Dissection of affinity captured LINE-1 macromolecular complexes

M.S. Taylor[1,*], I. Altukhov[2,*], K.R. Molloy[3,*], P. Mita[4], H. Jiang[5], E.M. Adney[4,6], A. Wudzinska[4], S. Badri[7], D. Ischenko[2], G. Eng[1], K.H. Burns[6,8], D. Fenyö[4], B.T. Chait[3], D. Alexeev[9], M.P. Rout[5], J.D. Boeke[4], and J. LaCava[4,5,#]

[1]Massachusetts General Hospital, Department of Pathology (Boston, MA, United States); [2]Moscow Institute of Physics and Technology, (Dolgoprudny, Russian Federation); [3]The Rockefeller University, Laboratory of Mass Spectrometry and Gaseous Ion Chemistry (New York, NY, United States); [4]Institute for Systems Genetics, NYU Langone Health, Department of Biochemistry and Molecular Pharmacology (New York, NY, United States); [5]The Rockefeller University, Laboratory of Cellular and Structural Biology (New York, NY, United States); [6]Johns Hopkins University School of Medicine, McKusick-Nathans Institute of Genetic Medicine (Baltimore, MD, United States); [7]NYU Langone Health, Department of Pathology (New York, NY, United States); [8]Johns Hopkins University School of Medicine, Department of Pathology (Baltimore, MD, United States); [9]Novosibirsk State University, (Novosibirsk, Russian Federation).

[*]Equal contribution
[#]Corresponding Author: J.L. – jlacava@rockefeller.edu

1. Summary

Long Interspersed Nuclear Element-1 (LINE-1, L1) is a mobile genetic element active in human genomes. L1-encoded ORF1 and ORF2 proteins bind L1 RNAs, forming ribonucleoproteins (RNPs). These RNPs interact with diverse host proteins, some repressive and others required for the L1 lifecycle. Using differential affinity purifications, quantitative mass spectrometry, and next generation RNA sequencing, we have characterized the proteins and nucleic acids associated with distinctive, enzymatically active L1 macromolecular complexes. Among them, we describe a cytoplasmic intermediate that we hypothesize to be the canonical ORF1p/ORF2p/L1-RNA-containing RNP, and we describe a nuclear population containing ORF2p, but lacking ORF1p, which likely contains host factors participating in target-primed reverse transcription.

2. Keywords

LINE-1, L1, retrotransposon, ribonucleoprotein, RNP, interactome, mass spectrometry, affinity proteomics, biochemistry, protein interactions

3. Introduction

Sequences resulting from retrotransposition constitute more than half of the human genome and are considered to be major change agents in eukaryotic genome evolution (Kazazian, 2004). L1 retrotransposons have been particularly active in mammals (Furano *et al*, 2004), comprising ~20% of the human genome (Lander *et al*, 2001); somatic retrotransposition has been widely implicated in cancer progression (Lee *et al*, 2012; Tubio *et al*, 2014) and may even play a role in neural development (Muotri *et al*, 2005). Despite the magnitude of their contributions to mammalian genomes, L1 genes are modest in size. A full-length L1 transcript is ~6 knt long and functions as a bicistronic mRNA that encodes two polypeptides, ORF1p and ORF2p (Ostertag & Kazazian, 2001), which respectively comprise a homotrimeric RNA binding protein with nucleic acid chaperone activity (Martin & Bushman, 2001) and a multifunctional protein with endonuclease and reverse transcriptase activities (Mathias *et al*, 1991; Feng *et al*, 1996). Recently, a primate-specific third ORF, named *ORF0*, has been identified on the Crick strand of the L1 gene; this ORF encodes a 71 amino acid peptide and may generate insertion site-dependent ORFs via splicing (Denli *et al*, 2015). ORF1p and ORF2p are thought to interact preferentially with the L1 RNA from which they were translated (in *cis*), forming a ribonucleoprotein (RNP) (Kulpa & Moran, 2006; Taylor *et al*, 2013) considered to be the canonical direct intermediate of retrotransposition (Hohjoh & Singer, 1996; Kulpa & Moran, 2005; Martin, 1991; Kulpa & Moran, 2006; Doucet *et al*, 2010). L1 RNPs also require host factors to complete their lifecycle (Suzuki *et al*, 2009; Peddigari *et al*, 2013; Dai *et al*, 2012; Taylor *et al*, 2013) and, consistent with a fundamentally parasitic relationship (Beauregard *et al*, 2008), the host has responded by evolving mechanisms that suppress retrotransposition (Goodier *et al*, 2013; Arjan-Odedra *et al*, 2012; Goodier *et al*, 2012; Niewiadomska *et al*, 2007). It follows that as the host and the parasite compete, L1 expression is likely to produce a multiplicity of RNP forms engaged in discrete stages of retrotransposition, suppression, or degradation.

Although L1 DNA sequences are modestly sized compared to typical human genes, L1 intermediates are nevertheless RNPs with a substantially sized RNA component; e.g. larger than the ~5 knt 28S rRNA (Gonzalez *et al*, 1985) and approximately three to four times the size of a "typical" mRNA transcript (Lander *et al*, 2001; Sommer & Cohen, 1980). Therefore, it is likely that many proteins within L1 RNPs form interactions influenced directly and indirectly by physical contacts with the L1 RNA. We previously reported that L1 RNA comprised an estimated ~25% of mapped RNA sequencing reads in ORF2p-3xFLAG affinity captured fractions (Taylor *et al*, 2013). We also observed that the retention of ORF1p and UPF1 within affinity captured L1 RNPs was reduced by treatment with RNases (Taylor *et al*, 2013). In the same study we observed that two populations of ORF2p-associated proteins could be separated by split-tandem affinity capture (ORF2p followed by ORF1p), a two-dimensional affinity enrichment procedure (Caspary *et al*, 1999; Taylor *et al*, 2013). Initial characterization of these two L1 populations by western blotting suggested that discrete L1 populations were likely primed for function in different stages of the lifecycle. We therefore expected additional uncharacterized complexity in the spectrum of L1-associated complexes present in our affinity enriched fractions.

In this study, we have used quantitative mass spectrometry (MS) to investigate the proteomic characteristics of endogenously assembled ectopic L1-derived macromolecules present in an assortment of affinity-enriched fractions. We revisited RNase treatment and split-tandem affinity capture approaches and complemented them with RNA sequencing, enzymatic analysis, and in-cell localization of ORF proteins by immunofluorescence microscopy (see also the companion manuscript by **Mita *et al* [co-submission]**). We additionally explored proteomes associated with catalytically-inactivated ORF2p point mutants and monitored the rates of protein exchange from L1 macromolecules *in vitro.* Taken together, our data support the existence of a variety of putative L1-related protein complexes.

4. Results

Affinity proteomic experiments conducted in this study use quantitative MS based upon metabolic labeling (Oda *et al*, 1999). Two main experimental designs (and modifications thereof) facilitating quantitative cross-sample comparisons have been used: SILAC (Ong *et al*, 2002; Wang & Huang, 2008) and I-DIRT (Tackett *et al*, 2005; Taylor *et al*, 2013). In these approaches, cells are grown for several doublings in media containing amino acids composed either of naturally-occurring 'light' isotopes or biologically identical 'heavy' isotopes (e.g. ^{13}C, ^{15}N lysine and arginine), such that the proteomes are thoroughly labeled. Protein fractions derived from the differently labeled cell populations, obtained e.g. before and after experimental manipulations are applied, are mixed and the relative differences in proteins contributed by each fraction are precisely measured by mass spectrometry. In addition to the above cited studies, these approaches have been adapted to numerous biological questions using a variety of analytical frameworks e.g. (Byrum *et al*, 2011; Luo *et al*, 2016; Trinkle-Mulcahy *et al*, 2008; Ohta *et al*, 2010; Kaake *et al*, 2010; Geiger *et al*, 2011). Because it is challenging to speculate on the potential physiological roles of protein interactions that form after extraction from the cell, we often use I-DIRT, which allows the discrimination of protein-protein interactions formed in-cell from those occurring post-extraction. Our prior affinity proteomic study, based on I-DIRT, identified 37 putative *in vivo* interactors (Taylor *et al*, 2013), described in **Table 1**. In this study we primarily analyze the behaviors of these "I-DIRT significant" L1 interactors, in order to determine their molecular associations and ascertain the variety of distinctive macromolecular complexes formed in-cell that copurify with affinity-tagged ORF2p. The complete lists of proteins detected in each experiment are presented in the supplementary information (see **Supplementary File 1**). We have represented any ambiguous protein group, which occurs when the same peptides identify a group of homologous protein sequences, with a single, consistently applied gene symbol and a superscript 'a' in all figures. **Supplementary File 1** contains the references to other proteins explaining the presence of the same

peptides. For example, RPS27A, (ubiquitin) UBB, UBC, and (ribosomal Protein L40) UBA52 can be explained by common ubiquitin peptides shared by these genes. RPS27A-specific peptides were not identified in this study, but we retained the nomenclature for consistency with our previous work; HSPA1A is reported in this study, but cannot be distinguished from the essentially identical protein product of HSPA1B.

Table 1

Gene Symbol	Uniprot Symbol	Protein	co-IP with
L1RE1	Q9UN81	ORF1p	ORF1/2
N/A	O00370	ORF2p	ORF1/2
MOV10	Q9HCE1	Putative helicase MOV-10	ORF1/2
PABPC1	P11940	Polyadenylate-binding protein 1	ORF1/2
PABPC4	Q13310	Polyadenylate-binding protein 4	ORF1/2
UPF1	Q92900	Regulator of nonsense transcripts 1	ORF1/2
ZCCHC3	Q9NUD5	Zinc finger CCHC domain-containing protein 3	ORF1/2
FKBP4	Q02790	Peptidyl-prolyl cis-trans isomerase FKBP4	ORF2
HAX1	O00165	HCLS1-associated protein X-1	ORF2
HMCES	Q96FZ2	Embryonic stem cell-specific 5-hydroxymethylcytosine-binding protein	ORF2
HSP90AA1	P07900	Heat shock protein HSP 90-alpha	ORF2
HSP90AB1	P08238	Heat shock protein HSP 90-beta	ORF2
HSPA1A	P0DMV8	Heat shock 70 kDa protein 1A	ORF2
HSPA8	P11142	Heat shock cognate 71 kDa protein	ORF2
IPO7	O95373	Importin-7	ORF2
NAP1L1	P55209	Nucleosome assembly protein 1-like 1	ORF2
NAP1L4	Q99733	Nucleosome assembly protein 1-like 4	ORF2
PARP1	P09874	Poly [ADP-ribose] polymerase 1	ORF2
PCNA	P12004	Proliferating cell nuclear antigen	ORF2
PURA	Q00577	Transcriptional activator protein Pur-alpha	ORF2
PURB	Q96QR8	Transcriptional activator protein Pur-beta	ORF2
RPS27A	P62979	Ubiquitin-40S ribosomal protein S27a	ORF2
TIMM13	Q9Y5L4	Mitochondrial import inner membrane translocase subunit Tim13	ORF2
TOP1	P11387	DNA topoisomerase 1	ORF2
TOMM40	O96008	Mitochondrial import receptor subunit	ORF2

		TOM40 homolog	
TUBB	P07437	Tubulin beta chain	ORF2
TUBB4B	P68371	Tubulin beta-4B chain	ORF2
YME1L	Q96TA2	ATP-dependent zinc metalloprotease YME1L1	ORF2
CORO1B	Q9BR76	Coronin-1B	ORF1
DDX6	P26196	Probable ATP-dependent RNA helicase DDX6	ORF1
ERAL1	O75616	GTPase Era, mitochondrial	ORF1
HIST1H2BO	P23527	Histone H2B type 1-O	ORF1
LARP7	Q4G0J3	La-related protein 7	ORF1
MEPCE	Q7L2J0	7SK snRNA methylphosphate capping enzyme	ORF1
PABPC4L	P0CB38	Polyadenylate-binding protein 4-like	ORF1
TROVE2	P10155	60 kDa SS-A/Ro ribonucleoprotein	ORF1
YARS2	Q9Y2Z4	Tyrosine--tRNA ligase, mitochondrial	ORF1

Except where noted otherwise, the presented experiments were conducted in suspension-cultured HEK-293T_{LD} cells, using a synthetic L1 construct - *ORFeus*-HS - driving the expression 3xFLAG-tagged L1 (*ORF1*; *ORF2::3xFLAG*; 3'-UTR) from a tetracycline inducible minimal-CMV promoter, harbored on a mammalian episome (pLD401 (Taylor *et al*, 2013; An *et al*, 2011; Dai *et al*, 2012)). All L1-related macromolecules described in this study were obtained by affinity capture of ORF2p-3xFLAG before further experimental manipulations were applied. We consider macromolecules containing L1 RNA (L1 RNPs, discussed throughout) and/or an L1 cDNA (i.e. L1 coding potential) to be L1s, as are their ectopic plasmid-borne and endogenous gDNA counterparts, reflecting the complexity and diversity of L1 forms arising from its lifecycle. In an effort to characterize this complexity, we have carried out RNA sequencing and enzymatic activity analyses on several affinity captured fractions, complementing the proteomic analyses.

4.1. RNase-sensitivity exhibited by components of affinity captured L1 RNPs

Figure 1 (panels A-C) illustrates the approach and displays the findings of our assay designed to reveal which proteins depend upon the presence of intact L1 RNA for retention within the obtained

L1 RNPs. Briefly, metabolically-labeled affinity captured L1s were treated either with a mixture of RNases A and T1 — thus releasing proteins that require intact RNA to remain linked to ORF2p and the affinity medium — or BSA, as an inert control. After removing the fractions released by the RNase or BSA treatments, the proteins remaining on the affinity media were eluted with lithium dodecyl sulfate (LDS), mixed together, and then analyzed by MS. Proteins released, and so depleted, by RNase treatment were thus found to be more abundant in the BSA-treated control. The results obtained corroborate and extend our previous findings: ORF1p and UPF1 exhibited RNase-sensitivity (Taylor *et al*, 2013). We also observed that ZCCHC3 and MOV10 exhibited RNase-sensitivity to a level similar to ORF1p. The remaining I-DIRT significant proteins were RNase-resistant in this assay. With the exception of the PABPC1/4 proteins (and ORF2p itself, see **Discussion**), the I-DIRT significant proteins (**colored nodes, Fig. 1C**) that were resistant to RNase treatment (nearest the origin of the graph) classify ontologically as nuclear proteins (GO:0005634, $p \approx 3 \times 10^{-4}$, see **Methods**). These same proteins were previously observed as specific L1 interactors in I-DIRT experiments targeting ORF2p but not in those targeting ORF1p; in contrast, the proteins that demonstrated RNase-sensitivity: ORF1, MOV10, ZCCHC3, and UPF1 were observed in both ORF1p and ORF2p I-DIRT experiments (**Table 1**). Stated another way, the proteins released upon treating an affinity captured ORF2p fraction with RNases are among those that can also be obtained when affinity capturing ORF1p directly, while those that are RNase-resistant are not ORF1p interactors (Taylor *et al*, 2013). The ORF1p-linked, I-DIRT significant, RNase-sensitive proteins were too few to obtain a high confidence assessment of ontological enrichment; but, when combined with remaining proteins exhibiting sensitivity to the RNase treatment (**black nodes, Fig. 1C**), they together classified as 'RNA binding' (GO:0003723, $p \approx 1 \times 10^{-11}$). This analysis also revealed a statistically significant overrepresentation of genes associated with the exon junction complex (EJC, GO: 0035145, $p \approx 1 \times 10^{-6}$, discussed below). Hence, the overlapping portion of the ORF1p- and ORF2p-associated interactomes appeared to depend upon intact L1 RNA. Host-encoded proteins segregated into groups that responded differentially to RNase treatment, with a

substantial population of RNase-resistant interactors linked to both ORF2p and the nucleus. This observation led to the hypothesis that our ORF2p-3xFLAG affinity captured L1s constitute a composite purification of at least, but not limited to, (1) a population of L1-RNA-dependent, ORF1p/ORF2p-containing L1 RNPs, and (2) an ORF1p-independent nuclear population associated with ORF2p.

While effects of PABPC1, MOV10, and UPF1 on L1 activity have been described (Arjan-Odedra *et al*, 2012; Taylor *et al*, 2013; Dai *et al*, 2012), effects of ZCCHC3 on L1 remained uncharacterized. ZCCHC3 is an RNA-binding protein associated with poly(A)+ RNAs (Castello *et al*, 2012) but otherwise little is known concerning its functions. Notably, in a genome-wide screen, small interfering (si)RNA knockdown of ZCCHC3 was observed to increase the infectivity of the Hepatitis C, a positive sense RNA virus (Li *et al*, 2009); and ZCCHC3 was observed to copurify with affinity captured HIV, a retrovirus, at a very high SILAC ratio (>10), supporting the specificity of this interaction (Engeland *et al*, 2014). We therefore explored the effects on L1 mobility both of over-expression and siRNA knockdown of ZCCHC3. Over-expression of ZCCHC3 reduced L1 retrotransposition to ~10% that observed in the control, consistent with a negative regulatory role for ZCCHC3 in the L1 lifecycle; small interfering RNA (siRNA) knockdown of ZCCHC3 induced a modest increase in retrotransposition compared to a scrambled control siRNA (~1.9x ±0.1; **Supplementary File 2**). Moreover, although not among our I-DIRT hits (see **Discussion**), the presence of EJC components (MAGOH, RBM8A, EIF4A3, UPF1) among the RNase-sensitive fraction of proteins intrigued us, given that L1 genes are intronless. We speculated that L1s may use EJCs to enhance nuclear export, evade degradation by host defenses, and/or aggregate with mRNPs within cytoplasmic granules. For this reason we carried out a series of siRNA knockdowns of these EJC components and other physically or functionally related proteins found in the affinity captured fraction (listed in **Supplementary File 2**). siRNA knockdowns of RBM8A and EIF4A3 caused inviability of the cell line. We found that knocking-down MAGOH or the EJC-linked protein IGF2BP1 (Jønson *et al*, 2007) reduced retrotransposition by ~50%, consistent with a role in L1

proliferation; although these knockdowns also caused a reduction in viability of the cell line (see **Discussion**).

4.2. Split-tandem separation of compartment-specific L1 ORF-associated complexes

To further test our hypothesis and better characterize the components of our L1 fraction, we conducted split-tandem affinity capture. **Figure 1 (panels D-F)** illustrates the approach and displays the findings of the assay, which physically separated ORF1p/ORF2p-containing L1 RNPs from a presumptive 'only-ORF2p-associated' population. Briefly, metabolically-labeled L1s were affinity captured by ORF2p-3xFLAG (1st dimension) and the obtained composite was subsequently further fractionated by α-ORF1p affinity capture (2nd dimension, or split-tandem capture), resulting in α-ORF1p-bound and unbound fractions. The bound fraction was eluted from the affinity medium with LDS. The bound and unbound fractions were then mixed and analyzed by MS to ascertain proteomic differences between them. The fraction eluted from the α-ORF1p medium contained the population of proteins physically linked to *both* ORF2p and ORF1p, whereas the supernatant from the α-ORF1p affinity capture contained the proteins associated *only* with ORF2p (and, formally, those which have dissociated from the ORF1p/ORF2p RNP). The results corroborated our previous observations that: i) almost all of the ORF1p partitioned into the elution fractions, ii) a quarter of the ORF2p (~26%) followed ORF1p during the α-ORF1p affinity capture, iii) roughly half of the UPF1 (~55%) followed ORF1p, and iv) most of the PCNA (~87%) remained in the ORF1p-depleted supernatant fraction (**Fig. 1F**, and consistent with prior estimates based on protein staining and western blotting (Taylor *et al*, 2013)); thus v) supporting the existence of at least two distinct populations of L1-ORF-protein-containing complexes in our affinity purifications.

The population eluted from the α-ORF1p affinity medium (**Fig. 1D,** far right gel lane, and nodes located in the upper right of the graph, **panel F)** is consistent with the composition of the ORF1p/ORF2p-containing L1 RNP suggested above. Our split-tandem separation segregated the

constituents of the L1 fraction comparably to the RNase-sensitivity assay, both in terms of which proteins co-segregated with ORF1p/ORF2p (compare **Fig. 1C** and **F,** blue nodes, upper right of graphs) as well as those which appear to be linked only to ORF2p (compare Figs. **1C** and **F**, magenta nodes, lower left of the graphs). The ORF1p/ORF2p RNPs obtained by split-tandem capture included putative *in vivo* interactions associated with both α-ORF1p and α-ORF2p I-DIRT affinity capture experiments; whereas the unbound, ORF1p-independent fraction includes proteins previously observed as significant only in α-ORF2p I-DIRT experiments (**Table 1**). Analysis of the nodes whose degree of ORF1p association was similar to that of UPF1 (**blue nodes exhibiting ≥ 55% ORF1p co-partitioning**, **Fig. 1F**) revealed that they map ontologically to a 'cytoplasmic ribonucleoprotein granule' classification (GO:0036464, $p \approx 6 \times 10^{-8}$; **see Discussion**). In contrast, all sixteen proteins exhibiting ORF1p co-partitioning approximately equal to or less than that of ORF2p were predominantly found in the supernatant fraction and were enriched for cell-compartment-specific association with the nucleus (GO:0005634, $p \approx 4 \times 10^{-5}$; **Fig. 1F: all magenta nodes ≤ 36%**). These two fractions therefore appear to be associated with different cell compartments, reaffirming our postulate: the ORF1p/ORF2p-containing population is a cytoplasmic intermediate related to the canonical L1 RNP typically ascribed to L1 assembly in the literature, and the predominantly ORF2p-associated population comprises a putative nuclear interactome; each therefore referred to, respectively, as cytoplasmic and nuclear L1 interactomes hereafter.

From the same analysis, we noted that PURA, PURB, PCNA, and TOP1 which all partition predominantly with nuclear L1, exhibited an ontological co-enrichment (termed 'nuclear replication fork,' GO:0043596, $p \approx 3 \times 10^{-4}$). The nodes representative of PURA, PURB, and PCNA appeared to exhibit a striking proximity to one another, suggesting highly similar co-fractionation behavior potentially indicative of direct physical interactions. In an effort to examine this possibility, we graphed the frequency distribution of the proximities of all three-node-clusters observed within **Figure 1F**, revealing the likelihood of the PURA/PURB/PCNA cluster to be $p = 3.2 \times 10^{-7}$ (see **Appendix 1**). We therefore concluded that PURA, PURB, PCNA, and (perhaps at a lower affinity)

TOP1, likely constitute a physically associated functional module interacting with L1. In further support of this assertion, we noted that known functionally linked protein pairs PABPC1/PABPC4 (cytoplasmic) (Jønson *et al*, 2007; Katzenellenbogen *et al*, 2007) and HSPA8/HSPA1A (nuclear) (Jønson *et al*, 2007; Nellist *et al*, 2005) also exhibited comparable co-partitioning by visual inspection, and statistical testing of these clusters revealed the similarity of their co-partitioning to be significant at $p \approx 0.001$ for the former, and $p \approx 0.0002$ for the latter. The observed variation in co-partitioning behavior between the different proteins comprising the nuclear L1 fraction might reflect the presence of multiple distinctive (sub)complexes present within this population.

To validate our hypothesis that these proteins are associated with ORF2p in the nucleus, possibly engaged with host genomic DNA, we carried out ORF2p-3xFLAG affinity capture from chromatin-enriched sub-cellular fractions and found that the co-captured proteins we identified (**Supplementary File 3**) overlapped with those described above as nuclear interactors, including: PARP1, PCNA, UPF1, PURA, and TOP1. We previously demonstrated that silencing PCNA expression adversely affects L1 retrotransposition (Taylor *et al*, 2013), in this study we found that knocking down TOP1 approximately doubled retrotransposition frequency, while a more modest 1.4x increase effect was observed for PURA, and no substantial effect was observed for PURB, compared to a scrambled siRNA control. In contrast, over-expression of PURA reduced retrotransposition to ~20% of the expected level (**Supplementary File 2**). IPO7 was also observed among the putative ORF2p co-factors within the chromatin enriched fraction, congruent with its matching behavior in Figures 1C and 1F. Notably, IPO7 functions as a nuclear import adapter for HIV reverse transcription complexes (Fassati *et al*, 2003). Several other proteins were observed that did not previously exhibit I-DIRT specificity (**Supplementary File 3**).

