Enrichment of processed pseudogene transcripts in L1-ribonucleoprotein particles

Prabhat K. Mandal1, Adam D. Ewing3, Dustin C. Hancks4 and Haig H. Kazazian Jr1,2,∗

1McKusick-Nathans Institute of Genetic Medicine and 2Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD, USA 3Center for Biomolecular Science and Engineering, University of California at Santa Cruz, Santa Cruz, California, USA 4Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT, USA

Received May 11, 2013; Revised and Accepted May 15, 2013

Long INterspersed Elements (LINE-1s, L1s) are responsible for over one million retrotransposon insertions and 8000 processed pseudogenes (PPs) in the human genome. An active L1 encodes two proteins (ORF1p and ORF2p) that bind with L1 RNA and form L1-ribonucleoprotein particles (RNPs). Although it is believed that the RNA-binding property of ORF1p is critical to recruit other mobile RNAs to the RNP, the identity of recruited RNAs is largely unknown. Here, we used crosslinking and immunoprecipitation followed by deep sequencing to identify RNA components of L1-RNPs. Our results show that in addition to retrotransposed RNAs [L1, Alu and SINE-VNTR-Alu (SVA)], L1-RNPs are enriched with cellular mRNAs, which have PPs in the human genome. Using purified L1-RNPs, we show that PP-source RNAs preferentially serve as ORF2p templates in a reverse transcriptase assay. In addition, we find that exogenous ORF2p binds endogenous ORF1p, allowing reverse transcription of the same PP-source RNAs. These data demonstrate that interaction of a cellular RNA with the L1-RNP is an inside track to PP formation.

INTRODUCTION

The human genome is littered with active and inactive non-long terminal repeat (non-LTR) retrotransposons. Over 500 000 Long Interspersed Elements (LINE or L1) and one million Alus occupy 17 and 11% of human genome sequence mass, respectively (1,2). An active L1 is 6.0 kb in length, containing a 900-nt 5′-untranslated region (UTR) with internal promoter (3,4), two open-reading frames (ORFs), designated ORF1 and ORF2, separated by a small inter-ORF spacer sequence and followed by a ~200-bp 3′-UTR. ORF2 encodes a 150-kDa protein (ORF2p) with reverse transcriptase (RT) (5) and endonuclease (EN) activity (6) whereas ORF1 encodes a 40-kDa protein (ORF1p) (7) with demonstrated nucleic acid chaperone activity (8). Although the functions of the ORF-encoded proteins are poorly understood, both proteins are critical for the process of retrotransposition (9). It is hypothesized that following transcription, L1 RNA is exported to the cytoplasm where both ORFs are translated. At the ribosome, the newly synthesized ORF1 and ORF2 proteins are thought to interact with their encoding RNA, a phenomenon known as cis preference (10–13), to form a ribonucleoprotein particle (L1-RNP). L1-RNP, the proposed functional intermediate, then enters the nucleus and inserts a new L1 copy into the genome via a coupled reverse-transcription and integration mechanism termed target-primed reverse transcription (TPRT) (14,15). Here, the ORF2p EN nicks the bottom-strand DNA target at an A/T-rich consensus site (5′-TTTT/AA-3′) (6) that generates a free 3′-OH that acts as a primer for reverse transcription of the L1 RNA. This results in a new insertion that ends in a polyA sequence and is usually flanked by a duplication of the target sequence (target-site duplication, TSD) at the 5′ and 3′ ends. L1 is active in present-day humans with ~2000 polymorphic insertions known (16–19) and is responsible for almost 100 de novo retrotransposition events resulting in genetic disease (20).

L1 proteins are also able to retrotranspose other RNAs in trans (12,21–25). Some of these RNAs, Alu, SINE–VNTR–Alu (SVA) and U6 small nuclear RNA (snRNA) may be preferential targets for L1 as inferred from the high copy number of these sequences in the genome. Additionally, sequence characteristics [variable TSD and poly A tail at the 3′ end] indicate that L1-encoded proteins are responsible for the multiple copies of other highly structured small RNAs such as yRNAs (hY1, hY3) (26) that are part of the Ro/SS-A autoantigen and

∗To whom correspondence should be addressed. Tel: +1 4105026660; Fax: +1 4105022006; Email: kazazian@jhmi.edu

© The Author 2013. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com
snRNAs (U1, U2, U4 and U5) (22, 25, 27, 28). Finally, L1 proteins drive processed pseudogene (PP) formation (12). PPs, also known as retropseudogenes, are copies of cellular mRNAs that have been reverse transcribed and inserted into the genome by the L1 machinery. A recent estimate suggests that the human genome contains over 8000 PPs that are derived from 2000 to 3000 protein-coding genes (29). In silico data indicate that some genes, for example glyceraldehyde-3-phosphate dehydrogenase (GAPDH), heterogeneous nuclear ribonucleoprotein A1 (HNRNPA1), actin beta (ACTB) and ribosomal protein L31 (RPL31) have a large number of PPs whereas ~2071 parent genes have just one PP present (29). Recent studies have shown that in some cases (~600), PPs are expressed and perform crucial regulatory roles through their RNA products (29, 30). A growing body of evidence strongly suggests their potential roles in regulating cognate wild-type gene expression by serving as a source of endogenous siRNA (31, 32). PP transcription has also been shown to regulate cognate wild-type gene expression by sequestering miRNAs (33). Why some RNAs are selected as templates for L1-mediated reverse transcription and others are not is unknown, although highly expressed and immunoprecipitation (PAR-CLIP) (41) followed by high-throughput cDNA sequencing to identify RNA targets of ORF1p in the L1-RNP. In brief, we transfected HEK293T cells with a primer that contains a unique 20-nt linker sequence at the 5’ end followed by a 12-nt poly(T) sequence. If ORF2p is present, elongation will occur and can be detected by carrying out PCR with an L1 3′ end-specific forward primer and a linker-specific reverse primer (Fig. 1B; panel 5). Analysis of PCR reactions resolved on a agarose gel demonstrated LEAP products were present as a diffuse band. Sanger sequencing of topoisomerase-cloned PCR products verified that the amplicons came from the transfected L1 and contained polyA tails of variable length. These assays confirmed that the IP complex purified using an ORF1 tag in full-length L1 contains basal L1-RNPs.

To identify which RNAs and where on these RNAs L1 proteins bind in vivo, we carried out PAR-CLIP (41) using the epitope-tagged L1 constructs. In brief, we transfected HEK293T cells with FL-O1F (Supplementary Material, Fig. S2). The cells were grown in the presence of the photoactivatable nucleoside 4-thiothymidine (4-SU), which is subsequently incorporated into nascent RNA to provide strongly enhanced crosslinking efficiency at a relatively short and low energy pulse of UV light. Cellular lysate was prepared after irradiating cells at 365 nm. Efficient crosslinking leads to specific nucleotide conversion of 4-SU to 6-iodo deoxyuridine (I-U) (44). Flow cytometry at Day 3 revealed that the presence of the FLAG epitope tag at the C-terminus of ORF1p did not greatly alter the retrotransposition activity (2.55% without tag versus 2.3% with tag) (Supplementary Material, Fig. S1B).

To determine whether these engineered L1s form functional L1-RNPs, cytoplasmic lysate was prepared for RNA analysis following transient transfection of FL-O1F. L1-RNPs were purified by anti-FLAG agarose beads (Supplementary Material, Fig. S1C). Using the anti-FLAG antibody for ORF1p and the anti-N- and anti-C terminus antibody for ORF2p (44), both L1 proteins were detected by western blot analysis in the FLAG-purified RNPs (Fig. 1B; panels 1, 2 and 3). To test for the presence of exogenous L1 RNA, FLAG-RNPs were resolved in a denatured agarose gel followed by northern blot analysis. Note that the engineered L1 transcription terminates at the Bovine Growth Hormone (BGH) polyadenylation (polyA) signal. This sequence provides a unique 200-bp tag with which to distinguish transfected L1 RNA from endogenous L1 RNA. Using a BGH anti-sense RNA probe (Fig. 1A, thick black line) a ~6.2-kb band representing full-length L1 RNA was detected in the purified RNPs (Fig. 1B; panel 4).

