

RESEARCH

Open Access

RIP-seq reveals LINE-1 ORF1p association with p-body enriched mRNAs



Erica M. Briggs^{1†}, Wilson McKerrow^{1,2†}, Paolo Mita^{1,2}, Jef D. Boeke^{1,2}, Susan K. Logan^{1,3*} and David Fenyo^{1,2*} 

Abstract

Background: Long INterspersed Element-1 (LINE-1) is an autonomous retroelement able to “copy-and-paste” itself into new loci of the host genome through a process called retrotransposition. The LINE-1 bicistronic mRNA codes for two proteins, ORF1p, a nucleic acid chaperone, and ORF2p, a protein with endonuclease and reverse transcriptase activity. Both proteins bind LINE-1 mRNA in cis and are necessary for retrotransposition. While LINE-1 transcription is usually repressed in most healthy somatic cells through a plethora of mechanisms, ORF1p expression has been observed in nearly 50% of tumors, and new LINE-1 insertions have been documented in a similar fraction of tumors, including prostate cancer.

Results: Here, we utilized RNA ImmunoPrecipitation (RIP) and the L1EM analysis software to identify ORF1p bound RNA in prostate cancer cells. We identified LINE-1 loci that were expressed in parental androgen sensitive and androgen independent clonal derivatives. In all androgen independent cells, we found higher levels of LINE-1 RNA, as well as unique expression patterns of LINE-1 loci. Interestingly, we observed that ORF1p bound many non-LINE-1 mRNA in all prostate cancer cell lines evaluated, and polyA RNA, and RNA localized in p-bodies were especially enriched. Furthermore, the expression levels of RNAs identified in our ORF1p RIP correlated with RNAs expressed in LINE-1 positive tumors from The Cancer Genome Atlas (TCGA).

Conclusion: Our results show a significant remodeling of LINE-1 loci expression in androgen independent cell lines when compared to parental androgen dependent cells. Additionally, we found that ORF1p bound a significant amount of non-LINE-1 mRNA, and that the enriched ORF1p bound mRNAs are also amplified in LINE-1 expressing TCGA prostate tumors, indicating the biological relevance of our findings to prostate cancer.

Keywords: LINE-1, L1EM, ORF1p, Prostate cancer, Processing bodies, RNA, Retrotransposition

Introduction

Many factors have played a role in shaping the evolution of the human genome. In particular, the Long INterspersed Element 1 (LINE-1) retrotransposon, has significantly contributed in shaping the size, structure, and expression of the human genome over millions of years [1, 2]. Since LINE-1 mobilizes through retrotransposition, a copy and paste mechanism that utilizes an RNA

intermediate, LINE-1 sequences have accumulated in the genome of virtually all organisms, and today over 17% of the modern human DNA is comprised of LINE-1 copies [3]. An estimated 500,000 copies of LINE-1 exist in the human genome, however, the majority of these sequences are unable to mobilize due to truncations or incapacitating mutations and inversions. However, an estimated 80–100 full length LINE-1 sequences, belonging to the subfamily L1Hs, are retrotransposition competent. Among these full-length LINE-1 s, only a few “hot” loci contribute to the bulk of LINE-1 mRNA expression [4, 5].

Full length LINE-1 sequences consist of a 5'UTR/promoter, two open reading frames, coding for ORF1p and

* Correspondence: Susan.Logan@nyulangone.org; David@FenyoLab.org

[†]Erica M. Briggs and Wilson McKerrow contributed equally to this work.

¹Departments of Biochemistry and Molecular Pharmacology, New York University, 450 East 29th Street, Room 321, New York, NY 10016, USA

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

ORF2p proteins, and a 3' UTR with a polyA signal [6]. Following transcription by RNA polymerase II, LINE-1 mRNA is exported from the nucleus. ORF1p and ORF2p are translated in the cytoplasm, where they bind LINE-1 mRNA and form LINE-1 ribonucleoproteins (RNP) composed of LINE-1 mRNA coated by many ORF1p trimers and presumably one or a few ORF2p [7, 8]. In dividing cells, the LINE-1 RNPs can enter the nucleus upon breakdown of the nuclear membrane during mitosis [9]. ORF2p then nicks the DNA in A/T rich regions (AA/TTTT consensus) using its endonuclease domain, and inserts a new copy of LINE-1 through its reverse transcriptase domain [10, 11]. While LINE-1 has demonstrated strong cis preference in mobilizing its own mRNA, its proteins have also been shown to bind and mobilize non-LINE-1 mRNA such as SINEs and other mRNAs that produced processed pseudogenes [12–14].

Studies have demonstrated that both LINE-1 proteins, ORF1p and ORF2p, may be necessary for LINE-1 retrotransposition [15]. ORF1p is a nucleic acid chaperone and is composed of a coiled coil domain, an RNA recognition motif (RRM), a carboxy-terminal domain (CTD) and an unstructured N-terminal region [16, 17]. The coiled coil domain is responsible for the formation of ORF1p homotrimers. The RRM, CTD, and ORF1p trimerization facilitates ORF1p binding to single stranded nucleic acids [16]. In studies of LINE-1 overexpression, ORF1p has been shown to have a strong cis preference to LINE-1 mRNA [18]. In addition to binding LINE-1 mRNA, ORF1p has also been shown to bind ssDNA. ORF1p has also been shown to aid in strand exchange and annealing, suggesting a possible role in directly facilitating LINE-1 reverse transcription [19, 20]. While ORF1p seems to be needed for LINE-1 and pseudogene retrotransposition, it is not crucial for SINE retrotransposition [12] although, increased levels of ORF1p were shown to promote higher SINE retrotransposition [21]. ORF1p predominantly localizes to the cytoplasm, and has been found in stress granules (SG) and in close proximity to processing bodies (p-bodies) [22–24]. Stress granules and processing bodies are both cytoplasmic RNA granules that play a role in RNA metabolism and translation regulation. Stress granules form in response to cellular stressors such as heat shock and oxidative stress, and have been shown to accumulate mRNA and proteins that are stalled during translation initiation. Processing bodies on the other hand are present in non-stressed cells and contain machinery involved in RNA decay, RNA mediated gene silencing, RNA storage, and translational repression [25–27]. Since ORF1p can be translated with high efficiency and specific antibodies are commercially available, its endogenous expression and localization have been widely studied. In contrast, ORF2p expression is restricted due to an unconventional

translation, leading to almost undetectable endogenous levels [28, 29]. Yet, evidence of ORF2p endonuclease and reverse transcriptase activity is clear through its impact on genome evolution and, as more current findings show, through the identification of many new LINE-1 insertions in several cancers [15, 30, 31].

DNA methylation, histone modifications, and RNA interference all limit LINE-1 expression and function in somatic cells [32–35]. However, mechanisms that limit LINE-1 expression are often dysfunctional in cancers, allowing for the expression and mobilization of LINE-1 [31, 36, 37]. LINE-1 expression has the potential to disrupt genomic stability, making it a likely component in cancer progression [38, 39]. LINE-1 ORF1p expression has been observed in around 47% of tumors tested, and roughly in 40% of prostate tumors [36]. In some cancers, such as pancreatic ductal adenocarcinomas and breast cancer, ORF1p expression patterns correlated with poorer clinical outcome [40, 41]. New LINE-1 insertions have been identified in 53% of tumors sequenced, including around 60% of prostate tumors [31]. Additionally, the LINE-1 insertion rate in prostate tumors was higher in metastatic tumors when compared to primary tumors, suggesting a correlation between LINE-1 retrotransposition and tumor progression [31].

The repetitiveness of LINE-1 sequences poses a challenge to identifying actively transcribed LINE-1 loci [42, 43]. However, a newly developed analysis software, LINE-1 Expectation Maximization (L1EM), utilizes the expectation maximization algorithm to overcome this obstacle and identify loci of actively transcribed LINE-1 [44]. Here, we performed RNA-immunoprecipitation of endogenous ORF1p to identify ORF1p bound transcripts and utilized L1EM to identify specific LINE-1 loci expressed in androgen sensitive prostate cancer cells, and androgen independent clones. Among the cell lines tested, we found a high degree of variation in the specific loci expressed, with LNCaP derived cell lines, LNCaP-95 and LNCaP-abl, having LINE-1 expression patterns that are more similar to each other and to LNCaP than to an unrelated prostate cancer cell line (22Rv1). Surprisingly, we found ORF1p bound significant levels of non-LINE-1 mRNA, including enrichments of circRNA, polyadenylated RNAs, and RNAs associated with p-bodies. Interestingly, we show that the expression of the non-LINE-1 RNAs enriched in the ORF1p immunoprecipitations correlates with LINE-1 mRNA expression in TCGA prostate cancer samples, suggesting a possible role of ORF1p in RNA accumulation and processing of these specific transcripts.