4.3. L1 RNA and LEAP activity in affinity captured fractions

Because the L1 RNA is an integral component of proliferating L1s, and because we observed that interactions between ORF2p, ORF1p, and some host proteins were sensitive to treatment with

RNases, we sought to characterize the RNAs present in our samples. We extracted RNAs from each of the three fractions produced by split-tandem affinity capture (**Fig. 1D**) and carried out RNA sequencing; **Figure 2A** displays the sequence coverage observed across the entirety of our synthetic L1 construct in each fraction, revealing a normalized ~2-fold difference in abundance between the elution and supernatant fractions. Synthetic L1s constituted ~60% of the mapped, annotated sequence reads in the fractions eluted from the α-FLAG and α-ORF1p affinity media, and ~30% of the reads in the ORF1p-depleted supernatant fraction; sequencing reads mapping to protein coding genes made up the majority of the remaining annotated population in all fractions. We observed that a substantial number of reads mapped to unannotated regions of the human genome, in particular in the supernatant fraction, enriched for putative nuclear L1 complexes; the breakdown of mapped and annotated sequencing reads is summarized in **Figure 2B** and expanded in **Supplementary File 4**.

Retrotransposition-competent L1 RNPs form in *cis*, with ORF proteins binding to the L1 RNA that encoded them ("*cis* preference"), presumably at the site of translation in the cytoplasm (Kulpa & Moran, 2006; Wei *et al*, 2001). Given that ORF1/2p partitioned to the split-tandem elution fraction along with the greater fraction of L1 RNA, yet only ORF2p and a lesser portion of the L1 RNA were observed in the supernatant, an important consideration regarding these fractions is: to what extent they contain L1 macromolecules capable of proliferation. To address this question, we performed the LINE-1 element amplification protocol (LEAP) on split-tandem affinity captured fractions (**Fig. 2C; Supplementary File 4**), including a *ΔORF1* construct (pLD561) as a control (Taylor *et al*, 2013). LEAP is currently the best biochemical assay for functional co-assembly of L1 RNA and proteins (Kulpa and Moran 2006); it measures the ability of ORF2p to amplify its associated L1 RNA by reverse transcription. To execute LEAP on the α-ORF1p affinity captured fraction, we developed a competitive di-peptide elution reagent based on the linear peptide sequence used to generate the α-ORF1p 4H1 monoclonal antibody: residues 35-44 in ORF1p ((Khazina *et al*, 2011; Taylor *et al*, 2013); see **Methods**). We were thus able to assay the partitioning of enzymatic activity within the different populations of copurifying proteins in a split-

tandem affinity capture experiment. Our data showed robust LEAP activity in split-tandem supernatant and elution fractions. We note that our 3xFLAG eluted fractions have been shown to possess ~70-fold higher specific activity than L1 RNPs obtained by sucrose cushion velocity sedimentation (Taylor *et al*, 2013), hence the activity levels detected far exceed those obtained by sedimentation.

4.4. ORF1p/ORF2p immunofluorescence protein localization

Although our proteomic and biochemical analyses supported the existence of distinctive nuclear and cytoplasmic L1 populations, our prior immunofluorescence (IF) analyses did not reveal an apparent nuclear population, leading us to revisit IF studies. Previously, IF of ORF1p and ORF2p in HeLa and HEK-293T cells yielded two striking observations: i) ORF2 expression was seemingly stochastic, with ORF2p observed in ~30% of cells; and ii) while ORF1p and ORF2p co-localized in cells that exhibited both, we did not observe an apparent nuclear population of either protein (Taylor *et al*, 2013). Subsequently, we noted an absence of mitotic cells from these preparations. Reasoning that these cells were lost due to selective adherence on glass slides, and noting that cell division has been reported to promote L1 transposition (Xie *et al*, 2013; Shi *et al*, 2007), we repeated the assays using puromycin-selected Tet-on HeLa cells grown on fibronectin coated coverslips. The results are shown in **Figure 3**.

The modified IF assay corroborated our prior results in that nearly all the cells exhibited cytoplasmic ORF1p and a minority subset of ~1/3rd also exhibited co-localized cytoplasmic ORF2p (Fig. 3A, top row). We also observed a rare and previously unrecognized subpopulation of cells, consisting of pairs exhibiting nuclear localized ORF2p (Fig. 3A, middle row); because these cells occurred in proximal pairs, were presumed them to have recently gone through mitosis. Statistical analysis of microscopy images displaying cells with nuclear localized ORF2p confirmed their proximities to be significantly closer than those of randomly selected cells (**Fig. 3B; Supplementary File 5**). Expression of *ORF2* in the absence of *ORF1* (*ΔORF1*; pLD561) resulted in

the majority of cells exhibiting cytoplasmic ORF2p, consistent with our previous work (Taylor *et al*, 2013). We did not observe instances of nuclear ORF2p using the *ΔORF1* construct (**Fig. 3A, bottom row**), suggesting that ORF1p is required for ORF2p nuclear localization (see **Discussion**). In a separate study, including more detailed analyses of ORF protein localization, **Mita *et al* (co-submission)** observed that both ORF proteins enter the nucleus of HeLa cells during mitosis, however, nuclear ORF1p does not seem to be physically associated with nuclear ORF2p (see **Discussion**). Taken together, the data obtained from the modified IF experiments aligned well with our proteomic and biochemical data; L1 expression resulted in *at least* two distinct populations: cytoplasmic complexes containing both ORF1p and ORF2p, and nuclear complexes containing ORF2p while potentially lacking ORF1p.

4.5. The effects of retrotransposition-blocking point mutations on the interactomes of affinity captured L1 RNPs

Based on the hypothesis that our composite purifications contain bona fide nuclear intermediates, we decided to explore the effects of catalytic point mutations within the ORF2p endonuclease and reverse transcriptase domains, respectively. We reasoned that such mutants may bottleneck L1 intermediates at the catalytic steps associated with host gDNA cleavage and L1 cDNA synthesis, potentially revealing protein associations that are important for these discrete aspects of target-primed reverse transcription (TPRT), the presumed mechanism of L1 transposition (Luan *et al*, 1993; Feng *et al*, 1996; Cost *et al*, 2002). For this we used an H230A mutation to inactivate the endonuclease activity (EN^- / pLD567), and a D702Y mutation to inactivate the reverse transcriptase activity (RT^- / pLD624) (Taylor *et al*, 2013). **Figure 4** illustrates the approach and displays the findings of our assay. Broadly, while we observed comparable RNA-level properties between samples (**Fig. 4B**, **Supplementary File 4**), our findings revealed several classes of distinctive protein-level behaviors (**Fig. 4C**). Two classes of behavior appeared to be particularly striking: (1) the yield of constituents of cytoplasmic L1s was reduced, relative to WT, by the EN^- mutation, yet

elevated by the RT$^-$ mutation (**Fig. 4C, left side**); and (2) numerous constituents of nuclear L1s were elevated in yield by the EN$^-$ mutation but reduced or nominally unchanged, relative to WT, by the RT$^-$ mutation (**Fig. 4C, right side**). With respect to the second group, IPO7, NAP1L4, NAP1L1, FKBP4, HSP90AA1, and HSP90AB1 were all elevated in the EN$^-$ mutants, potentially implicating these proteins as part of an L1 complex (or complexes) immediately preceding DNA cleavage. Notably, there is a third class of proteins, including PURA/B, PCNA, TOP1, and PARP1, that all respond similarly to both EN$^-$ and RT$^-$ mutants compared to WT, exhibiting reduced associations with the mutant L1s; although, the RT$^-$ mutant showed a larger effect size on the PURA/B proteins. These data suggest that cleavage of the host genomic DNA fosters associations between L1 and this third class of proteins, but that interactions with PURA/B may be further enhanced by L1 cDNA production. Other nuclear L1 proteins: HSPA8, HAX1, HSPA1A, TUBB, and TUBB4B were increased in both mutants. To better visualize the range of behaviors exhibited by our proteins of interest, and the population at large, we cross-referenced the relative enrichments of each protein detected in both experiments, shown in **Figure 4D**. We noted the same striking trend mentioned above, that two seemingly opposite behavioral classes of interactors could also be observed globally among all proteins associating with ORF2p catalytic mutants (see **Fig. 4C, left side and right side,** and **Fig. 4D**), creating the crisscross pattern displayed (see also **Fig. 4-S1**). Notably, the pattern observed appears to track with the relative behavior of ORF1p, which, along with other cytoplasmic L1 factors is elevated in RT$^-$ mutants and reduced in EN$^-$ mutants. We therefore speculate that the sum of observed interactomic changes include effects attributable directly to the catalytic mutations as well as indirect effects resulting from the response of ORF1p to the mutations.

4.6. Dynamics of L1 RNPs *in vitro*

We next decided to measure the *in vitro* dynamics of proteins copurifying with affinity captured L1s, reasoning that proteins with comparable profiles are likely candidates to be physically linked to one another or otherwise co-dependent for maintaining stable interactions with L1s. To achieve this, we first affinity captured heavy-labeled, affinity-tagged L1s and subsequently incubated them,

while immobilized on the medium, with light-labeled, otherwise identically prepared cell extracts from cells expressing untagged L1s (Luo *et al*, 2016). In this scenario, heavy-labeled proteins present at the zero time point are effectively "infinitely diluted" with light-labeled cell extract. The exchange of proteins, characterized by heavy-labeled proteins decaying from the immobilized L1s and being replaced by light-labeled proteins supplied by the cell extract, was monitored by quantitative MS. These experiments were conducted using constructs based on the naturally occurring $L1_{RP}$ sequence (Dai *et al*, 2014; Taylor *et al*, 2013; Kimberland *et al*, 1999). **Figure 5** illustrates the approach and displays the findings of our assay. We observed three distinctive clusters of behaviors (**Fig. 5B, C**). Notably, ORF1p, ZCCHC3, and the cytoplasmic poly(A) binding proteins clustered together, forming a relatively stable core complex. Exhibiting an intermediate level of relative *in vitro* dynamics, UPF1 and MOV10 clustered with TUBB, TUBB4B, and HSP90AA1. A third, and least stable, cluster consisted of only nuclear L1 interactors.

4.7. Multidataset integration

Having observed coordinated and distinctive behaviors exhibited by groups of L1 interacting proteins across several distinctive biochemical assays, we then integrated the data and calculated the behavioral similarity of the I-DIRT-significant interactors, producing a dendrogram; **Figure 6** displays their relative similarities. A cluster containing the putative cytoplasmic L1 components (MOV10, UPF1, ZCCHC3, PABC1/4, ORF1p) was observed, as was a cluster containing PURA/B, PCNA, TOP1, PARP1, aligning with our assessments of the separated datasets (**Figs. 1, 4, 5**). In addition to these, we also observed three distinctive clusters derived from the nuclear L1 interactome. We believe that this is likely to reflect the presence of a collection of distinctive macromolecules.

5. Discussion

In this study we have characterized biochemical, interatomic, enzymatic, and cellular localization properties of ectopically expressed L1s. Through the assays explored, we observed discrete and coordinated behaviors, permitting us to refine our model of L1 intermediates, **diagrammed in Figure 7**. We propose a cytoplasmic L1, composed of ORF1/2p, L1 RNA, PABPC1/4, MOV10, UPF1, and ZCCHC3, that constitutes an abundant, canonical RNP intermediate often referred to in the literature. MOV10, UPF1, and ZCCHC3 are apparently substoichiometric to ORF2p in our preparations, therefore it may be that only a subset of cytoplasmic intermediates engages these host restriction factors. On the other hand, this apparent relative abundance may simply reflect a lower *in vitro* stability of UPF1 and MOV10 within this complex (**Fig. 5**). We also propose a second more complex population, lacking (or with significantly less) ORF1p, that constitutes a nuclear L1 intermediate or, more likely, a collection of ORF2p-associated macromolecules. We note that *Alu* elements exhibit ORF2p-dependent mobilization that does not require ORF1p, but appears to be enhanced by ORF1p in some contexts (Dewannieux *et al* 2003; Wallace *et al* 2008); this is not true for L1 or processed pseudogenes, and we conclude *Alu* RNPs likely exploit an alternate mechanism of nuclear entry. The nuclear L1 population is enriched for factors linked to DNA replication and repair, including PURA, PURB, PCNA, TOP1, and PARP1; we propose that these proteins, along with ORF2p, form part of a direct intermediate of TPRT, although these components may not all act in synergy. Our proposals are broadly supported by the findings of **Mita *et al* (co-submission)**, who present data to support the hypothesis that PCNA-associated ORF2p is not appreciably associated with ORF1p, and also identified TOP1 and PARP1 in complex with ORF2p/PCNA.

Although the protein purification approach was the similar, we observed an apparently larger proportion of L1 RNA in our recent preparations than in our previous study. We reported that L1 constituted ~25% of mapped reads previously (Taylor *et al*, 2013); a comparable result was obtained when we reanalyzed that data using the pipeline described here (see **Methods**): ~93% of

reads in reanalyzed our 2013 dataset mapped to the human genome, and L1 constituted ~20% of reads mapped to annotated features ("annotated reads") in 3xFLAG eluates. In this study we report that ~60% of annotated reads mapped to synthetic L1 in 3xFLAG eluates (**Fig. 2A**). The higher proportion of L1 recovered may be due to the combination of higher fidelity RNA preparative methods and advanced sequencing technology used here; we observed ~10x more total reads mapping to L1 and comparatively improved, more uniform coverage across the entire L1 sequence, likely explaining the discrepancy. We also noted that the number of normalized reads mapped to L1 in our initial 3xFLAG elutions ("input") and subsequent tandem-purified α-ORF1p elutions were comparable, and yet ~1/2 as many were seen in the α-ORF1p supernatant fraction (**Fig. 2A, B**). We suspect that this is due to saturation in library preparations or sequencing steps for the "input" and "elution" fractions, but conclude that more L1 RNA is in the 'cytoplasmic' elution fraction than the 'nuclear' supernatant.

We observed substantial and comparable LEAP activity in both our tandem-purified ORF1p+ ("elution") and ORF1p– (immuno-depleted "supernatant") populations (**Fig. 2C, Supplementary File 4**). To our knowledge, these represent the simplest and purest endogenously assembled L1 RNPs yet reported that exhibit robust signal in the LEAP assay. We note that, our results demonstrating robust activity in the nuclear-enriched supernatant fraction (depleted of ORF1p) may contrast with previous reports of reduced LEAP activity in constructs where ORF1p RNA binding was compromised (Kulpa and Moran, 2006), but our fractions merit further study and comparisons on the basis of ORF2p and RNA levels to determine specific activity.

5.1. Cytoplasmic L1 macromolecules

ORF1p, MOV10, UPF1, and ZCCHC3 are released from L1 RNPs by treatment with RNases (**Fig. 1**), indicating the importance of the L1 RNA in the maintenance of these interactions. In this context, the L1 ORF and poly(A) binding proteins support L1 proliferation (Kulpa & Moran, 2006;

Dai *et al*, 2012; Wei *et al*, 2001), whereas ZCCHC3 (**Supplementary File 2**) and MOV10 (Goodier *et al*, 2012; Arjan-Odedra *et al*, 2012) function in repressive capacities. Although UPF1 might also be expected to operate in a repressive capacity through its role in nonsense mediated decay (NMD), we previously demonstrated that UPF1's role does not apparently resemble that of canonical NMD and it acts as an enhancer of retrotransposition despite negatively affecting L1 RNA and protein levels, supporting the possibility of repressive activity in the cytoplasm and proliferative activity in the nucleus (Taylor *et al*, 2013). Notably, MOV10 has been implicated in the recruitment of UPF1 to mRNA targets through protein-protein interactions (Gregersen *et al*, 2014). However, we observed that MOV10 exhibited a greater degree of RNase-sensitivity than UPF1, indicating that, if MOV10 directly modulates the UPF1 interactions with L1, a sub-fraction of UPF1 exhibits a distinct behavior (UPF1 is ~62% as sensitive to RNase treatment as MOV10, **Fig. 1C**). Bimodal UPF1 behavior can also be seen in split-tandem capture experiments, and UPF1 was recovered with L1s affinity captured from fractionated chromatin (further discussed below), and only about half of the UPF1 exhibits ORF1p-like partitioning with the canonical L1 RNP (**Fig. 1F**). Presumably, the RNase-sensitive fraction, released in concert with MOV10, is the same fraction observed in cytoplasmic L1s obtained by split-tandem capture. In contrast, PABPC1 and C4 exhibit strong ORF1p-like partitioning (comparable to MOV10), but appear wholly insensitive to RNase treatment. This is most likely due to the fact that neither RNase A nor T1 cleave RNA at adenosine residues (Volkin & Cohn, 1953; Yoshida, 2001); hence poly(A) binding proteins may not be ready targets for release from direct RNA binding by the assay implemented here (or generally, using these ribonucleases). Failure to release of ORF2p into the supernatant upon RNase treatment is expected due to its immobilization upon the affinity medium (Dai *et al*, 2014). However, we note that ORF2p binding to the L1 RNA has also been proposed to occur at the poly(A) tail (Doucet *et al*, 2015), raising the related possibility of close physical association on the L1 RNA between ORF2p and PABPC1/4 in cytoplasmic L1 RNPs. ORF1p, PABPC1/4, MOV10, ZCCHC3, and UPF1, all behaved comparably in response to EN^- and RT^- catalytic mutations, decreasing together

in EN$^-$ mutants, and increasing together in RT$^-$ mutants (**Fig. 4C**). Moreover, when the exchange of proteins within L1 RNPs was monitored directly, PABPC1/4 and ZCCHC3 exhibited nearly identical stability, well above the background distribution; UPF1 and MOV10 also exhibited comparable kinetics to one another, falling into an intermediary stability cluster (**Fig. 5B, C**).

RNase-sensitivity was displayed by numerous proteins not previously identified as putative L1 interactors (**Table 1, Fig. 1**; (Taylor *et al*, 2013)). A known limitation of I-DIRT (and many SILAC-based analyses) is that it cannot discriminate non-specific interactors from specific but rapidly exchanging interactors (Wang & Huang, 2008; Luo *et al*, 2016; Smart *et al*, 2009). Our samples likely contain rapidly exchanging, physiologically relevant factors that were not revealed by I-DIRT under the experimental conditions used. With this in mind, we note members of the exon junction complex (EJC), RBM8A (Y14), EIF4A3 (DDX48), and MAGOH, are among our RNase-sensitive constituents, with all exhibiting a similar degree of RNase-sensitivity (**Fig. 1C**, labeled black dots). Crucially, these proteins are physically and functionally connected to UPF1 (reviewed in (Schweingruber *et al*, 2013)), and physically to MOV10 (Gregersen *et al*, 2014), both validated L1 interactors. We therefore hypothesize that EJCs may constitute bona fide L1 interactors missed in our original screen. This may seem unexpected because canonical L1 RNAs are thought not to be spliced, but this assumption has been challenged by one group (Belancio *et al*, 2006), and splicing-independent recruitment of EJCs has also been demonstrated (Budiman *et al*, 2009). Perhaps more compelling, EJC proteins exhibited a striking similarity in RNase-sensitivity to MOV10 (**Fig. 1C**). EIF4A3 has been suggested to form an RNA-independent interaction with MOV10 (Gregersen *et al*, 2014), and MOV10 is a known negative regulator of L1, making it attractive to speculate that these proteins were recruited and released in concert with MOV10 and/or UPF1.

Ectopically expressed canonical L1 RNPs have been shown to accumulate in cytoplasmic stress granules (Doucet *et al*, 2010; Goodier *et al*, 2010), and our observation of UPF1, MOV10, and MAGOH in the RNase-sensitive fraction is consistent with this characterization (Jain *et al*, 2016). However, the additional presence of EIF4A3 and RBM8A suggested that our RNPs may instead

overlap with IGF2BP1 (IMP1) granules, reported to be distinct from stress granules (Jønson *et al*, 2007; Weidensdorfer *et al*, 2009). Consistent with this possibility, we observed IGF2BP1, YBX1, DHX9, and HNRNPU within the mixture of co-captured proteins (**Supplementary File 1**). We did not, however, observe canonical stress granule markers G3BP1 or TIA1 (Goodier *et al*, 2007; Jain *et al*, 2016; Doucet *et al*, 2010). Surprisingly, siRNA knockdown of IGF2BP1 substantially reduced L1 retrotransposition; however, we note that the cytotoxicity associated with knocking-down EJC components may confound interpretation (**Supplementary File 2**). Given the result obtained, IGF2BP1 appears to support L1 proliferation. Consistent with an established function (Bley *et al*, 2015; Weidensdorfer *et al*, 2009), IGF2BP1 granules may sequester and stabilize L1 RNPs in the cytoplasm, creating a balance of L1 supply and demand that favors proliferation over degradation. Although human L1 does not contain a known IRES, it is known that *ORF2* is translated by a non-canonical mechanism (Alisch *et al*, 2006), and IGF2BP1 may promote this (Weinlich *et al*, 2009).

5.2. Nuclear L1 macromolecules

The fraction eluted from the α-ORF1p medium contained the population of proteins physically linked to both ORF2p and ORF1p and greatly resembled the components released upon RNase treatment, hence these linkages primarily occur through the L1 RNA (or are greatly influenced by it). In contrast, the supernatant from the α-ORF1p affinity capture contained the proteins we speculate to be associated with ORF2p, but not ORF1p; moreover, fully intact RNA does not appear to be essential to the maintenance of these interactions. An exciting alternate interpretation to direct protein-protein linkage is that some of the L1 RNAs in this population may be at least partially hybridized to L1 cDNAs, which would render them RNase resistant: at the salt concentration used in our RNase assay (0.5 M; **Fig. 1C**), RNase A is unlikely to cleave the RNA component of DNA/RNA hybrids (Halász *et al*, 2017; Wyers *et al*, 1973), and such activity is not expected of RNase T1. This interpretation is supported by several pieces of indirect evidence: (1) the presence of well-known DNA binding factors (**Fig. 1**); (2) the presence of several of these same factors

(PARP1, PCNA, PURA, and TOP1) in ORF2p-3xFLAG affinity captured from enriched chromatin (**Supplementary File 3**); (3) The pronounced decrease in stable *in vivo* co-assembly of TOP1, PCNA, PARP1, PURA, and PURB in affinity captured L1 fractions harboring ORF2p EN$^-$ and RT$^-$ mutations (**Fig. 4**), with a greater effect in RT$^-$ mutations; and (4) our L1 preparations exhibit RT activity (**Fig. 2C**, *in vitro*; as well as *in vivo* (Taylor *et al*, 2013)). If true, linkage of subcomplexes via DNA/RNA hybrids would further support the nuclear origin of much of this fraction; further study is needed. Notable within this group of putative nuclear interactors was the PURA/PURB/PCNA cluster (**Fig. 1F**), with TOP1 also in close proximity, ontologically grouping to the nuclear replication fork (GO:0043596). Separately, a few physical and functional connections have been shown for PURA/PURB (Knapp *et al*, 2006; Kelm *et al*, 1999; Mittler *et al*, 2009), PCNA/TOP1 (Takasaki *et al*, 2001), and PURA/PCNA (Qin *et al*, 2013). Notably, PURA, PURB, and PCNA have been independently linked to replication-factor-C / replication-factor-C-like clamp loaders (Kubota *et al*, 2013; Havugimana *et al*, 2012). Given that we also observe tight clustering of protein pairs known to be physically and functionally linked, e.g. PABPC1/4 (Jønson *et al*, 2007; Katzenellenbogen *et al*, 2007) and HSPA8/1A (Jønson *et al*, 2007; Nellist *et al*, 2005), and because we have established PCNA as a positive regulator of L1 retrotransposition (Taylor *et al*, 2013), we propose that the [PURA/B/PCNA/TOP1] group is a functional sub-complex of nuclear L1. In addition, although it does not cluster as closely to the [PURA/B/PCNA/TOP1] group, PARP1 is found within the putative nuclear L1 population and is functionally linked with PCNA, specifically stalled replication forks (Bryant *et al*, 2009; Min *et al*, 2013; Ying *et al*, 2016). Further tying them together, these proteins all also exhibited substantial affinity capture yield decreases in response to mutations that abrogated ORF2p EN or RT activity (**Fig. 4**). This is compelling because these ORF2p enzymatic activities are required in order for it to manipulate DNA and traverse the steps of the L1 lifecycle that benefit from physical association with replication forks. One caveat to this interpretation is that, while knocking down PCNA reduced L1 retrotransposition (Taylor *et al*, 2013), no such effect was observed for TOP1 or PURA/B, which led instead to mild increases in L1

activity (**Supplementary File 2**). These proteins may be physically assembled within a common intermediate, but functionally antagonistic. HSP90 proteins were also observed in this fraction, and are also linked with stalled replication forks (Arlander *et al*, 2003; Ha *et al*, 2011), but exhibited a distinctive response to catalytic mutants, accumulating in EN$^-$ mutants while exhibiting a modest decrease in RT$^-$ mutants. The recruitment of the ORF2p/PCNA complex to stalled replication forks has been also proposed by **Mita *et al* (co-submission)**.