To test whether these samples contained functional L1-RNPs, we carried out a reverse transcriptase assay referred to as LINE-Element Amplification Protocol (LEAP) (45). Briefly, L1-RNPs were incubated with a primer that contains a unique 20-nt linker sequence at the 5’ end followed by a 12-nt poly(T) sequence. If ORF2p is present, elongation will occur and can be detected by carrying out PCR with an L1 3′ end-specific forward primer and a linker-specific reverse primer (Fig. 1B; panel 5). Analysis of PCR reactions resolved on an agarose gel demonstrated LEAP products were present as a diffuse band. Sanger sequencing of topoisomerase-cloned PCR products verified that the amplicons came from the transfected L1 and contained polyA tails of variable length. These assays confirmed that the IP complex purified using an ORF1 tag in full-length L1 contains basal L1-RNPs.

To identify which RNAs and where on these RNAs L1 proteins bind in vivo, we carried out PAR-CLIP (41) using the epitope-tagged L1 constructs. In brief, we transfected HEK293T cells with FL-O1F (Supplementary Material, Fig. S2). The cells were grown in the presence of the photoactivatable nucleoside 4-thiothymidine (4-SU), which is subsequently incorporated into nascent RNA to provide strongly enhanced crosslinking efficiency at a relatively short and low energy pulse of UV light. Cellular lysate was prepared after irradiating cells at 365 nm. Efficient crosslinking leads to specific nucleotide conversion of 4-SU to cytosine during reverse transcription of RNA and next-generation sequencing, thus marking the sites of bound protein. The crosslinked L1 RNA-protein complex was isolated by FLAG immunoprecipitation (Fig. 2B; panel 1) and checked for the presence of L1 proteins (ORF1p and ORF2p) by immunoblotting (Fig. 2B; panels 2 and 3). The RNP complex purified from FL-O1F contains ORF2p detected using an anti-N-terminus
ORF2p antibody (Fig. 2B; panel 3). Additionally, we detected RT activity of ORF2p in the same sample by employing the LEAP assay (Fig. 2B; panel 4) (45). The covalently bound RNA was used to prepare a cDNA library (Fig. 2B; panel 5).

The cDNA library from FL-O1F was then sequenced using the Illumina platform as described by Hafner et al. (41,46). A second cDNA library used a C-terminus FLAG-tagged ORF1p alone (ORF1F) (Fig. 2A and B) to determine ORF1p RNA binding in the absence of transfected ORF2p. Note that the RNA from the FL-O1F library likely represents two different populations: (1) RNA bound to free ORF1p that has no ORF2p present and (2) RNA bound to both ORF1p and ORF2p present in a complex representing true L1-RNPs. As a positive control for PAR-CLIP, we made an RNA library from HuR-bound RNA (Fig. 2A and B). HuR is a well-known RNA-binding protein whose RNA-binding sites have been identified using the PAR-CLIP method (47–49). pcDNA-FLAG, the empty vector, served as a negative control for specificity of RNA binding (Fig. 2A and B).

**Genome-wide RNA binding by ORF1p and L1-RNPs**

Our PAR-CLIP data provide a means to analyze the genome-wide binding profile of ORF1p both when alone and in the context of the L1 RNP. The L1-RNP mobilizes L1 RNA in cis and other RNAs in trans; thus, we first analyzed the RNA sequencing data for known non-LTR retrotransposons and other genomic repeats using the Repeatmasker annotations from the UCSC Genome Browser (see Materials and methods). Genome-wide binding of HuR protein served for comparison to another RNA-binding protein [see Materials and methods, see Supplementary Material, Fig. S6, for comparison with other HuR PAR-CLIP data (50)]. The ORF1p binding profiles for L1 RNA and examples of two active human non-autonomous elements (AluYa5 and SV40) are shown in Figure 3. The full results for additional element families are presented in Supplementary Material, Figs S8–64. As many Repeatmasker annotations are contained within exon or intron annotations, the binding profiles could be confounded by gene transcripts that include transposable element sequences. To determine whether this
was a factor, genome-wide alignment of PAR-CLIP reads was stratified into genic and non-genic regions based on UCSC Known Genes (51). Genic and non-genic alignments were then re-aligned to reference transposable element sequences independently, so the profiles could be compared. In general, read depth was higher for intergenic elements and profiles were similar between genic and intergenic reads re-aligning to repeat references, generally with higher read depth from intergenically-derived reads. One exception is U2 RNA, which appears to have a significantly different binding profile for genic reads versus intergenic reads (Supplementary Material, Figs S47 and S48). A complete list of read counts overlapping Repeatmasker annotations is present in Supplementary Material, Tables S2 and S3. It is noteworthy that many of the families bound specifically by FL-O1F/ORF1F (RNA bound in the context of L1-RNP and RNA bound with ORF1p alone) correspond to actively retrotransposed sequences including L1HS, AluYa5 and AluYb8, and SVA subfamilies D, E and F (Table 1).

Many highly structured RNAs such as Y-RNAs (hY1, hY3, hY4 and hY5), splicing factors (U1, U2, U4, U5 and U8) and the neuronally-expressed BC200 RNA appear to be bound by FL-ORF1/ORF1 far more frequently than by HuR in HEK293T cells (Supplementary Material, Table S2). Likewise, 7SL, the structured RNA from which Alus were derived (52), also appears to be bound specifically by FL-O1F/ORF1F (Supplementary Material, Figs S59–S61). Conversely, repeat transcripts bound in significantly higher quantity by HuR relative to FL-ORF1/ORF1 correspond to a wide variety of repeat elements not known to be actively mobilized (Supplementary Material, Table S3).

In addition to binding expressed repeat sequences, FL-O1F/ORF1F also bind expressed gene transcripts (1782 distinct genes) (Supplementary Material, Table S4). A substantial portion (392 of 1782, 22.0%) of those transcripts has one or more PPs in the human genome. Given that these pseudogene annotations [pseudogene.org build 69, (53)] are based on

**Figure 2.** Cloning of RNA associated with L1-RNPs by PAR-CLIP. (A) Constructs used to prepare the RNA library. Full-length L1 containing a FLAG epitope at the C-terminus of ORF1p (FL-O1F) was described in Figure 1. L15′-untranslated region (UTR) along with C-terminus FLAG-tagged ORF1p (ORF1F) and HuR containing FLAG at the C-terminus (HuRF) were cloned in pcDNA6. (B) Panel 1: the RNA–protein complex was separated on an SDS–PAGE gel and exposed to X-Ray film to detect the radiolabeled RNA-protein complex. Panel 2: Immunoblot probed with anti-FLAG antibody detects ORF1p (40 kDa) and HuR (37 kDa). Panel 3: western blot detection of ORF2p (150 kDa) by anti-ORF2 antibody. Panel 4: L1 ORF2 reverse transcriptase activity on L1 RNA detected by the LEAP assay. Panel 5: the RNA–protein complex containing radiolabeled RNA was excised from an SDS–PAGE gel; the RNAs were separated from the complex by electroelution, converted into a cDNA library and resolved on 5% PAGE gel. The cDNA library was sequenced using the Illumina platform.
Ensembl gene annotations and 2616/22532 Ensembl gene annotations (11.6%) are associated with PPs, this is a significant enrichment (Fisher’s exact test, $p < 2 \times 10^{-16}$). We sought to determine whether a certain region of the mRNA was preferentially bound (i.e. 5'-UTR, exons, 3'-UTR) by FL-O1F/ORF1. Of bound transcripts, 33% (593 of 1782) showed FL-O1F/ORF1 binding either at the last exon or 3'-UTR (Supplementary Material, Table S4). Of those 593 transcripts that were bound by FL-O1F/ORF1F at the last exon/3'-UTR, 139 (23.4%) have one or more PPs in the human genome. The overall distribution of reads mapping to exonic sequences (5'-UTR versus 3'-UTR versus Internal) is shown in Supplementary Material, Figure S4.

