, $P = 0.04$, Fig. 4*F*). There was no significant correlation between LINE-1 ORF1p expression and CNA burden in colon cancer ($\rho = 0.12$, $P = 0.31$). Even though the LINE-1 ORF1p/CNA correlations were highest in cancers that showed higher LINE-1 ORF1p expression in p53 mutant tumors, p53 mutation status alone was insufficient to explain these correlations. In endometrial cancer, restricting to p53 mutant tumors yielded a larger correlation ($\rho = 0.68$, $P = 0.002$, Fig. 4*D* teal points and letters), and in breast cancer the correlation was specific to p53 wild-type tumors ($\rho = 0.32$, $P = 0.017$, Fig. 4*E*, gold points and letters).

We then asked whether there are any classes of protein whose members are correlated with LINE-1 ORF1p expression across these four cancer types. To that end, we calculated Spearman correlations between LINE-1 ORF1p expression and the abundance of each protein (Dataset S2, correlations at false discovery rate [FDR] < 10% highlighted). Because we lacked the power to find genes correlated in all four cancer types, we performed gene set enrichment analysis (GSEA) (43) on the protein/protein correlation using Kyoto Encyclopedia of Genes and Genomes (KEGG) (44, 45), BioCarta (46), the Pathway Interaction Database (47), and Reactome (48) gene sets that encompass the encoded host cell proteins. This analysis yielded a single gene set that was enriched at FDR < 1% in all four cancers: DNA replication preinitiation (Reactome). Twelve additional gene sets were enriched in endometrial, ovarian, and colon cancers at FDR < 1%, but not in breast cancer. Ten of these were also enriched in the subset of breast cancers without p53 mutation—the subset that showed a LINE-1/CNA correlation (full GSEA results in Dataset S3). In particular, activation of ATR in response to replication stress (Reactome) was enriched for correlation with LINE-1 ORF1p expression in endometrial, ovarian, colon, and p53 wild-type breast cancers. This enrichment is driven largely by positive correlation with complexes involved in DNA replication initiation (Fig. 4*G*): Minichromosome maintenance (MCM) (previously shown to directly interact with LINE-1) (8), origin recognition complex (ORC), and replication factor C (RFC). These correlations may indicate that cells expressing more ORF1p fire more origins of replication and/or spend more time replicating their DNA. This could indicate either 1) slower DNA replication due to DNA damage or replication stress, or 2) cells are replicating their DNA more frequently (i.e., dividing more rapidly).

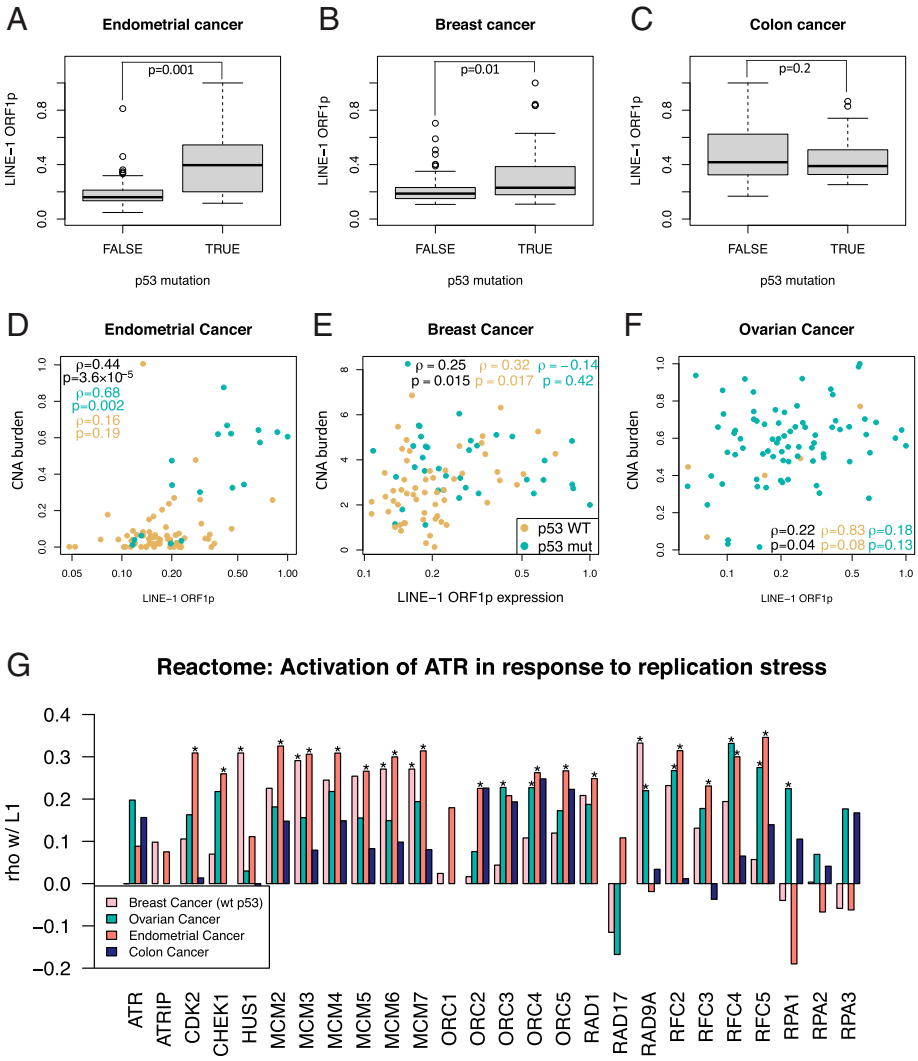

Fig. 4. (*A–C*) Box plots comparing LINE-1 expression in p53 wild-type (WT) versus mutated tumors for endometrial, breast, and colon tumors, respectively. (*D–F*) Spearman correlation between LINE-1 expression and CNA burden in endometrial, breast, and ovarian cancer, respectively. For the listed correlations and *P* values: black, all tumors; teal, p53 mutant tumors only; and gold, p53 wild-type tumors only. (*G*) Spearman correlation between LINE-1 expression and protein in the Reactome "Activation of ATR in response to replication stress" gene set. $*P < 0.05$.