As mentioned above, we previously speculated that an RNase-insensitive fraction of L1-associated UPF1 may support retrotransposition in conjunction with PCNA in the nucleus ((Azzalin & Lingner, 2006; Taylor *et al*, 2013) and **Mita *et al* [co-submission]**). In contrast to other PCNA-linked proteins, catalytic inactivation of ORF2p did not robustly affect the relative levels of co-captured UPF1, and UPF1 behaved in a distinct manner during tandem capture. The equivocal behavior of UPF1 in several assays (**Figs. 1, 4, & 5**) supports UPF1's association with both the putative cytoplasmic and nuclear L1 populations, the latter being additionally supported by the association of UPF1 with ORF2p-3xFLAG captured from chromatin (**Supplementary File 3**). NAP1L4, NAP1L1, FKBP4, HSP90AA1, and HSP90AB1 (Baltz *et al*, 2012; Castello *et al*, 2012; Simon *et al*, 1994; Rodriguez *et al*, 1997; Peattie *et al*, 1992) are associated with RNA binding, involved in protein folding and unfolding, and function as nucleosome chaperones. An interesting possibility is that they have a nucleosome remodeling activity that may be required to allow reverse transcription to begin elongating efficiently, or for assembly of nucleosomes on newly synthesized DNA.

5.3. Future Studies

An obvious need is the continued validation of putative interactors by *in vivo* assays. Genetic knockdowns coupled with L1 insertion measurements by GFP fluorescence (Ostertag *et al*, 2000) provides a powerful method to detect effects on L1 exerted by host factors. However, this approach can sometimes be limited by cell viability problems associated with important genes; it is therefore

critical to control for this (**Supplementary File 2**). IF and high-resolution microscopy may be useful to demonstrate co-localization of putative L1-associated proteins and may also be informative, warranting effort to identify appropriate antibodies and assay conditions. Bolstered by our analytical successes, RNA-sequencing, LEAP, and RNase-based affinity proteomics appear as notably high-value assays for further application-specific expansion and refinement.

Throughout this and our prior study (Taylor *et al*, 2013) we have used comparable *in vitro* conditions for the capture and analysis of L1 interactomes. However, we are aware that this practice has enforced a single biochemical "keyhole" through which we have viewed L1-host protein associations. It is important to expand the condition space in which we practice L1 interactome capture and analysis in order to expand our vantage point on the breadth of L1-related macromolecules (Hakhverdyan *et al*, 2015). In concert with this, we must develop sophisticated, automated, reliable, low-noise methods to integrate biochemical, proteomic, genomic, and ontological data; the first stages of which we have attempted in the present study. Although we have used I-DIRT to increase our chances of identifying bona fide interactors (Tackett *et al*, 2005; Taylor *et al*, 2013), it is clear, and generally understood, that some proteins not making the significance cut-off will nevertheless prove to be critical to L1 activity (Byrum *et al*, 2011; Luo *et al*, 2016; Joshi *et al*, 2013), such as demonstrated by our unexpected findings with IGF2BP1 (**Supplementary File 2**). Through further development, including reliable integration with diverse, publicly available interactome studies, we hope to enable the detection of extremely subtle physical and functional distinctions between (sub)complexes and their components, considerably enhancing reliable exploration and hypothesis formation. Furthermore, it is striking that no structures of assembled L1s yet exist; these are missing data that are likely to provide a profound advance for the mechanistic understanding of L1 molecular physiology. However, we believe that with the methods presented here, endogenously assembled ORF1p/ORF2p/L1-RNA-containing cytoplasmic L1 RNPs can be prepared at sufficiently high purity and yield **Fig. 1F**) to enable electron microscopy studies. Importantly, we have shown that our affinity captured fractions are enzymatically active for reverse

transcription of the L1 RNA (**Fig. 2C**; (Taylor *et al*, 2013)), providing some hope that cryo-electron microscopy could be used to survey the dynamic structural conformations of L1s formed during its various lifecycle stages (Takizawa *et al*, 2017).

6. Methods

Key Resources Table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
gene (human) LINE-1	*ORFeus*-Hs; L1RP	10.1016/j.cell.2013.10.021; 10.1186/1759-8753-2-2		
cell line (human)	HEK-293T_LD	10.1016/j.cell.2013.10.021; 10.1128/MCB.06785-11		Mycoplasma testing was done regularly and was negative. We received an authenticated cell line from the ATCC and subsequently made them blastomycin resistant so we validated cells by blastomycin resistance.
transfected construct (human)	pLD401; pLD561; pLD567; pLD624; pMT302; pMT289	10.1016/j.cell.2013.10.021		
antibody	anti-FLAG; anti-ORF1p	10.1016/j.cell.2013.10.021	Sigma-Aldrich Cat# F1804, RRID:AB_262044; custom made, Abmart: 4H1	

peptide, recombinant protein	ORF1p N-terminal di-peptide	this paper		
software, algorithm	Scripts for IF (Fig. 3); formal analysis used custom R code throughout	this paper		Scripts are in **Supplementary File 5**; R code in – **https://bitbucket.org/altukhov/line-1/**

The preparation of L1 RNPs was carried out essentially as previously described (Taylor *et al*, 2013, 2016), with modifications described here. Briefly, HEK-293T_{LD} cells (Dai *et al*, 2012) transfected with L1 expression vectors were cultured as previously described or using a modified suspension-growth SILAC strategy described below. L1 expression was induced with with 1µg / ml doxycycline for 24 hours, and the cells were harvested and extruded into liquid nitrogen. In all cases the cells were then cryogenically milled (LaCava *et al*, 2016) and used in affinity capture experiments and downstream assays. Custom computer code written in the R programming language was used in the analysis of mass spectrometry and RNA sequencing data; it has been published on https://bitbucket.org (Altukhov, 2017).

6.1. Modified SILAC Strategy

Freestyle-293 medium lacking Arginine and Lysine was custom-ordered from Life Technologies, and heavy or light amino acids plus proline were added at the same concentrations previously described (Taylor *et al*, 2013), without antibiotics. Suspension-adapted HEK-293T_{LD} were spun down, transferred to SILAC medium and grown for >7 cell divisions in heavy or light medium. On day 0, four (4) 1L square glass bottles each containing 200 ml of SILAC suspension culture at ~2.5×10^6 cells/ml were transfected using 1 µg/ml DNA and 3 µg/ml polyethyleneimine "Max" 40 kDa (Polysciences, Warrington, PA, #24765). A common transfection mixture was made by pre-mixing 800 µg DNA and 2.4mL of 1 mg/ml PEI-Max in 40 ml Hybridoma SFM medium (Life Technologies, Grand Island, NY, #12045-076) and incubating for 20 min at room temperature (RT);

10 ml of the mixture was added to each bottle. On day 1, cells (200 ml) were split 1:2.5 (final two bottles each containing 250 mL) without changing the medium. On day 3, the cells were induced with 1 µg/ml doxycycline, and on day 4 the cells were harvested and extruded into liquid nitrogen. Aliquots were tested by western blot and the per-cell expression of both ORFs was indistinguishable from puromycin-selected material described previously (**Appendix 1**); transfection efficiency was assessed at >95% by indirect immunofluorescence of expressed ORF proteins. The median lysine and arginine heavy isotope incorporation levels for cell lines presented in this study were > 90%, determined as previously described (Taylor *et al*, 2013).

6.2. RNase-Sensitivity Affinity Capture

Four sets of 200 mg of light (L) and heavy (H) pLD401 transfected cell powders, respectively, were extracted 1:4 (w:v) with 20 mM HEPES-Na pH 7.4, 500 mM NaCl, 1% (v/v) Triton X-100 (extraction solution), supplemented with 1x protease inhibitors (Roche, Indianapolis, IN, #11836170001). After centrifugal clarification, all of the L and H supernatants were pooled, respectively, and then split, resulting in two sets of cleared L and H extracts equivalent to duplicate 400 mg samples from each SILAC cell powder. These four samples were each subjected to affinity capture upon 20 µl α-FLAG magnetic medium. After binding and washing, one set of L and H samples were treated with a control solution consisting of 2 µl of 2 mg/ml BSA (Thermo Fisher Scientific, Waltham, MA, #23209) and 50 µl extraction solution, v:v (Ctrl); the other set of L and H samples was treated with a solution of 2 µl 2 mg/ml RNase A / 5000 u/ml RNase T1 (Thermo Fisher Scientific #EN0551) and 50 µl extraction solution, v:v (RNase). Samples were then incubated 30 min at RT with agitation, the supernatant was removed, and the medium was washed three times with 1 ml of extraction solution. The retained captured material was eluted from the medium by incubation with 40 µl 1.1x LDS sample loading buffer (Life Technologies #NP0007). To enable quantitative comparisons of fractions, the samples were combined, respectively, as follows: 30ul each of the ^{LL}RNase with ^{HH}Ctrl, and 30ul each of the Ctrl with RNase. These samples were reduced, alkylated and run until the dye front progressed ~6 mm on a 4-12% Bis-Tris

NuPAGE gel (Life Technologies, as per manufacturer's instructions). The gels were subsequently subjected to colloidal Coomassie blue staining (Candiano *et al*, 2004) and the sample regions ("gel-plugs") excised and processed for MS analyses, as described below.

6.3. Split-Tandem Affinity Capture

400 mg of light (L) and heavy (H) pLD401 transfected cell powders, respectively, were extracted and clarified as above. These extracts were subjected to affinity capture on 20 µl α-FLAG magnetic medium, 30 min at 4°C, followed by native elution with 50 µl 1 mg/ml 3xFLAG peptide (15 min, RT). 45 µl of the elution were subjected to subsequent affinity capture upon 20 µl α-ORF1 magnetic medium, resulting in a 45 µl supernatant (Sup) fraction depleted of ORF1p. Finally, the material was eluted (Elu) from the α-ORF1p medium in 45 µl 2.2x LDS sample loading buffer by heating at 70°C for 5 min with agitation. To enable quantitative comparisons of fractions the samples were combined, respectively, as follows: 28 µl each of the ^{L}Sup with ^{H}Elu, and 28 µl each of the ^{L}Elu with ^{H}Sup. These samples were then prepared as gel-plugs (as above) and processed for MS analyses, as described below.

6.4. Mass Spectrometry Sample Preparation and Data Acquisition

Gel plugs were excised, cut into 1 mm cubes, de-stained, and digested overnight with enough 3.1 ng/µl trypsin (Promega, Madison, WI, #V5280) in 25 mM ammonium bicarbonate to cover the pieces. In RNase-sensitivity and split-tandem SILAC analyses based on pLD401, as well as *in vitro* protein exchange experiments based on pMT302 and pMT289, an equal volume of 2.5 mg/ml POROS R2 20 µm beads (Life Technologies #1112906) in 5% v/v formic acid, 0.2% v/v TFA was added, and the mixture incubated on a shaker at 4°C for 24 hr. Digests were desalted on Stage Tips (Rappsilber *et al*, 2007), eluted, and concentrated by vacuum centrifuge to ~10 µl. ~3 µl were injected per LC-MS/MS analysis. RNase-sensitivity and split-tandem samples were loaded onto a PicoFrit column (New Objective, Woburn, MA) packed in-house with 6 cm of reverse-phase C18 material (YMC* Gel ODS-A, YMC, Allentown, PA). Peptides were gradient-eluted (Solvent A =

0.1 M acetic acid, Solvent B = 0.1 M acetic acid in 70% v/v acetonitrile, flow rate 200 nl/min) into an LTQ-Orbitrap-Velos or an LTQ-Orbitrap-XL mass spectrometer (Thermo Fisher Scientific) acquiring data-dependent CID fragmentation spectra. *In vitro* exchange samples were loaded onto an Easy-Spray column (ES800, Thermo Fisher Scientific) and gradient-eluted (Solvent A = 0.1% v/v formic acid in water, Solvent B = 0.1% v/v formic acid in acetonitrile, flow rate 300 nl/min) into an Q Exactive Plus mass spectrometer (Thermo Fisher Scientific) acquiring data-dependent HCD fragmentation spectra. In SILAC experiments comparing inactivated ORF2p catalytic mutants to WT (based on pLD401 [WT], pLD567 [EN^-], and pLD624 [RT^-]) peptides were extracted from the gel in two 1 hr incubations with 1.7% v/v formic acid, 67% v/v acetonitrile at room temperature with agitation. Digests were partially evaporated by vacuum centrifugation to remove acetonitrile, and the aqueous component was desalted on Stage Tips. Peptides were loaded onto an Easy-Spray column (ES800, Thermo Fisher Scientific) and gradient-eluted (Solvent A = 0.1% v/v formic acid in water, Solvent B = 0.1% v/v formic acid in acetonitrile, flow rate 300 nl/min) into an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific) acquiring data-dependent fragmentation spectra (either CID spectra alone, or CID and HCD spectra).

6.5. Mass Spectrometry Data Analysis

Raw files were submitted to MaxQuant (Cox & Mann, 2008) version 1.5.2.8 for protein identification and isotopic ratio calculation. Searches were performed against human protein sequences (UP000005640, April 2016), custom L1 ORF1p and ORF2p protein sequences, common exogenous contaminants, and a decoy database of reversed protein sequences. Search parameters included fixed modification: carbamidomethyl (C); variable modification: Arg10, Lys8, methionine oxidation; razor and unique peptides used for protein quantitation; requantify: enabled. Contaminants, low-scoring proteins and proteins with one razor+unique peptides were filtered out from the MaxQuant output file "proteingroups.txt". The list of contaminants was uploaded from the MaxQuant web-site (http://www.coxdocs.org/; "contaminants"). Additionally, proteins with the "POTENTIAL CONTAMINANT" column value "+" were filtered out. Proteins with at least 2

razor+unique peptides were retained for the analysis. H/(H+L) and L/(H+L) values were derived from unnormalized "ratio H/L" values and were used for plotting label-swapped RNase-sensitivity and split-tandem data. Unnormalized "ratio H/L" values were used to calculate H/(H+L) in ORF2p catalytic mutant comparisons and in vitro exchange experiments. These values have been referred to as "affinities" within the **Supplementary Materials**. Normalization and clustering procedures applied to data presented in the figures (**Supplementary File 1**) are detailed below and also in **Appendix 1**. Raw and processed data are available via ProteomeXchange with identifier PXD008542.

To plot RNase-sensitivity affinity capture results (**Fig. 1C**), these data were normalized such that proteins that did not change upon treatment with RNases are centered at the origin. The mean value and standard deviation were calculated using the distribution of distances from the origin. The distance threshold for p-value = 0.001 was calculated using the R programming language. A circle with radius equal to the threshold was plotted. Points with distances higher than the threshold were marked as black. To plot split-tandem affinity capture results (**Fig. 1F**), these data were normalized such that the ORF1p affinity was set to 1 and the distribution median was maintained. Probabilities associated with selected clusters were calculated based on the frequency distributions of 2- and 3-node clusters present in the data. To plot EN^- and RT^- mutant affinity capture results (**Fig. 4C**), the matrix of detected proteins for each experiment (EN^- and RT^-) was filtered to retain only proteins detected in at least two replicate experiments. The difference between the affinity value of ORF2p and 0.5 value was calculated for each experiment. The affinities of each protein were shifted by the calculated difference. To determine the statistical significance of differentially co-captured proteins between EN^- or RT^- and WT, respectively, we used a 1-sample t-test and applied Benjamini-Hochberg p-value correction. To determine the statistical significance of differentially co-captured proteins between EN^- and RT^- we used an unpaired t-test and applied Benjamini-Hochberg p-value correction. To plot *in vitro* dynamics (**Fig. 5B, C**), only proteins which were identified at all time points were used. The cosine similarity method was used to calculate distances between proteins,

and hierarchical clustering was used to visualize these distances. To integrate and plot the combined data (**Fig. 6**), we calculated Euclidean and cosine distances for each I-DIRT-significant protein pair present in each experiment. Euclidean distances were rescaled to the range (0, 0.9). Proteins not detected in any common experiments were assigned a Euclidian distance of 1 after rescaling. The total distance between protein pairs was calculated as d = log((rescaled Euclidean distance) * (cosine distance)). This distance was rescaled to the range (0, 1). Hierarchical clustering was used to visualize the calculated distances.

6.6. Gene Ontology (GO) Analysis

Genes corresponding to the proteins previously reported as significant by I-DIRT (Taylor *et al*, 2013) were tested for statistical overrepresentation using the default settings provided by http://www.panthnerdb.org (Mi *et al*, 2017, 2013), searches were conducted using GO complete molecular function, biological process, and cellular compartment: all results are compiled in **Supplementary File 6**.

6.7. RNA Sequencing Sample Preparation and Data Acquisition

RNA fractions were obtained from fractions of L1 macromolecules isolated from pLD401 expressing cells by split-tandem affinity capture (**Fig. 1D**) and from pLD567 and pLD624 expressing cells by affinity capture (**Fig. 4**). The fractions were produced as described above, except few adjustments to favor RNA extraction. Identical stock solutions were used for making buffers but were diluted to working concentration with nuclease-free water (Thermo Fisher Scientific #4387936) and supplemented with RNasin (Promega, Cat.# N2511) – 1:250 during sample extraction and 3xFLAG peptide elution, and 1:1000 during affinity media washing. 600 mg of cell powder was used per preparation, extracted as 3 x 200 mg and pooled after centrifugal clarification, producing ~3 ml of extract. The pooled extracts were combined magnetic affinity medium from 30 µl of slurry. 75 µl of 1 mg/mL 3xFLAG peptide was used for elution. ½ of the sample was saved for RNA extraction (input) and the other ½ was carried forward to split-tandem

IP, using 15 µl α-ORF1 affinity medium slurry. RNAs were extracted from input, α-ORF1 supernatant fractions, as well as directly from the α-ORF1 affinity medium (elution) with 500 µl of TRIzol (Thermo Fisher Scientific #15596026), following the manufacturer's instructions. Aqueous TRIzol extracts were re-extracted in an equal volume of chloroform, and the aqueous phase was again removed; 1 µl (~15 ug) of GlycoBlue (Thermo Fisher Scientific #AM9516) and 2ul of RNasin were added to this and mixed before combining with 250 µl of isopropanol and incubating for 10' on ice to precipitate RNA. Alcohol precipitates were centrifuged at 20k RCF for 30' @ 4°C and the pellets were washed twice with 500 µl of cold 70% ethanol, then air dried for 5' at RT and re-solubilized in 100 µl of nuclease-free water. Extracted RNAs in water were then further purified and concentrated using a Qiagen RNeasy MinElute Cleanup Kit (#74204) following the manufacturer's instructions, and eluted in 14 µl of nuclease-free water. 5 µl of purified RNA was used directly in RNA fragmentation. Libraries were prepared with unique barcodes and were pooled at equimolar ratios. The pool was denatured and sequenced on Illumina NextSeq 500 sequencer using high output V2 reagents and NextSeq Control Software v1.4 to generate 75 bp single reads, following manufacturer's protocols (#15048776, Rev.E).

6.7. RNA Sequencing Data Analysis

Human genome hg19 GRCh37.87 (FASTA) and annotation (GTF file) were downloaded from ENSEMBL (ftp://ftp.ensembl.org/pub/grch37/release-90) and reference FASTA and GTF files were created by combining the human genome and *ORFeus*-Hs from pLD401 (Taylor *et al.* 2013; **Supplementary File 7**: ORFeus-Hs_pLD401.gbk). To map sequencing reads onto the reference genome and produce differential gene expression analysis**:** (1) FASTAQ files were trimmed via trimmomatic (Bolger *et al*, 2014) using the following parameters: -phred33 -threads 8, LEADING:3 TRAILING:3 SLIDINGWINDOW:4:16 MINLEN:25; (2) mapping was performed via STAR (Dobin *et al*, 2013) version 2.5.3a (https://github.com/alexdobin/STAR) using the following parameters: -runThreadN 8, --quantMode GeneCounts, --outSAMtype BAM SortedByCoordinate, --outFilterMatchNmin 30; (3) the results were output to one binary alignment map file for each

sample matched to the reference; (4) genes with the coverage of 10 or more reads in at least 3 experiments were selected; and (5) data was normalized using the 'DESeq2' (Love *et al*, 2014) R package version 1.14.1. Raw and normalized mapped, annotated reads are described in **Supplementary File 4**. FASTAQ files are available through Gene Expression Omnibus at NCBI: GSE108270.

6.8. L1 element amplification protocol (LEAP)

We generated an N-terminally acetylated, C-terminally amidated version of the ORF1p peptide (MENDFDELRE) as a di-peptide composed of repeats of the same sequence linked by a four-unit polyethylene glycol moiety; which was used to elute ORF1p-containing complexes from α-ORF1p medium at a concentration of approximately 2 mM (**Appendix 1; Supplementary File 4**). Peptides were synthesized by standard Fmoc solid-phase synthesis methods (Kates & Albericio, 2000); the incorporation of a PEG spacer into the peptide sequence was accomplished using N-Fmoc-amido-(PEG) n-acid building blocks. 400 mg of cryogenically milled L1-expressing cells (pLD401 and pLD561) were subjected to split-tandem affinity capture as described above, but with native elution from α-ORF1p medium and included the addition of RNasin (Promega #N2515) at 1:500 v/v to the extraction buffer; 1x protease inhibitors and 1:200 v/v RNasin were also added to the 3xFLAG peptide and ORF1p-derived di-peptide solutions. For α-FLAG affinity capture, competitive elution was achieved using 60 µl of 1 mg/ml 3xFLAG peptide. Of this, 20 µl were held aside (Input), 40 µl were carried forward to α-ORF1p affinity capture. The ORF1p-depleted fraction was retained (Sup) and the captured material was eluted with 40 µl ORF1p di-peptide (Elu). Half of each fraction (Input, Sup, Elu) was set aside for protein analysis (**Supplementary File 4**) and to the other half, glycerol was added to 25% v/v (using a 50% v/v glycerol solution); the latter were subsequently analyzed for enzymatic activity by LEAP. Raw data resulting from these assays is located in **Supplementary File 4**. For LEAP, 2 µl from each of the above-described fractions were used in a 50 µl reaction, and 1 µl of each LEAP assay was used in SYBR Green qPCR (carried out in

triplicate) as previously described (Taylor *et al.* 2013). As controls, (1) an untagged L1RP construct was used in a "mock purification," and (2) pLD401-derived "Input" was heated at 100°C for 5 min and then added to the reaction mix, respectively. Neither produced detectable activity (**Supplementary File 4**). A second LEAP analysis was later carried out on an independently prepared set of fractions, prepared as above, stored frozen -80°C in 25% v/v glycerol.