Results

LINE-1 loci expressed in prostate cancer cell lines

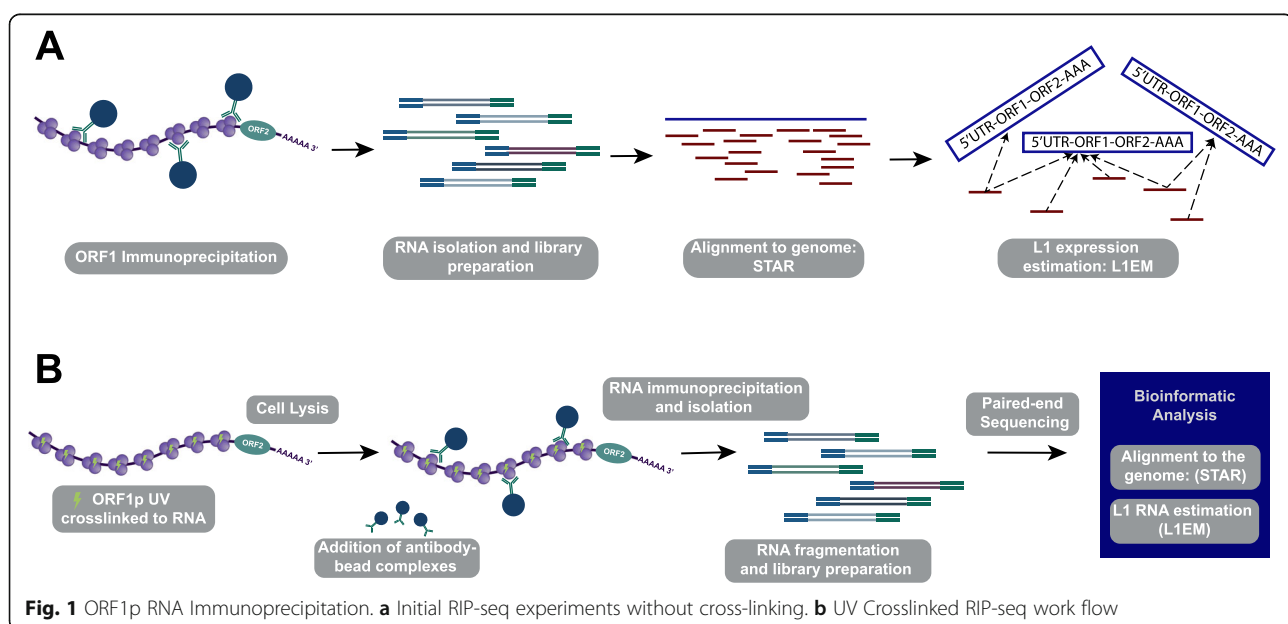
We recently showed that our tool, L1EM, is able to quantify RNA expression at specific LINE-1 loci from

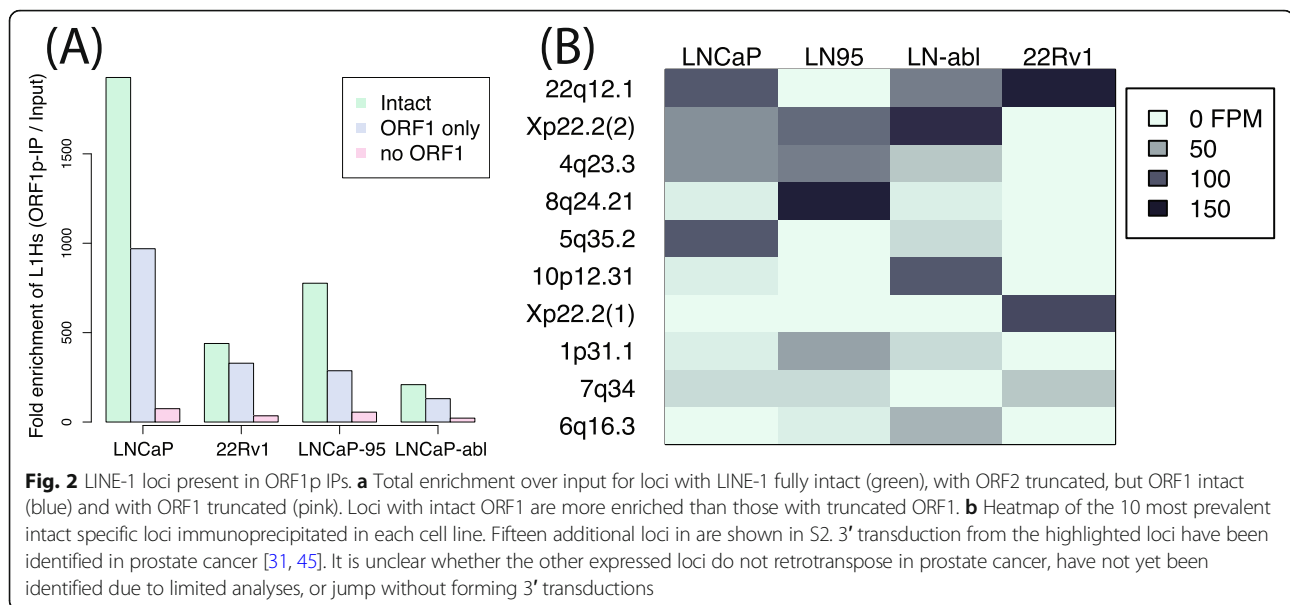
RNA-seq data [44]. We also found that accurately predicting this expression requires significantly deeper coverage compared to a standard RNA-seq analysis. Using a monoclonal antibody (4H1) [36], we performed RNA immunoprecipitation (RIP-seq) of endogenous ORF1p in four prostate cancer cell lines (LNCaP, LNCaP-95, LNCaP-abl and 22Rv1) to enrich for LINE-1 reads and more accurately measure the expressed loci (Fig. 1a). LNCaP cells are representative of an earlier stage, androgen sensitive prostate cancer, while its androgen independent clones, LNCaP-95 and LNCaP-abl, are more representative of treatment resistant prostate cancer, a later stage in disease progression. These cell lines were used to address the question of whether the LINE-1 expressed loci in parental LNCaP cells would change after undergoing the selective pressure of androgen deprivation and progression to androgen independent sub-clones (LNCaP-95 and LNCaP-abl). 22Rv1 cells are also an androgen independent prostate cancer cell line, but are not clonally related to the LNCaP cell line. While LINE-1 comprises a substantial percentage of the human genome, many loci contain mutations resulting in truncated LINE-1 mRNA. We found strong enrichment for LINE-1 RNAs in all cell lines, confirming our previous results [23]. As expected, the greatest enrichment was measured for loci that retain a full length ORF1 compared to those that have acquired an ORF1 nonsense mutation (Fig. 2a). Enrichment was even greater at loci that also have a full length ORF2. Given that ORF1p is assumed to have a cis-preference, this result validates our ability to immunoprecipitate ORF1p bound to RNA and to assign the immunoprecipitated LINE-1 mRNA to specific genomic loci. Lower

enrichment of LINE-1 mRNA in the androgen independent cell lines (LNCaP-95, LNCaP-abl and 22Rv1) is primarily due to higher levels of LINE-1 in the input RNA (LINE-1 quantifications before and after ORF1p-IP are shown in Figure S1). We then looked at the specific intact (full-length and containing no nonsense mutation) LINE-1 loci that are expressed in each cell line. Figure 2b shows a heat map of the ten most highly expressed intact LINE-1 loci; an extended heatmap with 25 loci is provided in Figure S2, and the full list of LINE-1 loci can be found in Table S1. Overall, we find that the expressed loci differ widely between the cell lines. The 22q12.1 locus is highly expressed in most of the cell lines, but was not detected in LNCaP-95. A locus at 8q24.21 is most highly expressed in LNCaP-95, but it was not detected in any other cell line. Overall, the LINE-1 expression pattern is more different than similar between cell lines, but has greater overlap among the LNCaP and LNCaP-derived cell lines than between the LNCaP “family” cell lines and 22Rv1. The overlap in expressed LINE-1 loci ranges from 35% overlap between LNCaP and LNCaP-95 to 42% between LNCaP and LNCaP-abl compared to a 6% overlap between LNCaP-95 and 22Rv1 and 28% for LNCaP and 22Rv1 (one-sided t-test $p = 0.03$). This likely reflects the fact that LNCaP-95 and LNCaP-abl are derived from LNCaP through androgen deprivation, whereas 22Rv1 is clonally independent.