In Endometrial Cancer, LINE-1 ORF1p Expression Is Associated with ATM-MRN-SMC S Phase Checkpoint Signaling. We next wanted to identify phosphorylation signaling pathways that are correlated with LINE-1 ORF1p expression. In particular, if LINE-1 expression is correlated with replication stress as potentially indicated by the protein and CNA correlations, we would expect to see a correlation between LINE-1 expression and checkpoint signaling. Therefore, for each of the four cancer types, we calculated Spearman correlations between LINE-1 ORF1p expression and each phosphosite identified in at least 50% of samples (Dataset S4). Because phosphoproteomics data are sparse, it is disappointing, but not surprising that we were only able to identify correlations that were significant after multiple hypothesis correction in one single tumor type: Endometrial cancer. The endometrial cancer data were generated as part of the newer CPTAC3 project and are of higher quality than the breast, ovarian, and colon data generated as part of the CPTAC2 project. Furthermore, some endometrial tumors express a high level of LINE-1, while others express little, making it easier to identify significant correlations in endometrial cancers. In total, 95 phosphosites were correlated (positively or negatively) with LINE-1 ORF1p expression in endometrial cancer at FDR < 5%. Six of these sites (all positively correlated) have known downstream functional consequences listed in PhosphoSitePlus (49). RAD50-S635 (29) and MDC1-S453 (50) phosphorylation are both required for the activation of S phase checkpoint in response to double-strand DNA breaks. CDC20-T70 phosphorylation is involved in mitotic exit (51) and RB1-T373 is involved in E2F release and entry into S phase (52). There were also two phosphosites on TOP2A correlated with LINE-1 ORF1p expression: S1106 phosphorylation improves TOP2A enzymatic activity (53) and S1247 is involved in targeting TOP2A to the centromere during mitotic prophase (54). MDC1, CDC20, and TOP2A protein levels were all positively correlated with LINE-1 ORF1p expression, so correlations with phoshphosites on these proteins could reflect expression rather than signaling. However, RAD50 and RB1 expression were not significantly correlated with LINE-1 ORF1p, so correlations with RAD50-S635 and RB1-T373 likely represent true changes in cell signaling.

Given the correlations with RAD50-S635 and MDC1-S453, we wanted to know whether there were additional phosphosites involved in double-strand break-induced S phase checkpoint that were correlated with LINE-1 ORF1p expression, but were

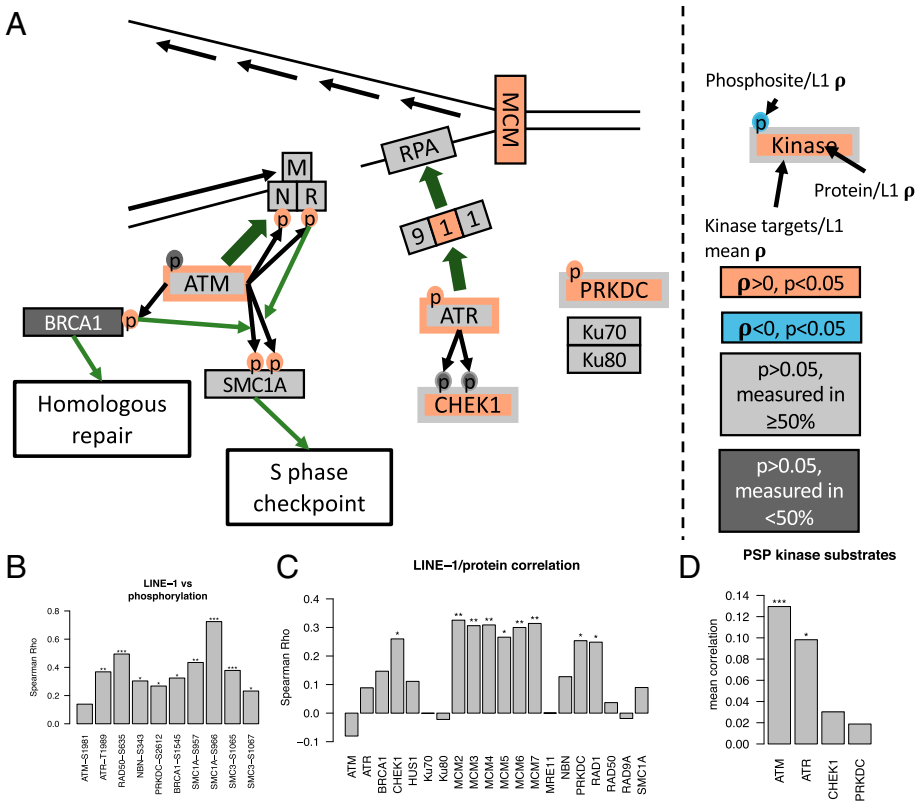

Fig. 5. LINE-1/DDR correlations in endometrial cancer. (*A*) A candidate model to explain the observed LINE-1/DDR correlations. At the *Top* is a broken replication fork, on the *Left* is the ATM-MRN-SMC pathway, and on the *Right* is the ATR-CHEK1 pathway. A protein or phosphosite filled in red indicates positive correlation at $P < 0.05$, gray indicates no significant correlation, with dark gray indicating at least 50% missing values. Kinases are outlined in red if there is significant enrichment for correlation between ORF1p and kinase targets, and gray if not. Big green arrows indicate recruitment; small green arrows indicate downstream effects. See *B–D* for the specific phosphosite/protein expression/kinase target correlations that support this model. (*B*) Spearman correlation between ORF1p and phosphosites supporting the ATM-MRN-SMC pathway being correlated with ORF1p expression. *$0.01 < P < 0.05$, **$0.001 < P < 0.01$, ***$P < 0.001$. (*C*) As in *D*, but for ORF1p/protein correlation. For the most part, phosphosite correlation cannot be explained by changes in expression. (*D*) KSEA enrichment for correlation between ORF1p- and DDR-related kinase targets. ATM activity is most closely correlated with LINE-1 expression.