6.9. ORF protein immunofluorescence analysis in HeLa cells

Tet-on HeLa M2 cells (Hampf & Gossen, 2007) (a gift from Gerald Schumann), were transfected and selected with 1 µg/ml puromycin for three days. Puromycin-resistant cells were plated on coverslips pre-coated for 1-2 hr with 10 µg/ml fibronectin in PBS (Life Technologies). 8-16 hr after plating, L1 was induced with 1 µg/ml doxycycline. 24 hr later, cells were fixed in 3% paraformaldehyde for 10 min. Fixative was then quenched using PBS containing 10 mM glycine and 0.2% w/v sodium azide (PBS/gly). The cells were permeabilized for 3 min in 0.5% Triton X-100 and washed twice with PBS/gly. Staining with primary and secondary antibodies was done for 20 min at room temperature by inverting coverslips onto Parafilm containing 45ml drops of PBS/gly supplemented with 1% BSA, mouse α-FLAG M2 (Sigma, 1:500), rabbit α-ORF1 JH73 (1:4000) (Taylor *et al*, 2013), Alexa Fluor 488 conjugated α-mouse IgG (Life Technologies, 1:1000), and Alexa Fluor 568 conjugated α-rabbit IgG (Life Technologies, 1:1000). DNA was stained prior to imaging with Hoechst 33285 (Life Technologies, 0.1 µg/ml). Epifluorescent images were collected using an Axioscop microscope (Zeiss, Jena, Germany) equipped for epifluorescence using an ORCA-03G CCD camera (Hamamatsu, Japan).

6.10. ORF2p+ nuclei proximity analysis

For each microscope field, nuclei were identified and spatially located using a custom script in ImageJ, consisting of Otsu thresholding and watershed transformation of DAPI signal to segment each of the nuclei. ORF2p positive nuclei were differentiated from ORF2p negative nuclei by using

another thresholding script for the ORF2p fluorescence channel and cross-registering the associated nuclei; all ORF2p positive nuclei were then hand-verified and then coordinates were converted into microns. The number of ORF2p+ nuclei per field, x, and a corresponding random distribution of x nuclei was calculated by randomly and repeatedly (n=1000) selecting x nuclei among all nuclei. The random distribution was used to calculate Bonferroni corrected p-values for the pairwise distances between ORF2p+ nuclei. The distribution of ORF2p+ inter-nuclei distances was then compared to the distribution of random inter-nuclei distances using Welch's t-test. The custom scripts used to select nuclei and calculate statistics, extracted data, calculated distances, p-values, and raw images are presented in the supplement **(Supplementary File 5;** original data: **Supplementary_IF_images.zip (Figure 3-source data 1))**.

7. Author Contributions

Conceptualization, J.L., M.S.T., J.D.B.; Methodology, J.L., K.R.M., M.S.T., I.A.; Software, I.A., S.B., G.E.; Formal Analysis, I.A., D.A., S.B., D.F., D.I.; Investigation, J.L., K.R.M., M.S.T., P.M., H.J., E.M.A., A.W.; Resources, M.P.R., B.T.C., D.A., J.D.B., D.F., K.H.B., J.L.; Data Curation, K.R.M., I.A., J.L., D.A.; Writing - Original Draft, J.L., M.S.T.; Writing - Review & Editing, J.L., M.S.T., P.M., M.P.R., J.D.B., K.H.B.; Visualization, I.A., J.L., M.S.T., D.I.; Supervision, J.L., D.A.; Project Administration, J.L.; Funding Acquisition, M.P.R., B.T.C., J.D.B., D.A., K.H.B.

8. Acknowledgements

This work was supported in part by National Institutes of Health (NIH) grants P41GM109824 (to M.P.R.), P41GM103314 (to B.T.C.), P50GM107632 (to J.D.B), and R01GM654321 (J.L.), and by the 5-100 Russian Academic Excellence Program. The mass spectrometric analysis of proteins co-captured with chromatin associated ORF2p (**Supplementary File 3**) was conducted within the NYU School of Medicine Proteomics Resource Lab, which is partially supported by the Laura and Isaac Perlmutter Cancer Center Support Grant, NIH P30CA16087, and NIH 1S10OD010582. RNA

sequence library preparation and next-generation sequencing was carried out by The Rockefeller University Genomics Resource Center. Peptide synthesis was performed by Henry Zebroski at The Rockefeller University Proteomics Resource Center. We thank Carolyn Machamer for advice and resources supporting fluorescence microscopy, and Lixin Dai for assistance with LEAP assays. This paper is subject to the NIH Public Access Policy. The authors declare no conflicts of interest.

9. Figure and Table Legends

Table 1. *Putative L1 interactors*: Through a series of affinity capture experiments (co-IP) using I-DIRT, we characterized a set of putative host-encoded L1 interactors (Taylor *et al*, 2013). The proteins observed were associated with both ORF1p and ORF2p (highlighted in blue), or only with only one ORF protein. Proteins only observed in association with ORF2p are highlighted in magenta. The two highlighted populations are the central focus of this study.

Figure 1. (**A)** *On-bead RNase-sensitivity assay*: L1 complexes were affinity captured by ORF2p-3xFLAG. The magnetic media were then treated with a solution containing either a mixture of RNases A and T1 or BSA. After treatment the supernatants were removed and the remaining bound material was released by treatment with LDS. Proteins requiring intact RNA to maintain stable interactions with immobilized ORF2p were released from the RNase-treated medium, while the BSA-treated sample provided a control for the spontaneous release of proteins from the medium over the time of the assay. Representative SDS-PAGE / Coomassie blue stained gel lanes are shown for each fraction. (**B**) The experiments described above was carried out in duplicate, once with light isotopically labeled cells (L) and once with heavy isotopically labeled cells (H), resulting in four label-swapped, SILAC duplicates (one light set & one heavy set). The four fractions were cross-mixed and the differential protein retention upon the affinity medium during the treatments (BSA vs. RNase) was assessed by quantitative MS. (**C**) Results from the RNase-sensitivity assay graphed as the fraction of each detected protein present in the BSA-treated sample (RNase-sensitive proteins are *more* present in the BSA treated sample), normalized such that proteins that did not change

upon treatment with RNases are centered at the origin. A cut-off of $p = 10^{-3}$ for RNase-sensitivity is indicated by a light gray circle; proteins that are RNase-sensitive with a statistical significance of $p < 10^{-3}$ are outside the circle. Proteins previously ranked significant by I-DIRT analysis (**Table 1**) are labeled and displayed in blue or magenta (as indicated); black nodes were not found to be significant by I-DIRT but were labeled if found to be RNase-sensitive; gray, unlabeled nodes were not found to be significant by I-DIRT. (**D**) *Split-tandem affinity capture*: L1 complexes were affinity captured by ORF2p-3xFLAG. After native elution with 3xFLAG peptide, this fraction was subsequently depleted of ORF1p containing complexes using an α-ORF1 conjugated magnetic medium, resulting in a supernatant fraction depleted of ORF1p-containing complexes. The α-ORF1 bound material was then released with LDS, yielding an elution fraction enriched for ORF1p-containing complexes. Representative SDS-PAGE / Coomassie blue stained results for each fraction are shown. (**E**) SILAC duplicates, two supernatants and two elution, were cross-mixed to enable an assessment of the relative protein content of each fraction by quantitative MS. (**F**) The results from split-tandem affinity capture graphed as the fraction of each protein observed in the elution sample. In order to easily visualize the relative degree of co-partitioning of constituent proteins with ORF1p, these data were normalized, setting the fraction of ORF1p in the elution to 1. Proteins which were previously ranked significant by I-DIRT analysis are labeled and displayed in blue or magenta (as indicated); gray, unlabeled nodes were not found to be significant by I-DIRT. MOV10 is marked with a dagger because in one replicate of this experiment it was detected by a single unique peptide, whereas we have enforced a minimum of two peptides (see **Methods**) for all other proteins, throughout all other proteomic analyses presented here.

Figure 2. (**A**) *RNA sequencing affinity captured L1s*: L1 complexes were obtained by split-tandem affinity capture, as in Figure 1D (simplified schematic shown); RNA extracted from these three fractions was subjected to next-generation sequencing. The results are summarized with respect to coverage of the synthetic L1 sequence (see schematic with nucleotide coordinates) as well as the relative quantities of mapped, annotated reads (pie charts; the mean of duplicate experiments is

displayed). (**B**) *Summary of sequencing reads*: displays the total number of sequencing reads that mapped to our reference library, the subset of mapped reads carrying a genome annotation, and the number of reads that corresponding to L1, both raw and normalized (see **Methods** and **Supplementary File 4**). The mean of duplicate experiments is displayed; +/- indicates the data range. (**C**) *LINE-1 element amplification protocol (LEAP) of affinity captured L1s*: L1 complexes were obtained from full length synthetic L1 (pLD401) and an otherwise identical *ΔORF1* construct (pLD561) following the same experimental design as in (A), except that elution from α-ORF1p affinity medium was done natively, by competitive elution. In this assay, L1 cDNAs are produced, in *cis*, by ORF2p catalyzed reverse transcription of L1 RNAs; the resulting cDNAs by were measure by quantitative PCR and presented as relative quantities normalized to pLD401 input (**Supplementary File 4**). The mean of duplicate experiments is displayed; error bars indicate the data range.

Figure 3. (**A**) *Immunofluorescent imaging reveals ORF1p expression is required for nuclear ORF2p staining*: Puromycin-selected HeLa-M2 cells containing pLD401 (Tet promoter, [*ORFeus*-Hs] full L1 coding sequence, ORF2p-3xFLAG, top two rows) or pLD561 (Tet promoter, *ΔORF1*, ORF2p-3xFLAG, bottom row) were plated on fibronectin-coated coverslips and induced for 24 hr with doxycycline prior to fixation and staining. With pLD401, the previously-observed pattern of cytoplasmic-only ORFs (top row) and a new pattern of pairs of cells displaying ORF2p in the nucleus (middle row) were apparent. When ORF1p is omitted from the construct (pLD561, bottom row), nuclear ORF2p was not apparent. **Scale bars:** 10 μm. (**B**) *Statistical analysis of the distances between pairs of ORF2p+ nuclei as compared to random*: Violin plots of the distributions of shortest distances between 1,000 pairs of randomly selected nuclei ('no') and the observed pairs of ORF2p+ nuclei ('yes') in cells transfected with pLD401; n=262 cells, 47 nuclear ORF2+. ***: $p=3.955 \times 10^{-11}$ (Welch's t-test).

Figure 4. *Catalytic inactivation of ORF2p alters the L1 interactome*: L1s were affinity captured from cells expressing enzymatically active ORF2p-3xFLAG sequences (pLD401, WT), a catalytically inactivated endonuclease point mutant (pLD567; H230A, EN^-), and a catalytically inactivated reverse transcriptase point mutant (pLD624; D702Y, RT^-). These were analyzed next-generation RNA sequencing and quantitative MS. (**A**) *Proteomic workflow*: WT L1s were captured from heavy-labeled cells, EN^- and RT^- L1s were captured from light-labeled cells. WT and either EN^- or RT^- fractions were mixed after affinity capture, in triplicate, and the relative abundance of each co-captured protein in the mixture was determined by quantitative MS. (**B**) *L1 RNA yield and coverage between different preparations*: As in Figure 2A, RNA extracted from 3xFLAG eluates originating from pLD401, pLD567, and pLD624 were subjected to next-generation sequencing. The results are summarized with respect to coverage of the synthetic L1 sequence (see schematic with nucleotide coordinates) as well as the relative quantities of mapped, annotated reads. The mean of duplicate experiments is displayed. (**C**) I-DIRT significant proteins displayed were detected in at least two replicates. All values were normalized to ORF2p. Data are represented as mean ± SD. Triangles (△) mark proteins whose levels of co-capture did not exhibit statistically significant differences in the mutant compared to the WT. A single or double asterisk denotes a statistically significant difference between the relative abundances of the indicated protein in EN^- and RT^- mutants: p-values of between 0.05 - 0.01 (*) and below 0.01 (**), respectively. Gray horizontal bars on the plot mark the 2x (upper) and 0.5x (lower) effect levels. (**D**) The double histogram plot displays the distributions of all proteins identified in at least two replicates, in common between both EN/WT (TOP) and RT/WT (LOWER) affinity capture experiments. The x-axis indicates the relative recovery of each copurifying protein and the y-axis indicates the number of proteins at that value (binned in 2 unit increments). The data are normalized to ORF2p. The relative positions of ORF2p and ORF1p are marked by colored bars. Differently colored lines illustrate the relative change in positions of the proteins within the two distributions (as indicated). Colored lines denote I-DIRT significance, with magenta lines indicating a statistically significant shift in position (p

≤0.05) within the two distributions and green lines indicating that statistical significance was not reached (entities labeled in **Fig. 4-S1**). A cluster of magenta lines can be seen to track with ORF1p (red line, upper and lower histogram), and another cluster can be seen to behave oppositely, creating a crisscross pattern in the center of the diagram. A similar crisscross pattern is exhibited by many gray lines.

Figure 4-S1. Double histogram plot with entities labeled. In the top plot (**A**), entities crossing from left to right, increasing between EN^- and RT^- mutants (and TOP1, in green), are labeled. In the bottom plot (**B**), entities crossing from right to left, decreasing between EN^- and RT^- mutants (and TROVE2 and YMEL1, in green).

Figure 5. *Monitoring coordinated dissociation and exchange exhibited by L1 interactors in vitro*: L1s were affinity captured from heavy-labeled cells expressing ORF2p-3xFLAG in the context of the naturally occurring L1RP sequence (pMT302); the stabilities of the protein constituents of the captured heavy-labeled L1 population were monitored *in vitro* by competitive exchange with light-labeled cell extracts containing untagged L1s (pMT298) (Taylor *et al*, 2013). (**A**) 3xFLAG-tagged L1s were captured from heavy-labeled cells and then, while immobilized on the affinity medium, were treated with an otherwise identically prepared, light-labeled, untagged-L1-expressing cell extract. Untreated complexes were compared to independently prepared complexes incubated for 30 sec, 5 min, and 30 min, (respectively) to determine the relative levels of exchange of *in vivo* assembled heavy-labeled interactors with *in vitro* exchanged light-labeled interactors using quantitative MS. (**B**) The results were plotted to compare the percentage of heavy-labeled protein versus time. I-DIRT significant proteins from **Table 1** are highlighted if present. Three clusters were observed (as indicated). (**C**) The cosine distance between the observed I-DIRT significant proteins was plotted along with time.

Figure 6. *Interactomic data integration*:(**A**) All MS-based affinity proteomic experiments presented were combined and analyzed for similarities across all I-DIRT significant proteins, producing five

groupings. Distance are presented on a one-unit arbitrary scale (**see Methods**: Mass Spectrometry Data Analysis). (**B**) The traces of each protein in each cluster, across all experiments, are displayed. The y-axis indicates the raw relative-enrichment value and the x-axis indicates the categories of each experiment-type. Each category is as wide as the number of replicates or time-point samples collected.

Figure 7. *Refined interatomic model*: Our results support the existence of distinct cytoplasmic and nuclear L1 interactomes. Affinity capture of L1 via 3xFLAG-tagged ORF2p from whole cell extracts results in a composite purification consisting of several macromolecular (sub)complexes. Among these, we propose a canonical cytoplasmic L1 RNP (**depicted**) and one or more nuclear macromolecules. UPF1 exhibited equivocal behavior within our fractionations and was also co-captured with chromatin associated ORF2p, suggesting it participates in both cytoplasmic and nuclear L1 interactomes. Within the nuclear L1 interactome, our data support the existence of a physically linked entity consisting of (at least) PCNA, PURA/B, TOP1, and PARP1 (**depicted**).

Supplementary File 1. *Supporting and supplemental data for the figures and experiments*: RNase – Figure 1C; Tandem – Figure 1F; Mutants – Figure 4; Exchange – Figure 5; Exchange distance matrix – Figure 5C; Integration – Figure 6; Integration distance matrix – Figure 6B; Raw – unnormalized values extracted from the MaxQuant proteinGroups.txt file for the experiments presented in this study.

Supplementary File 2. *Supporting and supplemental data for overexpression and siRNA knockdown experiments.*

Supplementary File 3. *Supporting and supplemental data for affinity capture of ORF2p-3xFLAG from fractionated chromatin.*

Supplementary File 4. *Supporting and supplemental data for figures and experiments*: RNA sequencing and LEAP assays – Figure 2.

Supplementary File 5. *Supporting and supplemental data for Nuclear Location and ORF2p status*: Figure 3.

Supplementary File 6. *Supporting and supplemental data for GO analysis.*

Supplementary File 7. *ORFeus-Hs sequence from pLD401 included in our reference FASTA file*: Figure 2 and Supplementary File 4.

Figure 3-source data 1. *Source data used in the analysis of ORF2p+ inter-nuclear distance analysis:* Figure 3 and Supplementary File 5.

10. Appendix 1

All normalized affinity values, derived from H/(H+L) and L/(H+L) isotopic ratios, can be found in **Supplementary File 1** on the appropriate sheet; pre-normalization values are located on the sheets named "Integration" and "Raw."

10.1. Modified SILAC Strategy

[Appendix 1-figure 1]

SILAC suspension expression of L1 constructs

Western blotting of cells grown in adherent culture with puromycin selection (A) or suspension culture with transient transfection (S). Cells were grown in heavy isotope-supplemented media (^{13}C ^{15}N lysine and arginine) (H), light isotope-supplemented media (L) or conventional commercial media (C) supplemented with tetracycline-free serum and L-glutamine. Note that serum used for heavy and light growth is dialyzed to remove amino acids; that with conventional commercial media is not. Construct LD401: synthetic ORFeus-HS, full L1 coding sequence (both ORFs and 3'UTR) with ORF2-3xFlag. Construct LD561: identical except for the absence of ORF1.

10.2. RNase sensitivity affinity capture

Data normalization

The RNase sensitivity data were rescaled and normalized such that proteins that did not change upon treatment with RNases were centered at the origin and those that were completely sensitive would give a value of 1.0. In a perfect experiment, unchanging proteins would yield a ratio of 0.5 when comparing the fraction of each protein present in the BSA-treated sample to the sum of both the BSA- and RNase-treated samples; i.e. 1 / (1+1). However, our data show some variability (below, left and also **Supplementary File 1**), with one replicate centering on ~0.4 (red) and another ~0.6 (blue). Therefore, we normalized the data such that the peaks at ~0.4 and ~0.6 were both re-centered at 0.5. From this set, 0.5 was subtracted from the data (centering insensitive proteins at the origin, and completely sensitive proteins at 0.5), followed by multiplication by 2 to expand the data to cover the range from 0 (insensitive) to 1 (completely sensitive); depicted below, right. These latter two transformations are encompassed by the functions: $g(x) = x + b$ [where $b = -0.5$] and $f(g(x)) = a(x+b)$ [where $a = 2$].

[Appendix 1-figure 2]

RNase normality test

The distances from the (0,0) point to protein coordinates were calculated. Proteins with distance less than 2 median distances were selected. The Shapiro-Wilk normality test (the null-hypothesis of this test is that the population is normally distributed) was applied for the distances (p-value = 0.29). The distribution of the distances was plotted as a histogram displaying the frequency (y-axis) versus RNase sensitivity (x-axis) of a simulation of normally distributed data (shown in black) and the actual data (**Supplementary File 1**) shown in blue. A Q-Q plot was also drawn.

[Appendix 1-figure 3]

10.3. Split-tandem affinity capture

Data normalization

The data were treated as follows: a and b coefficients were calculated as solutions of equation 1; the normalized values were calculated using the equation 2.

(1) $$\begin{pmatrix} a \\ b \end{pmatrix} \begin{pmatrix} \text{median} & 1 \\ \text{ORF1} & 1 \end{pmatrix} = \begin{pmatrix} \text{median} \\ 1 \end{pmatrix}$$

(2) $$x_{normalized} = (a^{*}x_{initial}) + b$$

[Appendix 1-figure 4]

Calculate the distances between node pairs

Distance between two points A and B with coordinates (A_x, A_y) and (B_x, B_y) was calculated as:

$$\sqrt{(\boldsymbol{A_x} - \boldsymbol{B_x})^2 + (\boldsymbol{A_y} - \boldsymbol{B_y})^2}$$

For each three points, the mean paired distance was calculated. The distributions of mean values are presented in the histograms below.

[Appendix 1-figure 5]

[Appendix 1-figure 6]

Associated likelihoods of selected clusters

Here, likelihood is defined as the frequency with which the same mean distance or less is observed within the distribution of clusters with the same number of nodes (above).

1046 **PURA/PURB/PCNA**: Likelihood = 3.2×10^{-7}

1047 **PABPC1/PABPC4**: Likelihood = 0.0008388427

1048 **HSPA8/HSPA1A**: Likelihood = 0.0001991309

1049 **NAP1L1/IPO7**: Likelihood = 0.0075885198

1050 10.4. Efficacy elution from α-ORF1 4H1 affinity medium using ORF1p peptides

1051 **[Appendix 1-figure 7]**

1052 ORF1p-FLAG was purified from 25 mg of cryo-milled HEK-293T_{LD} expressing pLD288 using α-
1053 ORF1 affinity medium, essentially as previously described (Taylor *et al*, 2013), and then eluted
1054 either eluted directly with 15 µl of 1x LDS, 70°C for 5 min (Ctrl LDS), with 2 mM monomeric ORF1
1055 peptide (Mono pep), or 2 mM dimeric ORF1 peptide (Di pep) (in both cases for 15 min at room
1056 temperature). After elution with peptide, the affinity medium was further eluted with 1x LDS at
1057 70°C for 5 min (Mono and Di LDS, respectively).

1058 10.5. Retrotransposition mutants affinity capture

1059 The distributions of normalized affinities for the two sets of experiments are shown below.

1060 **[Appendix 1-figure 8]**

1061 10.6. Protein *in vitro* exchange

1062 The distributions of H/(H+L) values present at each time point are shown.

1063 **[Appendix 1-figure 9]**

1064 **10.7. Affinity capture of ORF2p-3xFLAG L1 from fractionated chromatin and MS**

1065 **analyses**

1066 **Cell Culture**: Briefly, suspension grown HEK-293TLD cells were seeded at 1 x 106 cells/ml in 100

1067 ml of medium and transfected with pLD401 (ORF2p-3xFLAG L1 construct) or pLD259 (untagged

1068 L1 control construct) plasmid DNA. The transfection mixture consisted of Hybridoma serum free

1069 media (1/20 of final volume), PEI (3 µg/ml final volume) and plasmid DNA (1 µg/ml final volume).

1070 The mixture was incubated for 15 min at room temperature before adding to cell suspension. 24 hr

1071 post transfection, cells were split 1:3 into 1 µg/ml puromycin media. Expression was induced 48 hr

1072 post transfection by the addition of doxycycline (1 µg/ml) and maintained for 48 hr before

1073 collection for chromatin fractionation. A total of 900 ml final cell suspension per construct (≈ 3x106

1074 cells/ml) were prepared as follows.