ORF1p IPs are enriched for polyA RNAs

While the enrichment for LINE-1 RNAs with intact ORF1 indicates that we were successful in immunoprecipitating LINE-1 RNPs, we were surprised by the fact that





LINE-1 RNA was such a small fraction of the immunoprecipitated transcripts. Others have also noted widespread associations between ORF1p and host mRNAs [8, 46], and previous LINE-1 overexpression studies have shown that ORF1 bound mRNA consisted of 8.3–10.3% LINE-1 mRNA [8]. Our findings examining endogenous ORF1p found much lower levels of ORF1p bound LINE-1 mRNA.

Reads from younger LINE-1 families (L1Hs, L1PA2, L1PA3, L1PA4) ranged from 0.1% of all reads in LNCaP to 0.15% of all reads in LNCaP-abl. We therefore wondered whether LINE-1 mRNA was dissociating from ORF1p during the IP procedure, freeing ORF1p to interact with other mRNA species present in the cell lysates. To test this possibility, we performed UV crosslinking of LNCaP cells prior to a 1-h or 3-h immunoprecipitation. These IPs yielded slightly less LINE-1 RNA: < 0.1% of reads in all experiments, indicating that ORF1p was unlikely to disassociate from the LINE-1 mRNA during the IP and that most of the identified interactions of ORF1p and LINE-1 mRNAs formed in the cells, before lysis. Analysis of the non-LINE-1 RNA present in the ORF1p immunoprecipitates showed that, in all experiments, the majority of reads represented ribosomal RNA, which was not depleted from the pool of mRNAs in our experimental design. However, we also found a much greater fraction of reads aligning to exonic sequences in ORF1p IP compared to control (input or IgG IP) (Fig. 3a), suggesting an overall enrichment for mature (spliced) gene transcripts in the ORF1p IP. We then calculated the ORF1p co-IP to input RNA ratio for all genes in all four cell lines, finding that most genes are enriched in the ORF1p IP. The median gene enrichment ranged from 5-fold for 22Rv1 to 22-fold for LNCaP (all without cross-

linking) (Fig. 3b-e). However, in all cases, the enrichment for LINE-1 RNA with intact ORF1 (ranging from 51-fold enrichment over input samples for LNCaP-abl cells to 331 fold enrichment for LNCaP) eclipses the enrichments of all but a few dozen genes. In LNCaP there were 10 genes whose enrichment exceed LINE-1, in LNCaP-95 there were 20, 68 in LNCaP-abl and 24 in 22Rv1. While we did find certain classes of RNA to be enriched by ORF1p IP (see below), we did not identify RNAs that were consistently enriched more strongly than LINE-1. Seven mRNAs did exceed LINE-1 RNA enrichment values in two of the four considered cell lines: *HLTF*, *PCDHGA9*, *PCDHGB5*, *RELN*, *SMC2*, *SI*, and *SPI10*. The full list of mRNAs enriched in each experiment can be found in Table S2.

The global enrichment for exonic sequences led us to hypothesize that ORF1p is promiscuously binding polyadenylated RNAs. We therefore looked at histone mRNAs, which rely on a unique expression mechanism and are not polyadenylated, expecting a low enrichment for these transcripts in our ORF1p IP samples. We found this to be the case in the LNCaP cross-linking experiments (Fig. 4b-c, mean enrichment 1.1x), but not in the experiments without cross-linking (Fig. 4a). This discrepancy indicates that, without cross-linking, there may be some RNAs that do in fact exchange during the IP. Interestingly, as noted above, this does not appear to be the case for LINE-1 RNA itself, enrichment of which was not increased by cross-linking. The consistency of LINE-1 transcripts pulled down in cross-linked and non-cross-linked ORF1 RIPs may be due to ORF1p binding LINE-1 RNA with a higher affinity, or it may be that a higher number of ORF1p trimers are bound to each LINE-1 mRNA transcript, making RNA/protein

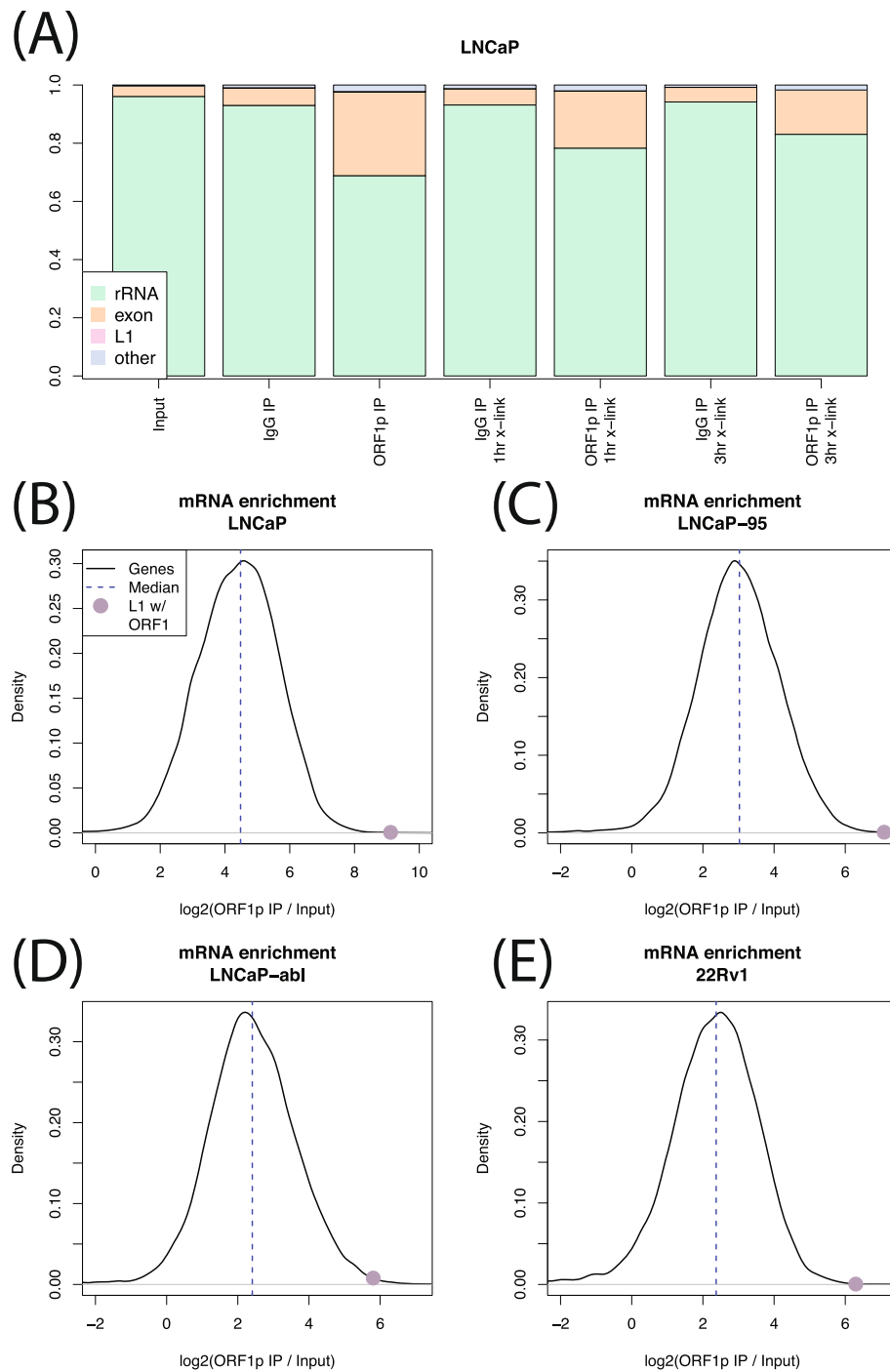
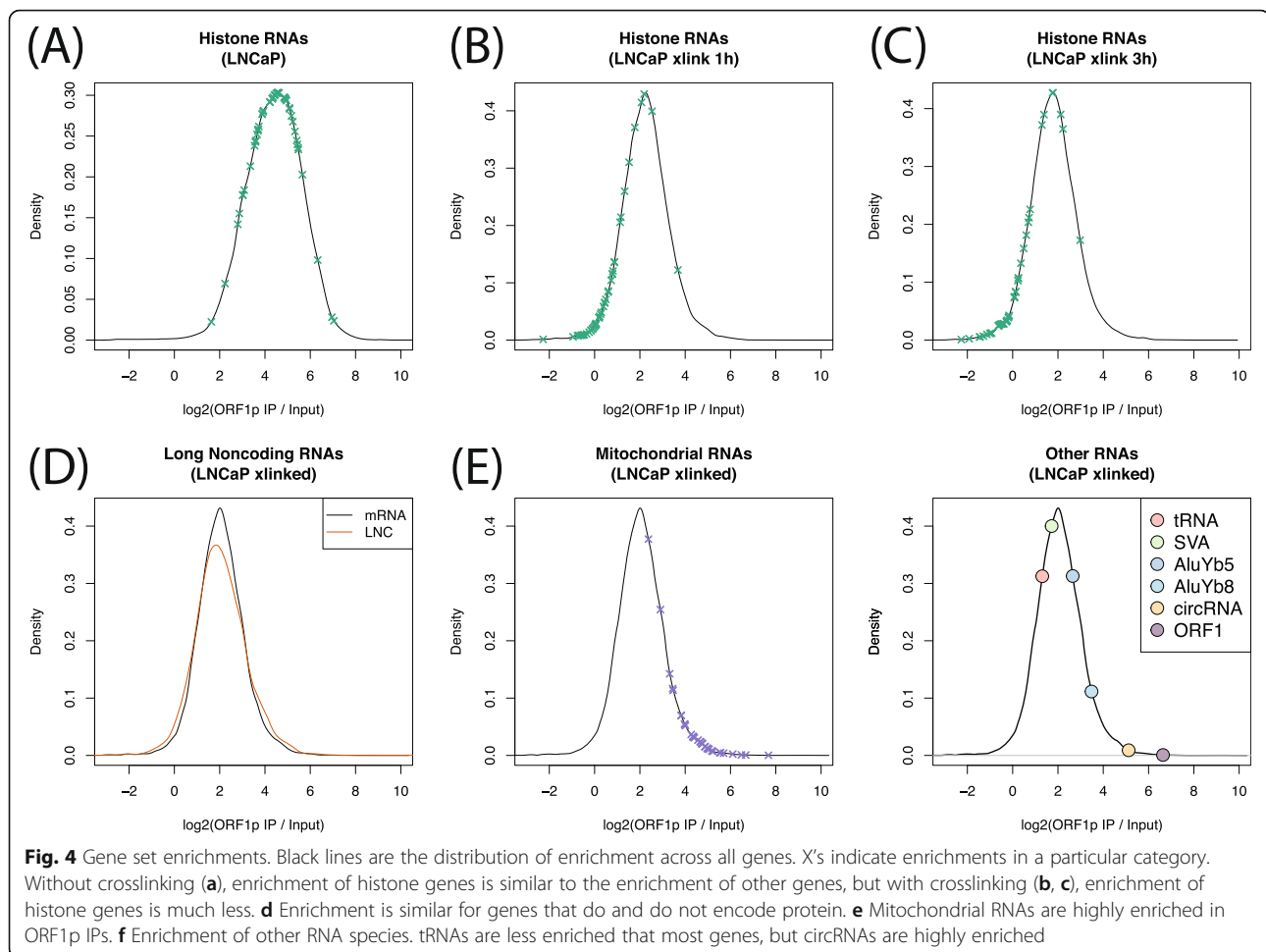