not detected after multiple hypothesis correction. As part of the MRN complex, RAD50 is recruited to DSBs, where it is phosphorylated at S635 by ATM in response to ionizing radiation (IR). RAD50-S635 phosphorylation is necessary for ATM phosphorylation of the cohesin subunit SMC1A/SMC1 (29), which in turn is necessary for the activation of S phase checkpoint in response to IR (55, 56) (pathway shown in Fig. 5 *A*, *Left*). Indeed, ORF1p was strongly correlated with both known ATM-targeted phosphorylation sites on SMC1A ($P < 0.003$) (Fig. 5*B*) and with four additional [ST]Q (ATM consensus motif) core cohesin residues: Two on SMC1A and two on SMC3 (all $P < 0.05$). ATM phosphorylation of SMC1A also requires BRCA1 and NBN/NBS1 (56, 57). At $P < 0.03$, we observed a correlation between ORF1p and ATM-targeted phosphosites on both proteins: BRCA1-S1545 (S1524 in uniprot) (58) and NBN-S343 (59) (Fig. 5*B*). LINE-1 ORF1p expression was not correlated with substrate abundance in any of these cases, consistent with the conclusion that these correlations are indicative of increased S phase checkpoint signaling (Fig. 5*C*). Despite this strong evidence for an association between LINE-1 ORF1p expression and ATM signaling, we did not observe a significant correlation between ORF1p and ATM-S1981 autophosphorylation ($\rho = 0.14$, $P = 0.29$). ATM-S1981 autophosphorylation is a popular measure of ATM activation (60), although it may not be essential (61). We did however find enrichment for ATM activity using kinase set enrichment analysis (KSEA). The mean ORF1p/ATM site correlation was 0.13 ($P = 7.8 \times 10^{-5}$, Fig. 5*D*).

We next wanted to know whether there was any evidence for activation of the canonical ATR-CHEK1 S phase checkpoint pathway. In response to replication stress, the RAD9A-HUS1-RAD1 (9-1-1) complex recruits ATR, which activates S phase checkpoint by phosphorylating CHEK1/CHK1 (Fig. 5 *A*, *Right*) (62). We found a strong correlation between ORF1p and ATR-T1989 autophosphorylation ($\rho = 0.37$, $P = 0.0025$). Given the overlapping motifs of ATM and ATR and the presence of a significant correlation with ATR but not ATM autophosphorylation, the MRN/SMC sites described above may be phosphorylated by ATR rather than ATM in this context. However, we found a much more modest enrichment for correlation with ATR target sites (Fig. 5*D*), and other analyses do point to a role for ATM in the response to LINE-1 (24, 63). We were not able to assay ATR phosphorylation of CHEK1 due to a lack of phosphosite detection, but we did not find an enrichment in the phosphorylation of CHEK1 targets (Fig. 5*D*). We also looked at PRKDC signaling in the nonhomologous end joining (NHEJ) pathway. While we did identify a modest correlation between LINE-1 ORF1p and PRKDC autophosphorylation at S2612 ($\rho = 0.27$, $P = 0.015$), we did not see enrichment of correlation with PRKDC target sites. We also did not see correlations between ORF1p and the NHEJ promoting Ku complex (Fig. 5*C*). Together these data are consistent with the hypothesis that

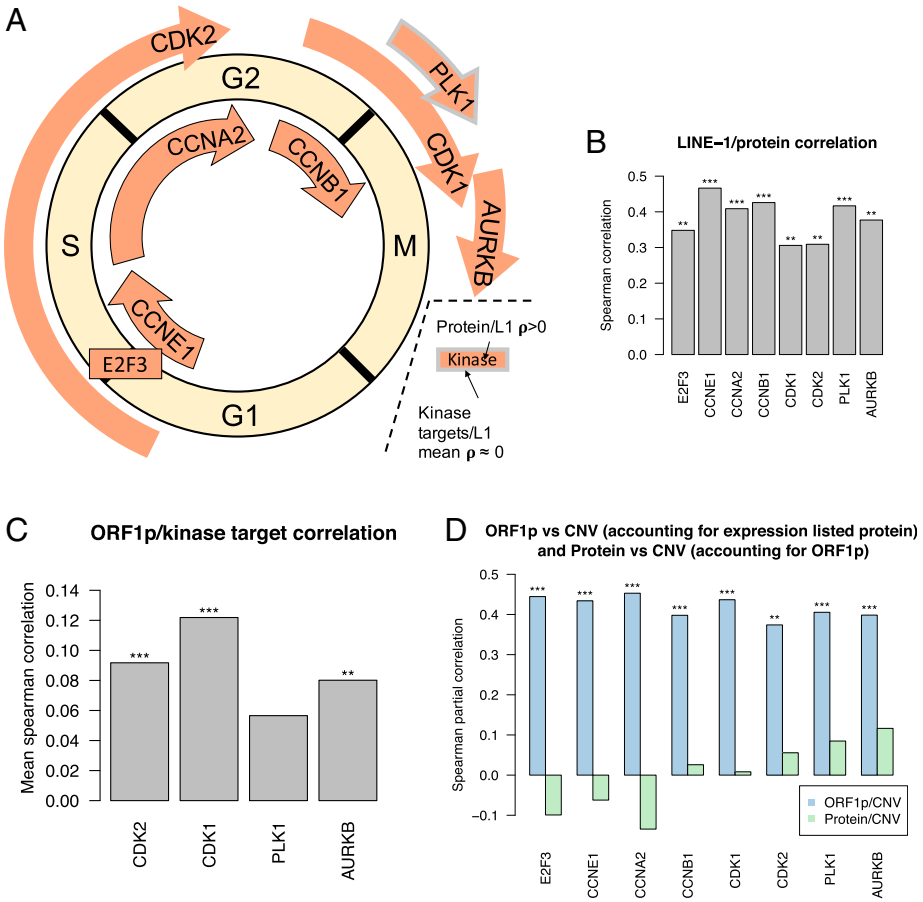

Fig. 6. LINE-1/cell cycle correlations in endometrial cancer. (*A*) Drivers of cell cycle progression correlated with ORF1p, all at $P < 0.01$. Kinases are outlined in red if there is significant enrichment for correlation between ORF1p and kinase targets, and gray if not. (*B*) Spearman correlation values for proteins shown in *A*. **$0.001 < P < 0.01$, ***$P < 0.001$. (*C*) KSEA for kinases shown in *A*. (*D*) Partial correlation between ORF1p and CNA accounting for the level of these proteins (separately) and partial correlation between these proteins and CNA accounting for ORF1p levels. Correlation between ORF1p and CNA does not depend on these proteins, but correlation between these proteins and CNA does depend on ORF1p.

ATM-MRN-SMC is the primary DNA damage response (DDR) pathway that is associated with LINE-1 expression in endometrial cancer.