Figure 3. L1-RNP and ORF1 binding profiles on active retroelements based on PAR-CLIP read depth. The left columns of panels represent L1-RNP mappings (FL-O1F library) to intergenic regions of GRCh37 re-aligned to the retroelement consensus as indicated on the left side of the figure. The right columns of panels are the same as the left, but for ORF1 binding profiles. For each panel, the horizontal axis indicates the position in the reference element consensus and the vertical axis indicates the read depth normalized to the number of mapped reads for the given element. A full collection of plots for a wider variety of repeat annotations is presented as Supplementary Material, Figures S8 through S64.
Alu, SVA, U snRNAs and hYRNAs are present in functional L1 RNPs

RNA species (Alu and SVA) that are successfully retrotransposed by the L1 machinery were identified in the ORF1p binding profile by PAR-CLIP. Thus, we sought to determine whether those RNAs are present in the L1-RNP. Previous studies demonstrate that ORF2p alone is sufficient for Alu retrotransposition (21). However, those studies do not consider the effects of endogenous ORF1p. Other studies show that supplementation of ORF1p can increase Alu retrotransposition more than 5-fold (54). Therefore, ORF1p might have some role in Alu retrotransposition. Aligned reads obtained from FL-O1F and ORF1p libraries with Alu RNA show a significant number of reads matching Alu sequence, specifically in the single-stranded linker region between the AluL and AluR monomer (Supplementary Material, Fig. S3A). We also note significant binding of ORF1p in the stretch of poly A at the 3′ end of Alu sequence (Supplementary Material, Fig. S3A). Analysis of HuR-bound RNA shows substantially less binding at this location in the HuR PAR-CLIP library, each normalized against the number of number of unique aligned reads with T-to-C transitions for the corresponding libraries (Supplementary Material, Table S1). Put another way, the data indicate the relative abundance of reads in the ORF1/L1-RNP library relative to the HuR library, normalized for total aligned read count. The ‘Min. chi square’ column contains the chi-squared statistic obtained by performing a 2 × 2 chi-squared test with Yates’s continuity correction. The 2 × 2 table in this case consists of the number of reads in the ORF1/L1-RNP library with the least reads for the corresponding repeat family, and the number of reads in the HuR library for that family, together with the number of unique aligned reads with T-to-C changes (Supplementary Material, Table S1). Max P contains the P-value associated with the value of the chi-squared statistic for 1 degree of freedom.

Table 1. Active human retrotransposons bound predominantly by ORF1p/L1-RNP

<table>
<thead>
<tr>
<th>Repeat name</th>
<th>Min. normalized ratio</th>
<th>Min. chi square</th>
<th>Max. P</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1HS</td>
<td>14.06</td>
<td>1508.90</td>
<td>0</td>
</tr>
<tr>
<td>AluYb8</td>
<td>4.27</td>
<td>700.15</td>
<td>2.77E-154</td>
</tr>
<tr>
<td>AluYa5</td>
<td>3.35</td>
<td>525.46</td>
<td>2.75E-116</td>
</tr>
<tr>
<td>SVA_D</td>
<td>2.06</td>
<td>144.02</td>
<td>3.51E-33</td>
</tr>
<tr>
<td>SVA_E</td>
<td>2.83</td>
<td>49.10</td>
<td>2.43E-12</td>
</tr>
<tr>
<td>SVA_F</td>
<td>1.46</td>
<td>34.13</td>
<td>5.15E-09</td>
</tr>
</tbody>
</table>

The column titled ‘Min. normalized ratio’ contains the ratio between the minimum number of reads across both replicates of the ORF1 and L1-RNP PAR-CLIP libraries and the number of reads in the same repeat family in the HuR PAR-CLIP library, each normalized against the number of number of unique aligned reads with T-to-C transitions for the corresponding libraries (Supplementary Material, Table S1). Put another way, the data indicate the relative abundance of reads in the ORF1/L1-RNP library relative to the HuR library, normalized for total aligned read count. The ‘Min. chi square’ column contains the chi-squared statistic obtained by performing a 2 × 2 chi-squared test with Yates’s continuity correction. The 2 × 2 table in this case consists of the number of reads in the ORF1/L1-RNP library with the least reads for the corresponding repeat family, and the number of reads in the HuR library for that family, together with the number of unique aligned reads with T-to-C changes (Supplementary Material, Table S1). Max P contains the P-value associated with the value of the chi-squared statistic for 1 degree of freedom.

LEAP products to the reference human genome GRCh37 using BLAT (56) revealed that seven belong to the AluY subfamily whereas the other two matched the AluSx and AluSg subfamilies (Supplementary Material) (57). In a similar experiment, we observed ORF2p RT activity on SVA and 7SL RNA templates (Fig. 4) and mapped SVA LEAP products to the reference human genome (Supplementary Material). Characterization of six individual SVA LEAP products revealed three matched with SVAD, two with SVA_A and one with SVA_E subfamilies (58).

Highly structured U snRNAs and hYRNAs were also identified in the PAR-CLIP data set. Thus, we tested whether those RNAs are also part of the L1-RNP. Primers were designed to check the LEAP activity on U1, U2, U4, U5, U6, hY1 and hY4 RNA templates. All those templates except hY4 showed LEAP activity compared with Alu and L1 templates (Fig. 4; panel 2). RNA analysis by RT-PCR showed absence of U5 and U6 RNAs from L1-RNPs purified after transfecting FL-O1F and treated with DNase. The RT reaction was performed using an oligo-dT primer. PCR was performed using transcript-specific primers. U5 and U6 RNA were not detected in L1-RNPs by RT-PCR analysis. Panel 2: LEAP assay with purified L1-RNPs was performed as described in Materials and methods. A diffuse band was detected when separated in a 2.0% agarose gel. Panels 3 and 4: LEAP assay obtained from pcDNA6 RNP (vector control) and FL-O1FD702YRT RNP did not show LEAP activity.

Figure 4. ORF2-mediated RT activity on Alu, SVA, 7SL, U snRNAs and hYRNAs. (A) Panel 1: detection of RNA in the L1-RNP by RT-PCR. RNA was purified from L1-RNPs following transfection with FL-O1F and treated with DNase. The RT reaction was performed using an oligo-dT primer. PCR was performed using transcript-specific primers. U5 and U6 RNA were not detected in L1-RNPs by RT-PCR analysis. Panel 2: LEAP assay with purified L1-RNPs was performed as described in Materials and methods. A diffuse band was detected when separated in a 2.0% agarose gel. Panels 3 and 4: LEAP assay obtained from pcDNA6 RNP (vector control) and FL-O1FD702YRT RNP did not show LEAP activity.

[Figure 4]
length poly A tract (Supplementary Material). However, LEAP sequences obtained from U4, U5 and hY1 RNAs showed only PCR artifacts. These data further indicate that some highly structured small RNAs may be dispersed in the genome by L1-encoded proteins.

**Source transcripts of PPs are enriched in L1-RNPs**

Multiple RNA species, including source transcripts of PPs, were identified in L1-RNPs by PAR-CLIP. Indeed, we find a greater than 3-fold enrichment for FL-O1F and ORF1F binding to PPs over HuR (Supplementary Material, Fig. S5). To confirm the results of PAR-CLIP, we selected twenty transcripts from Supplementary Material, Table S4, which showed significant reads in FL-O1F and ORF1F libraries, as well as PPs in the human reference genome (Fig. 5B) to determine whether these RNAs were used as ORF2p RT templates in the LEAP assay. These include PABPC1, HSPE1-MOB4, RPLP1, GTF2I, NPM1, RPL5, RPLP0, RPS7, HMGN1, RPL12, PCBP2, TPT1, GAPDH, HRNRPA1, RPSA, ASNS, ACTB, RPL31, GNAs and PSMD14. The full names of the transcripts are listed in Supplementary Material. Note that multiple ribosomal proteins are among those twenty selected transcripts (7 of 20). LEAP was carried out by the method described for L1 RNA; except that in the PCR step, we used a primer positioned either in the last exon or in the 3′-UTR for the 20 selected genes. The LEAP reaction was conducted using purified L1-RNPs following transfection of FL-O1F (Fig. 2A). A primer designed for the engineered L1 transcript served as a positive control. LEAP assays showed 17 of the 20 transcripts served as efficient templates for ORF2-mediated RT activity (Fig. 5C; panel 1). DNA sequence analysis confirmed that these products had features of L1 ORF2p-mediated reverse transcription (Supplementary Material). These data suggest that all 17 transcripts are present in the basal L1 RNP. For negative controls, LEAP reactions were conducted following purification of RNPs using construct FL-O1F(RT′), ORF1F and vector alone. Consistent with no reverse transcriptase activity encoded in these constructs, no LEAP activity was observed (Fig. 5C; panels 5, 6 and 7) for those 20 PP-source transcripts. However, RNPs purified following ORF1F transfection (ORF1 only) showed less but detectable LEAP activity on the exogenous L1 template (Fig. 5C; panel 6), indicating ORF1F transfection (ORF1 only) showed less but detectable LEAP activity on the exogenous L1 template (Fig. 5C; panel 6) and ORF1F transfection products served as negative controls (Fig. 6; panels 3, 4 and 5). Of 20 non-PP RNP-associated RNAs, 13 produced RT-PCR products (Fig. 6; panel 1) while only 4 of the 13 demonstrated LEAP activity (Fig. 6; panel 2). Agarose gel analysis showed much less intense LEAP products compared with a GAPDH control. Sequence analysis revealed that three of the four LEAP products had authentic features of ORF2p reverse transcription (Supplementary Material). Sequences obtained for the fourth, AP2M1, showed PCR artifacts. Thus, RNAs present in L1-RNPs that have PP counterparts are more permissive to ORF2p-mediated reverse transcription, and of 20 RNAs binding to ORF1p that do not form PPs, only 3 were found in the basal L1-RNP by LEAP.

