


Review

Elucidating the Role of Host Long Non-Coding RNA during Viral Infection: Challenges and Paths Forward

David J. Lemler ¹ , Hayden N. Brochu ^{1,2}, Fang Yang ¹, Erin A. Harrell ¹ and Xinxia Peng ^{1,2,3,*}

¹ Department of Molecular Biomedical Sciences, College of Veterinary Medicine, North Carolina State University, Raleigh, NC 27607, USA; djlemle@ncsu.edu (D.J.L.); hnathan@ncsu.edu (H.N.B.); fyang6@ncsu.edu (F.Y.); eaegan@ncsu.edu (E.A.H.)

² Bioinformatics Graduate Program, North Carolina State University, Raleigh, NC 27695, USA

³ Bioinformatics Research Center, North Carolina State University, Raleigh, NC 27695, USA

* Correspondence: xpeng5@ncsu.edu; Tel.: +1-919-515-4481

Academic Editors: Angela L. Rasmussen and Juliet Morrison

Received: 26 September 2017; Accepted: 17 October 2017; Published: 20 October 2017

Abstract: Research over the past decade has clearly shown that long non-coding RNAs (lncRNAs) are functional. Many lncRNAs can be related to immunity and the host response to viral infection, but their specific functions remain largely elusive. The vast majority of lncRNAs are annotated with extremely limited knowledge and tend to be expressed at low levels, making ad hoc experimentation difficult. Changes to lncRNA expression during infection can be systematically profiled using deep sequencing; however, this often produces an intractable number of candidate lncRNAs, leaving no clear path forward. For these reasons, it is especially important to prioritize lncRNAs into high-confidence “hits” by utilizing multiple methodologies. Large scale perturbation studies may be used to screen lncRNAs involved in phenotypes of interest, such as resistance to viral infection. Single cell transcriptome sequencing quantifies cell-type specific lncRNAs that are less abundant in a mixture. When coupled with iterative experimental validations, new computational strategies for efficiently integrating orthogonal high-throughput data will likely be the driver for elucidating the functional role of lncRNAs during viral infection. This review highlights new high-throughput technologies and discusses the potential for integrative computational analysis to streamline the identification of infection-related lncRNAs and unveil novel targets for antiviral therapeutics.

Keywords: lncRNA; next-generation sequencing; high-throughput analysis; transcriptomics; RNA-seq

1. Introduction

Viral infections remain a major world-wide concern. Even though we have all but eradicated some once deadly viruses, many still elude effective treatment. HIV-1 is becoming a chronic disease since current anti-retroviral therapy can suppress the infection but cannot clear the virus. Influenza is still a major global concern that evolves rapidly and kills thousands of individuals every year. As shown by recent examples, like Zika and Middle East Respiratory Syndrome Coronavirus (MERS-CoV), the threat of emerging viral infections is constant. A better understanding of virus and host interactions is needed to accurately define viral pathogenesis and to rapidly develop new therapies.

Long non-coding RNAs (lncRNAs), a new class of transcripts, have recently garnered interest in the field of infection, as studies of the host response to viral infection typically focus on protein-coding genes. lncRNAs are defined as RNAs greater than 200 nucleotides with insignificant coding potential, i.e., noncoding, but they can have very diverse regulatory functions ranging from active transcriptional regulation to epigenetics. Compared to the small number of proteins that are druggable, a large number of lncRNAs offer many new potential targets as they may be easily and accurately targeted using sequence specific oligonucleotides, and they are more cell type- and tissue- specific than coding

genes in general. We and others have documented the association of lncRNAs in viral infections and innate immunity [1–5]. The functions of numerous lncRNAs in host immune responses have also been extensively reviewed [6–8]. However, most lncRNAs, including those associated with viral infections, still lack detailed functional characterization.

First, we summarize our current understanding of the diverse functions of lncRNAs and their potential relevance to viral infection (Section 2). The analysis of lncRNA functions in viral infections comes with unique challenges, including their overwhelmingly low expression abundance, lack of sequence conservation, and limited annotation. To better address these challenges, we broadly outline the key steps to be taken to investigate infection-associated lncRNAs: discovery (Section 3.1), prioritization (Sections 3.2–3.4), and validation (Section 3.5). We discuss strategies, including existing and emerging technologies, for addressing challenges that researchers may encounter during each step. We conclude with suggestions for new developments needed to rapidly identify infection-related lncRNAs and potentially novel anti-viral targets. While our primary interest is viral infection, the general ideas we present here can also facilitate the studies of lncRNAs in other areas.

2. Functional Diversity of lncRNAs and Their Involvement in Viral Infections

lncRNAs are generated using the same machinery and components as mRNA (and other ncRNA), may possess 5' caps, and may also be polyadenylated at their 3' ends. Their main distinction from mRNA is the lack of a significant open reading frame. Unlike coding genes, lncRNAs tend to be poorly conserved at the primary sequence level [9,10]. lncRNAs are also known to have lower and more tissue- or cell-type specific expression than coding genes [11]. lncRNAs permeate the human genome and interdigitate coding regions, but are often dismissed as transcriptional noise due to low expression levels and the lack of primary sequence conservation. Expressed lncRNA genes have typical histone modifications, exhibit canonical splice site signals, and can produce alternative transcripts [11,12]. They localize both in the cytoplasm and the nucleus, suggesting roles in the epigenetic modification of chromatin and regulation of gene expression. Their structural architecture has been described as modular and multi-domained, allowing lncRNAs to form conformational switches and simultaneously interact with mRNAs, DNA, and proteins [13]. The functional roles of these RNAs are diverse and, due to their typically low sequence conservation, are believed to rely on secondary and tertiary structures.

Much of the initial interest in lncRNAs stems from the discovery of Xist and its involvement in X chromosome inactivation (XCI), a process that has been detailed in a number of extensive reviews over the years [14–17]. Though the majority of lncRNAs were considered “junk RNA” and transcriptional noise, lncRNAs have now been shown to regulate methylation [18,19], chromatin remodeling [20], and alternative splicing [21], as well as imprinting and dosage compensation [17]. Despite the lack of functional characterization of most lncRNAs, many have been functionally implicated in viral infections [3]. Of the many functions described above, several are known to be exhibited by lncRNAs associated with viral infections, e.g., paraspeckle formation [4] and pre-mRNA splicing [22]. Such lncRNAs are often involved in the host immune response, but many are also virally-encoded. In Table 1, we list several better characterized lncRNAs that are associated with commonly studied viruses, including influenza and herpes. Recent reviews by Ding et al. [23], Mumtaz et al. [6], Valadkhan et al. [7], and Liu et al. [8] highlight many of the extensively annotated host- and virally-encoded lncRNAs.