Fig. 3 Exonic sequences enriched in IPs. **a** Relative fraction of rRNA, exon aligned RNA, L1 RNA and other RNA in each LNCaP experiment. In all experiments, L1 abundance is too low to be seen. **b-e** Black lines show the distribution of IP enrichment across all genes in each cell line. Purple dot shows LINE-1 mRNA enrichment. In all cell lines, most genes are enriched, but ORF1 is among the strongest enrichments

dissociation less likely. Alternatively, it could be that these RNAs are bound to ORF1p indirectly, leading to a weaker association (see below.)

We then wanted to know whether the IP enrichment is specific to protein coding mRNAs. lncRNAs, which

are polyadenylated, but not translated, showed similar enrichment to mRNA sequences. Median mRNA enrichment was 2.05 fold over input, and median lncRNA enrichment was 2.03x (Wilcox $p = 0.09$, Fig. 4d). We also found a strong and surprising enrichment for



mitochondrial genome encoded RNA (MT-RNA) in the crosslinking experiments (Fig. 4e). This result was unexpected as ORF1p has not been observed to localize inside of mitochondria, but might reflect an association between mitochondria and p-bodies or a colocalization of ORF1p with RNA released from dysfunctional mitochondria presumably into p-bodies (see discussion). Finally, we calculated enrichment for other RNA species, including tRNA, active non-autonomous retrotransposons (SVA, AluYb5, AluYa8) and circular RNAs (circRNA), (Fig. 4f). Of these, only circRNA showed strong enrichment (34x enrichment on average). AluYb8 also showed some enrichment (8x on average). No individual circRNA was supported by 10 or more reads, but total circRNA was strongly enriched in the ORF1p IPs. This result likely reflects the known fact the circRNAs locate to p-bodies in the cytoplasm [47] (see next section).

Finally, we asked whether particular classes of mRNA are specifically enriched in ORF1p IPs. We chose to focus on the cross-linking experiments as those appeared more robust (see above). We therefore performed Gene Set Enrichment Analysis (GSEA) [48] using

Reactome [49], KEGG [50] and GO [51] gene sets. For all three gene sets, the most highly enriched category contained a large number of ribosomal proteins: ribosome (KEGG), translation initiation (Reactome), and cytosolic ribosome (GO). Ribosomal protein RNAs are some of the most highly expressed mRNAs and are frequently processed into pseudogenes [52], which, given the involvement of ORF1p in retrotransposition of pseudogenes, indicates that these mRNAs can bind directly to ORF1p. Only a few of the gene sets enriched at FDR < 5% are not dominated by ribosomal proteins. These are: propanoate metabolism (KEGG), apoptotic cleavage of cellular proteins (Reactome), apoptotic activation phase (Reactome), cilium or flagellum dependent cell motility (GO), and homophilic cell adhesion via plasma membrane adhesion molecules (GO). Enrichment in this last category is driven by protocadherins.

A similar enrichment of many host mRNAs was previously reported by Mandal et al. (2013) in the context of ectopic LINE-1 expression [46]. We found a significant overlap between the mRNAs enriched at FDR < 5% in our data and the mRNA reported to bind ORF1p by

Mandal et al., with 24% of our genes appearing in their list and 25% of their genes appearing in our list (odds ratio = 1.97, $p = 5.9 \times 10^{-24}$). Mandal et al. hypothesized that ORF1p bound mRNAs are poised for pseudogene retrotransposition. However, we only found a weak correlation between the number of times a gene has been processed into a pseudogene [46, 53] and enrichment in the ORF1p IP (Spearman $\rho = 0.11$, $p = 1.1 \times 10^{-6}$). Thus, pseudogenes alone are likely an insufficient explanation for the prevalence of mRNAs in the ORF1p IP.

ORF1p associated mRNA is enriched in p-body RNA

We reasoned that ORF1p localization to some phase separated RNP granules may explain why the RNAs that immunoprecipitated with ORF1p were particularly enriched for polyA RNAs. If ORF1p localizes to such a granule, we would expect to IP not only mRNAs directly bound ORF1p, but also those that indirectly interact with ORF1p through extended RNA/protein interactions. Because ORF1p is primarily cytoplasmic [9], two types of granules are natural candidates: stress granules (SG) and processing bodies (p-bodies). We therefore compared our ORF1p enrichments from the LNCaP crosslinking experiments to published SG [54] and p-

body RNA enrichments [55]. These comparisons are not perfect as our experiments were not done in chemically stressed cells and the cell lines used are different, but they can still provide a sense of whether the immunoprecipitated mRNAs are SG- or p-body-like. We found a strong correlation between the genes enriched by ORF1p IP and those enriched by p-body purification (Spearman $\rho = 0.4$, $p \approx 0$, Fig. 5a), with more than half (63%) of the genes that are significantly (BH FDR < 5%) enriched in the ORF1p IP also enriched in the p-bodies (odds ratio = 2.5, $p = 2.6 \times 10^{-77}$, Figure S3A). There was also a correlation with SG enrichment (Spearman $\rho = 0.15$, $p = 7.1 \times 10^{-43}$, Fig. 5b) and a significant overlap between genes enriched in ORF1p IP and those enriched in SG (odds ratio = 1.3, $p = 1.1 \times 10^{-5}$, Figure S3B), but the correlation becomes negative when using partial correlation to account for the similarity between p-bodies and SGs ($\rho = -0.08$, $p = 1.4 \times 10^{-12}$), indicating that the positive relationship between ORF1p bound and SG localized RNA can be explained by the fact that many RNAs are enriched in both p-bodies and SG. The Mandal et al. data also has stronger overlap with p-body RNA (odds ratio = 5.3, $p = 9.9 \times 10^{-209}$) than with SG RNA (odds ratio = 1.4, $p = 5.2 \times 10^{-9}$) [46]. We also compared our