**Analyzing RNA in ORF1 mutants RNPs**

Next, we sought to determine whether ORF1 mutants known to be retrotransposition defective are able to bind RNA in vivo. We chose known RNA-binding mutant RRR206.210.211AAA (in the loop region between beta 2 and beta 3 in the RRM domain) (40) and the most extensively studied ORF1 mutant JM111 (RR261.262AA in the CTD region) (9) (Fig. 7A). We also made two truncated ORF1 mutants, ORF1F(RRM) (5′-UTR plus LC-CTD) and ORF1F(CC) (5′-UTR plus CC) (Fig. 7A) to determine whether RRM-domain-deleted ORF1p supports RNA binding in vivo. Employing a PAR-CLIP experiment using mutant ORF1 constructs ORF1F(RRR206.210.211AAA) and ORF1F(RR261.262AA), we found similar patterns of radioactive RNA–protein complex in an SDS–PAGE gel compared with the wild-type ORF1 construct (Fig. 7B; lanes 2 and 3). Reduced RNA binding was detected using ORF1F(ΔRRM) (Fig. 7B; lane 4) whereas no binding was observed for construct ORF1F(CC) (Fig. 7B; lane 5). Negative controls [FLAG-epitope-tagged β-actin, lactate dehydrogenase (LDHA) (59) and empty expression vectors] showed no RNA binding by PAR-CLIP (Fig. 7B; lanes 6, 7 and 8). These results suggest that ORF1 mutants defective in retrotransposition can still bind RNA in vivo. These results also suggest that RRM-domain-deleted CC-CTD-fused ORF1p can support RNA binding to some extent.

To determine whether RNPs purified from ORF1F (RRR206.210.211AAA) and ORF1F (RR261.262AA) ORF1p mutants contain full-length L1 RNA, we performed northern blot analysis. RNA was purified from RNPs following transfection of either FL-O1F(RRR206.210.211AAA) or FL-O1F(RR261.262AA) construct. Using a BGH anti-sense RNA probe, we detected a 6.2-kb RNA in the RNPs purified using ORF1p mutants (Fig. 7C). RNAs purified from FL-O1F and FL-O1F(RT′) were used as positive controls whereas RNPs from untransfected lysate was used as a negative control (Fig. 7C). These results suggest that ORF1 mutants that do not support retrotransposition are able to bind full-length L1 RNA, thus maintaining *cis* preference.

Next, to determine whether other RNAs identified by PAR-CLIP are present in L1-RNPs purified from constructs FL-O1F(RRR206.210.211AAA) and FL-O1F(RR261.262AA), we carried out LEAP reactions on those 20 selected genes including engineered L1. LEAP assays using these ORF1 mutants showed that the same 17 of 20 selected transcripts and L1 served as
Figure 5. ORF2p-mediated RT activity on processed pseudogene-source transcripts. (A) Detection of ORF1p and ORF2p in RNPs purified from FL-O1F(RRR206.210.211AAA), FL-O1F(RR261.262AA) and FL-O1F(RT−). L1-RNPs were purified from one 100-mm culture disc and finally eluted in 100 μl elution buffer. For western analysis 20 and 5 μl was loaded to detect ORF2p (anti-ORF2p) (44) and ORF1p (anti-FLAG), respectively. RNPs obtained from FL-O1F and pcDNA were used as positive and negative controls, respectively. (B) Number of processed pseudogenes in the human reference genome (GRCh37) based on pseudogene.org release 69. (C) L1-RNPs were purified from ~6 × 10⁶ cells (one 100-mm culture disc) following transient transfection of FL-O1F, ORF1F, ORF2F(EN−) [3xFLAG tag at the N-terminus of ORF2ENmut (D205G) sequence] and vector control (detailed in Materials and methods) and finally eluted in 100 μl elution buffer. The LEAP reaction was conducted using 1.0 μl of purified RNPs in a 20 μl total volume. PCR was performed using 0.5 μl of
efficient templates in ORF2-mediated reverse transcription (Fig. 5A; panels 3 and 4). Western blot analysis showed RNPs purified from ORF1p mutants contain ORF1p and ORF2p levels comparable with those of wild-type L1 with FL-O1F(RRR206.210.211AAA) showing a reduction in ORF2p levels (Fig. 5A, lanes 3 and 4). DNA sequence analysis on LEAP products of GAPDH, RPLP1 and L1 confirmed that the sequences were those expected for ORF2 reverse transcription on those RNAs. These data suggest that all 17 transcripts found in wild-type L1-RNPs are also present in RNPs purified using ORF1 mutant constructs.

An over-expressed ORF2p construct associates with ORF1p in an L1-RNP

Endogenous ORF1p is present in high amounts in many cell lines that support L1 retrotransposition, including HEK293T (35,36), whereas endogenous ORF2p escapes physical detection in those cells. Whether or not endogenous and exogenous L1s interact to form retrotransposition competent (RC)-L1-RNPs to form chimeric RNPs is unknown. We sought to determine whether hybrid or chimeric L1 RNPs can support LEAP activity for those transcripts that have multiple PPs in the human genome by using an over-expressed ORF2p clone and testing if it could associate with endogenous ORF1p to form RC-L1-RNPs. Here, we show that an ORF2 alone EN catalytic mutant, D205G, construct ORF2F(EN−), can produce significant amounts of ORF2p detectable in Coomassie-stained SDS–PAGE gels following transient transfection and affinity purification of L1-RNPs from 293 T cells (Fig. 8A; panel 1). From these L1-RNPs, ORF2p (EN−) was easily detected (Fig. 8A; panel 2) by western blot with an anti-FLAG antibody along with a significant amount of endogenous ORF1p detected by immunoblotting with an anti-ORF1 antibody (Fig. 8A; panel 3) (35). In the same complex, we observed robust ORF2p RT activity by the LEAP assay on those transcripts (17 of 20) that have multiple PPs in the genome (Fig. 5C; panel 2). Cloning and sequencing of individual LEAP products confirmed that the sequences were the source transcripts containing variable poly-A tail lengths.

Next to determine what fraction of endogenous ORF1p co-immunoprecipitated with exogenous over-expressed ORF2p, we quantified the amount of ORF1p in total lysate, the unbound fraction (flow through) and in the ORF2p immunopurified complex (Fig. 8B). No qualitative difference was observed in ORF1p amount between total and flow through fractions, suggesting very little ORF1p co-immunoprecipitated with 3xFLAG tagged ORF2p. Band intensity measurement showed that 1.9% of ORF1p co-immunoprecipitated with ORF2p (Fig. 8B). Quantification of ORF2p in total, unbound and RNP fractions revealed that 65% of ORF2p bound with anti-FLAG conjugated agarose beads of which 27% of ORF2p eluted during FLAG peptide competition (Fig. 8B). These data suggest L1-RNPs purified using over-expressed ORF2F(EN−) clone contain a large excess of ORF2p compared with ORF1p in the immunopurified complex. These results also suggest that a very limited amount of endogenous ORF1p (∼2%) participates in forming the chimeric L1-RNP complex.