Table 1. LncRNAs involved in viral infection.

lncRNA	Encoding Organism	General Function	Specific Function	Infection Type	ID Method	Citation
NEAT1	Host	Scaffold	Nuclear localization, paraspeckle formation	HIV-1, HTNV	1 of 83 lncRNAs profiled in HIV-1-infected Jurkat and MT4 cells	[2,4,24]
NRON	Host	Scaffold	Latency via inhibition of NFAT nuclear translocation	HIV-1	1 of 90 lncRNAs profiled in two human T cell lines	[25,26]
Tmevpg1 (NeST, IfngAS1)	Host	Epigenetics	IFN-gamma-mediated regulation of adaptive immunity	Theiler's murine encephalo myelitis (TMEV)	Candidate gene from Tmevp3 locus	[27] and reviews by [6,23]
NRAV	Host	Epigenetics	Modulates transcription of ISGs, promotes IAV replication	Influenza A Virus (IAV)	1 of 907 differentially expressed lncRNAs from microarray analysis	[28] and reviews by [6,8]
HIV-expressed antisense lncRNA (ASP-L)	Pathogen	Epigenetics	Epigenetic transcriptional regulation	HIV-1	qPCR	[29]
PAN RNA	Pathogen	Epigenetics	Required for KSHV gene expression, repression of IFN-alpha, IFN-gamma, ISGs	Kaposi's Sarcoma-associated Herpes-virus (KSHV)	Northern Blot	[30] and reviews by [7,8]

2.1. Epigenetic Regulation and Promotion of Viral Latency

LncRNAs have been established as significant players in epigenetic regulation. Several studies have characterized associations between lncRNAs and DNA methyltransferases (DNMTs): DNMT1 [31,32], DNMT3a [33], and DNMT3b [34]. The interactions between lncRNAs, various histone modifiers, and chromatin remodeling complexes have been extensively reviewed, among which, the HOXA and HOXC clusters are particularly well-studied [35,36]. The biological functionality of lncRNAs from the HOXA and other loci as recruiting factors of major complexes is well-described [37–39]. A comprehensive review by Betancur et al. [20] underscores the pervasive interactions between lncRNAs and complexes such as PRC1, PRC2, MLL1, and BAF. The regulation of expression via chromatin looping and nucleosome positioning is also an established function of lncRNAs [12]. LncRNA methylation functionality and their involvement in developmental processes and disease are highlighted in several reviews [18,19,40].

Several host- and virally-encoded lncRNAs have also been shown to function epigenetically. For example, PAN RNA (Table 1) has been implicated in the transcriptional inhibition of the IFN cascade and is known to alter gene expression by adding or removing H3K27me3 marks through interaction with the PRC2 complex and histone methyltransferase MLL2 [8]. Virally-encoded lncRNAs have also been shown to interact with DNMTs and chromatin remodeling complexes. Rossetto et al. demonstrated that a human cytomegalovirus (HCMV) lncRNA interacts with Polycomb Repressive Complex 2 (PRC2) during latency to inhibit host transcription [41]. Saayman et al. [29] identified an HIV-1-encoded antisense lncRNA (Table 1) that regulates viral transcription via direct interaction with a chromatin remodeling complex consisting of DNMT3a, EZH2, and HDAC-1. The HIV-1-encoded antisense lncRNA also interacts with PRC2, inhibiting viral transcription and affecting nucleosome assembly [42]. Both of these viral lncRNAs have been implicated in the promotion and maintenance of early latency of their respective viral infections [41,42]. Moving forward, the extent of the relationship between epigenetic regulation and viral latency resulting from lncRNA expression needs to be further investigated.

The role of lncRNAs in epigenetic regulation has also been described through the concept of extra-coding RNA (ecRNA). The term was first introduced by Di Ruscio et al. [32] upon the identification of a ncRNA (ecCEBPA) in humans that contained the entire CEBPA pre-mRNA sequence along with “extra” up- and down-stream regions. ecCEBPA blocks the methylation of the CEBPA promoter region via a direct interaction with DNA methyltransferases (DNMTs). The expression of ecCEBPA was later associated with neuronal gene promoter methylation [33]. While the validity of this class of RNA is still open for debate due to similarities with pre-mRNA, their potential involvement with the host immune response might be of great interest.

2.2. Scaffolding and Nuclear Localization

Biological scaffolding is a function typically reserved for proteins; however, it has recently been identified as a function of lncRNAs. In plants, RNA polymerase V-dependent lncRNAs are known to recruit AGO4 in the canonical RNA-directed DNA methylation (RdDM) pathway [43]. The functional role of lncRNAs has been expanded to include stepwise binding to IDN2 and DRM2 in addition to AGO4, characterizing lncRNAs as a lynchpin of the RdDM pathway [12]. The function of lncRNAs as scaffolds is also described in the formation of oddly-formed nuclear compartments in mammalian cells, known as paraspeckles. Two examples are the HIV-associated lncRNAs NEAT1 (Table 1) and MALAT1. While NEAT1 has been established as an essential structural scaffold in the maintenance of nuclear paraspeckles, MALAT1 is not required [44]. The role of these lncRNAs in paraspeckle formation has been extensively reviewed [44–47]. The knockdown of NEAT1 in HIV-1-infected T cells demonstrated that the maintenance of nuclear paraspeckles by NEAT1 is associated with HIV-1 replication [2]. More recently, NEAT1 was shown to promote IFN production in response to Hantaan virus (HTNV) infection by localizing SFPQ (splicing factor proline- and glutamine-rich protein) to nuclear paraspeckles [24]. In vitro and in vivo inhibition of NEAT1 expression suppressed

the host immune response. Conversely, ectopic expression increased IFN- β production and inhibited HTNV replication. These results indicate that NEAT1 can modulate the host immune response by the localization of RNA-binding proteins to paraspeckles. Whether or not NEAT1 exhibits this functionality in response to other viral infections needs to be investigated.

2.3. Transcriptional Regulation of mRNA via miRNA Sponges

LncRNAs can also act as sponges for miRNA and prevent the miRNA-mediated degradation of target transcripts. The mode of this function has been described through the competing endogenous RNA (ceRNA) hypothesis and has been extensively reviewed [48,49]. LncRNAs have been functionally described as ceRNAs in the progression of gastric cancer [50] and hepatocellular carcinoma [51]. The involvement of ceRNAs, such as lincROR, HOTAIR, and BARD1 9'L, in the initiation and progression of various cancers has also been well-documented and reviewed [52,53]. There is also evidence that virally-encoded lncRNAs function as miRNA sponges. When studying Herpesvirus saimiri (HVS), Cazalla et al. [54] identified sequence complementarity between HVS-encoded lncRNAs (HSUR1 and HSUR2) and host miRNAs expressed in T cells, and confirmed the downregulation of miRNA-27 as a result. Guo et al. [55] recently analyzed the functional roles of miRNAs in HVS-transformed T cells, revealing that miRNA-27 directly downregulates several proteins in the T cell receptor signaling pathway. A review by Tavanez et al. [56] further describes the function of HSUR1 and HSUR2 and highlights the susceptibility of host miRNAs in viral infections. This interplay of virally-encoded lncRNAs and host miRNAs may prove to be an essential function of viral pathogenesis.

2.4. Alternative Splicing

The notion that lncRNA can affect alternative splicing stems from the fact that these RNAs are often found in complex loci that also contain protein-coding genes. A recent study has shown that the highly conserved human h5S-OT lncRNA expressed within the 5S rDNA locus can impact alternative splicing by interacting with U2AF65, a core splicing factor that binds to intron-exon junctions [57]. An anti-Alu element in the 3' end mimics the polypyrimidine tract necessary for the recruitment of splicing factors, resulting in the targeting of introns containing Alu elements and downstream exon inclusion [57]. Previously associated with HIV-1 infection [3], MALAT1 has been shown to regulate trans-acting pre-mRNA splicing factors in cancer cells, specifically those in the SR protein family [22]. Extensive reviews on MALAT1 [58], RNA-guided mechanisms in alternative splicing [21], and several other disease-related lncRNAs experimentally linked to alternative splicing [59,60] are available for further reading.