1075 **Chromatin fractionation**: Cell suspensions were centrifuged at 200 RCF for 10 min and washed

1076 with 20 ml PBS. Cell pellets were resuspended in 5 ml of Buffer A (100 mM HEPES, 1.5 mM MgCl2,

1077 0.34 M sucrose, 10% (v/v) glycerol; with 1 mM DTT and protease inhibitors freshly added). Triton

1078 X-100 was added to 0.1% (v/v) final concentration and cells were allowed to swell on ice for 10 min.

1079 Nuclei were pelleted for 5 min at 1300 RCF, 4°C and the supernatant (cytoplasmic fraction) was

1080 discarded. Nuclei were resuspended in 2.5 ml Buffer B (3 mM EDTA, 0.2 mM EGTA; with 1 mM

1081 DTT and protease inhibitors freshly added) and incubated on ice for 30min before centrifuging at

1082 1700 RCF for 5 min. The soluble nuclear fraction was discarded and the insoluble material was

1083 washed twice with Buffer B. The remaining chromatin fraction was resuspended in 5 ml MNase

1084 buffer (a Tris buffered 10 mM KCl, 1 mM CaCl2 solution) supplemented with 5 U/ml micrococcal

1085 nuclease and incubated at 37°C for 5 min with agitation. The reaction was quenched by adding
1086 EGTA to 1 mM final concentration and incubating for 2 min. The solution was centrifuged for 5
1087 min at max speed and supernatant (chromatin fraction) transferred to a fresh tube.

1088 **Immunoprecipitation**: The chromatin fractions were normalized by Bradford Assay and equal
1089 amounts of proteins were used for the IP. The chromatin fractions were diluted in concentrated
1090 buffer to a final concentration of 500 mM NaCl, 20mM HEPES, pH 7.4, and 1% (v/v) Triton X-100
1091 (same formula used as washing buffer, below). 50 µl of magnetic beads (Life Technologies 14311D)
1092 conjugated to FLAG-M2 antibody (Sigma F1804) were added to the fractions incubated for 1 hr at
1093 4°C under end-over-end rotation. The affinity media were washed 10 times with washing and twice
1094 with 500 mM NaCl, 20mM HEPES, pH 7.4, and 0.1% (v/v) Triton X-100. Proteins were eluted for
1095 30min at room temperature under continuous shaking in 50 µl of 1 mg/ml 3xFLAG peptide (Sigma
1096 F4799) diluted in washing buffer with 0.1% Triton X-100. The eluates were collected and combined
1097 with NuPAGE® 4x LDS Sample Buffer (Novex) to a final concentration of 1x.

1098 **Preparation for Mass Spectrometry**: The samples were reduced with 2µl of 0.2M dithiothreitol
1099 (Sigma) for one hour at 57 °C at pH 8.0. Next the samples were alkylated with 2µl of 0.5M
1100 iodoacetamide (Sigma) for 45 minutes at room temperature in the dark. The samples were loaded
1101 on a NuPAGE® 4-12% Bis-Tris Gel 1.0 mm (Life Technologies) and run for 6 minutes at 200V. The
1102 gel was stained with GelCode Blue Stain Reagent (Thermo). The gel plugs were excised and
1103 destained for 15 minutes in a 1:1 (v/v) solution of methanol and 100mM ammonium bicarbonate.
1104 The buffer was exchanged and the samples were destained for another 15 minutes. This was
1105 repeated for another 3 cycles. The gel plugs were dehydrated by washing with acetonitrile, and then
1106 further dried by placing in a SpeedVac for 20 minutes. The gel plugs were treated with 250ng of

1107 sequencing grade modified trypsin (Promega) by adding directly on top of the dried gel plugs, and
1108 then enough 100mM ammonium bicarbonate was added in order to cover the gel pieces. The gel
1109 plugs were allowed to shake at room temperature and digestion proceeded overnight. The digestion
1110 was halted by adding a slurry of R2 50 µm Poros beads (Applied Biosystems) in 5% formic acid and
1111 0.2% trifluoroacetic acid (TFA) to each sample at a volume equal to that of the ammonium
1112 bicarbonate added for digestion. The samples were allowed to shake at 4°C for three hours. The
1113 beads were loaded onto C18 ziptips (Millipore), equilibrated with 0.1% TFA, using a
1114 microcentrifuge for 30 s at 6,000 rpm. The beads were washed with 0.5% acetic acid. Peptides were
1115 eluted with 40% acetonitrile in 0.5% acetic acid followed by 80% acetonitrile in 0.5% acetic acid. The
1116 organic solvent was removed using a SpeedVac concentrator and the sample reconstituted in 0.5%
1117 acetic acid.

1118 **Mass Spectrometry Analysis – Thermo Orbitrap Elite instrument**: An aliquot of each sample was
1119 loaded onto an Acclaim PepMap100 C18 75-µm x 15-cm column with 3µm bead size, coupled to an
1120 EASY-Spray 75-µm x 50-cm PepMap C18 analytical HPLC column with a 2µm bead size, using the
1121 auto sampler of an EASY-nLC 1000 HPLC (ThermoFisher) and solvent A (2% acetonitrile, 0.5%
1122 acetic acid). The peptides were eluted into a ThermoFisher Scientific Orbitrap Elite Hybrid Ion Trap
1123 Mass Spectrometer increasing from 2% to 30% solvent B (90% acetonitrile, 0.5% acetic acid) over 60
1124 minutes, followed by an increase from 30% to 40% solvent B over 30 minutes. Solvent B was then
1125 put to 100% and held at 100% for 20 minutes. High resolution full MS spectra were obtained with a
1126 resolution of 60,000 at 400m/z, an AGC target of 1e6, with a maximum ion time of 200ms, and a
1127 scan range from 300 to 1500m/z. Following each full MS scan, fifteen data-dependent MS/MS
1128 spectra were acquired. The MS/MS spectra were collected in the ion trap, with an AGC target of 1e4,

maximum ion time of 150ms, one microscan, 2m/z isolation window, fixed first mass of 150 m/z, and Normalized Collision Energy (NCE) of 35.

Mass Spectrometry Analysis – Thermo Fusion instrument: An aliquot of each sample was loaded onto an Acclaim PepMap100 C18 75-µm x 15-cm column with 3µm bead size, coupled to an EASY-Spray 75-µm x 50-cm PepMap C18 analytical HPLC column with a 2µm bead size, using the auto sampler of an EASY-nLC 1000 HPLC (ThermoFisher) and solvent A (2% acetonitrile, 0.5% acetic acid). The peptides were eluted into a ThermoFisher Scientific Orbitrap Fusion Mass Spectrometer increasing from 2% to 30% solvent B (90% acetonitrile, 0.5% acetic acid) over 60 minutes, followed by an increase from 30% to 40% solvent B over 30 minutes. Solvent B was then put to 100% and held at 100% for 20 minutes. High resolution full MS spectra were obtained with a resolution of 120,000, an AGC target of 400,000, with a maximum ion time of 50ms, and a scan range from 400 to 1500m/z. The MS/MS spectra were collected in the ion trap, with an AGC target of 100, maximum ion time of 250ms, one microscan, 2m/z isolation window, fixed first mass of 150 m/z, and Normalized Collision Energy (NCE) of 27.

Data Processing: All acquired MS2 spectra were searched against a UniProt human database using Sequest within Proteome Discoverer (ThermoScientific). The search parameters were as follows: precursor mass tolerance ±10 ppm, fragment mass tolerance ± 0.4 Da, digestion parameters allowing trypsin 2 missed cleavages, fixed modification of carbamidomethyl on cysteine, variable modification of oxidation on methionine, and variable modification of deamidation on glutamine and asparagine. The results were filtered to only include proteins identified by at least two peptides.

11. Appendix 1 Figure Legends

Appendix 1-Figure 1 .SILAC suspension expression of L1 constructs: western blotting

Appendix 1-Figure 2. RNase sensitivity affinity capture: data normalization

Appendix 1-Figure 3. RNase sensitivity affinity capture: normality test

Appendix 1-Figure 4. Split-tandem affinity capture: data normalization

Appendix 1-Figure 5. Distances between two-node groups

Appendix 1-Figure 6. Distances between three-node groups

Appendix 1-Figure 7. Efficacy of elution using ORF1p peptides: Coomassie blue stained gel

Appendix 1-Figure 8. Retrotransposition mutants affinity capture: distributions of normalized affinities

Appendix 1-Figure 9. Protein in vitro exchange: the distributions of H/(H+L) values

12. References

Alisch RS, Garcia-Perez JL, Muotri AR, Gage FH & Moran JV (2006) Unconventional translation of mammalian LINE-1 retrotransposons. *Genes & development* **20:** 210–224

An W, Dai L, Niewiadomska AM, Yetil A, O'Donnell KA, Han JS & Boeke JD (2011) Characterization of a synthetic human LINE-1 retrotransposon ORFeus-Hs. *Mobile DNA* **2:** 2

Altukhov I. 2017. LINE-1. Bitbucket. **https://bitbucket.org/altukhov/line-1/.** 1e90c62

Arjan-Odedra S, Swanson CM, Sherer NM, Wolinsky SM & Malim MH (2012) Endogenous MOV10 inhibits the retrotransposition of endogenous retroelements but not the replication of exogenous retroviruses. *Retrovirology* **9**

Arlander SJH, Eapen AK, Vroman BT, McDonald RJ, Toft DO & Karnitz LM (2003) Hsp90 inhibition depletes Chk1 and sensitizes tumor cells to replication stress. *The Journal of biological chemistry* **278:** 52572–52577

Azzalin CM & Lingner J (2006) The human RNA surveillance factor UPF1 is required for S phase progression and genome stability. *Current Biology* **16:** 433–439

Baltz AG, Munschauer M, Schwanhäusser B, Vasile A, Murakawa Y, Schueler M, Youngs N, Penfold-Brown D, Drew K, Milek M, Wyler E, Bonneau R, Selbach M, Dieterich C & Landthaler M (2012) The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. *Molecular cell* **46:** 674–690

Beauregard A, Curcio MJ & Belfort M (2008) The take and give between retrotransposable elements and their hosts. *Annual review of genetics* **42:** 587–617

Belancio VP, Hedges DJ & Deininger P (2006) LINE-1 RNA splicing and influences on mammalian gene expression. *Nucleic acids research* **34:** 1512–1521

Bley N, Lederer M, Pfalz B, Reinke C, Fuchs T, Glaß M, Möller B & Hüttelmaier S (2015) Stress granules are dispensable for mRNA stabilization during cellular stress. *Nucleic acids research* **43:** e26

Bolger AM, Lohse M & Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30:** 2114–2120

Bryant HE, Petermann E, Schultz N, Jemth A-S, Loseva O, Issaeva N, Johansson F, Fernandez S, McGlynn P & Helleday T (2009) PARP is activated at stalled forks to mediate Mre11-dependent replication restart and recombination. *The EMBO journal* **28:** 2601–2615

Budiman ME, Bubenik JL, Miniard AC, Middleton LM, Gerber CA, Cash A & Driscoll DM (2009) Eukaryotic initiation factor 4a3 is a selenium-regulated RNA-binding protein that selectively inhibits selenocysteine incorporation. *Molecular cell* **35:** 479–489

Byrum S, Mackintosh SG, Edmondson RD, Cheung WL, Taverna SD & Tackett AJ (2011) Analysis of Histone Exchange during Chromatin Purification. *Journal of integrated OMICS* **1:** 61–65

Candiano G, Bruschi M, Musante L, Santucci L, Ghiggeri GM, Carnemolla B, Orecchia P, Zardi L & Righetti PG (2004) Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis. *Electrophoresis* **25:** 1327–1333

Caspary F, Shevchenko A, Wilm M & Séraphin B (1999) Partial purification of the yeast U2 snRNP reveals a novel yeast pre-mRNA splicing factor required for pre-spliceosome assembly. *The EMBO journal* **18:** 3463–3474

Castello AA, Fischer BB, Eichelbaum KK, Horos RR, Beckmann BMB, Strein CC, Davey NEN, Humphreys DTD, Preiss TT, Steinmetz LML, Krijgsveld JJ & Hentze MWM (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149:** 1393–1406

Cost GJ, Feng QH, Jacquier A & Boeke JD (2002) Human L1 element target-primed reverse transcription in vitro. *The EMBO journal* **21:** 5899–5910

Dai L, LaCava J, Taylor MS & Boeke JD (2014) Expression and detection of LINE-1 ORF-encoded proteins. *Mobile Genetic Elements* **4:** e29319

Dai L, Taylor MS, O'Donnell KA & Boeke JD (2012) Poly(A) binding protein C1 is essential for efficient L1 retrotransposition and affects L1 RNP formation. *Molecular and cellular biology* **32:** 4323–4336

Denli AM, Narvaiza I, Kerman BE, Pena M, Benner C, Marchetto MCN, Diedrich JK, Aslanian A, Ma J, Moresco JJ, Moore L, Hunter T, Saghatelian A & Gage FH (2015) Primate-Specific ORF0 Contributes to Retrotransposon-Mediated Diversity. *Cell* **163:** 583–593

Dewannieux M, Esnault, C & Heidmann (2003) T LINE-mediated retrotransposition of marked Alu sequences. *Nature Genetics* **35:** 41–48.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M & Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29:** 15–21

Doucet AJ, Hulme AE, Sahinovic E, Kulpa DA, Moldovan JB, Kopera HC, Athanikar JN, Hasnaoui M, Bucheton A, Moran JV & Gilbert N (2010) Characterization of LINE-1 ribonucleoprotein particles. *PLoS genetics* **6:** 1–19

Doucet AJ, Wilusz JE, Miyoshi T, Liu Y & Moran JV (2015) A 3' Poly(A) Tract Is Required for LINE-1 Retrotransposition. *Molecular cell* **60:** 728–741

Engeland CE, Brown NP, Börner K, Schümann M, Krause E, Kaderali L, Müller GA & Kräusslich H-G (2014) Proteome analysis of the HIV-1 Gag interactome. *Virology* **460-461:** 194–206

Fassati A, Görlich D, Harrison I, Zaytseva L & Mingot J-M (2003) Nuclear import of HIV-1 intracellular reverse transcription complexes is mediated by importin 7. *The EMBO Journal* **22:** 3675–3685.

Feng Q, Moran JV, Kazazian HH & Boeke JD (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87:** 905–916

Furano AV, Duvernell DD & Boissinot S (2004) L1 (LINE-1) retrotransposon diversity differs dramatically between mammals and fish. *Trends in genetics* **20:** 9–14

Geiger T, Wiśniewski JR, Cox J, Zanivan S, Kruger M, Ishihama Y & Mann M (2011) Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics. *Nature protocols* **6:** 147–157

Gonzalez IL, Gorski JL, Campen TJ, Dorney DJ, Erickson JM, Sylvester JE & Schmickel RD (1985) Variation among human 28S ribosomal RNA genes. *Proceedings of the National Academy of Sciences of the United States of America* **82:** 7666–7670

Goodier JL, Cheung LE & Kazazian HH (2012) MOV10 RNA Helicase Is a Potent Inhibitor of Retrotransposition in Cells. *PLoS genetics* **8:** e1002941

Goodier JL, Cheung LE & Kazazian HH (2013) Mapping the LINE1 ORF1 protein interactome reveals associated inhibitors of human retrotransposition. **41:** 7401–7419

Goodier JL, Mandal PK, Zhang L & Kazazian HH (2010) Discrete subcellular partitioning of human retrotransposon RNAs despite a common mechanism of genome insertion. *Human Molecular Genetics* **19:** 1712–1725

Goodier JL, Zhang L, Vetter MR & Kazazian HH (2007) LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex. *Molecular and cellular biology* **27:** 6469–6483

Gregersen LH, Schueler M, Munschauer M, Mastrobuoni G, Chen W, Kempa S, Dieterich C & Landthaler M (2014) MOV10 Is a 5’ to 3' RNA Helicase Contributing to UPF1 mRNA Target Degradation by Translocation along 3' UTRs. *Molecular cell* **54:** 573-585

Ha K, Fiskus W, Rao R, Balusu R, Venkannagari S, Nalabothula NR & Bhalla KN (2011) Hsp90 inhibitor-mediated disruption of chaperone association of ATR with hsp90 sensitizes cancer cells to DNA damage. *Molecular Cancer Therapeutics* **10:** 1194–1206

Hakhverdyan Z, Domanski M, Hough LE, Oroskar AA, Oroskar AR, Keegan S, Dilworth DJ, Molloy KR, Sherman V, Aitchison JD, Fenyö D, Chait BT, Jensen TH, Rout MP & LaCava J (2015) Rapid, optimized interactomic screening. *Nature methods* **12:** 553–560

Halász L, Karányi Z, Boros-Oláh B, Kuik-Rózsa T, Sipos É, Nagy É, Mosolygó-L Á, Mázló A, Rajnavölgyi É, Halmos G & Székvölgyi L (2017) RNA-DNA hybrid (R-loop) immunoprecipitation mapping: an analytical workflow to evaluate inherent biases. *Genome research* **27:** 1063–1073

Hampf M & Gossen M (2007) Promoter crosstalk effects on gene expression. *Journal of molecular biology* **365:** 911–920

Havugimana PC, Hart GT, Nepusz T, Yang H, Turinsky AL, Li Z, Wang PI, Boutz DR, Fong V, Phanse S, Babu M, Craig SA, Hu P, Wan C, Vlasblom J, Dar V-N, Bezginov A, Clark GW, Wu GC, Wodak SJ, et al (2012) A census of human soluble protein complexes. *Cell* **150:** 1068–1081

Hohjoh H & Singer MF (1996) Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *The EMBO journal* **15:** 630–639

Love MI, Huber W & Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15:** 550

Jain S, Wheeler JR, Walters RW, Agrawal A, Barsic A & Parker R (2016) ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure. *Cell* **164:** 487–498

Jønson L, Vikesaa J, Krogh A, Nielsen LK, Hansen T vO, Borup R, Johnsen AH, Christiansen J & Nielsen FC (2007) Molecular composition of IMP1 ribonucleoprotein granules. *Molecular & cellular proteomics* **6:** 798–811

Joshi P, Greco TM, Guise AJ, Luo Y, Yu F, Nesvizhskii AI & Cristea IM (2013) The functional interactome landscape of the human histone deacetylase family. *Molecular systems biology* **9:** 672

Kaake RM, Wang X & Huang L (2010) Profiling of Protein Interaction Networks of Protein Complexes Using Affinity Purification and Quantitative Mass Spectrometry. *Molecular & cellular proteomics* **9:** 1650–1665

Kates SA & Albericio F (2000) Solid-Phase Synthesis. *CRC Press*

Katzenellenbogen RA, Egelkrout EM, Vliet-Gregg P, Gewin LC, Gafken PR & Galloway DA (2007) NFX1-123 and poly(A) binding proteins synergistically augment activation of telomerase in human papillomavirus type 16 E6-expressing cells. *Journal of virology* **81:** 3786–3796

Kazazian HH (2004) Mobile elements: drivers of genome evolution. *Science* **303:** 1626–1632

Kelm RJ, Cogan JG, Elder PK, Strauch AR & Getz MJ (1999) Molecular interactions between single-stranded DNA-binding proteins associated with an essential MCAT element in the mouse smooth muscle alpha-actin promoter. *The Journal of biological chemistry* **274:** 14238–14245

Kimberland ML, Divoky V, Prchal J, Schwahn U, Berger W & Kazazian HH (1999) Full-length human L1 insertions retain the capacity for high frequency retrotransposition in cultured cells. *Human Molecular Genetics* **8:** 1557–1560

Khazina E, Truffault V, Büttner R, Schmidt S, Coles M & Weichenrieder O (2011) Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition. *Nature Structural & Molecular Biology* **18:** 1006–1014.

Knapp AM, Ramsey JE, Wang S-X, Godburn KE, Strauch AR & Kelm RJ (2006) Nucleoprotein interactions governing cell type-dependent repression of the mouse smooth muscle alpha-actin promoter by single-stranded DNA-binding proteins Pur alpha and Pur beta. *The Journal of biological chemistry* **281:** 7907–7918

Kubota T, Nishimura K, Kanemaki MT & Donaldson AD (2013) The Elg1 replication factor C-like complex functions in PCNA unloading during DNA replication. *Molecular cell* **50:** 273–280

Kulpa DA & Moran JV (2005) Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. *Human Molecular Genetics* **14:** 3237–3248

Kulpa DA & Moran JV (2006) Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nature structural & molecular biology* **13:** 655–660

LaCava J, Jiang H & Rout MP (2016) Protein Complex Affinity Capture from Cryomilled Mammalian Cells. *Journal of Visualized Experiments***:** e54518

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, Fitzhugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, Levine R, McEwan P, McKernan K, et al (2001) Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921

Lee E, Iskow R, Yang L, Gokcumen O, Haseley P, Luquette LJI, Lohr JG, Harris CC, Ding L, Wilson RK, Wheeler DA, Gibbs RA, Kucherlapati R, Lee C, Kharchenko PV, Park PJ & Network CGAR (2012) Landscape of Somatic Retrotransposition in Human Cancers. *Science* **337:** 967–971

Li Q, Brass AL, Ng A, Hu Z, Xavier RJ, Liang TJ & Elledge SJ (2009) A genome-wide genetic screen for host factors required for hepatitis C virus propagation. *Proceedings of the National Academy of Sciences of the United States of America* **106:** 16410–16415

Luan DD, Korman MH, Jakubczak JL & Eickbush TH (1993) 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

Luo Y, Jacobs EY, Greco TM, Mohammed KD, Tong T, Keegan S, Binley JM, Cristea IM, Fenyö D, Rout MP, Chait BT & Muesing MA (2016) HIV–host interactome revealed directly from infected cells. *Nature Microbiology* **1:** 16068

Martin SL (1991) Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. *Molecular and cellular biology* **11:** 4804–4807

Martin SL & Bushman FD (2001) Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. *Molecular and cellular biology* **21:** 467–475

Mathias SL, Scott AF, Kazazian HH, Boeke JD & Gabriel A (1991) Reverse transcriptase encoded by a human transposable element. *Science* **254:** 1808–1810

Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D & Thomas PD (2017) PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic acids research* **45:** D183

Mi H, Muruganujan A, Casagrande JT & Thomas PD (2013) Large-scale gene function analysis with the PANTHER classification system. *Nature protocols* **8:** 1551–1566

Min W, Bruhn C, Grigaravicius P, Zhou Z-W, Li F, Krüger A, Siddeek B, Greulich K-O, Popp O, Meisezahl C, Calkhoven CF, Bürkle A, Xu X & Wang Z-Q (2013) Poly(ADP-ribose) binding to Chk1 at stalled replication forks is required for S-phase checkpoint activation. *Nature communications* **4:** 2993

Mittler G, Butter F & Mann M (2009) A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements. *Genome research* **19:** 284–293

Muotri AR, Chu VT, Marchetto MCN, Deng W, Moran JV & Gage FH (2005) Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* **435:** 903–910

Nellist M, Burgers PC, van den Ouweland AMW, Halley DJJ & Luider TM (2005) Phosphorylation and binding partner analysis of the TSC1-TSC2 complex. *Biochemical and biophysical research communications* **333:** 818–826

Niewiadomska AM, Tian C, Tan L, Wang T, Sarkis PTN & Yu XF (2007) Differential inhibition of long interspersed element 1 by APOBEC3 does not correlate with high-molecular-mass-complex formation or P-body association. *Journal of virology* **81:** 9577–9583

Oda Y, Huang K, Cross FR, Cowburn D & Chait BT (1999) Accurate quantitation of protein expression and site-specific phosphorylation. *Proceedings of the National Academy of Sciences of the United States of America* **96:** 6591–6596

Ohta S, Bukowski-Wills J-C, Sanchez-Pulido L, Alves F de L, Wood L, Chen ZA, Platani M, Fischer L, Hudson DF, Ponting CP, Fukagawa T, Earnshaw WC & Rappsilber J (2010) The protein composition of mitotic chromosomes determined using multiclassifier combinatorial proteomics. *Cell* **142:** 810–821

Ong S-E, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A & Mann M (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & cellular proteomics* **1:** 376–386