Quantification of mRNAs in the L1-RNP

Next we used quantitative RT-PCR (qRT-PCR) to determine the relative amount of other RNAs in comparison with L1 RNA in RNPs purified using FL-O1F construct. We chose ribosomal protein large P1 (RPLP1) and GAPDH RNAs as they showed a similar amount of LEAP product after 35 cycles of PCR when resolved in a 2% agarose gel (Fig. 5C). We also chose β-actin as it showed significantly less LEAP compared with L1 LEAP (Fig. 5C). qRT-PCR analysis showed that the relative amount of RPLP1 and β-actin transcripts are 14.5- and 0.126-fold compared with the L1 transcript, respectively (Fig. 9A). The GAPDH transcript is similar (1.16-fold) to L1 RNA. Next we used a semi-quantitative time-course RT-PCR to measure the amount of LEAP products for those four transcripts because a qRT-PCR cannot be used because of poly-A length
heterogeneity of the PCR products. We found that L1 and GAPDH were equally amplified in the LEAP assay (Fig. 9A). L1, GAPDH and RPLP1 were visible at 29 cycles of PCR, whereas β-actin became visible at 31 cycles.

Similarly, to determine the relative abundance of those other RNAs compared with L1 RNA in RNPs purified using the over-expressed ORF2p clone, qRT-PCR was carried out (Fig. 9B). Note that the amount of total RNA obtained from ORF2(EN) RNPs (4.5 ng/μl) is ~8-fold less than the RNA recovered from FL-O1F RNPs (35 ng/μl). qRT-PCR analysis showed RPLP1, GAPDH and β-actin are 33-, 2.22- and 0.126-fold more concentrated in the RNP than the L1 transcript, respectively. Semi-qRT-PCR to measure the amount of LEAP products showed GAPDH product amplified very early (29 cycles), whereas β-actin was visible at 33 cycles (Fig. 9B). Note that the amount of L1 LEAP obtained from ORF2F(EN) RNPs was less than that from FL-O1F RNPs (Fig. 9A and B).

Retrotransposon RNAs along with several PP transcripts are concentrated in L1-RNPs. To determine what fraction of the total cellular RNA is concentrated in L1-RNPs, we measured the absolute amount of RNA in total lysate, unbound lysate and the RNP prep for construct FL-O1F. The analysis showed very little difference in the RNA content between total and unbound lysate and only ~0.358% of the total RNA in the L1-RNPs (amount of total RNA—324 μg; amount of RNP RNA—1.16 μg). Similar analysis using the over-expressed ORF2p clone showed as low as 0.05% of the total RNA concentrated in the L1-RNPs (amount of total RNA—300 μg; amount of RNP RNA—0.150 μg).

In addition to L1, multiple PP transcripts are present in purified L1-RNPs from FL-O1F or ORF2F(EN). Next, we used qRT-PCR to determine what fraction of some of those transcripts (RPLP1, GAPDH and β-actin) and L1 are concentrated in L1-RNPs. RPLP1, GAPDH and β-actin showed ~2-fold
enrichment, whereas L1 transcript showed ~3.7-fold enrichment compared with total lysate for construct FL-O1F (Fig. 10; panel A). Similar analysis using ORF2(EN−) construct showed 4- and 2-fold enrichment for β-actin and GAPDH, whereas RPLP1 and L1 lacked noticeable enrichment (Fig. 10; panel B). RNPs purified from a control (empty vector) showed 6.67- and 11-fold reduced RPLP1 and GAPDH RNA, respectively, whereas no β-actin RNA was detected in these RNPs (Fig. 10; panel C). Thus, for both the FL-ORF1 and ORF2(EN−) constructs, most RNAs studied that tend to form high numbers of PPs are enriched in the L1 RNP.

Next, we analyzed fold enrichment of those transcripts (RPLP1, GAPDH and β-actin) in FL-O1F RNPs owing to excess potentially free ORF1p compared with enrichment in ORF2(EN−) RNPs. Quantitative PCR analysis showed RPLP1, GAPDH and β-actin transcripts in ORF2(EN−) RNPs were 9, 5 and 3%, respectively, of their levels in FL-O1F RNPs (Supplementary Material, Fig. S6). These levels are similar to the level of ORF1p in ORF2(EN−) RNPs (~2%). These data suggest that free ORF1p in the FL-O1F RNP complex binds most of the RNA, and only a small fraction of cellular RNAs are part of the actual L1 RNPs. Note that the amount of total RNA in ORF2(EN−) RNPs (4.5 ng/μl) was comparable with that purified from untransfected control RNPs (5.0 ng/μl) and 8-fold less than the total RNA obtained from FL-O1F RNPs (35 ng/μl). This suggests that although the total amounts of RNA are the same in both RNPs [ORF2(EN−) and empty vector], RNAs that are part of RNPs (RPLP1, GAPDH and β-actin) are more concentrated in ORF2(EN−) RNPs (compare Fig. 10B and C). This also suggests that ORF2p has very weak general RNA-binding affinity and co-immunoprecipitated
endogenous ORF1p is likely responsible for RNA enrichment in ORF2(EN-) RNPs.

**DISCUSSION**

**Composition of L1-RNPs**

In this study, we used epitope-tagged L1 proteins followed by affinity purification to analyze the RNA components of L1-RNPs, the L1 retrotransposition functional intermediate, (45,61). Our assay builds on previous L1-RNP studies (45,60–63) including those using epitope-tagged L1 proteins, a field that has been severely hampered by the inability to detect ORF2p (44). It is noteworthy that this current approach significantly reduces the time required to isolate RNP preps from 16 to 4 days. This is mainly because we use HEK293T cells, which grow extremely fast, are very efficient in transient transfection and display a high level of L1 retrotransposition. We observe the ability of the protein of either transfected ORF (ORF1p or ORF2p) to immunoprecipitate the protein of the other ORF. Transfected L1 RNA and ORF2p-mediated RT activity are also detected in the L1-RNP. Likewise, as previously reported (44), more ORF2p is observed and can be detected on Coomassie-stained SDS–PAGE gels when the ORF contains a mutation (D205G) that abolishes EN activity (Fig. 8C; panel A). As previously postulated, the EN mutations may lead to increased protein stability because of reduced cellular toxicity of the protein (64) or make the protein more accessible to the antibody.

**Endogenous L1 proteins may complement exogenous L1 proteins**

Use of constructs containing only ORF1 or ORF2 indicates that the over-expressed L1 proteins interact with endogenously
expressed L1 proteins. Of particular interest is that the ORF2p EN mutant protein derived from the ORF2 alone construct can form functional L1-RNPs that contain ORF1p from endogenous loci (Fig. 8; panel 3). These L1-RNPs also support LEAP activity on L1 RNA (Fig. 5C; panel 2). Although it has been demonstrated that ORF1 and ORF2 expressed from separate plasmids can drive trans-complementation (23, 54, 65), this is the first report of endogenous L1 protein interacting with L1 protein from exogenous sources to form a functional L1-RNP complex. Surprisingly, much less ORF1p co-immunoprecipitates (~2% of total ORF1p) with over-expressed ORF2pEN mutant protein. Why so little endogenous ORF1p participates in forming complex with over-expressed ORF2p and ~98% remains free and requires further study. The association of endogenous ORF1p suggests that this over-expressed ORF2 EN mutant construct can be used to study other interacting proteins in L1 RNPs.

It has been thought that ORF1p is dispensable for ORF2p RNA-binding and RT activity; however, the EN mutant suggests that endogenous ORF1p might play a role in over-expressed ORF2p function (61). Furthermore, although both RNP preparations have the same amount of LEAP activity on an L1 RNA template, more ORF1p was observed by western analyses when L1-RNPs were purified from full-length L1 tagged ORF1 (FL-O1F) than in L1-RNPs purified from the full-length L1 tagged ORF2 (FL-O2F) (data not shown). This suggests that a significant fraction of immunoprecipitated complex using FL-O1F likely contains free ORF1p that is not part of the L1-RNPs. These data support a model where the amount of ORF1p required in a functional L1-RNP is much less than previously anticipated (66).

To address the role of endogenous ORF1p in ORF2 activity, one approach would be purification of L1-RNPs from an ORF2 alone construct following complete RNAi knockdown of endogenous ORF1p by L1 ORF1p-specific siRNAs and then assaying for LEAP activity. We tried this experiment but were unsuccessful in completely knocking down endogenous ORF1p in 293 T cells (PKM and HHK, unpublished data). A fraction of cells appears always to escape L1 ORF1p-specific
siRNA, potentially because of multiple source loci contributing to the ORF1p cellular pool.