3. Discovery, Prioritization, and Validation of lncRNAs

In the previous section, we discussed some of the characteristics and functions of lncRNAs. Due to their unique nature, methodologies developed for the interrogation of coding genes may not be applicable or may require modification to adequately investigate lncRNAs. This section will discuss selected methods for the discovery, prioritization, and validation of lncRNAs. We begin with a discussion of the modifications to RNA-seq that allow for the detection of lncRNAs (Section 3.1), which represents the discovery phase for determining the association of lncRNAs with viral infections. The prioritization phase focuses on computational methods for ranking lncRNAs based on differential expression and genomic context (Section 3.2). Section 3.3 delves further into the genomic context of lncRNAs by incorporating evolutionary analyses to aid in the identification of function. An alternative or complementary strategy for identifying lncRNA function employs large scale in vitro screening studies (Section 3.4). These screens are capable of directly associating genes with the phenotype of interest while providing support for computationally derived results. Finally, Section 3.5 highlights important considerations for validation that arise as a result of the unique features of lncRNAs. Together, these methods represent a comprehensive strategy for identifying the involvement of novel lncRNAs in viral infection while simultaneously utilizing all available data to ascertain function.

3.1. Discovering Viral Infection-Related lncRNAs: The Different Flavors of Transcriptome Deep Sequencing

During viral infection there are a multitude of changes throughout the transcriptome, which includes lncRNAs. Therefore, a natural choice for discovering specific lncRNAs important for infections is to systematically profile lncRNA changes in response to infection. While multiple methods for the detection of lncRNAs are currently available, as highlighted in [61], transcriptome deep sequencing (RNA-seq) is likely the most preferred and widely used, in part due to its broad applicability and accessibility. More importantly, RNA-seq directly sequences actual transcripts, regardless of whether or not the reference genomic sequences exist [62]. This is especially relevant as many viral infections occur in non-model organisms that do not have fully sequenced genomes or well-annotated with lncRNAs. Though RNA-seq is available as a standard service from core facilities, there are still specific considerations in terms of studying lncRNAs.

3.1.1. Total RNA vs. mRNA

Standard RNA-seq focuses on mRNAs (mRNA-seq) by enriching transcripts with poly(A) tails using approaches like poly(dT) coated magnetic beads. This would be less desirable for lncRNA analysis as many lncRNAs lack poly(A) tails. Alternatively, the expression of both poly(A+) (mostly mature transcripts of coding genes) and poly(A-) (mostly non-coding RNAs) can be captured by the sequencing of total RNAs (Total RNA-seq), as shown by our studies [1,3] and others. Limitations to Total RNA-seq, on the other hand, include the requirement of millions of additional sequencing reads to achieve a comparable coverage as mRNA-seq, and the difficulties in quantifying individual isoforms due to the coverage of immature pre-mRNAs. In spite of this, we have used this technique to identify non-coding RNAs differentially expressed during HIV-1 [3], SARS coronavirus [1], and influenza infections [63]. Further, total RNAs may be separated into mRNA enriched and mRNA depleted fractions [32], allowing for coding and noncoding transcripts to be analyzed independently. In any case, special attention should be paid to the removal of extremely abundantly transcripts like ribosomal RNAs in total RNAs and hemoglobin transcripts in whole blood samples.

lncRNAs are frequently expressed at very low levels. Total RNA-seq or the sequencing of mRNA depleted RNAs may still be insufficient for the detection of lowly expressed lncRNAs [61,64]. Instead, RNA capture sequencing (CaptureSeq, CAP-seq) uses custom-designed, hybridization-based oligonucleotide probes to capture and enrich genes or regions of interest [64]. The probes are designed to target the transcripts of interest, while untargeted species are washed away [64]. The enriched fraction is subjected to sequencing, resulting in greatly increased target coverage [64]. When sufficient reference sequence information is available for probe design, this technique may ease detailed transcript assembly and abundance quantification [64]. In particular, this method might be effective for profiling a list of target lncRNAs across a large number of samples.

3.1.2. Computational Considerations for Identifying lncRNAs from RNA-seq Data

Like any other study, the described RNA-seq analysis requires necessary computational infrastructure and bioinformatics expertise for data storage, processing, and statistical analysis. Ideally, this is planned out before the start of the study, and may be accomplished in-house, in-collaboration, or through core services. Particularly, the annotation of lncRNA genes and its completeness should be examined carefully. For viral infections in well-studied systems like humans or mice, there are large numbers of annotated lncRNAs. For example, there are 14,720 human lncRNA genes and 8980 mouse lncRNA genes in the current Ensembl annotation (release 90.38). In these cases, lncRNA expression analysis can be carried out simultaneously just as all other coding genes. However, after quality control, it would be very beneficial to align cleaned sequencing reads to viral reference sequences to verify the identity of infected samples and the percentage of total reads mapped to viruses.

For other organisms with draft genome assemblies, it is likely that the annotation of lncRNA genes is largely incomplete. Therefore, it would be important to annotate lncRNA transcripts *ab initio*,

using the collected RNA-seq data in combination with other available sequencing data. For example, we were able to expand the ferret genome annotation with about 40,000 intergenic loci, which were enriched with polyadenylated and non-polyadenylated intergenic ncRNAs [3]. Similarly, we recovered thousands of ncRNA enriched intergenic loci in both rhesus macaque and cynomolgus macaque genomes [3]. Typically, this lncRNA annotation can be accomplished by first aligning sequencing reads to the species matched reference genome assembly using gapped aligners such as TopHat [65] or STAR [66]. The aligned reads can then be assembled into transcripts using tools such as Cufflinks [67], Scripture [68], or StringTie [69]. Depending on the completeness of the reference genome assembly, unmapped short reads may also be assembled de novo into full-length transcripts or longer transcript fragments using tools like Trinity [70]. These assembled transcript sequences can now be aligned back to the original genome assembly or simply added to the existing reference sequences and updated annotations.

Transcripts assembled from RNA-seq data consist of known as well as novel transcripts. Tools like Cuffcompare, part of the Cufflinks package [67], may be used to compare transcripts against reference annotations to separate novel from known genes. In order to focus on lncRNAs, it is important to filter out unwanted transcripts such as single-exon transcripts and transcripts of less than 200 nucleotides [71]. Next, transcripts with protein-coding potential can be assessed using various conservation-based computational tools, such as the phylogenetic codon substitution frequency (PhyloCSF) [72]. This can be done alone or in combination with non-conservation-based computational tools, such as a coding potential calculator (CPC) [73], coding potential assessment tool (CPAT) [74], and LncRNApred [71,75,76]. To further filter out potential protein coding transcripts, tools like HMMER software can be used to sort out transcripts that encode protein domains [71,77]. The remaining transcripts are putative lncRNAs and can be incorporated into further RNA-seq data analysis. It is also possible to perform lncRNA analysis for species without a reference genome assembly by complete de novo transcript assembly using various tools like Trinity [70] or Trans-ABYSS [78] followed by similar filtering steps as described. Since there is no reference genome, lncRNA quantification should use transcript alignment-based methods like RSEM [79] or alignment-free methods like Salmon [80].