Ostertag EM & Kazazian HH (2001) Biology of mammalian L1 retrotransposons. *Annual review of genetics* **35:** 501–538

Ostertag EM, Prak ET, DeBerardinis RJ, Moran JV & Kazazian HH (2000) Determination of L1 retrotransposition kinetics in cultured cells. *Nucleic acids research* **28:** 1418–1423

Peattie DA, Harding MW, Fleming MA, DeCenzo MT, Lippke JA, Livingston DJ & Benasutti M (1992) Expression and characterization of human FKBP52, an immunophilin that associates with the 90-kDa heat shock protein and is a component of steroid receptor complexes. *Proceedings of the National Academy of Sciences of the United States of America* **89:** 10974–10978

Peddigari S, Li PW-L, Rabe JL & Martin SL (2013) hnRNPL and nucleolin bind LINE-1 RNA and function as host factors to modulate retrotransposition. *Nucleic acids research* **41:** 575–585

Qin Y, Ouyang H, Liu J & Xie Y (2013) Proteome identification of proteins interacting with histone methyltransferase SET8. *Acta Biochimica Et Biophysica Sinica* **45:** 303–308

Rappsilber J, Mann M & Ishihama Y (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature protocols* **2:** 1896–1906

Rodriguez P, Munroe D, Prawitt D, Chu LL, Bric E, Kim J, Reid LH, Davies C, Nakagama H, Loebbert R, Winterpacht A, Petruzzi MJ, Higgins MJ, Nowak N, Evans G, Shows T, Weissman BE, Zabel B, Housman DE & Pelletier J (1997) Functional characterization of human nucleosome assembly protein-2 (NAP1L4) suggests a role as a histone chaperone. *Genomics* **44:** 253–265

Schweingruber C, Rufener SC, Zünd D, Yamashita A & Mühlemann O (2013) Nonsense-mediated mRNA decay - mechanisms of substrate mRNA recognition and degradation in mammalian cells. *Biochimica et biophysica acta* **1829:** 612–623

Shi X, Seluanov A & Gorbunova V (2007) Cell divisions are required for L1 retrotransposition. *Molecular and cellular biology* **27:** 1264–1270

Simon HU, Mills GB, Kozlowski M, Hogg D, Branch D, Ishimi Y & Siminovitch KA (1994) Molecular characterization of hNRP, a cDNA encoding a human nucleosome-assembly-protein-I-related gene product involved in the induction of cell proliferation. *The Biochemical journal* **297 (Pt 2):** 389–397

Smart SK, Mackintosh SG, Edmondson RD, Taverna SD & Tackett AJ (2009) Mapping the local protein interactome of the NuA3 histone acetyltransferase. *Protein science* **18:** 1987–1997

Sommer SS & Cohen JE (1980) The size distributions of proteins, mRNA, and nuclear RNA. *Journal of Molecular Evolution* **15:** 37–57

Suzuki J, Yamaguchi K, Kajikawa M, Ichiyanagi K, Adachi N, Koyama H, Takeda S & Okada N (2009) Genetic evidence that the non-homologous end-joining repair pathway is involved in LINE retrotransposition. *PLoS genetics* **5:** e1000461

Tackett AJ, DeGrasse JA, Sekedat MD, Oeffinger M, Rout MP & Chait BT (2005) I-DIRT, a general method for distinguishing between specific and nonspecific protein interactions. *Journal of proteome research* **4:** 1752–1756

Takasaki Y, Kogure T, Takeuchi K, Kaneda K, Yano T, Hirokawa K, Hirose S, Shirai T & Hashimoto H (2001) Reactivity of anti-proliferating cell nuclear antigen (PCNA) murine monoclonal antibodies and human autoantibodies to the PCNA multiprotein complexes involved in cell proliferation. *Journal of immunology* **166:** 4780–4787

Takizawa Y, Binshtein E, Erwin AL, Pyburn TM, Mittendorf KF & Ohi MD (2017) While the revolution will not be crystallized, biochemistry reigns supreme. *Protein science* **26:** 69–81

Taylor MS, LaCava J, Dai L, Mita P, Burns KH, Rout MP & Boeke JD (2016) Characterization of L1-Ribonucleoprotein Particles. *Transposons and Retrotransposons* **1400:** 311–338

Taylor MS, LaCava J, Mita P, Molloy KR, Huang CRL, Li D, Adney EM, Jiang H, Burns KH, Chait BT, Rout MP, Boeke JD & Dai L (2013) Affinity Proteomics Reveals Human Host Factors Implicated in Discrete Stages of LINE-1 Retrotransposition. *Cell* **155:** 1034–1048

Trinkle-Mulcahy L, Boulon S, Lam YW, Urcia R, Boisvert F-M, Vandermoere F, Morrice NA, Swift S, Rothbauer U, Leonhardt H & Lamond A (2008) Identifying specific protein interaction

partners using quantitative mass spectrometry and bead proteomes. *The Journal of cell biology* **183:** 223–239

Tubio JMC, Li Y, Ju YS, Martincorena I, Cooke SL, Tojo M, Gundem G, Pipinikas CP, Zamora J, Raine K, Menzies A, Roman-Garcia P, Fullam A, Gerstung M, Shlien A, Tarpey PS, Papaemmanuil E, Knappskog S, Van Loo P, Ramakrishna M, et al (2014) Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science* **345:** 1251343

Wallace N, Wagstaff, BJ, Deininger PL & Roy-Engel AM (2008) LINE-1 ORF1 protein enhances Alu SINE retrotransposition. *Gene* **419:** 1–6

Wang X & Huang L (2008) Identifying dynamic interactors of protein complexes by quantitative mass spectrometry. *Molecular & cellular proteomics* **7:** 46–57

Weidensdorfer D, Stoehr N, Baude A, Lederer M, Koehn M, Schierhorn A, Buchmeier S, Wahle E & Huettelmaier S (2009) Control of c-myc mRNA stability by IGF2BP1-associated cytoplasmic RNPs. *RNA* **15:** 104–115

Weinlich S, Hüttelmaier S, Schierhorn A, Behrens S-E, Ostareck-Lederer A & Ostareck DH (2009) IGF2BP1 enhances HCV IRES-mediated translation initiation via the 3'UTR. *RNA* **15:** 1528–1542

Wei W, Gilbert N, Ooi SL, Lawler JF, Ostertag EM, Kazazian HH, Boeke JD & Moran JV (2001) Human L1 retrotransposition: cis preference versus trans complementation. *Molecular and cellular biology* **21:** 1429–1439

Wyers F, Sentenac A & Fromageot P (1973) Role of DNA-RNA hybrids in eukaryotes. Ribonuclease H in yeast. *European Journal of Biochemistry* **35:** 270–281

Xie Y, Mates L, Ivics Z, Izsvák Z, Martin SL & An W (2013) Cell division promotes efficient retrotransposition in a stable L1 reporter cell line. *Mobile DNA* **4**

Ying S, Chen Z, Medhurst AL, Neal JA, Bao Z, Mortusewicz O, McGouran J, Song X, Shen H, Hamdy FC, Kessler BM, Meek K & Helleday T (2016) DNA-PKcs and PARP1 Bind to Unresected Stalled DNA Replication Forks Where They Recruit XRCC1 to Mediate Repair. *Cancer research* **76:** 1078–1088

Figure 1

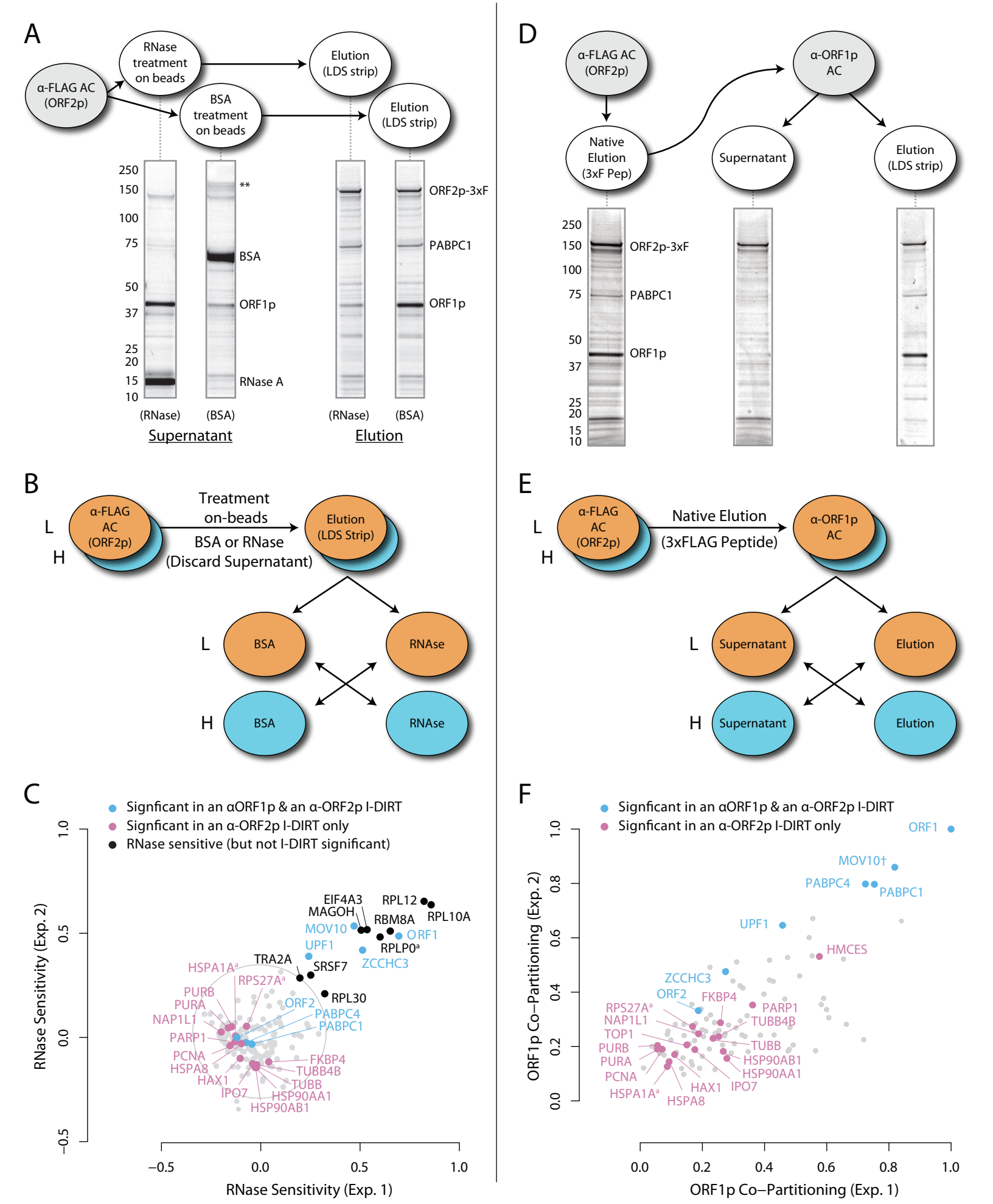
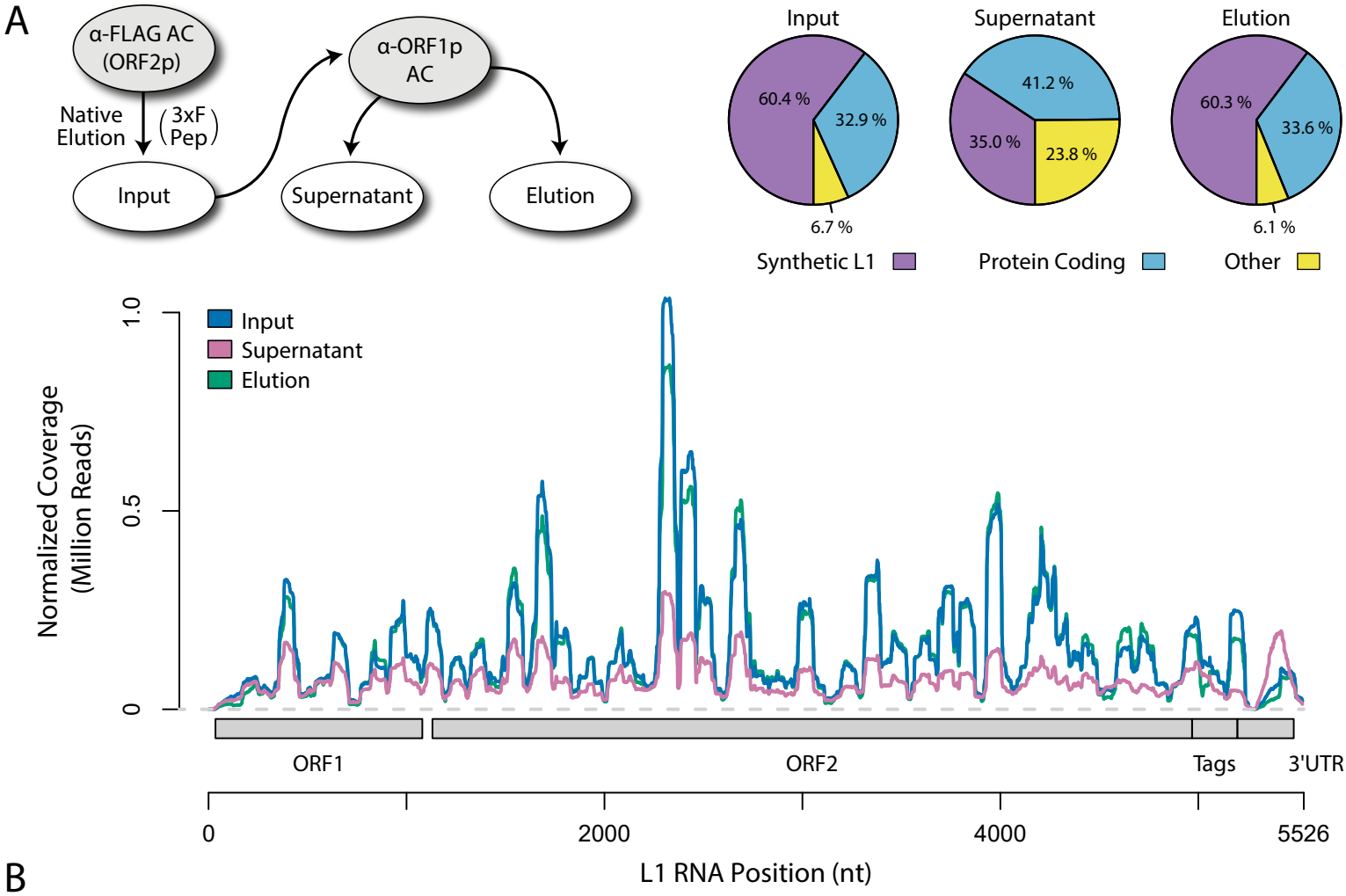

A
α-FLAG AC (ORF2p)
RNase treatment on beads
BSA treatment on beads
Elution (LDS strip)
Elution (LDS strip)
250
150
100
75
50
37
25
20
15
10
**
BSA
ORF1p
RNase A
ORF2p-3xF
PABPC1
ORF1p
(RNase)
(BSA)
(RNase)
(BSA)
Supernatant
Elution
B
L
H
α-FLAG AC (ORF2p)
Treatment on-beads
BSA or RNase
(Discard Supernatant)
Elution (LDS Strip)
L
H
BSA
RNAse
BSA
RNAse
C
Signficant in an αORF1p & an α-ORF2p I-DIRT
Signficant in an α-ORF2p I-DIRT only
RNase sensitive (but not I-DIRT significant)
RNase Sensitivity (Exp. 2)
RNase Sensitivity (Exp. 1)
−0.5
0.0
0.5
1.0
EIF4A3
MAGOH
RPL12
RPL10A
RBM8A
MOV10
ORF1
UPF1
RPLP0a
ZCCHC3
TRA2A
SRSF7
HSPA1Aa
RPS27Aa
PURB
PURA
NAP1L1
PARP1
PCNA
HSPA8
HAX1
IPO7
HSP90AB1
HSP90AA1
TUBB
TUBB4B
FKBP4
PABPC1
PABPC4
ORF2
RPL30
D
α-FLAG AC (ORF2p)
α-ORF1p AC
Native Elution (3xF Pep)
Supernatant
Elution (LDS strip)
250
150
100
75
50
37
25
20
15
10
ORF2p-3xF
PABPC1
ORF1p
E
L
H
α-FLAG AC (ORF2p)
Native Elution
(3xFLAG Peptide)
α-ORF1p AC
L
H
Supernatant
Elution
Supernatant
Elution
F
Signficant in an αORF1p & an α-ORF2p I-DIRT
Signficant in an α-ORF2p I-DIRT only
ORF1p Co−Partitioning (Exp. 2)
ORF1p Co−Partitioning (Exp. 1)
0.0
0.2
0.4
0.6
0.8
1.0
ORF1
MOV10†
PABPC4
PABPC1
UPF1
HMCES
ZCCHC3
ORF2
FKBP4
RPS27Aa
NAP1L1
TOP1
PURB
PURA
PCNA
HSPA1Aa
HSPA8
HAX1
IPO7
PARP1
TUBB4B
TUBB
HSP90AB1
HSP90AA1

Figure 2

A



B

	Mapped Reads x10^6		Mapped Annotated Reads x10^6	L1 Reads x10^6	Normalized L1 Reads x10^6
Input	34.8 ± 0.9	(92 ± 0.9%)	19.5 ± 1.1	11.8 ± 1.3	12.6 ± 1.5
Supernatant	30.1 ± 1.9	(86 ± 0.1%)	5.1 ± 0.1	1.7 ± 0.1	6.0 ± 0.2
Elution	33.3 ± 0.4	(94 ± 0.1%)	21.4 ± 0.2	12.9 ± 0.5	11.9 ± 1.1

C

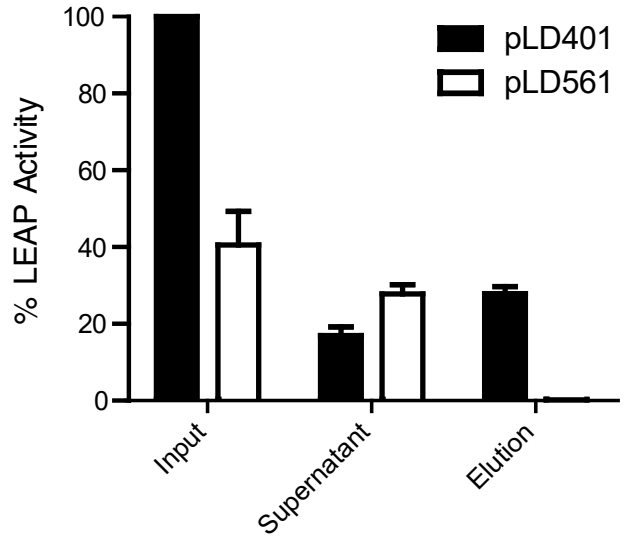
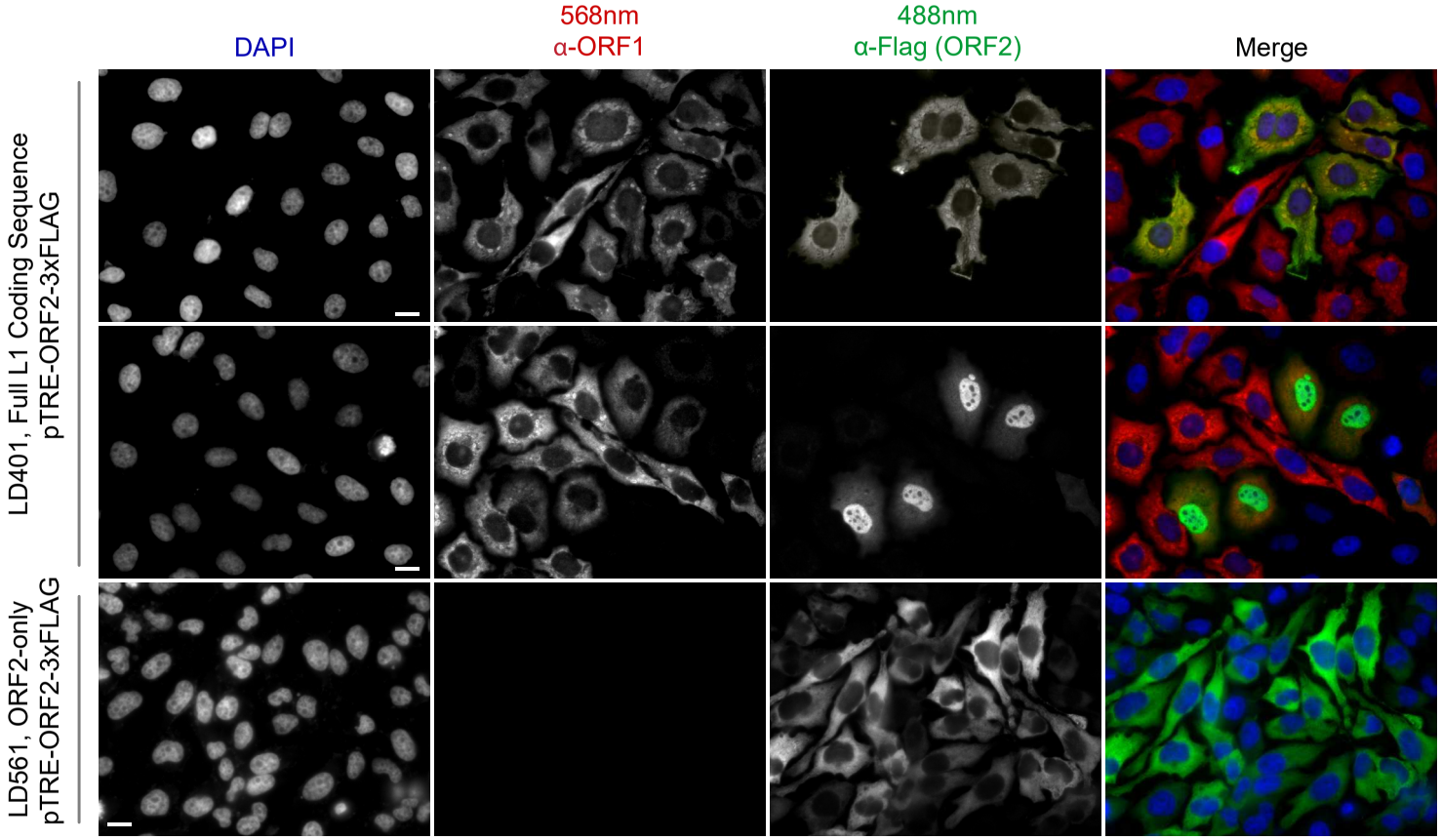


Figure 3

A


DAPI
568nm
α-ORF1
488nm
α-Flag (ORF2)
Merge
LD401, Full L1 Coding Sequence
pTRE-ORF2-3xFLAG
LD561, ORF2-only
pTRE-ORF2-3xFLAG