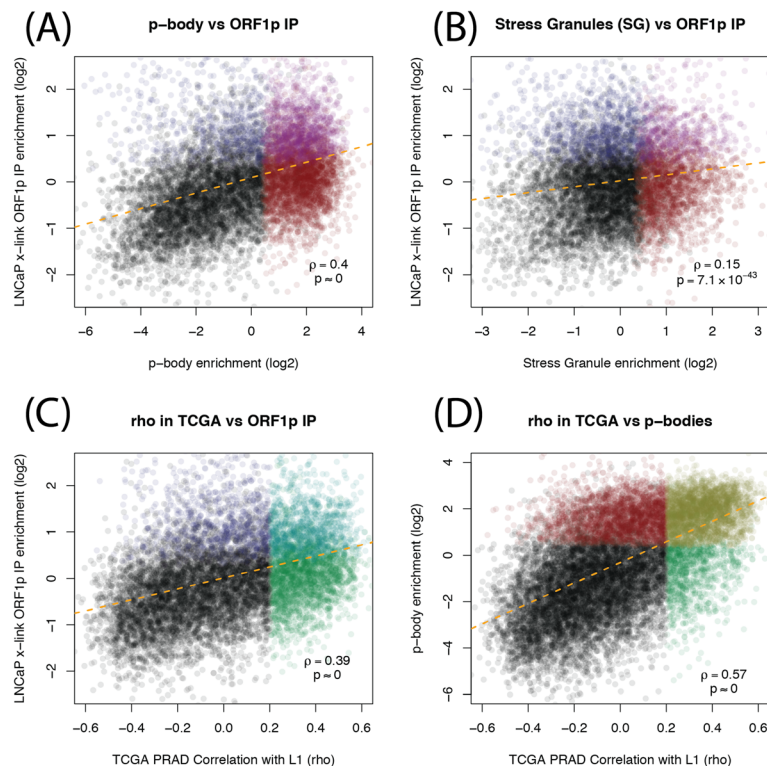


Fig. 5 Scatter plots comparing ORF1p IP RNAs, granules RNAs and RNAs correlated with LINE-1 in TCGA prostate cancer. **a** For each gene, enrichment in p-body purification study [55] (x-axis) vs enrichment in our ORF1p IPs, calculated by DESeq2 (y-axis). **b** Same y-axis as in **a**, but using SG enrichments [54] on the x-axis. **c** RNA expression correlation between host genes and intact LINE-1 in TCGA prostate cancer samples (x-axis) vs enrichment in our ORF1p IPs (y-axis). **d** Same x-axis as in **c**, but y-axis is p-body enrichment (same as x-axis in **a**)

results to a study of mRNAs that pellet under arsenite stressed and unstressed conditions and found correlations that were similar to those between ORF1p IP and stress granules RNAs: $\rho = 0.19$, $p = 1.2 \times 10^{-66}$ under unstressed conditions and $\rho = 0.18$, $p = 2.5 \times 10^{-56}$ under stress conditions. These pellets include stress granules and other RNAs, but not p-bodies [56]. These results indicate that ORF1p interacts with mRNAs that locate to p-bodies and that these interactions are more prevalent than interactions between ORF1p and mRNAs that localizes to stress granules, at least under unstressed conditions.

Because p-bodies are important regulators of mRNA processing, we hypothesized that LINE-1, or at least ORF1p expression, could have an impact on the abundance of certain mRNAs. To investigate this, we conducted an siRNA knockdown of LINE-1 and examined the mRNA levels of 12 genes enriched in our ORF1 IP and in p-bodies. Upon knockdown, we observed a reduction in mRNA levels in eleven out of twelve of the genes we measured (Figure S4). Furthermore, we used LIEM to quantify LINE-1 RNA in prostate tumor RNA-seq available through The Cancer Genome Atlas (TCGA). We found that genes significantly correlated with LINE-1 expression were more likely to be enriched in the ORF1p IP (odds ratio = 2.5, $p = 1.2 \times 10^{-73}$, Figure S3C), and that there is a positive relationship between correlation with LINE-1 RNA in TCGA and enrichment in our ORF1p IP experiments ($\rho = 0.39$, $p \approx 0$, Fig. 5c). More strikingly, most (72%) of the genes that were significantly correlated with LINE-1 expression were also enriched in p-bodies (odds ratio = 7.5, $p \approx 0$, Figure S3D), and there is an even stronger positive relationship between LINE-1 RNA expression in TCGA and enrichment in p-bodies.

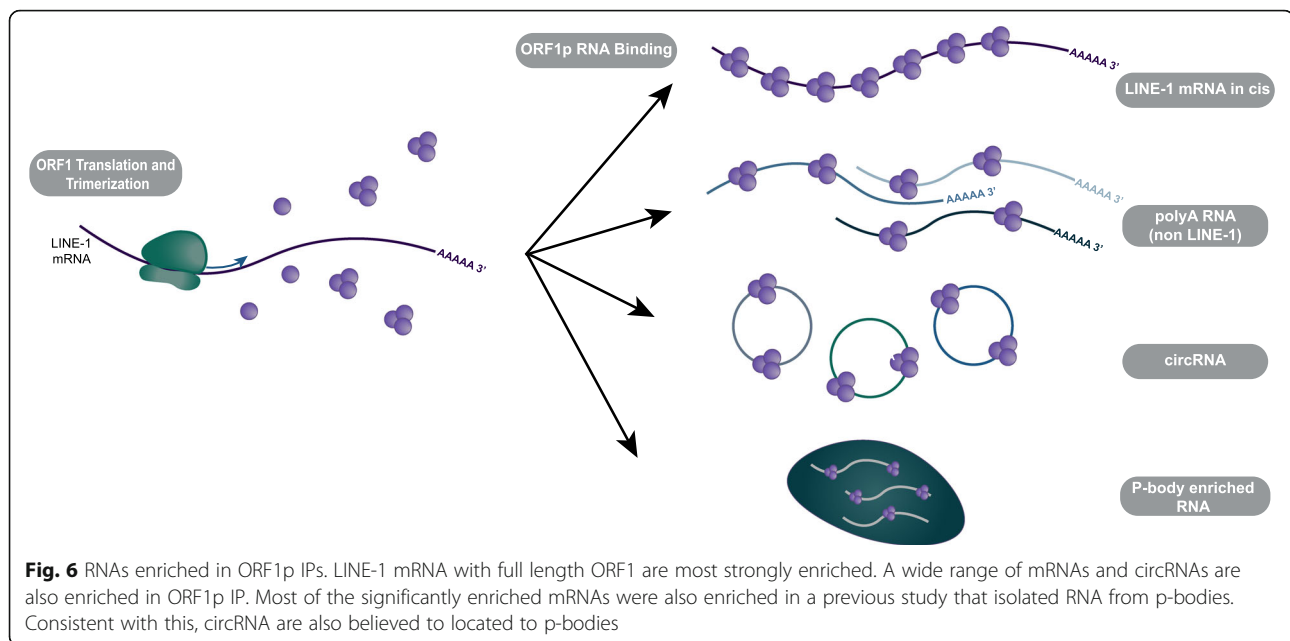
Discussion

The reactivation of LINE-1 expression in prostate cancer has raised new questions regarding its role in cancer initiation and progression. Sequencing studies have shown active retrotransposition in prostate tumors, and an increased rate of LINE-1 retrotransposition in metastatic prostate cancer compared to primary tumors [31]. However, the expression profiles of LINE-1 loci (mRNA) during prostate cancer progression has not been widely examined. We have previously shown an increase in ORF1p expression in androgen independent LNCaP cell lines, LNCaP-95 and LNCaP-abl, compared to parental LNCaP cells [23]. Here we show an overall increase in LINE-1 mRNA expression in both LNCaP-abl and LNCaP-95 cells compared to the parental LNCaP line. While LNCaP related cell lines (LNCaP, LNCaP-abl, LNCaP-95) show stronger similarities in LINE-1 loci expression when compared to non-related cell lines

(22Rv1), the profile of LINE-1 loci changes significantly in the androgen independent clones compared to androgen sensitive cells. Together, these findings may offer insight into the increased rate of retrotransposition in metastatic prostate cancer. Additionally, several of the loci expressed in these cell lines have generated new somatic insertions previously identified in actual human prostate tumors, indicating that at least some of the expressed loci are actually capable of retrotransposition.