While it is not covered here, we want to emphasize that as with any other experiments, proper experimental design should be carefully conducted before the start of infections and RNA-seq analysis.

3.1.3. Singling Out Cells in Lieu of Bulk Analysis

An exciting new development of RNA-seq is the analysis of large numbers of individual cell transcriptomes [81]. Transcriptome studies of viral infection are often conducted on samples of mixed cell populations, resulting in an average population-level expression for each transcript [82]. However, in doing so, the distinct functions of different cells may be overlooked. Lung tissue from an influenza virus infected individual can be a mixture of infiltrated immune cells, infected epithelial cells, uninfected epithelial cells, and many other types of cells. Similarly, a solid tumor is composed of various types of infiltrating immune cells in addition to a heterogeneous mixture of cancerous cells carrying a range of mutations and gene expression patterns [83]. To address this cell-to-cell heterogeneity on a genomic scale, researchers turn to single-cell RNA-seq (scRNA-seq). Aside from the isolation of single cells and amplification of small volumes of RNA, scRNA-seq is very similar to conventional bulk RNA-seq [82]. Although scRNA-seq has many advantages when compared with bulk RNA-seq, researchers should carefully consider multiple factors when they plan to use this technology for lncRNA analysis. First, existing scRNA-seq protocols can only capture polyadenylated transcripts, so many of the ploy(A-) lncRNAs will be missed. Second, due to the limited materials, the detection of lowly expressed transcripts like lncRNAs might be less sensitive in single cells than bulk samples [84]. Third, multiple technical factors like the biosafety measures for handling infectious materials and the breakdown of tissues with enzymes during cell dissociation used in most scRNA-seq protocols may impact transcriptional profiles [84]. Notwithstanding these caveats, scRNA-seq has the

potential to enable the identification of highly cell type specific lncRNAs that escape identification through current bulk sequencing strategies. The applications of scRNA-seq in large scale perturbation studies will be discussed further in Section 3.4.

3.2. Prioritization of Infection-Related lncRNAs by Computational Prediction

A typical RNA-seq analysis of virus infected samples may identify hundreds if not thousands of lncRNAs that respond to infection. One of the more challenging analytical steps is the prioritization of these identified lncRNAs, i.e., the identification of a small subset of lncRNAs that may play more important roles in the phenotype of interest. We see from Table 1 that many of the annotated lncRNAs have been researched through ad hoc experimentation, whereas potentially novel lncRNAs are overlooked in favor of those that have been previously studied. Due to the lack of lncRNA functional annotations, researchers may also simply limit downstream analysis to a few lncRNAs that have significant differential expression and/or are highly abundant. Given the limitations of this strategy, we recommend that additional computational analyses may help to guide this prioritization.

3.2.1. Prediction of lncRNA Function by “Guilt-by-Association”

The prediction of individual lncRNA functions is often based on “guilt-by-association” analysis, wherein tightly co-expressed lncRNA and coding transcripts are assumed to share similar functions [85–89]. Therefore, these lncRNAs may be inferred with the same functions as those annotated coding genes. This prediction starts with expression data that simultaneously profiles both coding genes and lncRNAs across multiple conditions. This expression data can be the same RNA-seq data used for identifying infection-related lncRNAs, or a complementary RNA-seq data relevant to the infection, or both. As an example, an independent human tissue compendium of RNA-seq data was used to infer the functions for a set of HIV infection-related human lncRNAs [3]. In another example, the functions of infection-related lncRNAs were directly inferred from the same RNA-seq data from mouse lung samples infected with either influenza A virus or severe acute respiratory syndrome coronavirus (SARS-CoV) [63]. Next, the choice of a suitable method for calculating co-expressions may be guided by an initial assessment of co-expressions of genes annotated with similar functions, as illustrated in [63].

There are multiple approaches for inferring putative lncRNA functions from co-expressed coding genes. As shown in [3,63], for each lncRNA of interest, all detected coding genes can be ranked by their correlation coefficients with the lncRNA. The enrichment of annotated gene sets, such as pathways and biological processes, in this ranked list of coding genes can be analyzed using commonly used methods, including gene set enrichment analysis (GSEA) [90], to obtain the functions that are highly correlated with the specific lncRNA.

Alternatively, co-expressed lncRNAs and coding genes can be determined by clustering methods. Genome-wide clustering of similar gene expression profiles extrapolates lncRNA function based on known gene functions in the same cluster, where transcripts in the same cluster are considered to be co-regulated [91]. The commonly used clustering methods include hierarchical clustering, k-means clustering, and self-organizing maps (SOMs) [87,91–93]. Similarly, a network-based approach can be utilized to predict the functions of lncRNAs from the known functions within the same network [91]. For example, a popular method called weighted gene co-expression analysis (WGCNA) [94] can be used to build gene co-expression networks using the RNA-seq data from virus infected samples. It first constructs a gene-gene network based on similar gene expression profiles, then divides genes with a similar expression into groups of genes or network modules. These modules of both lncRNAs and coding genes can then be annotated with enriched biological functions and associated with clinical traits. lncRNAs occupying key points (hubs or bottlenecks) in the derived networks may likely play important regulatory roles. As shown in [63], some of the identified lncRNAs may regulate the interferon response to viral infection.

3.2.2. Prediction of lncRNA Function Based on Local Genomic Context

Another strategy for inferring putative functions of lncRNAs is to examine protein-coding genes located near lncRNAs of interest, as lncRNAs may have cis-regulatory effects on flanking genes [86,95]. For example, in an early study we found that during SARS-CoV infection the changes in the expression of neighbor coding genes in mouse lungs were significantly associated with those of the corresponding lncRNAs, suggesting those lncRNAs may modulate host responses through neighboring coding genes [1]. Similarly, in a follow-up study with a larger number of samples, we also observed positive correlations between potential cis-regulatory lncRNAs and coding gene neighbors, indicating that some lncRNAs have transcriptional “enhancer-like” functions during viral infections [63]. Therefore, it is useful to prioritize lncRNAs proximal to coding genes, as they may have a higher potential for regulatory functions.

When portions of lncRNAs overlap a coding gene, it is expected that this region be conserved and under a strong selective pressure. A recent study found that the evolutionary age, overlapping configuration, and local genomic environment of an lncRNA-coding gene pair influence the expression correlation of the pair [96]. While there needs to be further research of lncRNA-coding pairs in more well-annotated genomes, it is clear that positional information is important in understanding their functional relationship. As we will discuss in Section 3.5, it is crucial to consider lncRNA baseline expression relative to its coding counterpart when performing the knockdown or overexpression of such an lncRNA.

3.3. Identifying Functional lncRNAs Using Evolutionary Analysis

When assessing the function of a gene, including lncRNAs, it is imperative that it is done in the context of evolution. Homology searches leverage annotation from closely related organisms in order to infer the function of novel lncRNAs. The existence of homology often indicates that there is some selective pressure for the lncRNA in question, thereby implying functional significance. However, lncRNAs pose computational challenges when it comes to identifying and ascertaining cross-species homology. lncRNAs have been previously categorized into subclasses, each with functions that reflect their evolutionary conservation [9]. For example, lncRNAs that occlude the transcription of coding genes through their own transcription typically lack sequence and structural conservation outside of those regions overlapping the promoter region of the coding gene. The heterogeneity of lncRNAs is clear when considering not just the variety of their functions, but also their wide range of conservation. Therefore, all forms of conservation: sequence, syntenic, and structural, must be considered to effectively conduct homology searches of lncRNAs.