B

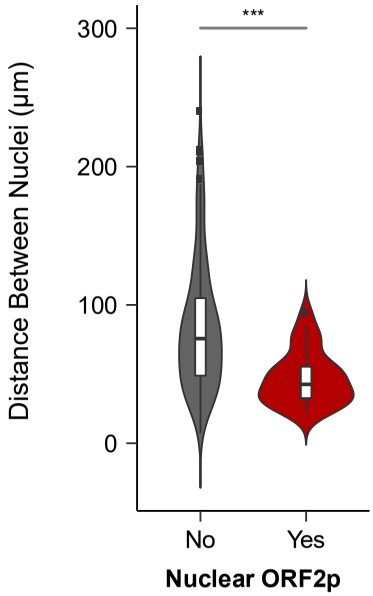

***
300
200
100
0
Distance Between Nuclei (µm)
No
Yes
Nuclear ORF2p

Figure 4

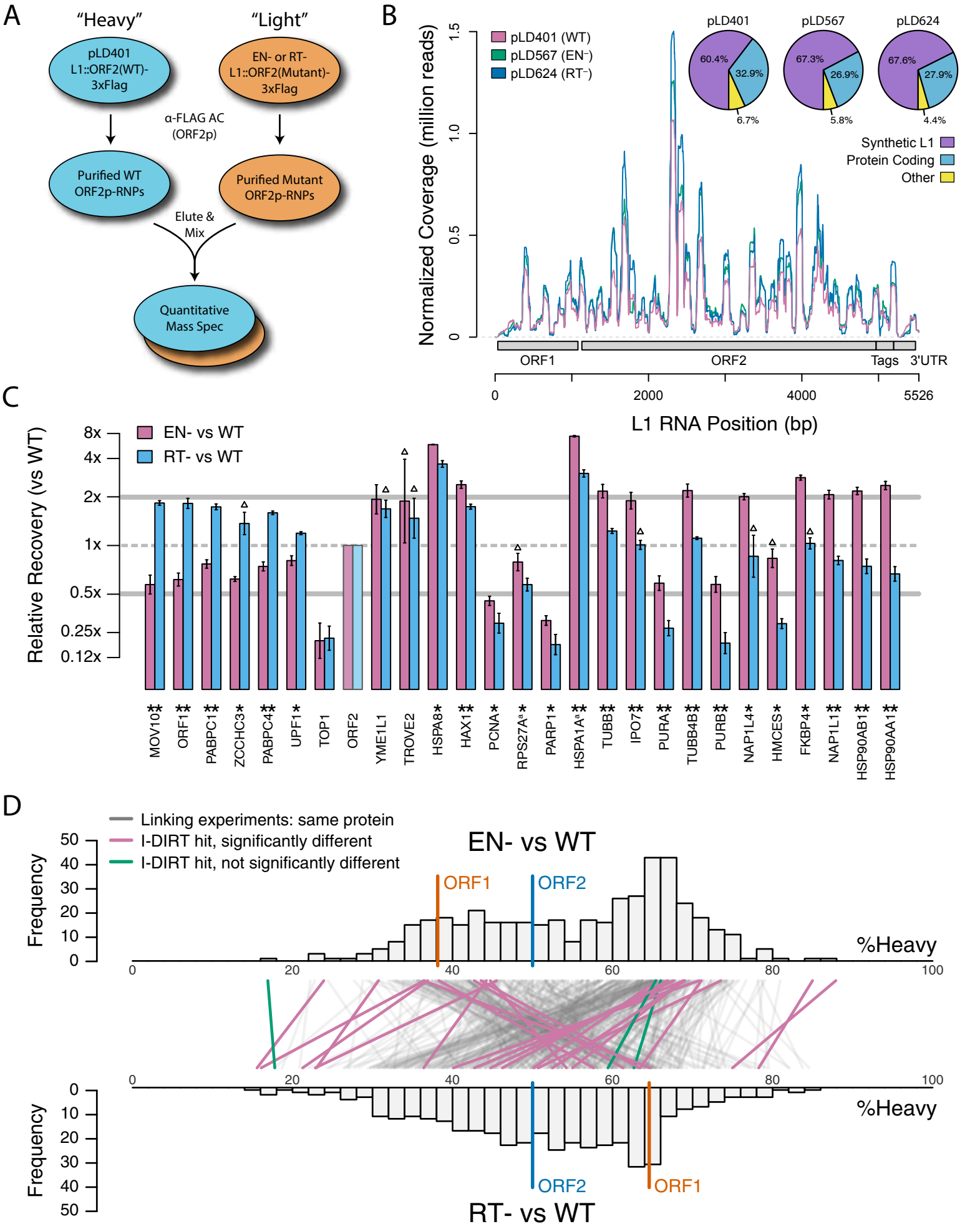

A
"Heavy"
"Light"
pLD401 L1::ORF2(WT)-3xFlag
EN- or RT- L1::ORF2(Mutant)-3xFlag
α-FLAG AC (ORF2p)
Purified WT ORF2p-RNPs
Purified Mutant ORF2p-RNPs
Elute & Mix
Quantitative Mass Spec
B
Normalized Coverage (million reads)
pLD401 (WT)
pLD567 (EN⁻)
pLD624 (RT⁻)
pLD401
pLD567
pLD624
60.4%
32.9%
6.7%
67.3%
26.9%
5.8%
67.6%
27.9%
4.4%
Synthetic L1
Protein Coding
Other
ORF1
ORF2
Tags
3'UTR
0
2000
4000
5526
L1 RNA Position (bp)
C
Relative Recovery (vs WT)
EN- vs WT
RT- vs WT
8x
4x
2x
1x
0.5x
0.25x
0.12x
MOV10
ORF1
PABPC1
ZCCHC3
PABPC4
UPF1
TOP1
ORF2
YME1L1
TROVE2
HSPA8
HAX1
PCNA
RPS27A
PARP1
HSPA1A
TUBB
IPO7
PURA
TUBB4B
PURB
NAP1L4
HMCES
FKBP4
NAP1L1
HSP90AB1
HSP90AA1
D
Linking experiments: same protein
I-DIRT hit, significantly different
I-DIRT hit, not significantly different
EN- vs WT
Frequency
ORF1
ORF2
%Heavy
RT- vs WT

Figure 4-S1

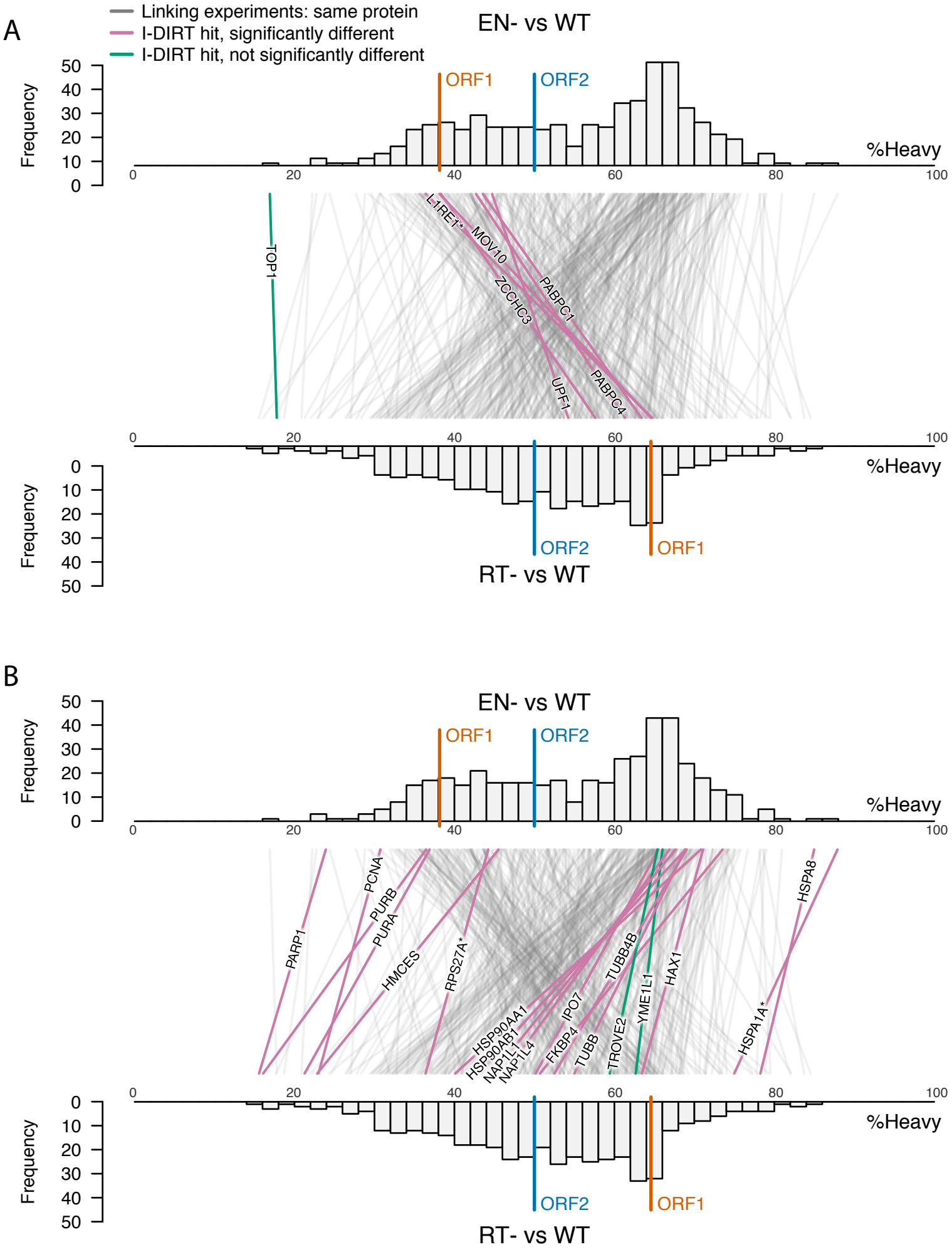

A
Linking experiments: same protein
I-DIRT hit, significantly different
I-DIRT hit, not significantly different
EN- vs WT
Frequency
ORF1
ORF2
%Heavy
TOP1
L1RE1*
MOV10
ZCCHC3
PABPC1
UPF1
PABPC4
RT- vs WT
B
EN- vs WT
PARP1
PCNA
PURB
PURA
HMCES
RPS27A*
HSP90AA1
HSP90AB1
NAP1L1
NAP1L4
IPO7
FKBP4
TUBB
TUBB4B
TROVE2
YME1L1
HAX1
HSPA8
HSPA1A*
RT- vs WT

Figure 5

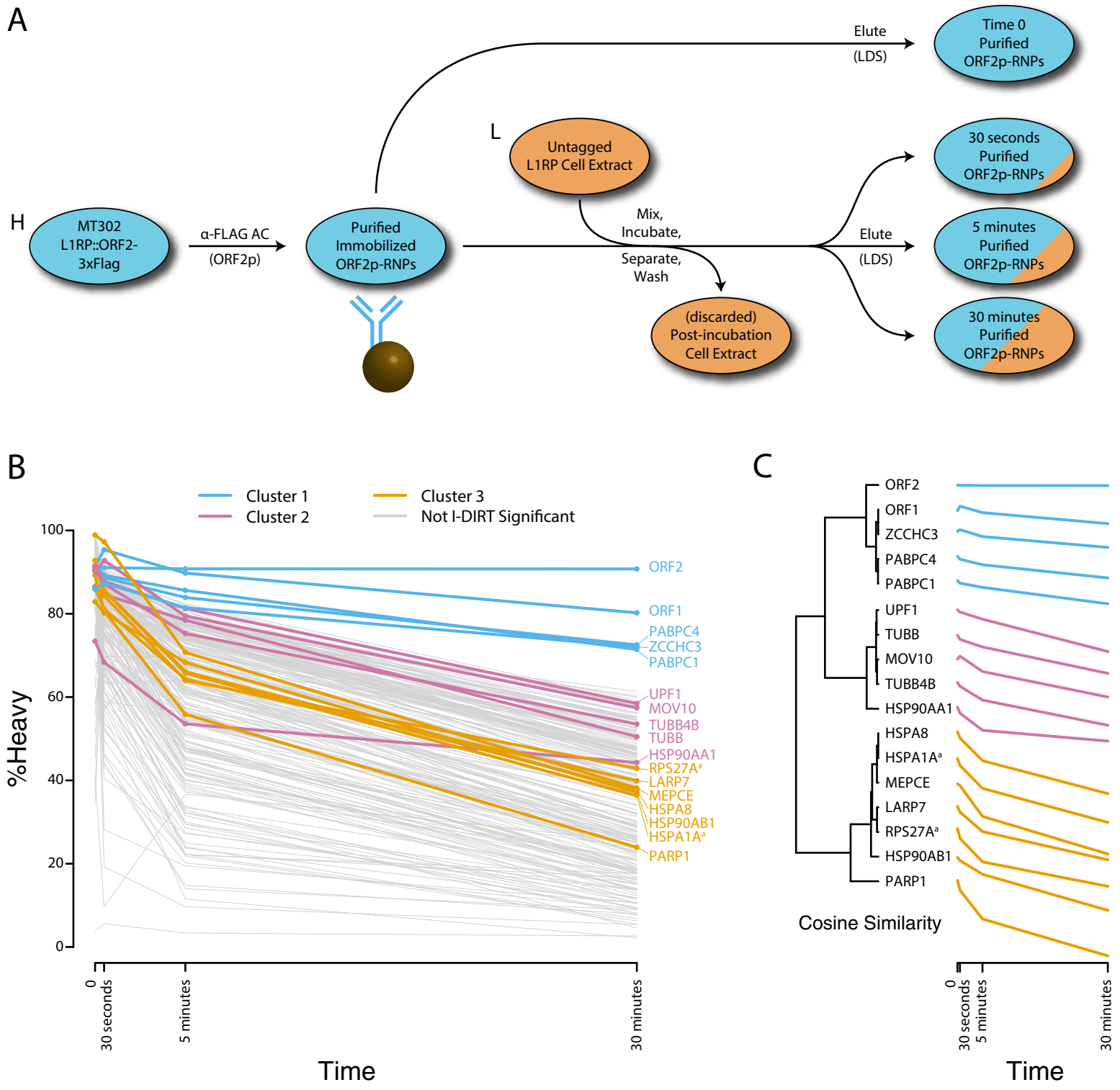

A
H
MT302
L1RP::ORF2-
3xFlag
α-FLAG AC
(ORF2p)
Purified
Immobilized
ORF2p-RNPs
Elute
(LDS)
Time 0
Purified
ORF2p-RNPs
L
Untagged
L1RP Cell Extract
Mix,
Incubate,
Separate,
Wash
(discarded)
Post-incubation
Cell Extract
Elute
(LDS)
30 seconds
Purified
ORF2p-RNPs
5 minutes
Purified
ORF2p-RNPs
30 minutes
Purified
ORF2p-RNPs
B
Cluster 1
Cluster 2
Cluster 3
Not I-DIRT Significant
%Heavy
100
80
60
40
20
0
ORF2
ORF1
PABPC4
ZCCHC3
PABPC1
UPF1
MOV10
TUBB4B
TUBB
HSP90AA1
RPS27A[a]
LARP7
MEPCE
HSPA8
HSP90AB1
HSPA1A[a]
PARP1
0
30 seconds
5 minutes
30 minutes
Time
C
ORF2
ORF1
ZCCHC3
PABPC4
PABPC1
UPF1
TUBB
MOV10
TUBB4B
HSP90AA1
HSPA8
HSPA1A[a]
MEPCE
LARP7
RPS27A[a]
HSP90AB1
PARP1
Cosine Similarity
0
30 seconds
5 minutes
30 minutes
Time

Figure 6

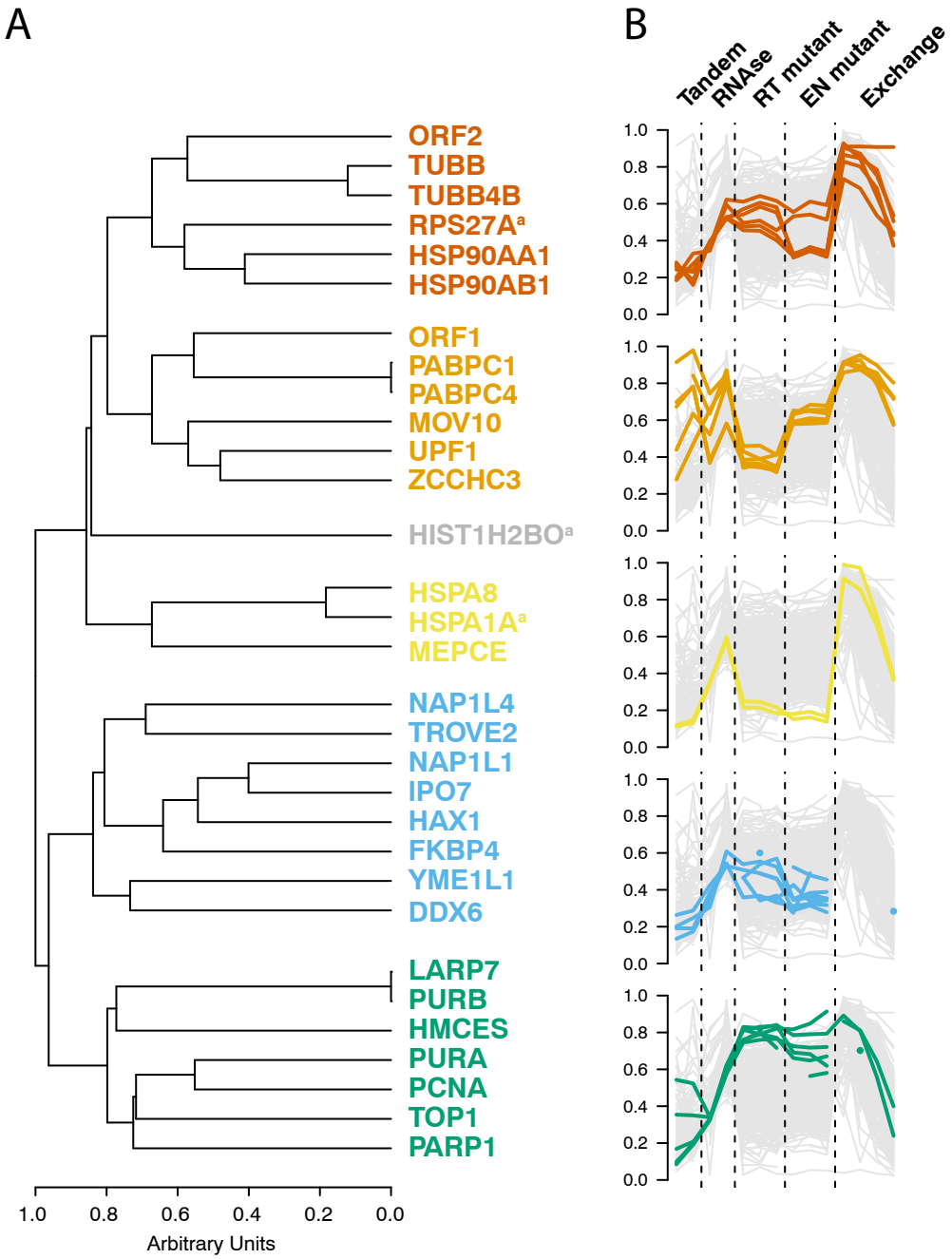
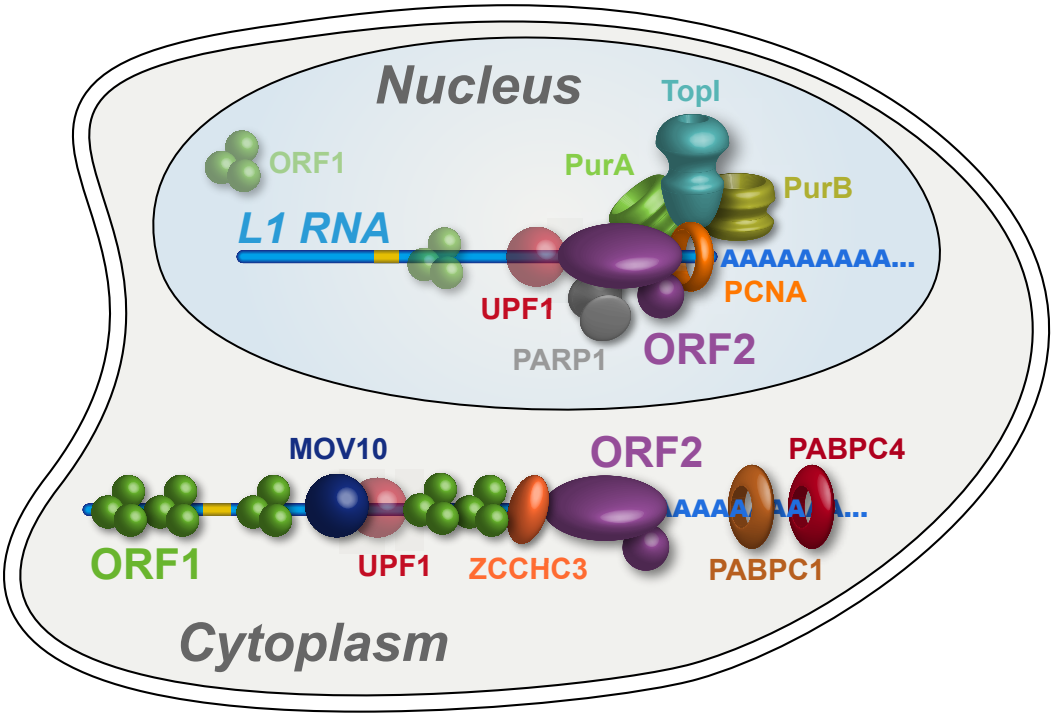
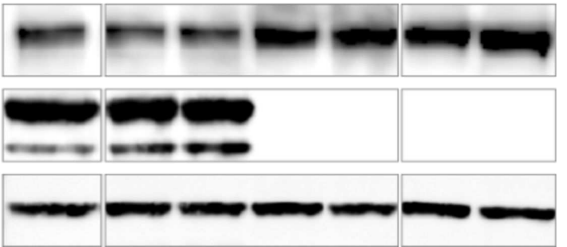
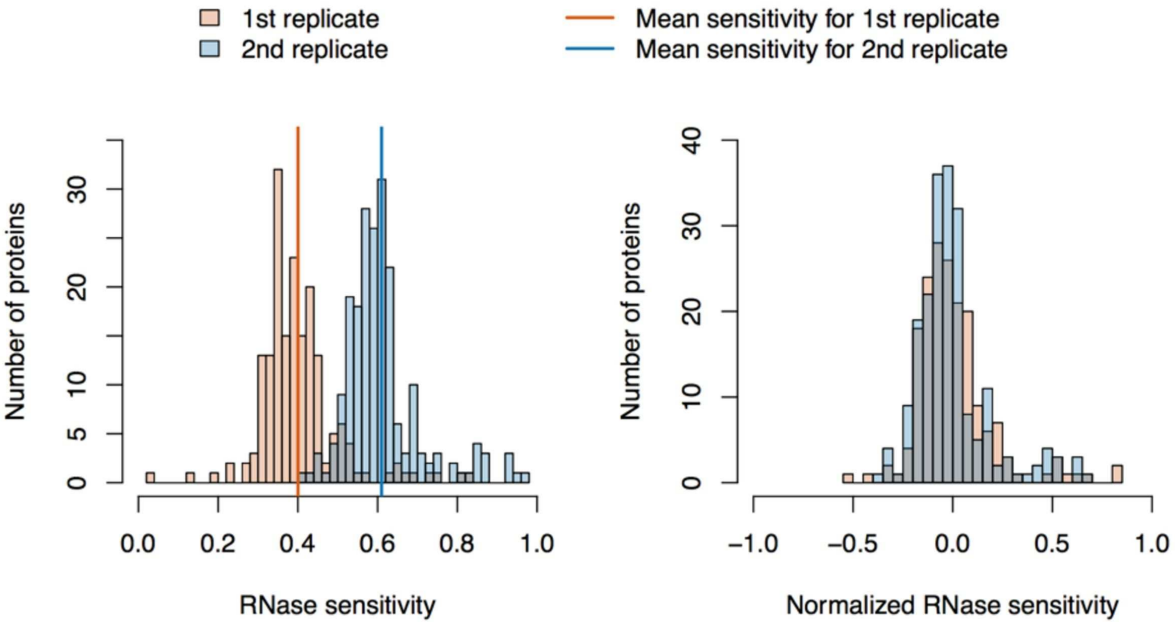