The above correlations suggest a connection between LINE-1 expression and DNA damage, replication stress, and CNA. However, the correlation with RB1 hints at an alternate explanation. Phosphorylation of RB1 causes it to detach from E2F transcription factors, allowing E2F to relocate to chromatin and activate its transcriptional program. Many DNA replication genes (including the MCM complex) are targeted by E2F (64). Thus, it is possible that LINE-1 expression is indirectly related to CNA through checkpoint loss. Indeed, cell cycle dysregulation/checkpoint loss has been suggested as a potential factor contributing to CNA in endometrial cancer (34).

We therefore wanted to know whether the LINE-1 ORF1p/CNA correlation can be explained by cell cycle–related factors. In endometrial cancer, we did find strong correlations between LINE-1 ORF1p expression and markers of cell cycle progression, including: Cyclins E1, A2, and B1, as well as CDK1/2 expression and phosphorylation, PLK1, and AURKB (Fig. 6 *A*–*C*). So, we applied partial correlation, which uses linear regression to subtract out the effect of a confounding variable (in this case the cell cycle–related proteins and phosphosites) from two variables (in this case LINE-1 ORF1p and CNA burden) and then tests the residual correlation. We found that the LINE-1 ORF1p/CNA correlation remained significant after separately accounting for cyclins E1, A2, or B1 or CDK1/2, but that these proteins were not correlated with CNA after taking ORF1p levels into account (Fig. 6*D*). This suggests that the relationship between LINE-1 and CNA burden is not mediated by the expression of these proteins, but the relationship between the expression of these proteins and CNA burden may be mediated by ORF1p expression.

Induction of LINE-1 Expression Leads to Increased RAD50-S635 Phosphorylation. The fact that phosphosites involved in ATM-MRN-SMC S phase checkpoint activation were correlated with LINE-1 expression in endometrial cancer suggests a connection between LINE-1 expression and DNA damage/replication stress. However, we wanted to address whether inducing LINE-1 expression in cells actually leads to increased phosphorylation of targets in the ATM-MRN-SMC S phase checkpoint pathway. We chose to focus on RAD50-S635 as it showed the highest correlation with LINE-1 expression. Namely, we performed immunoblotting analysis in p53-deficient RPE-1 cells harboring doxycycline (Dox)-inducible LINE-1 expression construct treated with dimethylsulfoxide (DMSO) vehicle control or Dox. As genomically normal diploid cells, RPE-1 cells are quite different from the endometrial cancer tumors in which we observed these correlations. However, RPE-1 cells provide a

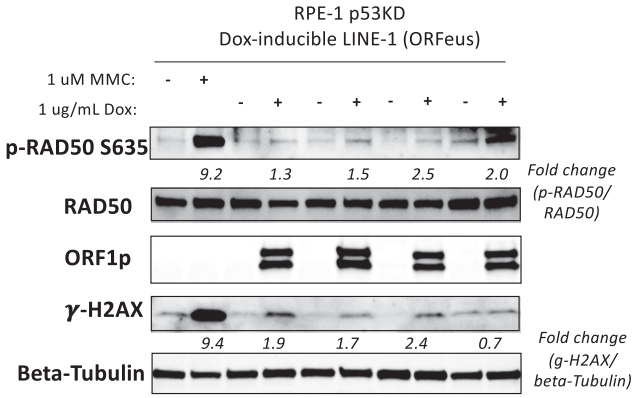

Fig. 7. RAD50-S635 phosphorylation in response to induction of LINE-1 expression. Shown are immunoblotting analyses of p-RAD50 S635, RAD50, ORF1p, and ɣ-H2AX in whole cell lysates from p53-deficient RPE-1 cells harboring Dox-inducible LINE-1 expression construct treated with DMSO control or Dox (1 μg/mL) for 5 d. Immunoblotting analysis of beta-tubulin is included as a loading control. Also, analysis of p53-deficient RPE-1 cells treated with DMSO or DNA damaging agent mitomycin C (MMC) are included as a positive control for DNA damage. The numbers shown are quantification of the fold change (versus DMSO) for p-RAD50/total RAD50 or ɣ-H2AX/beta-tubulin from two technical replicates.

background without LINE-1 expression and minimal DDR activity, allowing us to see modest effects that might be confounded in cancer cell lines with highly dysregulated DDR. These RPE-1 cells also overexpress hTERT. While a relationship between hTERT expression and LINE-1 expression has been suggested (65), we do not anticipate that this provides much effect in this context in which both hTERT and LINE-1 are overexpressed.

We found that cells treated with Dox versus control showed a modest increase (1.8×, *t* test on logarithm [log] fold change $P = 0.048$) in phosphorylated RAD50-S635 levels (Fig. 7), but no notable change in RAD50 substrate levels, supporting our model that induction of LINE-1 expression increases phosphorylated RAD50-S635 levels. We also found that in three out of four replicates, the cells with induced LINE-1 expression also showed an increase in ɣ-H2AX (phosphorylation of H2AX-S139), a marker of DNA double-strand breaks (66) that has been previously correlated with LINE-1 induction (24).

Discussion

Correlation between Measures of LINE-1. One might expect that because of the challenges associated with quantifying the expression of highly repetitive elements such as LINE-1, the LINE-1 RNA/protein correlation would be weaker than those for most host genes. Observing that the LINE-1 RNA/ORF1 protein correlation was larger than it is for most genes would then be surprising. However, ORF1p binds *in cis* to its own RNA (67), likely protecting it from degradation. If naked LINE-1 RNAs are highly susceptible to degradation, most LINE-1 RNA would be present in complex with ORF1p, potentially explaining the elevated RNA/protein correlation. Correlation between LINE-1 RNA and ORF2p or between ORF1p and ORF2p is less clear. Endogenous ORF2p is extremely difficult to measure (37) and was at best sparsely observed in our analysis. However, we did find that LINE-1 ORF1p expression is correlated with the number of somatic insertions in endometrial cancer. Because ORF2p is critical to retrotransposition, this suggests that it is also correlated with somatic insertions and thus with ORF1p, at least in endometrial cancer. Translational regulation of ORF2p is well reported (27, 68, 69) and WGS was limited, so such a correlation may not be universal. It is also possible to have ORF1p expression from loci that lack an intact ORF2. However, we find ORF1p to be similarly correlated with estimates of both intact and ORF1-only LINE-1 RNA expression (*SI Appendix*, Fig. S1).