3.3.1. Incorporating Synteny in Sequence Homology Searches

It is well-understood that lncRNAs often possess very little conservation at the primary sequence level [9,10]. For example, it has been estimated that there is only 22% sequence identity between orthologous human and mouse lncRNAs [97]. Furthermore, there are only a few examples of known lncRNAs with sequence conservation tantamount to coding genes, e.g., TUNA and MALAT1 [10]. The disparity between lncRNA exon and mRNA exon sequence conservation suggests that there is little selective constraint at this level [9]. There is, however, negative selection in promoter regions, which is one of the few commonalities across the lncRNA landscape [10,98]. Despite the lack of sequence conservation across species, there still exists many cases of functional conservation, suggesting the conservation of higher order structures.

Positional information can aid sequence homology searches, whereby the homology of lncRNAs is inferred by their relative positioning near orthologous genes. This syntenic conservation can help overcome the shortcomings of lncRNA sequence homology searches [9]. A recent study identified lncRNAs from *Arabidopsis thaliana* that were syntenic with high sequence divergence in two closely related plant species [99], which validates the assertion that synteny can be more informative than the

sequences themselves. Moreover, the existence of syntenic conservation confers functional significance, as it indicates cross-species selective pressure. This additional conserved feature has been employed in two recently developed tools, *slncky* [100] and *Evolinc* [101]. When tested on a set of known orthologous human-mouse lncRNAs, *slncky* successfully identified the vast majority as homologous and increased the size of this orthologous set by roughly 8%. Furthermore, these computational analyses may help investigators decide which lncRNA targets to study further, since conserved lncRNAs can potentially be further studied in other species, especially using *in vivo* models.

3.3.2. Using Structural Conservation to Functionally Annotate lncRNAs

The lack of sequence conservation among most lncRNAs has also prompted researchers to look at higher-order structures to infer functional significance. It is commonly understood that canonical RNAs, such as rRNA and tRNA, can form complex secondary and tertiary structures. There is now mounting evidence that lncRNAs, unlike shorter ncRNAs, assume higher-order structures with functional significance. For example, MALAT1 is known to have a highly conserved 3' cloverleaf structure, one of several structures described by Nitsche et al. [10]. The most common methods for predicting the secondary and tertiary structure of lncRNA are extensively reviewed by Yan et al. [102]. *Pfold* [103] and *Foldalign* [104] are two examples of methods that use multiple sequence alignments to predict the RNA secondary structure. *Mfold* [105] and *RNAfold* [106] are popular examples of minimum free energy models. Alternatively, *CMfinder* [107] is an expectation maximum algorithm that uses a heuristic approach to search for RNA motifs, which is not constrained to sequence-based alignments.

Many of these approaches fall short of identifying conserved lncRNA structures, as their focus lies primarily at the sequence level. To overcome the low sequence identity observed among many ncRNAs, a recent approach has made use of *CMfinder* in a computational pipeline that makes structural alignments to successfully capture conserved RNA structures (CRSs) [108]. The study also showed the functional significance of these CRSs, showing an enrichment in gene regulatory regions as well as overlap with RBP binding sites. The method predicted CRSs for a staggering 22% of lncRNAs annotated in GENCODE v25, showing a higher density of CRSs at the 5' end of lncRNAs. Interestingly, the majority of lncRNAs still lack CRSs. This could be a biological phenomenon of lncRNAs; however, it could also be due to the implementation of the tool. The search strategy encompasses many divergent species, making the search for conserved CRSs highly stringent. If the search space were limited to more closely related species, there could be a higher rate of CRS identification. Future studies may devise a tiered strategy consisting of an initial broad search followed by more narrow searches of less divergent species.

3.4. Large Scale Perturbation Studies for Probing lncRNA Functions

Considering the large number of lncRNAs that are lacking functional annotation, large scale perturbation will be another attractive strategy for probing their roles in viral infection. As shown by earlier studies [109–111], traditional large scale screening of host factors related to viral infection has been based on RNA interference (RNAi). Though RNAi-based perturbation can be applied to investigations of a large number of lncRNAs [112], targeting lncRNA in this manner may not be optimal. Many lncRNA exert their effect in the nucleus, NEAT1 and MALAT1 for example [113], whereas shRNAs are processed through DICER, which is mainly cytoplasmic. While nuclear DICER activity has been demonstrated, its activity may be cell type specific [113]. Additionally, the act of transcription is often sufficient to observe the downstream effects of lncRNAs [114]. In this case, targeting the transcript may not produce the desired functional effect. Finally, the off-target effects of RNAi are well-documented [115].

As summarized in [116], the CRISPR-Cas9 system has recently emerged as the leading technique for screening host factors important for different viral replications. These screens targeted coding genes based on the loss-of-functions induced by small indels created by the CRISPR-Cas9 system.

New modifications are rapidly advancing this approach for screening non-coding RNAs. For example, considering that indels caused by a single cut from Cas9 in non-coding regions are unlikely to produce a functional knockout, Zhu et al. [117] reported a high-throughput method to produce large deletions of non-coding DNA that is based on a lentiviral paired guide RNA (pgRNA) library. Using this screening method, they identified 51 human lncRNAs that can positively or negatively regulate human cancer cell growth and validated nine of the 51 lncRNA hits using multiple orthogonal techniques.

Alternatively, Cas9 is mutated to an endonuclease-deficient form (dCas9) and fused to a repressive Krüppel associated box (KRAB) motif preventing cytotoxicity due to DNA double-strand breaks [118] and inhibiting target gene expression [119]. The guide RNA (gRNA) library is produced via oligonucleotide printing and packaged into a lentiviral pool. Cell lines expressing dCas9-KRAB are transduced with the gRNA library in order to achieve a single gRNA copy per cell [120]. The resulting dCas9-KRAB inhibits transcription by directly blocking transcription machinery or through the activity of an effector domain, in this case KRAB [119]. At the conclusion of the treatment of interest, barcoded gRNAs are PCR amplified and deep sequenced to identify gRNA enrichment and thereby genes associated with susceptibility to the treatment [120].

Gilbert et al. [121] established a proof-of-concept by combining dCas9-KRAB with a gRNA library in K562 cells. The inhibition of coding genes in this manner resulted in high specificity and low off-target effects [121]. Additionally, a small library (six genes targeted by three gRNAs each) was used to target lncRNAs, resulting in an at least 80% knockdown of target genes as determined by qPCR [121]. The Weissman group expanded on that small library with a large scale CRISPRi library screen of lncRNA in seven different cell lines consisting of a gRNA library targeting over 16,000 lncRNA genes [122]. Between 28 and 438 lncRNAs were determined to cause a growth phenotype [122]. Among these lncRNAs was LINC00263, which was associated with a negative growth phenotype. LINC00263 was used to demonstrate the cell type specificity of lncRNA by showing that there is very low correlation between transcript abundance and the presence of a phenotype. Additionally, the treatment of all seven cell lines with LINC00263 antisense oligonucleotides retards proliferation differently in each cell line, further demonstrating the cell type specificity of LINC00263 and lncRNAs in general.