To study putative endogenous ORF2p, we used an ORF1 alone construct. Here, we have detected LEAP activity on an L1 template following purification of RNPs using ORF1 sequence alone from 293 T cells (Fig. 5C; panel 6, lane marked L1). This suggests that 293 T cells also express low levels of endogenous ORF2p that can form a complex with exogenously-derived ORF1p and ORF1 RNA. These data are consistent with trans-complementation assays where retrotransposition events are observed in the absence of a driver (65,67). They also indicate a very strong cis preference because no other transcripts show LEAP in the ORF1p RNP prep. Overall, it appears that the ORF2 EN mutant and solo ORF constructs should be useful in further interrogating L1-RNP dynamics.

PAR-CLIP reveals ORF1p RNA-binding profiles of human retrotransposons

PAR-CLIP was used to identify which RNAs and where in the RNA ORF1p binds in the context of functional L1-RNPs. Consistent with the cis-preference model (11–13), PAR-CLIP revealed L1-RNP and ORF1p direct binding on L1 RNA with specific enrichments at position 1997–2035 nucleotides and 4837–4871 nucleotides (Fig. 3). Likewise, PAR-CLIP and LEAP assays confirmed the presence of the other two known L1 templates (Alu and SVA RNAs) in the L1-RNPs. Alignment of sequenced LEAP PCR amplicons to the human genome reference sequence suggests that these Alu and SVA transcripts originate from different loci. We note that because of the high copy number of Alus and Svas and their abundance in the transcriptome, either as spliced exons or perhaps transcriptional readthroughs, some LEAP products may not represent ‘true’ Alu and SVA transcripts. However, the observation that most Alu and SVA RNA sequences present in the RNPs belong to active, disease-causing subfamilies (AluY, AluYα5 and SVA D, E, F) (20) indicates that at least some of these transcripts represent bona fide Alu and SVA RNAs.

In contrast to L1 RNA, ORF1p appears to bind Alu and SVA RNAs at specific domains and primarily in regions that would be predicted to be less structured than the unbound domains. For Alu, ORF1p binds primarily in the Alu A-rich polylinker and is depleted from the left and right monomers. Likewise, the PAR-CLIP data indicate that ORF1p binds SVA RNA outside of the VNTR domain, a sequence which is very GC rich and may be highly structured (68). Another possibility for an ORF1p binding determinant that is not exclusively RNA secondary structure is inaccessibility of specific Alu and SVA sequences because these sequences are already bound by other protein/s. It has been hypothesized that SRP 9/14 binds the Alu monomers (10,57) and that PABP binds the Alu polyA tail (69). Both may help explain the depletion of ORF1p at these Alu domains. To our knowledge, no examples have been reported of non-L1 proteins playing a function in SVA retrotransposition. Therefore, because many of the cell lines used in the cell culture retrotransposition assay express ORF1p at detectable levels (35,56,70) and transfected ORF1 can enhance engineered Alu (52) and SVA retrotransposition (23,25), these data may indicate a role for endogenous ORF1p in non-autonomous element mobilization in cell culture.

L1-RNPs contain many Pol III transcripts

Most retrotransposed RNAs that are not ‘retrotransposons’ are RNA Pol III transcripts and are known to be associated with the ribosome and nucleolus (22,24,26–28). Many of these sequences in the genome contain the hallmarks of L1-mediated retrotransposition (TSD, insertion at a predicted L1 EN cleavage site, 3′-polyA stretch). Here, we have identified the presence of significant quantities of highly structured RNAs in L1-RNPs, including the spliceosomal RNAs and hY transcripts. Consistent with the retrotransposed copies of these RNAs in the genome (22,24–28) and other functional studies (22,71), LEAP analysis confirms that these RNAs can serve as ORF2 RT templates (Fig. 5). These data expand the list of experimentally demonstrated ORF2 RT templates to include U1, U2 and hY4.

Although U6 is more abundant in the human genome than the DNA of other spliceosomal RNAs (22), we were unable to detect U6 sequences in the PAR-CLIP data set. RT-PCR confirmed also that U6 RNA was below the level of detection in these L1-RNPs. The inability to detect U6 RNA associated with L1-RNPs suggests either that U6 is expressed at low levels in HEK293T cells or an unknown technical issue such as suboptimal PCR conditions is preventing its detection. Likewise, U5 and U6 RNAs terminate with a 3′-TTTT that might limit priming by the LEAP adaptor. However, this issue likely has a minimal effect because more U6 insertions are present as U6/L1 chimeras (22,27,28). Alternatively, one biological possibility may be that the predominantly nuclear U6 RNA associates with the L1-RNP only transiently and during TPRT.

Incorporation of mRNAs into functional L1 RNPs is a potential step in PP formation

Another important observation of this study was identification of a wide range of cellular transcripts in L1-RNPs. A significant portion (22.0%) of those transcripts has one or more PPs in the human genome. High-throughput transcriptome sequencing has found that a large fraction of PPs are differentially expressed in certain tissues and cancer cells (30). Previous studies have demonstrated that the LINE-1 machinery is responsible for the formation of PPs in the human genome (12,13). Here, we demonstrate that mRNAs that produce PPs are enriched for LEAP activity (17/20, 85%) in purified L1-RNPs from FL-O1F where the amount of ORF1p is in large excess. Similar enrichment of mRNAs was observed using an over-expressed ORF2 EN mutant construct where the amount of ORF1p (from an endogenous source) is very limited. These data and the quantitative analysis of some of those mRNAs (Supplementary Material, Fig. S3) suggest that the fraction of each mRNA within the L1-RNP is real and not merely associated with over-expressed ORF1p.

Surprisingly, we also showed RNPs purified using ORF1p mutants exhibit a similar variety of mRNA enrichment (9,40,61). These data demonstrate that mutated ORF1p binds similar types of RNA as seen for wild-type ORF1p. Previous structural studies coupled with in vitro RNA–protein binding assays demonstrated that mutation of three arginines at positions 206, 210 and 211 to alanine (situated in the RRM domain of ORF1p) dramatically affects ORF1p RNA binding and hence was called an ORF1 RNA-binding mutant (40). Here by
employing an in vivo RNA-binding assay (PAR-CLIP), we did not find any changes in RNA binding of the mutant ORF1p compared with wild type. Affinity-purified RNPs using the above ORF1p mutant also showed significant amounts of both L1 proteins (ORF1p and ORF2p) and full-length L1 RNA. Similar findings were observed when another ORF1p mutant, RR261.262AA (situated in the CTD region of ORF1p), was studied (9,61). Previous studies using ORF1RR261.262AA mutant demonstrated a severe reduction of ORF1p, but not ORF2p, when RNPs were purified by ultracentrifugation (45,61). That study also showed the nature of LEAP products obtained from ORF1RR261.262AA mutant RNPs was different (more heterogeneous with multiple lower molecular weight bands) (45,61). We did not find those differences when we compared mutant versus wild-type ORF1 RNPs. These data suggest that the ORF1p RNA-binding property, basal RNP structure (purified from cytoplasmic lysate by affinity purification) and associated LEAP activity on those RNPs do not differ from RNPs made using wild-type ORF1p. We conclude that further study is required to determine why those ORF1p mutants are unable to drive retrotransposition.

Our results suggest that some kind of selection exists for the L1-mediated trans effect on transcripts that form PPs. One form of selection might be very high expression of those transcripts in the germ line, a site where the L1 machinery is also very active. RNPs purified from germ cells would be ideal to determine the types of RNA that bind with ORF1p and L1-RNPs. Currently, this experiment is not possible because the amount of RNPs is much less in germ cells, and more importantly a good ORF1p antibody with which to make an ORF1p affinity column to purify those RNPs is not available. However, the average expression percentile of genes from 293 T cells shows that a significant number of genes producing RNPs also have increased expression. On the other hand, there are transcripts that have significantly high expression that do not form PPs. Here, we demonstrate that of genes with high expression that do not have PPs, very few (3/20, 13%) show LEAP activity in purified L1-RNPs. This suggests that among those transcripts, some other selection process exist in PP formation.