Correlations with p53 Mutation, CNA, Replication Stress, and DNA Damage. Our identification of a correlation between p53 mutation and LINE-1 ORF1p expression reflects studies that have found higher LINE-1 expression and retrotransposition in p53 mutant tumors (19, 28, 42). Furthermore, LINE-1 overexpression is lethal to cells with intact p53 (26). Thus, there is likely a strong selective pressure against cells with derepressed LINE-1 and intact p53. Alternatively, or in addition, p53 may be directly involved in the silencing of LINE-1 (42, 70). We also found a positive correlation between LINE-1 ORF1p expression and CNA burden in endometrial, ovarian, and breast cancer. This reflects a previous study that found positive correlation between LINE-1 retrotransposition and structural genomic alteration burden in many cancer types (28). Several lines of evidence indicate that LINE-1 overexpression promotes DNA damage and replication stress (24, 26, 27). This could in turn be responsible for the increased CNA burden in LINE-1–expressing tumors.

However, because p53 mutant tumors have both higher LINE-1 expression and greater CNA burden, p53 is a potential confound in this analysis. We addressed this by showing that LINE-1 is still correlated with CNA burden in endometrial cancer if we restrict analysis to only p53 mutant tumors and in breast cancer if we restrict to only p53 wild-type tumors. This shows that the correlation between LINE-1 ORF1p expression and CNA burden cannot be explained by p53 mutation status alone. However, binary p53 mutation status is not a perfect measure of p53 pathway activity, so it is impossible to fully discount it as a confound. Additional confounds may also exist. Most endometrial tumors have low LINE-1 expression and low CNA burden, but a subset (dubbed CNV high) has high LINE-1 expression and high CNA burden. There are many potential third factors that could be the determinant(s) of these subtypes and thus be independently responsible for both LINE-1 expression and CNA burden. This includes several cell cycle–related genes with the potential to drive CNA. We addressed this by using partial correlation to show that none of these genes can fully explain the correlation between LINE-1 expression and CNA burden.

We then looked for the proteomic and phosphoproteomic correlates of LINE-1 ORF1p expression to identify the specific pathways associated with LINE-1 expression in these tumors. In endometrial cancer, we found the ATM-MRN-SMC S phase checkpoint pathway activation to be correlated with LINE-1 ORF1p expression. We were then able to show that induction of LINE-1 expression in RPE-1 cells with p53 knockdown leads to increased RAD-S635 phosphorylation, a key step in this pathway that is highly correlated with LINE-1 expression in endometrial cancer. This suggests that LINE-1 expression can lead to DNA damage signaling, at least in certain contexts.

Conclusion. Our study shows that LINE-1 expression can be consistently measured in large multiomic cancer datasets. We leveraged these measurements to find in vivo evidence that LINE-1 expression is correlated with replication stress and S phase checkpoint signaling and validated this relationship by inducing LINE-1 expression. This indicates that LINE-1 expression can promote DNA damage and replication stress, at least in certain contexts, and that this activity may have an impact on human tumors, potentially explaining correlations between LINE-1 and CNA. The correlations described in this study are unlikely to be exhaustive, especially for the phosphorylation data. Most

phosphosites have unknown functional consequences, and a larger cohort is likely to reveal additional pathways, especially in breast and ovarian tumors.

Methods

Quantification of LINE-1 RNA. LINE-1 RNA was quantified from available RNA-sequencing (RNA-seq) data using an implementation of L1EM (35) on the Cancer Genomics Cloud. L1EM employs the expectation maximization algorithm to estimate locus-specific LINE-1 expression and to separate proper LINE-1 expression from passive cotranscription that includes LINE-1 RNA, but does not support retrotransposition. For intact LINE-1 RNA quantifications, full-length LINE-1 loci with no stop codon in either ORF1 or ORF2 were considered expressed if at least two read pairs per million (FPM) were assigned to that locus and less than 10% of the RNA assigned to that locus was estimated to be passive cotranscription. Total intact LINE-1 RNA expression was estimated by adding together the FPM values for each such locus. Full-length loci with stop codons in ORF2 but not ORF1 were also included to generate an ORF1 RNA expression estimate. Samples were excluded if no locus was detected at 2 FPM as this may be due to data quality rather than a lack of LINE-1 RNA. Specifically, ovarian cancer samples that failed LINE-1 RNA quantification did not have lower ORF1p quantifications.

Quantification of LINE-1 ORF1p. LINE-1 ORF1p was quantified from isobaric labeled tandem mass spectrometry (MS/MS) proteomics data using a set of 20 proteotypic ORF1p peptides generated from analysis of older CPTAC breast (71) and ovarian cancer (72) data. X!Tandem (73) was used to search mass spectra against a combined database that includes both the standard ensembl human proteome and in silico translations of intact LINE-1 open reading frames in the human reference genome. Oxidation of methionine (+15.994915@M) was included as a potential modification. Carbamidomethylation of cysteine (+57.022@C) was set as a fixed modification in addition to modifications appropriate to the particular isobaric labeling used (+144.102063@[, +144.102063@K for iTRAQ4 and +229.162932@[, +229.162932@K for TMT10). Peptides were quantified by calculating the log ratio between the reporter intensity for each sample and the reporter intensity of an internal control. Peptide/peptide spearman correlations were calculated and peptides that had a correlation of at least 0.6 with two other peptides were retained (*SI Appendix*, Fig. S3). This led to the following list: DFVTTRPALK, EALNMER, EWGPIFNILK, LIGVPESDGENGTK, LIGVPESDVENGTK, LSFISEGEIK, LTADLSAETLQAR, NLEECITR, NVQIQEIQR, QANVQIQEIQR, REWGPIFNILK, RNEQSLQEIWDYVK, SNYSELR, SNYSELREDIQTK, VSAMEDEMNEMK, YQPLQNHAK, DFVTT RPALQELLK, LENTLQDIIQENFPNLAR, VSAMEDEMNEMKR, and NEQSLQEIWDYVK.

For the actual peptide quantifications, the X!Tandem search and peptide quantifications were performed as above. If a peptide were detected multiple times in the same sample, the median across peptide spectral matches was used. Then to get a protein quantification, the median was taken across all peptides in the preceding list that were identified in that sample. Finally, the quantification was translated by the median log ratio for all human proteins to account for variation in the size of the input sample. This pipeline was implemented and executed as a workflow on the Cancer Genomics Cloud.