3.4.1. Gain-Of-Function Library Screen

Often, lncRNA genes are not detectably expressed, or they may be repressed by a disease state. This limited expression of lncRNAs makes detecting meaningful changes by inhibitory assays difficult, if not impossible. Several CRISPR-Cas9 methods have been developed to induce gene expression [118,123–125], which might be able to address this issue. These CRISPR activation proteins rely on the same endonuclease-deficient Cas9 described above. dCas9-SunTag is fused to a peptide epitope which recruits co-expressed VP64 activation domains [123]. A high-throughput gRNA library screen using this system in K562 cells was able to identify gene activation affecting cell growth in K562 cells as proof-of-principle [121]. Synergistic Activation Mediator (SAM) and dCas9-VPR are fused to multiple activation domains [124,125]. SAM was used to screen code genes for gain-of-function resulting in BRAF inhibitor resistance in A375 melanoma cells [125]. This screen identified 13 genes that were then individually confirmed at the transcript and protein level. This system was also used successfully to identify the lncRNA AK023948 as a positive regulator of AKT [126]. dCas9-VPR was shown to enhance the activation of the non-coding gene MIAT, in addition to other genes [124]. These developments highlight the potential of using dCas9-based activators to study lncRNA functions during viral infection.

3.4.2. New Developments: Multiplexed Library Screen and Single cell Library Screen

An exciting development for these screening methods is to inhibit or activate two or more target genes within the same single cells [127,128], as it allows the interrogation of complex gene-gene interactions. Currently, the assay relies on the expression of multiple gRNAs from a single plasmid or

the expression of multiple plasmids, each with a single gRNA [127,128]. It is also possible to combine gRNAs such that genes are activated and inhibited in parallel [129]. Combining large libraries is a complex task. Not only does this generate a tremendous amount of complex data, but the cost increases considerably with the library size. One method employed to mitigate these issues is to conduct preliminary screening with a single unbiased library. The resulting hits from the initial screen can then be used to produce a smaller library for the second round screening. The second round of screening could involve comparing the smaller library against itself or against the larger unbiased library [120,127].

Another key consideration for designing large scale screens is the choice of readout for approximating the phenotypes of interest. Using single cell sequencing (scRNA-seq) as a readout represents a potential paradigm shift [128,130,131]. Typically, the readout is relatively simple such as cell proliferation or survival, or some sortable marker proteins, which can be very limiting. For example, a pooled screen only provides an association of the perturbation with the phenotype of interest, which may not be an appropriate readout of a complex system [132].

Several groups in the past year have made modifications to the dCas9/gRNA system, circumventing this shortcoming of pooled screens [128,130,131]. Each group also made use of single cell sequencing (scRNA-seq) in order to improve the robustness of the screen and downstream analysis. These scRNA-seq-based approaches allow the perturbation and the phenotype of each cell to be measured simultaneously. In particular, the full transcriptome sequencing analysis provides a 'one-size-fits-all' assay that can cover a wide range of phenotypes [133]. By removing the need for specific biomarkers, it may become more broadly applicable for investigating viral infections.

Overall, these large scale perturbation approaches are unbiased and especially suitable for identifying associations between diseases and genes with no known function such as many lncRNAs. The CRISPR-dCas9 system is scalable and provides sufficient throughput for handling the large number of lncRNAs that require functional interrogation. While these newer methodologies like gain-of-function, multiplexed library screens, and scRNA-seq have yet to be implemented in the interrogation of lncRNA functions in the context of viral infection, the utilization of this technology will ultimately hasten the discovery of infection-related lncRNAs.

3.5. Considerations for Experimentally Validating Specific lncRNAs

Cross-examining the results generated by different methods described above will likely produce a shorter list of lncRNAs of higher interest. The next challenge is to experimentally validate their associated phenotype, e.g., pathway modulation or resistance to viral infection. Validation is especially important when considering the results of *in silico* analyses; however, confirmation studies should also be performed with regard to large-scale library screening due to the size and complexity of those assays. Though many existing techniques for studying RNA are easily transferable for investigating lncRNAs, there are some specific considerations given the uniqueness of lncRNAs.

One consideration is the subcellular localization of specific lncRNAs. A nuclear lncRNA may function through interactions with DNA methyltransferases, histone modifying enzymes, or other epigenetic machinery [20,31–34]. Cytoplasmic localization, on the other hand, is often associated with miRNA sponges or protein trafficking [134,135]. NEAT1 and MALAT1 are particularly well-studied examples, which are amenable to fluorescent localization studies [22,136,137]. Fluorescent *in situ* hybridization (FISH) techniques have been used extensively to show their localization to nuclear paraspeckles. These techniques can be utilized broadly for the study of lncRNA localization, so long as the appropriate sequence information is available [5].

This subcellular localization information will also inform the choice of experimental techniques. As described, inhibition studies are useful for studying candidate lncRNAs. RNAi, in the form of siRNA, degrades target RNA via DICER and has been used in many instances to inhibit the expression of lncRNAs [112]. Another type of RNAi, antisense oligonucleotides (ASOs), binds target RNA and promotes degradation via RNase H. Unlike other RNAi, these small molecules pass freely through the

cell membrane, i.e., without the assistance of transfection reagents, and are capable of entering the nucleus. The ability to pass through the nuclear envelope makes them ideal for targeting lncRNAs, the majority of which localize in the nucleus. siRNA has been reported to be less effective for the knockdown of nuclear targets even though DICER and the RNAi machinery have been shown to be active in the nucleus [138]. A comprehensive analysis found that both siRNAs and ASOs are capable of inhibiting the expression of lncRNAs [113]. As expected, ASOs were more effective against nuclear lncRNA, while siRNAs were more effective against cytoplasmic lncRNAs, and each was effective against lncRNA that localized to both the cytoplasm and the nucleus [113]. Ultimately, the authors concluded that a combination of siRNA and ASOs may be the best method to ensure that the target is sufficiently inhibited [113]. Alternatively, CRISPRi (see Section 3.4) or the traditional CRISPR-Cas9 are effective methods for the inhibition of lncRNA expression [122,139].

The complexity of genomic regions of interest and the ability to target specific lncRNAs in a given region are also important considerations for validation studies. CRISPR inhibition or activation will potentially influence the expression of overlapping or neighboring coding genes, making it difficult to single-out the effects of a specific lncRNA. Goyal et al. [138] illustrate this point well by demonstrating that the lncRNA LOC389641, which arises from a bidirectional promoter of TNFRSF10A, can be specifically inhibited by RNAi. CRISPRi, in this case, down regulates the expression of both genes [138]. Alternatively, the effect of a non-coding RNA may be independent of the transcript, but rely on the act of transcription to modulate a target [114]. In such instances, RNAi would have no functional effect even though it may inhibit the expression of the transcript. To that end, CRISPR-Cas9 would be a more effective system. However, targeting lncRNA with incomplete genomic annotations with the CRISPR-Cas9 system may prove difficult, if not impossible [138]. This outlook will improve as more detailed annotations of non-coding RNA regions are gathered. On the other hand, it is likely that in many instances the specificity of CRISPR systems will not be sufficient and the development of novel methods will be required to adequately modulate lncRNA expression.