Surprisingly, we found that cellular mRNAs whose levels are similar or higher to the levels of L1 RNA are as good or better LEAP targets than L1 RNA. Previous data had demonstrated that during the LEAP assay, PP-source transcripts were reverse transcribed at very low levels compared with the L1 template, even though the amount of source RNAs for PPs is higher than L1 RNA for some mRNAs in the purified RNP (45). Our data showed a large fraction of LEAP products (10 of 17) from PP-source transcripts along with Alu RNA had the same or greater amplification intensity as the L1 template. Interestingly, these cellular RNAs are good LEAP targets even in the absence of exogenous ORF1 (RNPs purified using the ORF2F(EN−) construct). This suggests that in the LEAP assay, ORF2 RT has equal preference on L1, Alu and certain cellular mRNAs that are successful in forming PPs. This result also questions whether other factors are involved in preferential retrotransposition of L1 RNA over cellular mRNAs. Our results may differ from those of others because of differences in cell lines, constructs and/or methods used here to purify L1-RNPs (45). In sum, we dissected the RNA components in L1-RNPs and demonstrated that apart from retrotransposon RNAs (L1, Alu and SVA), L1-RNPs are also enriched in cellular RNAs that have one or more PPs in the human genome.

MATERIALS AND METHODS

Cell culture

HEK293T cells were maintained in a tissue culture incubator at 5% CO2, 37°C in high glucose Dulbecco’s modified Eagle medium (DMEM) without pyruvate (Invitrogen) supplemented with 10% fetal bovine calf serum, 2 mM l-glutamine and 100 U/ml penicillin–streptomycin.

Vectors and cloning

All primers used to make constructs are listed in the primer table (Supplementary Material).

ORF1F: To generate this construct, L15′-UTR plus ORF1 was amplified from pJCC5 (L1RP) (42) using primer pair 5UTR-NotF/ORF1-FlagApaR. The fragment was cloned into Not1-Apa1 sites of pcDNA6/myc-HisB (Invitrogen).

L1-01F: L1 inter-ORF spacer, ORF2 plus L1-3′UTR was amplified from pJCC5 (L1RP) using primer pair ORF2ApaF/ORF2ApaR. The fragment was cloned into the Apa I site of ORF1F.

ORF2F(EN−): A minigene containing three tandem FLAG sequences (3XFLAG) flanked by XhoI and NotI was synthesized (IDT). pEGFP-N1 was digested with XhoI and NotI to remove the entire EGFP coding sequence to clone 3XFLAG sequence that creates the pN1–3XFLAG vector backbone. ORF2ENmut(D205G) was created in pJCC5 by site-directed mutagenesis (SDM), amplified using primer pair ORF2modi_NotF/ORF2_NotR and cloned at the Not1 site of pN1–3XFLAG.

pcDNA6-FLAG: To generate this construct, Vec-FLAGF and Vec-FLAGR oligonucleotides were annealed and digested with NotI and Apa1. The fragment was cloned into Not1-Apa1 sites of pcDNA6/myc-HisB (Invitrogen).

L1-01F-EGFP: EGFP cassette was amplified from 99RPS-EGFP (43) and cloned at the Ale1 site present in the 3′-UTR of L1.

FL-01F(RT−) (D702Y): This construct was generated by site-directed mutagenesis on pJCC5(L1RP) template. The fragment containing the desired mutation was liberated by digesting with restriction enzyme BsrG1 and swapped into BsrG1 site of FL-01F.

ORF1F RR261.262AA: To generate this construct, 99 RPS JM111 (43) was digested with restriction enzymes NotI and AciI. The 1.9-kb fragment was cloned in NotI-Apa1 sites of ORF1F.

ORF1FRR260.210.211AAA: This construct was generated by site-directed mutagenesis on the ORF1F template.

ORF1CCF: To generate this construct, L15′-UTR plus CC domain was amplified from pJCC5(L1RP) (42) using primer pair CC-EcoF/CCFLAG-NotR. The fragment was cloned into EcoRI-Not1 sites of pcDNA6/myc-HisB (Invitrogen).

ORF1ΔRRMF: To generate this construct, ORF1F was digested with restriction enzymes Bsu36I and BstEII, which
will remove part of the CC domain and retain part of the RRM
domain (CC domain stretch 52–153 amino acids; RRM
domain stretch 157–239 amino acids; Bsu36I and BstEI digestion
removed 108–239 amino acids stretch). The vector part
(6694 bp) was gel purified, end filled and ligated. Four clones
were screened to get the correct ORF1 clone.
ACTBF: To generate this construct, the actin coding region
was amplified from ‘Ultimate ORF clones’ (obtained from the
HIT center, Johns Hopkins University) using PCR primers
HsACTB-EcoF/HsACTB-FLAGNotR. The fragment was
cloned into EcoRI-NotI sites of pcDNA6/myc-HisB.

LDHAF: To generate this construct, the LDHA coding region
was amplified from ‘Ultimate ORF clones’ (obtained from the
HIT center, Johns Hopkins University) using PCR primers
HsLDHA-EcoF/HsLDHA-FLAGNotR. The fragment was
cloned into EcoRI-NotI sites of pcDNA6/myc-HisB.

FL-O1FRRR261.262AA: This construct was generated by swap-
ing the NotI-AfeI fragment from construct
ORF1FRRR261.262AAA to construct FL-O1F.

FL-O1FRR261.262AA: This construct was generated by swap-
ing the NotI-AfeI fragment from construct ORF1FRR261.262-
AA to construct FL-O1F.

Alu Del linker: This construct was made by site-directed muta-
genesis using Alu-NeoF template (21). All constructs were
confirmed by DNA sequencing.

L1 retrotransposition in HEK293T cells
Approximately 1 × 10^5 HEK293T cells were plated in each well
of a six-well tissue culture plate. After 8–10 h incubation,
1.0 µg of L1-EGFP construct was transfected into cells. Retro-
transposition efficiency was assayed by FACS analysis at 60 h
post-transfection as described previously (43,65).

L1-RNP purification
The protocol described below is for a 100-mm dish and can be
scaled up or down for different plate formats based on the
culture vessel surface area (cm^2). At 8–12 h before transfection,
1.8 × 10^6 HEK293T cells in 10 ml of cell culture media without
antibiotics were plated to achieve 50% confluency at the time of
media was mixed with 12

LYSED in 1 ml lysis buffer [100 mM KCl, 5 mM MgCl2, 10 mM
HEPES (pH 7.0), 0.5% NP-40, 0.5 mM DTT] containing
RNase and protease inhibitors, for 20 min on ice. The super-
antant was collected by centrifugation at 12,000 g for 8 min at
4°C. Anti-FLAG agarose beads (Sigma) were prepared accord-
ing to manufacturer’s instructions (Sigma). Approximately
20 µl packed agarose beads were incubated with 1 ml lysate
for 1 h at 4°C. The beads were washed 5 times with wash
buffer 1 (25 mM HEPES (pH 7.0) 250 mM KCl, 5 mM MgCl2,
0.1% NP-40) and once with wash buffer 2 (25 mM HEPES, pH
7.0, 100 mM KCl, 5 mM MgCl2, 0.1% NP-40). The complex
was eluted by incubating at 4°C for 20 min in 100 µl wash
buffer 2, containing 150 µg/ml 3X FLAG peptide. RNPs were
stored in aliquots in 25% glycerol at −70°C.

LEAP assay
The assay was performed as described by Kulpa et al. (45) with
minor modification. The reaction was performed in a 20 µl reaction
volume containing 1 µl RNP, 4 µl 5X buffer (250 mM Tris–
HCl, pH 8.3 at room temperature; 400 mM KCl; 20 mM MgCl2),
1 µl dNTPs (10 mM), 1 µl LEAP primer (50 µM), 0.5 µl
RNasein (20 U/µl), 2 µl DTT (0.1M) and 10.5 µl nuclease-free
water at 37°C for 1 h. The PCR reaction was performed in 25 µl
total volume containing 1 µl LEAP reaction, 0.5 µl linker-
specific reverse primer (10 µM), 0.5 µl transcript-specific forward primer (10 µM), 12.5 µl 2X Go Taq green master mix
(Promega) and 10.5 µl water using PCR conditions of one
cycle at 94°C for 30 s followed by 35 cycles at 94°C for 20 s,
56°C for 20 and 72°C for 20 s and finally one cycle at 72°C for 2
min. The products were resolved in a 2.0% agarose gel. The
bands were excised, gel extracted and cloned in a TOPO TA
cloning vector. Clones were checked by colony PCR using
M13F and M13R. Clones containing inserts were gel excised and
sequenced using M13 Reverse Primer. The LEAP product
sequences are presented in a Supplementary Material.