Quantification of LINE-1 ORF1p Phosphorylation. ORF1p phosphopeptides were identified using the above X!Tandem search strategy with the addition of phosphorylation at serine, threonine, and tyrosine (79.966331@S, 79.966331@T, and 79.966331@Y) as potential modifications. Identified ORF1p phosphopeptides were aligned to the UniProt LINE-1 ORF1p sequence (L1RE1) using the pairwise2 method in Biopython to identify the phosphorylation coordinate. Phosphopeptides were quantified as above using the log ratio between the sample reporter intensity and the reference reporter intensity. If multiple phosphopeptide spectral matches indicated the same phosphosite, the median was taken.

Quantification of High-Confidence LINE-1 Somatic Insertions. Somatic insertions were identified using MELT v2.1.15 (36) on the Cancer Genomics Cloud. MELT was run on each pair of matched cancer/normal WGS datasets. Only insertions with the greatest evidence (ASSESS = 5) were considered. To be considered a high-confidence somatic insertion, the insertion had to be called heterozygous in cancer and homozygous absent in normal, with the log likelihood of both of these genotypes being at least 10 greater than the log likelihood of the next most probable genotype.

Estimation of CNA Burden. For breast and endometrial cancer, CNA burden scores were provided by the respective CPTAC working groups (30, 34). For colon cancer, the CNA score was recalculated according to the description provided (32). For ovarian cancer, a global CNA score was calculated from gene level CNA provided by the CPTAC working group (31). First, any gene overlapping another gene with a smaller leftmost coordinate was removed. Then, each remaining gene coordinate range was extended in the plus direction to reach the start of the next gene. The global CNA score was then a linear combination of the absolute value of the gene level CNA with the length of the extended gene ranges.

Enrichment Analysis. For GSEA, Spearman correlations were calculated between our LINE-1 ORF1p quantification and the log normalized quantification for each identified protein identified in at least half of the samples. Analysis was then performed using the GSEAPreranked option in GSEA 4.0.3.

For KSEA, we calculated partial Spearman correlation between ORF1p and each phosphosite that was identified in at least half of the samples, accounting for the substrate protein quantification. Enrichment was calculated using the KSEA web app (74).

Statistical Analysis. Statistical analysis was performed in R. Correlation was calculated using the cor.test function in the standard stats package. Partial correlation was calculated using pcor.test in the ppcor package. All correlation tests (partial and normal) were made using the Spearman (rank) method. For the ORF1p/CNV partial correlation analysis, each of the cell cycle genes were considered individually in separate partial correlation analyses. The Benjamini–Hochberg (BH) method was used to calculate FDR/q value when considering multiple hypothesis tests. Uncorrected *P* values were used when literature or analysis of another cancer type pointed to potential involvement of a specific gene/protein/phosphosite.

Validation of RAD50-S635 Phosphorylation in Response to LINE-1 Expression. We used hTERT-RPE-1^{PuroS}-Cas9 cells harboring Tet-inducible codon-optimized LINE-1 (ORFeus), which were previously characterized (26). For immunoblotting analysis, cells were treated with DMSO or Dox (1 μg/mL) for 5 d, or DMSO, or 1 μM mitomycin C (MMC, SC-3514) for 48 h, followed by protein extraction using radioimmunoprecipitation assay buffer (Boston BioProducts BP-115) supplemented with protease and phosphatase inhibitors (Cell Signaling, 5872S). Gel electrophoresis was performed on protein extracts using 4 to 20% Mini-PROTEAN TGX gels (Bio-Rad, 456-1095). Proteins were then transferred to low fluorescence polyvinylidene difluoride membranes using Trans-Blot Turbo (Bio-Rad). Membranes were blocked using EveryBlot Blocking Buffer (Bio-Rad, 12010020) or Intercept Blocking Buffer (Li-COR, 927-60001), and probed with primary antibodies for RAD50-S635 (Cell Signaling, 14223S), RAD50 (Cell Signaling, 3427T), ORF1p (Millipore Sigma, MABC1152), beta-tubulin (Cell Signaling, 2128S), and gH2AX (Cell Signaling, 2577S), followed by secondary antibodies (IRDye 800CW goat anti-mouse IgG, 925-32210); IRDye 680RD goat anti-rabbit IgG, 925-68071; and anti-rabbit IgG horseradish peroxidase (HRP), 7074S). ECL substrate (Thermo Scientific, 34580) was used to develop HRP signals. Immunoblotting signals were detected using the Chemi-Doc imaging system (Bio-Rad).

Data Availability. Proteomic and phosphoproteomic data are available through the Proteomic Data Commons (https://pdc.cancer.gov/pdc/). Transcriptomic and genomic data for CPTAC3 (endometrial and kidney cancers) can be found in the Genomic Data Commons (GDC) (https://gdc.cancer.gov/). Additional data for each cancer type can be found in the corresponding publication: Colon (32), breast (30), ovarian (31), endometrial (34), and kidney (33). L1EM code can be found at https://github.com/FenyoLab/L1EM. CGC implementation of L1EM and ORF1p quantification are available on request (please provide a CGC username). L1EM can also be accessed directly on the CGC through the public apps portal: https://cgc.sbgenomics.com/public/apps/whm240/l1em-commit/l1em-cptac3-workflow. MELT can be downloaded from its home page: https://melt.igs.umaryland.edu/. Version 2.1.5 was used in this study.

Previously published data were used for this work (genomic and transcriptomic data were accessed from the GDC: https://gdc.cancer.gov/. Proteomic and phosphoproteomic data were accessed from the CPTAC data portal: https://cptac-data-portal.georgetown.edu/).

ACKNOWLEDGMENTS. This project has been funded in whole or in part with federal funds from the NCI, NIH, under Contract No. HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government. Additional funding was provided by NIH grants P01AG051449 (National Institute on Aging subcontracts to J.D.B. and D.F.), U24CA210972 (NCI to D.F.), and 1R21CA235521 (NCI to J.D.B.).