Since lncRNA tends to have a lower expression abundance, overexpression might be better suited for lowly expressed lncRNAs and to gain information complementary to inhibition studies. CRISPRa or ectopic plasmid expression [126,140] are effective methods towards this end. The genomic layout of lncRNAs, such as overlap with coding genes, will prevent ectopic plasmid expression for certain genes. In addition, poorly annotated genes will be difficult to express if the transcription start site (TSS) or the bounds of the gene are unknown. Additionally, overexpression in this manner will not be effective if transcription is the inhibitory event, rather than some function of the lncRNA transcript.

The low expression of lncRNAs is in part due to their cell-type specific expression. Low expression levels can be particularly pronounced when an lncRNA is identified from a heterogeneous cell population, such as blood or tissues. In these cases, it is important to determine the cell types of origin for these lncRNAs. Elucidation of cell type may be done by combining cell staining and FISH techniques coupled with flow cytometry or microscopy. Knowledge of the cell type may also assist in hypothesizing the lncRNA function [141] and designing validation studies.

4. Conclusions and Future Direction

Our understanding of the biological relevance of lncRNAs has greatly increased over the past decade. Unfortunately, many of these genes still lack sufficient functional annotation, and their functional role in host-pathogen interactions remains largely unknown. In order to efficiently elucidate the role of these lncRNAs during viral infection, multiple high-throughput methodologies coupled with computational strategies are being utilized to process large quantities of poorly characterized lncRNAs.

RNA-seq has emerged as the most widely used technology for this purpose, which usually results in a large number of lncRNAs significantly associated with viral infection. In order to narrow down these lists of lncRNAs, complementary computational strategies like co-expression network analysis and evolutionary analysis may be leveraged to aid the annotation and prioritization of lncRNA genes. Alternatively, large-scale perturbation screens fueled by rapidly advancing CRISPR-Cas9 techniques are

providing novel tools for investigating lncRNA functions in specific areas including viral infection and quickly expanding our knowledge of lncRNA functions. These additional layers of information will reduce the long list of potential interesting lncRNAs to a short list of high-confidence ‘hits’. Ultimately, verifying the mechanism and function of candidate lncRNAs identified by high-throughput strategies requires orthogonal experimental confirmations.

Figure 1 shows a proposed workflow for the identification of key lncRNAs that are most relevant to specific viral infections. The proposed workflow includes three major phases: (1) broad discovery of infection associated lncRNAs; (2) annotation and prioritization of identified lncRNAs; and (3) experimental validation of specific candidate lncRNAs. The most challenging task is to devise efficient learning strategies for ranking high quality lncRNA “hits” by quantitatively combining multiple sources of information. However, to achieve this goal, it is necessary to establish high quality training datasets in which the functions of a sizable number of lncRNAs have been experimentally verified. Simultaneously, compatible high-throughput datasets are needed from the same experimental systems in order to extract predictive features for inferring lncRNA functions. We anticipate that as large scale perturbation techniques mature, some of these lncRNA screen studies will emerge as initial training datasets. Obviously, given the complexities of different high-throughput technologies, developing benchmark training datasets for this purpose requires community-based collaborative efforts.

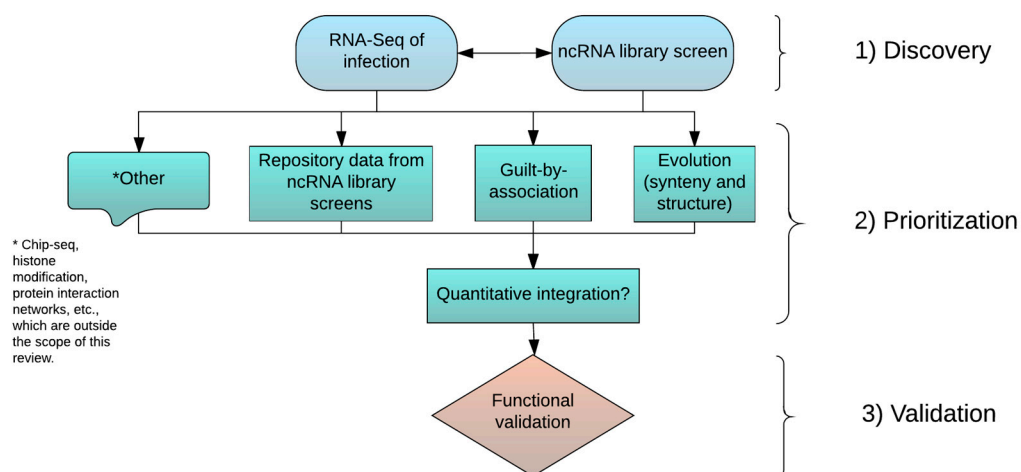


Figure 1. A proposed workflow of three major phases for identifying host lncRNAs that are involved in viral infections. Considering the lack of functional information for lncRNAs in general, the first step will be to survey lncRNAs associated with infection of interest (Discovery phase). Unbiased genome scale approaches like a transcriptome deep sequencing (RNA-seq) analysis of infected samples collected in vitro or in vivo is widely employed. Large-scale lncRNA screening is also emerging as a powerful alternative, but is less applicable due to technical constraints. Once a set of infection-related lncRNAs are identified, the next step is to narrow down the list to a small set of high interest lncRNAs (Prioritization phase), an extremely challenging task. There are multiple computational strategies for the annotation and prioritization of identified lncRNAs based on orthogonal information as indicated. Though not covered here, other types of information, like regulatory elements uncovered by Chip-Seq experiments, histone modification marks, and curated molecular interaction networks can all facilitate the prioritization. However, analytical methods for the quantitative integration of information from different sources need to be developed. This advancement may require community-based collaborative efforts. The last step is to experimentally validate specific candidate lncRNAs (Validation phase), while accounting for the unique characteristics of lncRNAs.

In summary, the large number of less studied lncRNAs represent a great opportunity for uncovering novel insights into virus-host interactions and potentially new targets for intervention.

To fully realize this potential, different high-throughput technologies can be leveraged. While efforts are underway to enable searching putative lncRNAs across large multimodal datasets [142], there is a pressing need to develop specialized computational strategies for prioritizing candidate lncRNAs for downstream experimental validations.

Acknowledgments: NIH/NIAID R21AI125040, NIH/NIAID R21AI120713, and startup fund from NC State College of Veterinary Medicine.