A
B
Tandem
RNAse
RT mutant
EN mutant
Exchange
ORF2
TUBB
TUBB4B
RPS27A[a]
HSP90AA1
HSP90AB1
ORF1
PABPC1
PABPC4
MOV10
UPF1
ZCCHC3
HIST1H2BO[a]
HSPA8
HSPA1A[a]
MEPCE
NAP1L4
TROVE2
NAP1L1
IPO7
HAX1
FKBP4
YME1L1
DDX6
LARP7
PURB
HMCES
PURA
PCNA
TOP1
PARP1
1.0
0.8
0.6
0.4
0.2
0.0
Arbitrary Units

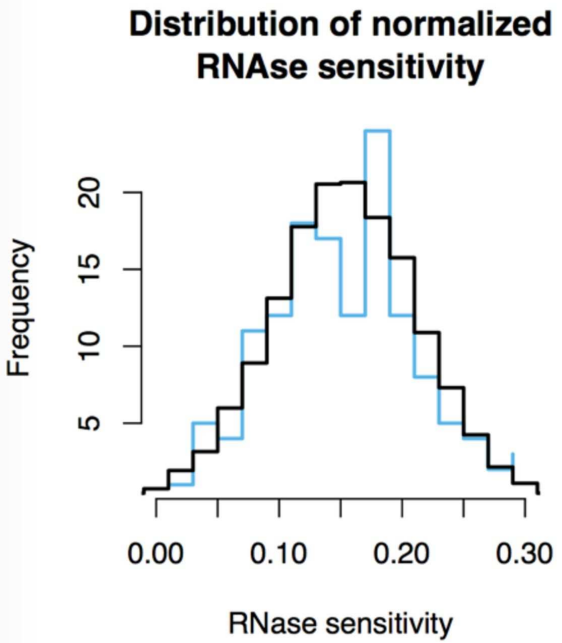
Figure 7

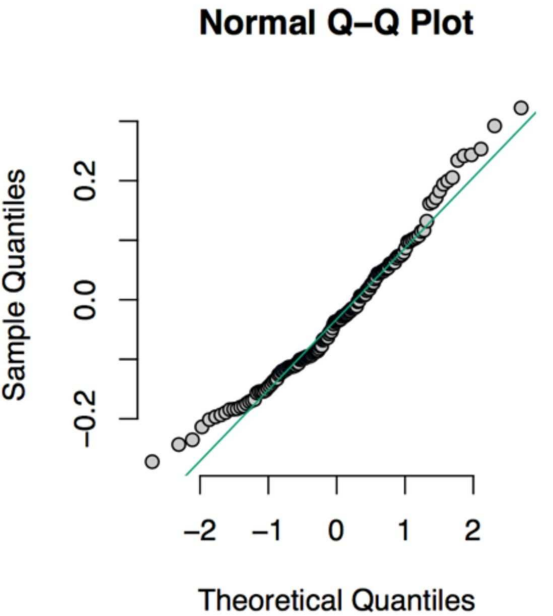

Nucleus
TopI
ORF1
PurA
PurB
L1 RNA
AAAAAAAAAA...
PCNA
UPF1
PARP1
ORF2
MOV10
ORF2
PABPC4
AAAAAAAAA...
ORF1
UPF1
ZCCHC3
PABPC1
Cytoplasm

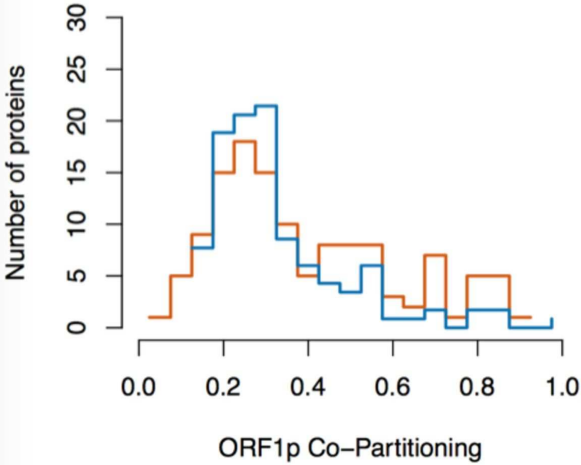
Construct	LD401			LD561			
Adherent/Suspension	A	S	S	S	S	S	S
ORF1	+	+	+	—	—	—	—
Heavy/Light/Conventional	C	L	H	L	H	C	C

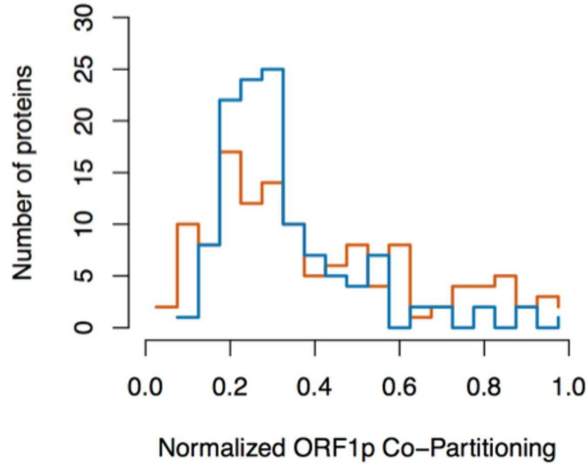


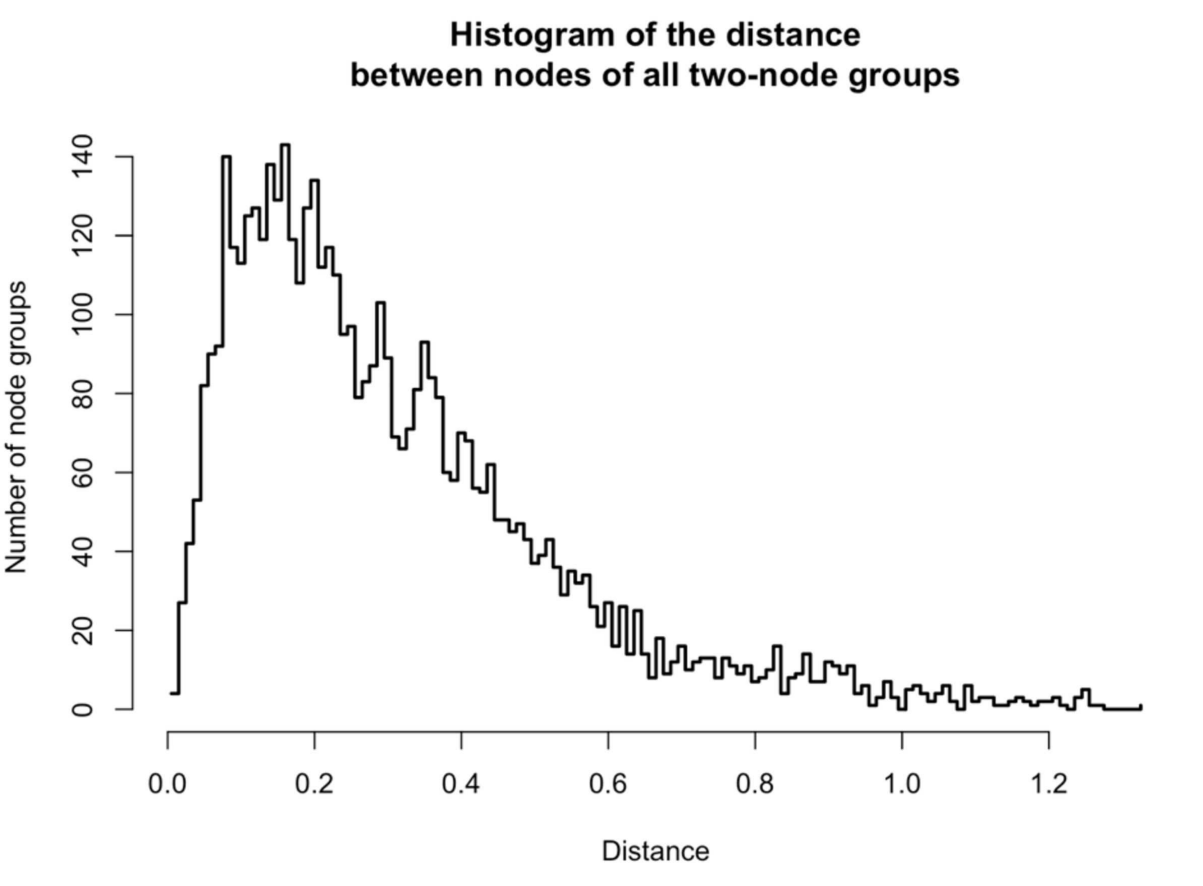

1st replicate
2nd replicate
Mean sensitivity for 1st replicate
Mean sensitivity for 2nd replicate
Number of proteins
0
5
10
20
30
0.0
0.2
0.4
0.6
0.8
1.0
RNase sensitivity
Number of proteins
0
10
20
30
40
−1.0
−0.5
0.0
0.5
1.0
Normalized RNase sensitivity

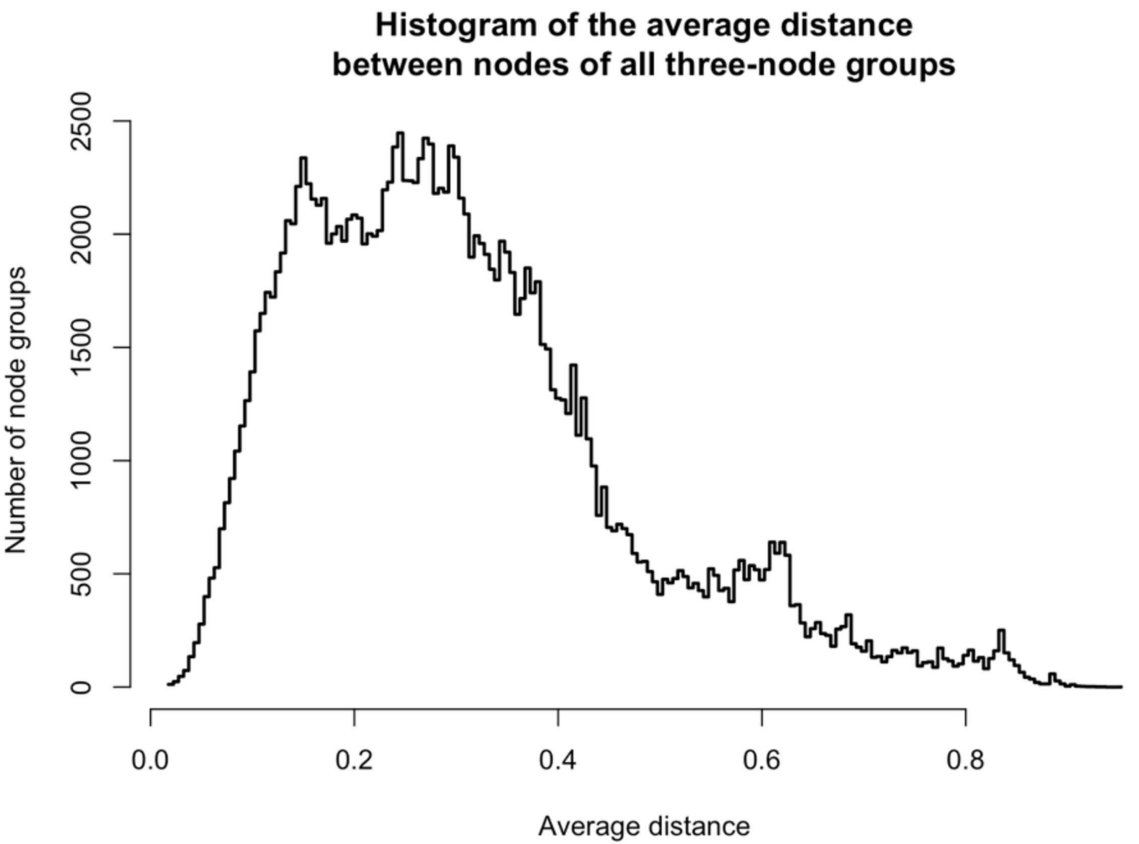

Distribution of normalized
RNAse sensitivity
Frequency
5
10
15
20
0.00
0.10
0.20
0.30
RNase sensitivity

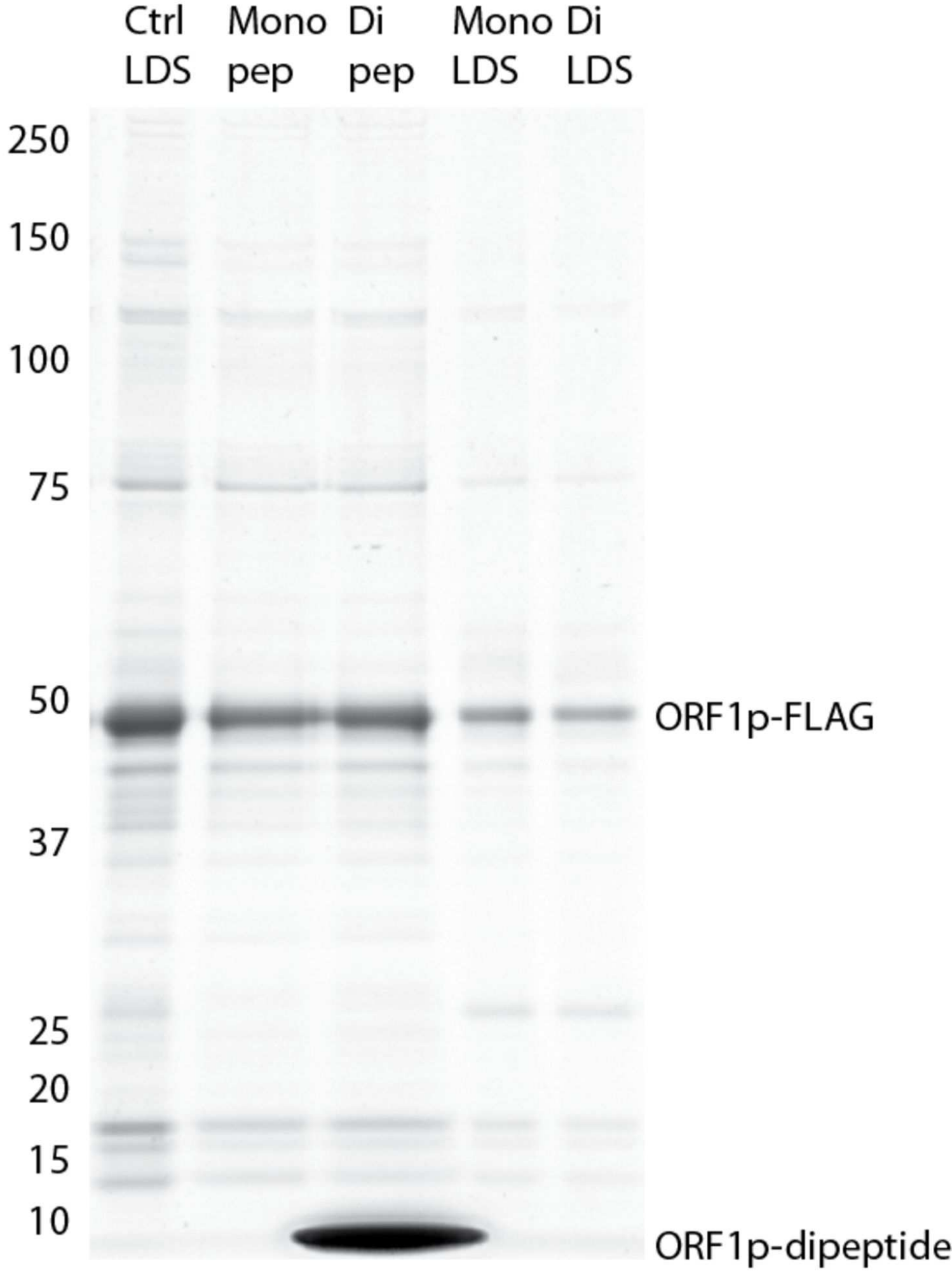

Normal Q–Q Plot
Sample Quantiles
−0.2
0.0
0.2
−2
−1
0
1
2
Theoretical Quantiles

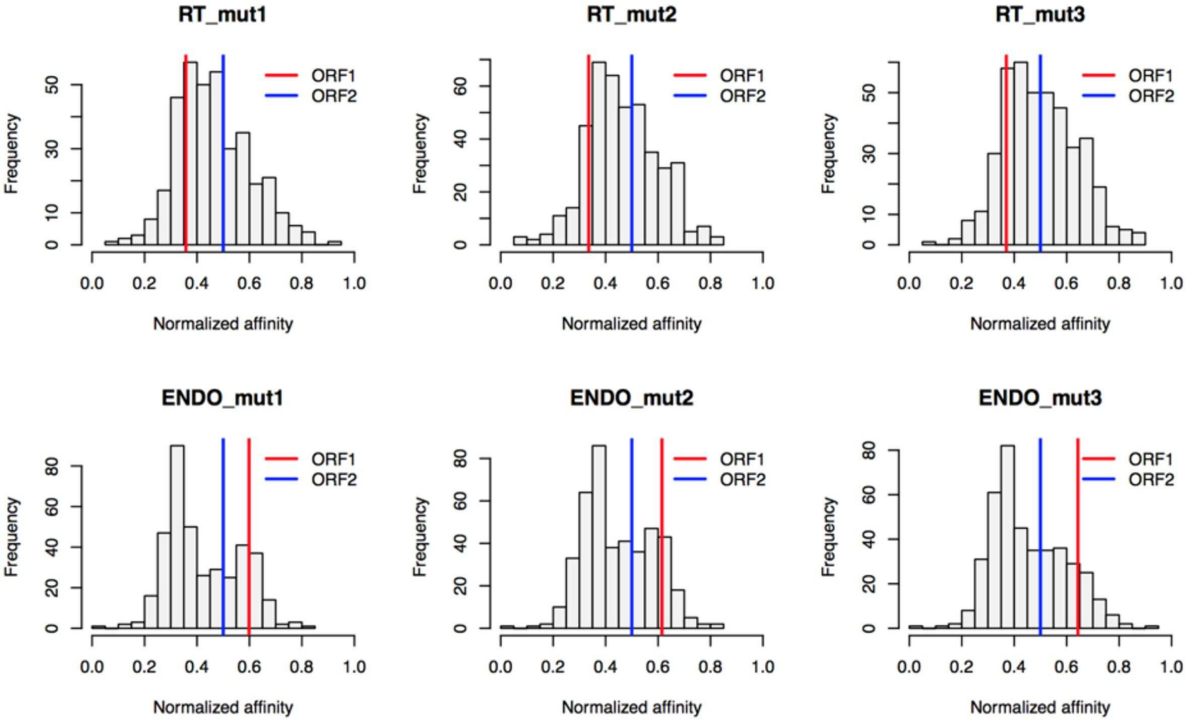

Number of proteins
0
5
10
15
20
25
30
0.0
0.2
0.4
0.6
0.8
1.0
ORF1p Co−Partitioning

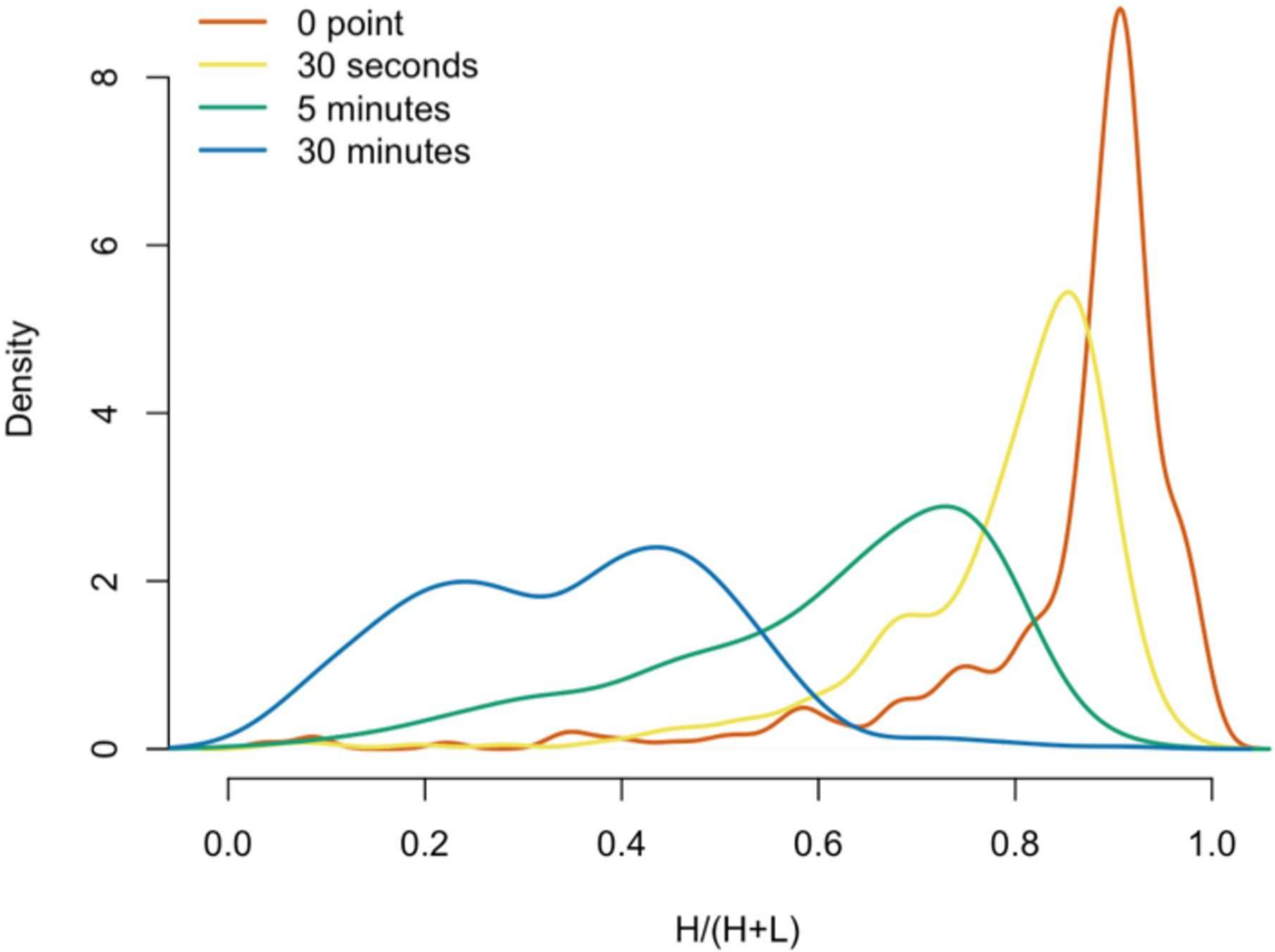

Number of proteins
0
5
10
15
20
25
30
0.0
0.2
0.4
0.6
0.8
1.0
Normalized ORF1p Co−Partitioning


Histogram of the distance
between nodes of all two-node groups
Number of node groups
0
20
40
60
80
100
120
140
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Distance


Histogram of the average distance
between nodes of all three-node groups
Number of node groups
0
500
1000
1500
2000
2500
0.0
0.2
0.4
0.6
0.8
Average distance


Ctrl LDS
Mono pep
Di pep
Mono LDS
Di LDS
250
150
100
75
50
37
25
20
15
10
ORF1p-FLAG
ORF1p-dipeptide


RT_mut1
RT_mut2
RT_mut3
ENDO_mut1
ENDO_mut2
ENDO_mut3
ORF1
ORF2
Frequency
Normalized affinity
0.0
0.2
0.4
0.6
0.8
1.0


0 point
30 seconds
5 minutes
30 minutes
Density
0
2
4
6
8
0.0
0.2
0.4
0.6
0.8
1.0
H/(H+L)