Unexpectedly, we found that among all the RNA bound by ORF1p only a small percentage is represented by LINE-1 mRNA. While our analysis showed that our ORF1p IP was enriched for full length and complete ORF1p LINE-1 mRNA, indicating successful LINE-1 ORF1p immunoprecipitation, the vast majority of RNA enriched in our ORF1 IP consisted of non-LINE-1 transcripts (Fig. 6). Interestingly, mature exonic mRNAs were highly enriched in the ORF1-IP, indicating an enrichment of polyadenylated RNAs and generally speaking cytoplasmic as opposed to nuclear location. The lower levels of the non-polyadenylated histone mRNA in the ORF1 IP confirmed ORF1p's propensity to associate with polyA mRNA. ORF1p has been shown to bind nucleic acids in a sequence independent manner [18]. These findings make it unlikely that ORF1p is binding polyA sequences due to a sequence preference. However, we do not know the exact nature of the association between ORF1p and polyA mRNAs. As an RNA binding protein, ORF1p may be directly bound to some or all of these mRNAs. It may also associate with polyA mRNA through its interaction with other proteins. For example, polyadenylate-binding protein 1 (PABPC1), a polyA binding protein necessary for efficient retrotransposition, has also been shown to interact with ORF1p [57]. This interaction, as well as ORF1p's localization in the cytoplasm, may all increase its tendency to associate with polyA mRNAs.

We were particularly surprised to find MT (mitochondrial DNA encoded)-RNA enrichment in the crosslinked ORF1p IP experiments. MT-mRNAs are polyadenylated, and it is possible that ORF1p could bind them directly. While no study has shown ORF1p localizing in the mitochondria, ORF1p may bind MT-RNA that has been released from dysfunctional mitochondria. However, we did also find that MT-rRNAs are also enriched (although to a lesser degree). We considered whether this enrichment might be the result of MT-DNA contamination as mtDNA contamination can be amplified by the many MT genome copies that exist in each cell. Globally, we find that the vast majority of reads align to rRNA or exons, indicating that DNA contamination is not likely to be a significant factor in our analysis. However, MT specific DNA contamination is difficult to rule as most of MT genome is transcribed as a single transcript. A



third possibility is that crosslinking connects mitochondria to nearby ORF1p containing RNPs. Mitochondria are involved in interferon signaling [58] and RNA interference [59], both of which are involved in LINE-1 repression [60], and p-bodies have been shown to locate near mitochondria [59]. It is feasible that cytoplasmic ORF1p and LINE-1 RNPs may localize in close proximity to mitochondria and can be connected by crosslinking, leading to the extraction of entire mitochondria during the IP step.

Given strong evidence from previous studies indicating that ORF1p localizes to stress granules (SG), we expected to find a large overlap between ORF1p bound mRNAs and SG enriched mRNAs. Instead, we found only a small overlap that can be explained by the overlap between p-body and SG RNA. In contrast, we found a much larger overlap between ORF1p bound mRNAs and p-body enriched mRNAs compared to SG RNAs. This may indicate that endogenously expressed ORF1p localizes to p-bodies in LNCaP cells, or it may indicate that ORF1p promotes the mislocalization of p-body RNAs to sites of ORF1p accumulation. Alternatively, the high levels of L1 sequences may lead to the formation of a third type of body/granule with intermediate properties.

The discrepancy between our results and the previous studies pointing to SG as sites of LINE-1 proteins/RNA localization, may be explained by the fact that LNCaP cells were “unstressed” in our study. Goodier et al. (2007) did find endogenous ORF1p localizing with SG markers in embryonal carcinoma cell lines [22]. However, this localization was most pronounced after exogenous stress. It may be that LNCaP cells are able to tolerate endogenous LINE-1 expression with minimal

stress, leading to a lack of SG for ORF1p to localize to. Another major difference between our study and this previous study is that our study used RNA sequencing – whereas immunofluorescence and microscopy approaches were used by Goodier et al. Overall, LINE-1 may localize to SG or to SG-like granules, but, at least in LNCaP cells, the mRNAs in those granules may be more similar to p-body mRNAs than to the mRNAs present in canonical arsenite-induced SGs.

Whether ORF1p localizes to p-bodies or causes p-body RNA mislocalization, it may be interfering with the normal processing of p-body RNAs. Our results show that ORF1p-bound mRNAs expression correlates with LINE-1 expression in prostate cancer. Additionally, upon LINE-1 knockdown, we observed a decrease in p-body mRNA levels that were enriched in our ORF1IP. Thus, LINE-1 ORF1p expression may be interfering, directly or indirectly, with the processing/degradation of certain mRNAs. In particular, almost all of the genes that are significantly correlated with LINE-1 RNA in prostate cancer were also enriched in p-bodies, indicating that LINE-1 may interfere with the degradation of p-body associated RNAs. In yeast, p-body proteins are involved in the regulation of gene expression related to DNA replication stress resistance [61], possibly indicating a link between our findings and the previously documented relationship between LINE-1 and replication stress [39, 62]. Further studies are required to better understand the impact of ORF1p binding on mRNA processing and stability.

Conclusions

Our study, summarized in Fig. 6, finds that the increased LINE-1 expression in the androgen independent LNCaP-

95 and LNCaP-abl cells over the parental LNCaP is accompanied by a large-scale remodeling of the expressed LINE-1 loci. We also find that ORF1p associates not only with LINE-1 mRNA but also with a wide range of non-LINE-1 transcripts, particularly polyA RNAs. ORF1p bound RNA transcripts are enriched for p-bodies localized RNAs. Notably, the ORF1p-bound and p-bodies localized RNA species also correlate with LINE-1 expression in prostate cancer raising the intriguing possibility that cytoplasmic ORF1p may affect RNA processing in prostate cancer cells.

Methods

Cell culture

LNCaP (CRL-1740) and 22Rv1 (CRL-2505) cells were purchased from ATCC and maintained in RPMI 1640 with 10% FBS. LNCaP-abl, and LNCaP-95 cell lines were generous gifts from Z. Culig, and J. Isaacs, respectively and maintained in RPMI 1640, phenol red free, and 10% charcoal dextran stripped FBS. Cells were regularly screened for mycoplasma.

RNA Immunoprecipitation

For crosslinked samples, cells were kept on ice, washed twice with cold PBS, and UV-crosslinked using a Stratalinker 2400 with 150 mJ/cm² at 254 nm. Four 15 cm plates of each cell type, ~70% confluency, were used for each immunoprecipitation. Cells were washed with cold PBS and centrifuged at 1500 rpm (485 x g) for 5 min at 4 °C. RNA immunoprecipitation was conducted using the Magna RIP RNA-Binding Protein Immunoprecipitation Kit (Millipore Sigma 17–700) as follows. Cells were resuspended in 400 µl RIP lysis buffer supplemented with protease inhibitor cocktail and RNase inhibitor, and snap frozen to –80 °C. Magnetic beads were incubated with 10 µg ORF1p (Millipore Sigma MABC1152) or mouse IgG (Santa Cruz sc-2025) antibody per 50 µl of magnetic beads (200 µl beads used per condition). Cells were thawed and centrifuged at 14,000 rpm (18,407 x g) for 10 minutes at 4 °C. Supernatant was collected and mixed with RIP Immunoprecipitation buffer, supplemented with EDTA and RNase Inhibitor, and antibody bound magnetic beads. Bead/antibody complexes and cell lysates were rotated at 4 °C for either 1 h or 3 h. Beads were then washed 7x with 500 µl RIP Wash Buffer, and then incubated with Proteinase K Buffer for 30 min at 55 °C. Supernatant was separated from magnetic beads and RNA was isolated using a phenol chloroform extraction. Lastly, an on-column DNase digestion and RNA cleanup was performed using a RNeasy MinElute Cleanup Kit according to manufacturer's protocol (Qiagen 74,204, 79,254).

RNA library preparation and sequencing

Sequencing library was prepared using the NEBNext Ultra II RNA Library Prep Kit for Illumina (NEB E7770S) according to manufacturer's protocol, each sample prepared with its own unique barcode (NEB E7600S). Prepared libraries were sequenced as paired end 36-cycle reads (20 M) on the NextSeq 500. Reads were demultiplexed with Illumina bcl2fastq v2.20 requiring a perfect match to indexing BC sequences.