PAR-CLIP
The protocol was performed as described in Hafner et al. (41)
with minor modification. HEK293T cells were transiently trans-
fected with plasmid constructs and grown for 24 h before adding
100 µM 4-SU and grown for another 12 h. UV-irradiated cells
were lysed and treated with RNaseT1, and the RNA–protein
complex was immunoprecipitated with anti-FLAG agarose
beads for 1 h at 4°C. RNA was labeled with [γ-32P] ATP, and the
crosslinked RNA–proteins were resolved in a SDS–PAGE
gel. ORF1 and HuR RNA–protein complexes were cut from
the gel. The RNA was recovered and converted into a CDNA
library using a small RNA cloning protocol and deeply
sequenced (46).

Synthesis of biotinylated RNA
Linearized, double stranded DNA with SP6 promoter sequence
at the 5’ end was used as a template for RNA synthesis. Biotin-
ylated RNA was synthesized by SP6 RNA polymerase using
100-ng DNA template and labeling mix containing an optimal
concentration of unlabeled nucleotides and biotin-16-UTP in a
20 µl reaction volume. The reaction was incubated at 37°C for
one hour. The template DNA was digested by incubating at 37°C
for 15 min after adding 1 µl of RNase-free DNase I. Biotinylated RNA was purified by ethanol precipitation.
The quality and quantity of the RNA was assayed by electrophoresis
on a neutral agarose gel, or on a polyacrylamide gel depending on
the size of the RNA.
**In vitro ORF1p Alu RNA interaction**

A construct containing a single FLAG epitope at the C-terminus of ORF1p was transiently transfected into HEK293T cells, and cytoplasmic lysate was prepared 48 h post-transfection. Cytoplasmic lysate (100 µg) containing FLAG-ORF1p was incubated with 1 µg of purified biotinylated Alu transcript for 30 min at 25°C. The presence of ORF1p in the pull down material was analyzed by immunoblotting using anti-FLAG antibody.

**Northern blot analysis**

L1-RNPs were purified from one 100-mm disc after transfecting FL-O1F into HEK293T cells as described in the L1-RNP purification section. The RNA was isolated from L1-RNPs by acid phenol: chloroform (5:1) (pH 4.5) extraction followed by ethanol precipitation. RNA was mixed with NorthernMax–Gly Sample Loading Dye (Ambion) at a 1:1 RNA: dye ratio, incubated at 65°C for 15 min, and chilled on ice for 5 min before separating on a 0.8% denatured agarose gel. The RNA was transferred to a nylon membrane, UV-crosslinked and pre-hybridized at 68°C for one hour, followed by overnight hybridization at 68°C with a BGH anti-sense riboprobe (200 bp), and labeled with digoxigenin (DIG)-11-UTP. The next day the membrane was washed, immunodetected with anti-DIG-Ap Fab fragments (Roche), visualized with the chemiluminescence substrate CDP-star (Roche) and exposed.

**qRT-PCR**

L1-RNPs were purified from 2 × 10^7 cells following transfection of FL-O1F. Total lysate (3.0 ml) was incubated with anti-FLAG-conjugated agarose beads, and the L1-RNPs were eluted in 150 µl volume by FLAG peptide competition. LEAP reaction was conducted using 2 µl of the FLAG elution in a 40 µl reaction volume that was used for semi-quantitative LEAP assay. The remainder (148 µl) was used to isolate total RNA from RNPs by Trizol (Ambion). Total RNA (1.05 µg) was dissolved in 30 µl RNa-free water, and 9 µl (∼280 ng) total RNA was treated with DNase (Promega) in 20 µl reaction volume. The DNase was heat-inactivated, and 9 µl DNase-treated RNA was used for cDNA synthesis in a 20 µl reaction volume using oligo-dT primer and superscript III reverse transcriptase (Invitrogen). For quantitative PCR, 0.5 µl of the RT reaction was used in a 20 µl reaction volume using SYBR Green PCR master mix (Qiagen) in Viia™ 7 Real-Time PCR System (Applied Biosystems). Three independent runs were used to calculate ΔΔ CT values. Similarly, for the ORF2(EN) construct, RNA was isolated from purified RNPs. The yield of RNA was ∼135 ng (4.5 ng/µl). Around 80 ng total RNA was treated with DNase in 20 µl reaction volume, and the same procedure was followed as described for FL-O1F.

**Analysis of PAR-CLIP sequencing results**

Five libraries were sequenced: two replicates for FL-O1F, two replicates for ORF1F and one replicate for HuR. Raw sequence data and alignment pileups in bigWig format are available through GEO (GSE43801). Residual adapter sequence (GATCTCGTATGCCCCTCTTCTGCCTTG) was removed using cutadapt (72), and the trimmed reads were aligned to the reference genome assembly GRCh37 using bowtie (73) with command-line options –best, –chunkmbs 512. Note that these settings allow for non-unique alignments—in cases where a read has more than one equally good alignment possibility one is chosen at random. This is necessary to consider mappings to repeat families whose members are very similar to one another (e.g. L1HS). Alignments were then filtered to keep only reads with T-to-C changes relative to the top strand (A-to-G for reads aligned to the bottom strand). Only unique sequences were retained to eliminate PCR duplicates, and the alignments were indexed using Tabix (74). Indices for each of the five libraries (1F replicates 1 and 2, ORF1F replicates 1 and 2, and HuR) were then compared against positions of RepeatMasker (A.F.A. Smit, R. Hubley & P. Green. RepeatMasker at http://repeatmasker.org) annotations obtained from the UCSC Genome Browser (75). Total aligned reads were counted for each repeat family represented in the RepeatMasker annotation of hg19/GRCh37 for all five libraries independently. The ORF1p (ORF1F) and L1-RNP (L1-O1F) counts were compared pairwise against HuR PAR-CLIP data to ascertain whether RNA binding for a given repeat was significantly greater for L1-RNP/ORF1. Significant differences between HuR and ORF1/L1-RNP counts for each repeat family were assessed through a Chi-squared test using the total number of aligned reads with T > C transitions and unique sequences for each sample (Supplementary Material, Table S1) versus the number of unique reads with T > C changes aligned to a given repeat annotation. To correct for multiple testing, a Bonferroni-adjusted threshold of p < 1e-06 was used for significance. To explore the binding of L1-RNP/ORF1 to transcribed genes, we only considered reads with a unique best alignment to the reference genome. The Tabix-indexed unique alignments with PAR-CLIP-induced T-to-C transitions were compared against a flattened list of exons. The flattened exon list consists of all non-overlapping exon annotations from UCSC Known Genes on hg19/GRCh37 (51). Where two exon annotations from alternate transcripts overlap, the longest combined interval was kept. Significant differences between HuR and ORF1/L1-RNP binding were assessed similarly to Repeatmasker annotations, but for each gene/exon combination, reads aligned to Repeatmasker annotations were excluded. The adjusted significance cut-off used for exons for differential binding to an exon annotation was p < 1e-08. The gene expression percentile for HEK293T cells was derived from data published by Takahashi et al. (76). Raw expression data from three HGU133Plus 2.0 arrays corresponding to untreated 293 T cells (GEO accession GSM711410, GSM711411 and GSM711412) were normalized by RMA (77) via the justRMA function in the affy package in Bioconductor (www.bioconductor.org). For each probe set, the expression level was averaged across the three normalized arrays and ranked to yield an expression percentile. PP counts for each parent gene were obtained based on annotations provided by pseudogene.org human pseudogenes build 69 (53).

**DATA ACCESSIBILITY**

Data are available in FASTQ format and as pileups in bigWig format (http://genome.ucsc.edu/goldenPath/help/bigWig.html) from GEO under accession GSE43801.
SUPPLEMENTARY MATERIAL
Supplementary Material is available at HMG online.

ACKNOWLEDGEMENTS

We thank M Gorospe and T Heidmann for the HuRF and Alu-Neos™ constructs, respectively; members of the Kazazian Lab for critical discussion; D Sigmon and L Cheung for technical assistance; the DNA sequencing cores at Johns Hopkins; HIT Center ChemCore of Johns Hopkins University for providing ACTB and LDHA clone; Dan Arking’s laboratory for use of his q-PCR machine [ViiA™ 7 Real-Time PCR System (Applied Biosystems)]; and Samantha Maragh and Tyler Creamer for helping in analyzing qRT-PCR data.

Conflict of Interest statement. None declared.

FUNDING
This work was supported by a grant from the NIH (RO1GM099875) and NIH (RC4MH092880) to H.H.K.

REFERENCES