1. S. L. Martin, The ORF1 protein encoded by LINE-1: Structure and function during L1 retrotransposition. *J. Biomed. Biotechnol.* **2006**, 45621 (2006).
2. E. Khazina, O. Weichenrieder, Human LINE-1 retrotransposition requires a metastable coiled coil and a positively charged N-terminus in L1ORF1p. *eLife* **7**, e34960 (2018).
3. Q. Feng, J. V. Moran, H. H. Kazazian Jr., J. D. Boeke, Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87**, 905–916 (1996).
4. S. L. Mathias, A. F. Scott, H. H. Kazazian Jr., J. D. Boeke, A. Gabriel, Reverse transcriptase encoded by a human transposable element. *Science* **254**, 1808–1810 (1991).
5. A. J. Doucet, J. E. Wilusz, T. Miyoshi, Y. Liu, J. V. A. Moran, A 3′ poly(A) tract is required for LINE-1 retrotransposition. *Mol. Cell* **60**, 728–741 (2015).
6. D. D. Luan, M. H. Korman, J. L. Jakubczak, T. H. Eickbush, Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: A mechanism for non-LTR retrotransposition. *Cell* **72**, 595–605 (1993).
7. G. J. Cost, Q. Feng, A. Jacquier, J. D. Boeke, Human L1 element target-primed reverse transcription in vitro. *EMBO J.* **21**, 5899–5910 (2002).
8. P. Mita *et al.*, LINE-1 protein localization and functional dynamics during the cell cycle. *eLife* **7**, e30058 (2018).
9. R. Beraldi, C. Pittoggi, I. Sciamanna, E. Mattei, C. Spadafora, Expression of LINE-1 retroposons is essential for murine preimplantation development. *Mol. Reprod. Dev.* **73**, 279–287 (2006).
10. H. Kano *et al.*, L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes Dev.* **23**, 1303–1312 (2009).
11. M. Percharde *et al.*, A LINE1-nucleolin partnership regulates early development and ESC identity. *Cell* **174**, 391–405.e19 (2018).
12. F. Yang, P. J. Wang, Multiple LINEs of retrotransposon silencing mechanisms in the mammalian germline. *Semin. Cell Dev. Biol.* **59**, 118–125 (2016).
13. V. P. Belancio, A. M. Roy-Engel, R. R. Pochampally, P. Deininger, Somatic expression of LINE-1 elements in human tissues. *Nucleic Acids Res.* **38**, 3909–3922 (2010).
14. F. C. Navarro *et al.*, TeXP: Deconvolving the effects of pervasive and autonomous transcription of transposable elements. *PLOS Comput. Biol.* **15**, e1007293 (2019).
15. M. Van Meter *et al.*, SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age. *Nat. Commun.* **5**, 5011 (2014).
16. M. De Cecco *et al.*, L1 drives IFN in senescent cells and promotes age-associated inflammation. *Nature* **566**, 73–78 (2019).
17. M. Simon *et al.*, LINE1 derepression in aged wild-type and SIRT6-deficient mice drives inflammation. *Cell Metab.* **29**, 871–885.e5 (2019).
18. V. Gorbunova *et al.*, The role of retrotransposable elements in ageing and age-associated diseases. *Nature* **596**, 43–53 (2021).
19. N. Rodić *et al.*, Long interspersed element-1 protein expression is a hallmark of many human cancers. *Am. J. Pathol.* **184**, 1280–1286 (2014).
20. D. Ardeljan, M. S. Taylor, D. T. Ting, K. H. Burns, The human long interspersed element-1 retrotransposon: An emerging biomarker of neoplasia. *Clin. Chem.* **63**, 816–822 (2017).
21. Y. Miki *et al.*, Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res.* **52**, 643–645 (1992).
22. E. C. Scott *et al.*, A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer. *Genome Res.* **26**, 745–755 (2016).
23. T. Cajuso *et al.*, Retrotransposon insertions can initiate colorectal cancer and are associated with poor survival. *Nat. Commun.* **10**, 4022 (2019).
24. S. L. Gasior, T. P. Wakeman, B. Xu, P. L. Deininger, The human LINE-1 retrotransposon creates DNA double-strand breaks. *J. Mol. Biol.* **357**, 1383–1393 (2006).
25. A. Haoudi, O. J. Semmes, J. M. Mason, R. E. Cannon, Retrotransposition-competent human LINE-1 induces apoptosis in cancer cells with intact p53. *J. Biomed. Biotechnol.* **2004**, 185–194 (2004).
26. D. Ardeljan *et al.*, Cell fitness screens reveal a conflict between LINE-1 retrotransposition and DNA replication. *Nat. Struct. Mol. Biol.* **27**, 168–178 (2020).
27. P. Mita *et al.*, BRCA1 and S phase DNA repair pathways restrict LINE-1 retrotransposition in human cells. *Nat. Struct. Mol. Biol.* **27**, 179–191 (2020).
28. B. Rodriguez-Martin *et al.*; PCAWG Structural Variation Working Group; PCAWG Consortium, Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. *Nat. Genet.* **52**, 306–319 (2020).
29. M. Gatei *et al.*, ATM protein-dependent phosphorylation of Rad50 protein regulates DNA repair and cell cycle control. *J. Biol. Chem.* **286**, 31542–31556 (2011).
30. K. Krug *et al.*; Clinical Proteomic Tumor Analysis Consortium, Proteogenomic landscape of breast cancer tumorigenesis and targeted therapy. *Cell* **183**, 1436–1456.e31 (2020).
31. J. E. McDermott *et al.*; Clinical Tumor Analysis Consortium, Proteogenomic characterization of ovarian HGSC implicates mitotic kinases, replication stress in observed chromosomal instability. *Cell Rep Med* **1**, 100004 (2020).
32. S. Vasaikar *et al.*; Clinical Proteomic Tumor Analysis Consortium, Proteogenomic analysis of human colon cancer reveals new therapeutic opportunities. *Cell* **177**, 1035–1049.e19 (2019).
33. D. J. Clark *et al.*; Clinical Proteomic Tumor Analysis Consortium, Integrated proteogenomic characterization of clear cell renal cell carcinoma. *Cell* **179**, 964–983.e31 (2019).
34. Y. Dou *et al.*; Clinical Proteomic Tumor Analysis Consortium, Proteogenomic characterization of endometrial carcinoma. *Cell* **180**, 729–748.e26 (2020).
35. W. McKerrow, D. Fenyö, L1EM: A tool for accurate locus specific LINE-1 RNA quantification. *Bioinformatics* **36**, 1167–1173 (2020).
36. E. J. Gardner *et al.*, The Mobile Element Locator Tool (MELT): Population-scale mobile element discovery and biology. *Genome Res.* **27**, 1916–1929 (2017).
37. D. Ardeljan *et al.*, LINE-1 ORF2p expression is nearly imperceptible in human cancers. *Mob. DNA* **11**, 1 (2019).
38. E. Helman *et al.*, Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. *Genome Res.* **24**, 1053–1063 (2014).
39. H. Jung, J. K. Choi, E. A. Lee, Immune signatures correlate with L1 retrotransposition in gastrointestinal cancers. *Genome Res.* **28**, 1136–1146 (2018).
40. P. R. Cook, C. E. Jones, A. V. Furano, Phosphorylation of ORF1p is required for L1 retrotransposition. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 4298–4303 (2015).
41. J. M. C. Tubio *et al.*; ICGC Breast Cancer Group; ICGC Bone Cancer Group; ICGC Prostate Cancer Group, Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science* **345**, 1251343 (2014).
42. A. Wylie *et al.*, p53 genes function to restrain mobile elements. *Genes Dev.* **30**, 64–77 (2016).
43. A. Subramanian *et al.*, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005).
44. M. Kanehisa, S. Goto, KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
45. M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, K. Morishima, KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45** (D1), D353–D361 (2017).
46. D. Nishimura, BioCarta. *Biotech Softw. Internet Rep.* **2**, 117–120 (2001).
47. C. F. Schaefer *et al.*, PID: The pathway interaction database. *Nucleic Acids Res.* **37**, D674–D679 (2009).
48. B. Jassal *et al.*, The reactome pathway knowledgebase. *Nucleic Acids Res.* **48** (D1), D498–D503 (2020).
49. P. V. Hornbeck *et al.*, PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations. *Nucleic Acids Res.* **43**, D512–D520 (2015).
50. L. Wu, K. Luo, Z. Lou, J. Chen, MDC1 regulates intra-S-phase checkpoint by targeting NBS1 to DNA double-strand breaks. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 11200–11205 (2008).
51. J. B. Hein, E. P. T. Hertz, D. H. Garvanska, T. Kruse, J. Nilsson, Distinct kinetics of serine and threonine dephosphorylation are essential for mitosis. *Nat. Cell Biol.* **19**, 1433–1440 (2017).
52. J. R. Burke, A. J. Deshong, J. G. Pelton, S. M. Rubin, Phosphorylation-induced conformational changes in the retinoblastoma protein inhibit E2F transactivation domain binding. *J. Biol. Chem.* **285**, 16286–16293 (2010).
53. K. Chikamori *et al.*, Phosphorylation of serine 1106 in the catalytic domain of topoisomerase II alpha regulates enzymatic activity and drug sensitivity. *J. Biol. Chem.* **278**, 12696–12702 (2003).
54. M. Antoniou-Kourounioti, M. L. Mimmack, A. C. G. Porter, C. J. Farr, The impact of the C-terminal region on the interaction of topoisomerase II alpha with mitotic chromatin. *Int. J. Mol. Sci.* **20**, 1238 (2019).
55. S.-T. Kim, B. Xu, M. B. Kastan, Involvement of the cohesin protein, Smc1, in Atm-dependent and independent responses to DNA damage. *Genes Dev.* **16**, 560–570 (2002).
56. P. T. Yazdi *et al.*, SMC1 is a downstream effector in the ATM/NBS1 branch of the human S-phase checkpoint. *Genes Dev.* **16**, 571–582 (2002).
57. R. Kitagawa, C. J. Bakkenist, P. J. McKinnon, M. B. Kastan, Phosphorylation of SMC1 is a critical downstream event in the ATM-NBS1-BRCA1 pathway. *Genes Dev.* **18**, 1423–1438 (2004).
58. M. Gatei *et al.*, Ataxia telangiectasia mutated (ATM) kinase and ATM and Rad3 related kinase mediate phosphorylation of Brca1 at distinct and overlapping sites. In vivo assessment using phospho-specific antibodies. *J. Biol. Chem.* **276**, 17276–17280 (2001).
59. D. S. Lim *et al.*, ATM phosphorylates p95/nbs1 in an S-phase checkpoint pathway. *Nature* **404**, 613–617 (2000).
60. C. J. Bakkenist, M. B. Kastan, DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature* **421**, 499–506 (2003).
61. M. Pellegrini *et al.*, Autophosphorylation at serine 1987 is dispensable for murine Atm activation in vivo. *Nature* **443**, 222–225 (2006).
62. E. R. Parrilla-Castellar, S. J. H. Arlander, L. Karnitz, Dial 9-1-1 for DNA damage: The Rad9-Hus1-Rad1 (9-1-1) clamp complex. *DNA Repair (Amst.)* **3**, 1009–1014 (2004).
63. N. G. Coufal *et al.*, Ataxia telangiectasia mutated (ATM) modulates long interspersed element-1 (L1) retrotransposition in human neural stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20382–20387 (2011).
64. B. Ren *et al.*, E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. *Genes Dev.* **16**, 245–256 (2002).
65. T. Aschacher *et al.*, LINE-1 induces hTERT and ensures telomere maintenance in tumour cell lines. *Oncogene* **35**, 94–104 (2016).
66. E. P. Rogakou, D. R. Pilch, A. H. Orr, V. S. Ivanova, W. M. Bonner, DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J. Biol. Chem.* **273**, 5858–5868 (1998).