Alignment and LINE-1 RNA identification

Reads were aligned to the hg38 human reference genome using the STAR aligner and TCGA mRNA analysis pipeline options [63]. Locus specific LINE-1 RNA expression was estimated using LIEM [44]. ORFs >300aa were translated from Repeatmasker annotated LINE-1 (L1Hs, L1PA2, L1PA3, L1PA4) sequences in hg38 using ORFfinder (NCBI) and then aligned to ORF1p and ORF2p consensus sequences from Dfam using BLAST [64]. ORFs were considered intact if the alignment covered at least 95% of the consensus sequence. Intact LINE-1 and ORF1 expression were calculated by adding the estimated expression for all loci with intact ORFs.

Quantification of non-LINE-1 RNA

rRNA was quantified using samtools [65] to count reads overlapping rRNA genes annotated in the UCSC genome browser repeat track [66]. These reads were then filtered out and reads overlapping UCSC genome browser annotated exons were counted by samtools. Other repetitive RNAs were quantified using bedtools and the UCSC genome browser repeat track [66]. Circular RNAs were identified and quantified using CIRIquant [67]. Because no circRNA was supported by a large number of reads, all circRNA reads were pooled into a single quantification. Reads aligning to each gene in GRCh38.96 were counted using featureCounts [68]. The LNCipedia database was used with featureCounts to quantify lncRNA reads [68].

ORF1p IP, p-body, SG and TCGA comparisons

ORF1p IP enrichment for each gene was estimated using DESeq2 [69]. Note that DESeq2 normalizes expression to exon aligned reads, rather than all aligned reads leading to the difference in scale for Fig. 5 compared to Figs. 3 and 4. P-body enrichments were obtained from Hubstenberger et al. [55] and SG enrichments were obtained from Khong et al. [54] LINE-1 expression was estimated in TCGA prostate cancers using LIEM as implemented on the Cancer Genome Cloud (CGC) [70]. Spearman correlation and significance between LINE-1 RNA and upper quantile normalized expression for each gene (available from the Genomic Data Commons (GDC)) was calculated using in R.

siRNA knockdown

7×10^5 LNCaP cells were treated twice with siRNA for 48 h using Lipofectamine RNAiMAX (Invitrogen; 13,778,030). LINE-1 siRNA contained a pool of siRNA (Silencer Select AAGCAAUGUUGAGAGAUUtt and Stealth GAAAuGAAGCGAGAAGGGAAGuuuA) custom made by Thermo Fisher Scientific (4,399,666, 10,650,006). Protein was collected by lysing cells in RIPA buffer ((50 mM Tris pH 8, 150 mM NaCl, 1% NP-40, 0.1% SDS, 10 mM EDTA, 10 μ g/mL aprotinin and leup-tin, 0.1 mM PMSF, and 0.1 mM Na_3VO_4) for 10 minutes, spun at 14,000 rpm for 5 min, and normalized using a Bradford Assay (Bio-Rad 5,000,006).

qPCR

RNA was collected using the Qiagen RNeasy Plus Mini Kit (74134) followed by a DNase digestion (Thermo Fisher Scientific AM1907) according to manufacturer's protocol. RNA was reverse transcribed using the Verso cDNA kit (Thermo Scientific- AB1453A). qPCR was conducted using SYBR Green Master Mix (Life Technologies 4,344,463) with gene specific primers (Table S3). LINE-1 primers have been previously published [71]. Relative mRNA levels were calculated using $\Delta\Delta\text{CT}$ with RPL19 as an internal control.

Western blot

Western blot was conducted as previously described [23]. Primary antibodies include ORF1 (EMD Millipore MABC1152), and Tubulin (Covance MMS-489P).

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13100-021-00233-3>.

Additional file 1: Figure S1. (A) LINE-1 RNA in the input for each cell LINE-1. (B) LINE-1 RNA after ORF1p-IP enrichment.

Additional file 2: Figure S2. Heatmap of the specific intact loci immunoprecipitated in each cell line. Expansion of Fig. 1b to include the top 25 loci.

Additional file 3: Figure S3. Overlap between transcripts enriched in ORF1p IP and in cytoplasmic granules. (A) Overlap between ORF1p IP enrichment and p-body enrichment [55]. A 5% FDR cutoff was used. (B) As in A, but comparing ORF1p IP to SG enrichment [54]. (C) Overlap between transcripts enriched in ORF1p IP and those whose expression is positively correlated with LINE-1 RNA in TCGA prostate cancer samples (i.e. these genes are more highly expressed in tumors that express more LINE-1). (D) Overlap between transcripts enriched in p-bodies and those correlated with LINE-1 RNA in TCGA prostate cancer.

Additional file 4: Figure S4. qPCR quantification of p-body and ORF1 IP RNA after siRNA knockdown of LINE-1. (A) qPCR of top ORF1 IP and p-body genes after 48 h siRNA LINE-1 knockdown. Normalized using RPL19. (B) Western blot of 48 h siRNA LINE-1 knockdown used for qPCR in (A). Demonstrates knockdown of ORF1p protein.

Additional file 5: Table S1. L1EM estimates of locus specific LINE-1 expression for each of the samples.

Additional file 6: Table S2. Enrichment of genes in ORF1p IP vs input estimated by DESeq2.

Additional file 7: Table S3. Primers used in qPCR experiments.

Abbreviations

LINE-1: Long Interspersed Element-1; L1EM: LINE-1 Expectation Maximization; SG: Stress granule; P-body: Processing body; IP: Immunoprecipitation; RNP: Ribonucleoprotein

Acknowledgements

Sequencing was performed by the NYU Langone Institute for Systems Genetics. We would like to thank Megan Hogan and Raven Luther for their technical assistance and expertise.

Conflict of interest

The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Authors' contributions

Experiments were performed by EMB with help from PM. Data and statistical analyses were done by WM and DF. Experiments were planned and analyzed by EMB, WM, PM, SKL, and DF. Experiments were supervised by SL, JDB and DF. WM, EMW, SKL and DF wrote the manuscript. The authors read and approved the final manuscript.

Funding

This work was supported by the National Institutes of Health Grants R01CA112226 (to S. K. L.), 1F31CA225053-01A1 (to E.M.B.) and P01AG051449 (subcontract to J.D.B. and D.F.). This project has been funded in part with Federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. HHSN261200800001E (subcontract to W.M.).

Availability of data and materials

Raw Illumina sequencing reads are available from the SRA database under bioproject PRJNA643489.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

David Fenyő is a Founder and President of The Informatics Factory, and serves or served on the Scientific Advisory Board or consults for: Spectragen Informatics, Protein Metrics, Preverna. Jef Boeke is a Founder and Director of CDI Labs, Inc., a Founder of Neochromosome, Inc, a Founder and SAB member of ReOpen Diagnostics, and serves or served on the Scientific Advisory Board of the following: Sangamo, Inc., Modern Meadow, Inc., Sample6, Inc. and the Wyss Institute. John Sedivy is a cofounder of Transposon Therapeutics, Inc., serves as Chair of its Scientific Advisory Board, and consults for Astellas Innovation Management LLC, Atropos Therapeutics, Inc. and Gilead Sciences, Inc.

Author details

¹Departments of Biochemistry and Molecular Pharmacology, New York University, 450 East 29th Street, Room 321, New York, NY 10016, USA.

²Institute for Systems Genetics, New York University Grossman School of Medicine, 435 East 30th St, 9th Floor, NY 10016 New York, USA. ³Urology, New York University Grossman School of Medicine, 450 East 29th Street, Room 321, New York, NY 10016, USA.

Received: 24 August 2020 Accepted: 27 January 2021

Published online: 09 February 2021

References

1. Konkel MK, Walker JA, Batzer MA. LINEs and SINEs of primate evolution. *Evol Anthropol.* 2010;19(6):236–49.

2. Han JS, Szak ST, Boeke JD. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature*. 2004;429(6989):268–74.
3. Lander ES, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409(6822):860–921.
4. Brouha B, et al. Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci U S A*. 2003;100(9):5280–5.
5. Philippe C, et al. Activation of individual L1 retrotransposon instances is restricted to cell-type dependent permissive loci. *Elife*. 2016;5:e13926.
6. Scott AF, et al. Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. *Genomics*. 1987;1(2):113–25.
7. Dai L, et al. Expression and detection of LINE-1 ORF-encoded proteins. *Mob Genet Elements*. 2014;4:e29319.
8. Taylor MS, et al. Affinity proteomics reveals human host factors implicated in discrete stages of LINE-1 retrotransposition. *Cell*. 2013;155(5):1034–48.
9. Mita P, et al. LINE-1 protein localization and functional dynamics during the cell cycle. *Elife*. 2018;7:e30058.
10. Feng Q, et al. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell*. 1996;87(5):905–16.
11. Cost GJ, et al. Human L1 element target-primed reverse transcription in vitro. *EMBO J*. 2002;21(21):5899–910.
12. Esnault C, Maestre J, Heidmann T. Human LINE retrotransposons generate processed pseudogenes. *Nat Genet*. 2000;24(4):363–7.
13. Dewannieux M, Esnault C, Heidmann T. LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet*. 2003;35(1):41–8.
14. Wei W, et al. Human L1 retrotransposition: cis preference versus trans complementation. *Mol Cell Biol*. 2001;21(4):1429–39.
15. Moran JV, et al. High frequency retrotransposition in cultured mammalian cells. *Cell*. 1996;87(5):917–27.
16. Khazina E, et al. Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition. *Nat Struct Mol Biol*. 2011;18(9):1006–14.
17. Khazina E, Weichenrieder O. Human LINE-1 retrotransposition requires a metastable coiled coil and a positively charged N-terminus in L1ORF1p. *Elife*. 2018;7:e34960.
18. Martin SL. The ORF1 protein encoded by LINE-1: structure and function during L1 retrotransposition. *J Biomed Biotechnol*. 2006;2006(1):45621.
19. Kolosha VO, Martin SL. In vitro properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition. *Proc Natl Acad Sci U S A*. 1997;94(19):10155–60.
20. Evans JD, et al. Paired mutations abolish and restore the balanced annealing and melting activities of ORF1p that are required for LINE-1 retrotransposition. *Nucleic Acids Res*. 2011;39(13):5611–21.
21. Wallace N, et al. LINE-1 ORF1 protein enhances Alu SINE retrotransposition. *Gene*. 2008;419(1–2):1–6.
22. Goodier JL, et al. LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex. *Mol Cell Biol*. 2007;27(18):6469–83.
23. Briggs EM, et al. Long interspersed nuclear element-1 expression and retrotransposition in prostate cancer cells. *Mob DNA*. 2018;9:1.
24. Doucet AJ, et al. Characterization of LINE-1 ribonucleoprotein particles. *PLoS Genet*. 2010;6(10):e1001150.
25. Anderson P, Kedersha N, Ivanov P. Stress granules, P-bodies and cancer. *Biochim Biophys Acta*. 2015;1849(7):861–70.
26. Decker CJ, Parker R. P-bodies and stress granules: possible roles in the control of translation and mRNA degradation. *Cold Spring Harb Perspect Biol*. 2012;4(9):a012286.
27. Ivanov P, Kedersha N, Anderson P. Stress Granules and Processing Bodies in Translational Control. *Cold Spring Harb Perspect Biol*. 2019;11(5):a032813.
28. Ardeljan D, et al. LINE-1 ORF2p expression is nearly imperceptible in human cancers. *Mob DNA*. 2020;11:1.
29. Alish RS, et al. Unconventional translation of mammalian LINE-1 retrotransposons. *Genes Dev*. 2006;20(2):210–24.
30. Tang Z, et al. Human transposon insertion profiling: analysis, visualization and identification of somatic LINE-1 insertions in ovarian cancer. *Proc Natl Acad Sci U S A*. 2017;114(5):E733–40.
31. Tubio JM, et al. Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science*. 2014;345(6196):1251343.
32. Chen L, et al. Naturally occurring endo-siRNA silences LINE-1 retrotransposons in human cells through DNA methylation. *Epigenetics*. 2012;7(7):758–71.
33. Goodier JL, Cheung LE, Kazazian HH Jr. MOV10 RNA helicase is a potent inhibitor of retrotransposition in cells. *PLoS Genet*. 2012;8(10):e1002941.
34. Yoder JA, Walsh CP, Bestor TH. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet*. 1997;13(8):335–40.
35. Brennecke J, et al. An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science*. 2008;322(5906):1387–92.
36. Rodic N, et al. Long interspersed element-1 protein expression is a hallmark of many human cancers. *Am J Pathol*. 2014;184(5):1280–6.
37. Sun X, et al. Transcription factor profiling reveals molecular choreography and key regulators of human retrotransposon expression. *Proc Natl Acad Sci U S A*. 2018;115(24):E5526–35.
38. Mckerrow W, Wang X, Mita P, Cao S, Grivainis M, Ding L, LaCava J, Boeke J, Fenyo D. LINE-1 expression in cancer correlates with DNA damage response, copy number variation, and cell cycle progression. *BioRxiv*. 2020; 26:174052.
39. Ardeljan D, et al. Cell fitness screens reveal a conflict between LINE-1 retrotransposition and DNA replication. *Nat Struct Mol Biol*. 2020;27(2):168–78.
40. Harris CR, et al. Association of nuclear localization of a long interspersed nuclear element-1 protein in breast tumors with poor prognostic outcomes. *Genes Cancer*. 2010;1(2):115–24.
41. Ting DT, et al. Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. *Science*. 2011;331(6017):593–6.
42. Steranka JP, et al. Transposon insertion profiling by sequencing (TIPseq) for mapping LINE-1 insertions in the human genome. *Mob DNA*. 2019;10:8.
43. Lee E, et al. Landscape of somatic retrotransposition in human cancers. *Science*. 2012;337(6097):967–71.
44. Mckerrow W, Fenyo D. L1EM: a tool for accurate locus specific LINE-1 RNA quantification. *Bioinformatics*. 2020;36(4):1167–73.
45. Rodriguez-Martin B, et al. Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. *Nat Genet*. 2020;52(3):306–19.
46. Mandal PK, et al. Enrichment of processed pseudogene transcripts in L1-ribonucleoprotein particles. *Hum Mol Genet*. 2013;22(18):3730–48.
47. Lasda E, Parker R. Circular RNAs: diversity of form and function. *RNA*. 2014; 20(12):1829–42.
48. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–50.
49. Croft D, et al. Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res*. 2011;39(Database issue):D691–7.
50. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27–30.
51. The Gene Ontology, C. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res*. 2019;47(D1):D330–8.
52. Zhang Z, Harrison P, Gerstein M. Identification and analysis of over 2000 ribosomal protein pseudogenes in the human genome. *Genome Res*. 2002; 12(10):1466–82.
53. Zhang Z, et al. PseudoPipe: an automated pseudogene identification pipeline. *Bioinformatics*. 2006;22(12):1437–9.
54. Khong A, et al. The stress granule Transcriptome reveals principles of mRNA accumulation in stress Granules. *Mol Cell*. 2017;68(4):808–20 e5.
55. Hubstenberger A, et al. P-body purification reveals the condensation of repressed mRNA regulons. *Mol Cell*. 2017;68(1):144–57 e5.
56. Matheny T, Rao BS, Parker R. Transcriptome-Wide Comparison of Stress Granules and P-Bodies Reveals that Translation Plays a Major Role in RNA Partitioning. *Mol Cell Biol*. 2019;39(24):e00313–19.
57. Dai L, et al. Poly(a) binding protein C1 is essential for efficient L1 retrotransposition and affects L1 RNP formation. *Mol Cell Biol*. 2012;32(21): 4323–36.
58. Vazquez C, Horner SM. MAVS coordination of antiviral innate immunity. *J Virol*. 2015;89(14):6974–7.
59. Huang L, et al. Mitochondria associate with P-bodies and modulate microRNA-mediated RNA interference. *J Biol Chem*. 2011;286(27):24219–30.
60. Pizarro JG, Cristofari G. Post-transcriptional control of LINE-1 Retrotransposition by cellular host factors in somatic cells. *Front Cell Dev Biol*. 2016;4:14.
61. Loll-Krippelber R, Brown GW. P-body proteins regulate transcriptional rewiring to promote DNA replication stress resistance. *Nat Commun*. 2017; 8(1):558.
62. Mita P, et al. BRCA1 and S phase DNA repair pathways restrict LINE-1 retrotransposition in human cells. *Nat Struct Mol Biol*. 2020;27(2):179–91.

63. Dobin A, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21.
64. Camacho C, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10:421.
65. Li H, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9.
66. Karolchik D, et al. The UCSC genome browser database. *Nucleic Acids Res*. 2003;31(1):51–4.
67. Zhang J, et al. Accurate quantification of circular RNAs identifies extensive circular isoform switching events. *Nat Commun*. 2020;11(1):90.
68. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923–30.
69. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
70. Cancer Genome Atlas Research, N. The Molecular Taxonomy of Primary Prostate Cancer. *Cell*. 2015;163(4):1011–25.
71. Goering W, Ribarska T, Schulz WA. Selective changes of retroelement expression in human prostate cancer. *Carcinogenesis*. 2011;32(10):1484–92.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