67. W. Wei *et al.*, Human L1 retrotransposition: Cis preference versus trans complementation. *Mol. Cell. Biol.* **21**, 1429–1439 (2001).
68. R. S. Alisch, J. L. Garcia-Perez, A. R. Muotri, F. H. Gage, J. V. Moran, Unconventional translation of mammalian LINE-1 retrotransposons. *Genes Dev.* **20**, 210–224 (2006).
69. M. S. Taylor *et al.*, Affinity proteomics reveals human host factors implicated in discrete stages of LINE-1 retrotransposition. *Cell* **155**, 1034–1048 (2013).
70. B. Tiwari *et al.*, p53 directly represses human LINE1 transposons. *Genes Dev.* **34**, 1439–1451 (2020).
71. P. Mertins *et al.*; NCI CPTAC, Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **534**, 55–62 (2016).
72. H. Zhang *et al.*; CPTAC Investigators, Integrated proteogenomic characterization of human high-grade serous ovarian cancer. *Cell* **166**, 755–765 (2016).
73. R. Craig, R. C. Beavis, TANDEM: Matching proteins with tandem mass spectra. *Bioinformatics* **20**, 1466–1467 (2004).
74. D. D. Wiredja, M. Koyutürk, M. R. Chance, The KSEA app: A web-based tool for kinase activity inference from quantitative phosphoproteomics. *Bioinformatics* **33**, 3489–3491 (2017).