Author Contributions: All authors contributed equally to this work by gathering references, outlining the manuscript, and providing their expertise to all sections as appropriate. Major contributions were provided by D.J.L. for Sections 3.4 and 3.5; H.N.B. for Sections 2 and 3.3; F.Y. for Sections 3.1 and 3.2. E.A.H contributed to Section 2 and provided editorial assistance for the manuscript. X.P. contributed to all Sections and guided the project.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Peng, X.; Gralinski, L.; Armour, C.D.; Ferris, M.T.; Thomas, M.J.; Proll, S.; Bradel-Tretheway, B.G.; Korth, M.J.; Castle, J.C.; Biery, M.C.; et al. Unique signatures of long noncoding RNA expression in response to virus infection and altered innate immune signaling. *mBio* **2010**, *1*, e00206-10. [[CrossRef](#)] [[PubMed](#)]
- Zhang, Q.; Chen, C.Y.; Yedavalli, V.S.; Jeang, K.T. Neat1 long noncoding RNA and paraspeckle bodies modulate HIV-1 posttranscriptional expression. *mBio* **2013**, *4*, e00596-12. [[CrossRef](#)] [[PubMed](#)]
- Peng, X.; Sova, P.; Green, R.R.; Thomas, M.J.; Korth, M.J.; Proll, S.; Xu, J.; Cheng, Y.; Yi, K.; Chen, L.; et al. Deep sequencing of HIV-infected cells: Insights into nascent transcription and host-directed therapy. *J. Virol.* **2014**, *88*, 8768–8782. [[CrossRef](#)] [[PubMed](#)]
- Jin, C.; Peng, X.; Xie, T.; Lu, X.; Liu, F.; Wu, H.; Yang, Z.; Wang, J.; Cheng, L.; Wu, N. Detection of the long noncoding RNAs nuclear-enriched autosomal transcript 1 (NEAT1) and metastasis associated lung adenocarcinoma transcript 1 in the peripheral blood of HIV-1-infected patients. *HIV Med.* **2016**, *17*, 68–72. [[CrossRef](#)] [[PubMed](#)]
- Winterling, C.; Koch, M.; Koepfel, M.; Garcia-Alcalde, F.; Karlas, A.; Meyer, T.F. Evidence for a crucial role of a host non-coding RNA in influenza A virus replication. *RNA Biol.* **2014**, *11*, 66–75. [[CrossRef](#)] [[PubMed](#)]
- Mumtaz, P.T.; Bhat, S.A.; Ahmad, S.M.; Dar, M.A.; Ahmed, R.; Urwat, U.; Ayaz, A.; Shrivastava, D.; Shah, R.A.; Ganai, N.A. LncRNAs and immunity: Watchdogs for host pathogen interactions. *Biol. Proced. Online* **2017**, *19*, 3. [[CrossRef](#)] [[PubMed](#)]
- Valadkhan, S.; Gunawardane, L.S. LncRNA-mediated regulation of the interferon response. *Virus Res.* **2016**, *212*, 127–136. [[CrossRef](#)] [[PubMed](#)]
- Liu, W.; Ding, C. Roles of lncRNAs in viral infections. *Front. Cell. Infect. Microbiol.* **2017**, *7*, 205. [[CrossRef](#)] [[PubMed](#)]
- Ulitsky, I. Evolution to the rescue: Using comparative genomics to understand long non-coding RNAs. *Nat. Rev. Genet.* **2016**, *17*, 601–614. [[CrossRef](#)] [[PubMed](#)]
- Nitsche, A.; Stadler, P.F. Evolutionary clues in lncRNAs. *Wiley Interdiscip. Rev. RNA* **2017**, *8*, e1376. [[CrossRef](#)] [[PubMed](#)]
- Derrien, T.; Johnson, R.; Bussotti, G.; Tanzer, A.; Djebali, S.; Tilgner, H.; Guernec, G.; Martin, D.; Merkel, A.; Knowles, D.G.; et al. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res* **2012**, *22*, 1775–1789. [[CrossRef](#)] [[PubMed](#)]
- Bohmdorfer, G.; Wierzbicki, A.T. Control of chromatin structure by long noncoding RNA. *Trends Cell Biol.* **2015**, *25*, 623–632. [[CrossRef](#)] [[PubMed](#)]
- Mercer, T.R.; Mattick, J.S. Structure and function of long noncoding RNAs in epigenetic regulation. *Nat. Struct. Mol. Biol.* **2013**, *20*, 300–307. [[CrossRef](#)] [[PubMed](#)]
- Plath, K.; Mlynarczyk-Evans, S.; Nusinow, D.A.; Panning, B. Xist RNA and the mechanism of X chromosome inactivation. *Annu. Rev. Genet.* **2002**, *36*, 233–278. [[CrossRef](#)] [[PubMed](#)]
- Wutz, A. Xist function: Bridging chromatin and stem cells. *Trends Genet. TIG* **2007**, *23*, 457–464. [[CrossRef](#)] [[PubMed](#)]
- Senner, C.E.; Brockdorff, N. Xist gene regulation at the onset of X inactivation. *Curr. Opin. Genet. Dev.* **2009**, *19*, 122–126. [[CrossRef](#)] [[PubMed](#)]

17. Mira-Bontenbal, H.; Gribnau, J. New xist-interacting proteins in X-chromosome inactivation. *Curr. Biol.* **2016**, *26*, R338–R342. [[CrossRef](#)] [[PubMed](#)]
18. Beckedorff, F.C.; Amaral, M.S.; Deocesano-Pereira, C.; Verjovski-Almeida, S. Long non-coding RNAs and their implications in cancer epigenetics. *Biosci. Rep.* **2013**, *33*, e00061. [[CrossRef](#)] [[PubMed](#)]
19. Zhao, Y.; Sun, H.; Wang, H. Long noncoding RNAs in DNA methylation: New players stepping into the old game. *Cell Biosci.* **2016**, *6*, 45. [[CrossRef](#)] [[PubMed](#)]
20. Betancur, J.G. Pervasive lncRNA binding by epigenetic modifying complexes—The challenges ahead. *Biochim. Biophys. Acta* **2016**, *1859*, 93–101. [[CrossRef](#)] [[PubMed](#)]
21. Zhou, H.L.; Luo, G.; Wise, J.A.; Lou, H. Regulation of alternative splicing by local histone modifications: Potential roles for RNA-guided mechanisms. *Nucleic Acids Res.* **2014**, *42*, 701–713. [[CrossRef](#)] [[PubMed](#)]
22. Pruszko, M.; Milano, E.; Forcato, M.; Donzelli, S.; Ganci, F.; Di Agostino, S.; De Panfilis, S.; Fazi, F.; Bates, D.O.; Bicciato, S.; et al. The mutant p53-ID4 complex controls VEGFA isoforms by recruiting lncRNA MALAT1. *EMBO Rep.* **2017**, *18*, 1331–1351. [[CrossRef](#)] [[PubMed](#)]
23. Ding, Y.Z.; Zhang, Z.W.; Liu, Y.L.; Shi, C.X.; Zhang, J.; Zhang, Y.G. Relationship of long noncoding RNA and viruses. *Genomics* **2016**, *107*, 150–154. [[CrossRef](#)] [[PubMed](#)]
24. Ma, H.; Han, P.; Ye, W.; Chen, H.; Zheng, X.; Cheng, L.; Zhang, L.; Yu, L.; Wu, X.; Xu, Z.; et al. The long noncoding RNA NEAT1 exerts antihantaviral effects by acting as positive feedback for RIG-I signaling. *J. Virol.* **2017**, *91*. [[CrossRef](#)] [[PubMed](#)]
25. Imam, H.; Bano, A.S.; Patel, P.; Holla, P.; Jameel, S. The lncRNA NRON modulates HIV-1 replication in a NFAT-dependent manner and is differentially regulated by early and late viral proteins. *Sci. Rep.* **2015**, *5*, 8639. [[CrossRef](#)] [[PubMed](#)]
26. Li, J.; Chen, C.; Ma, X.; Geng, G.; Liu, B.; Zhang, Y.; Zhang, S.; Zhong, F.; Liu, C.; Yin, Y.; et al. Long noncoding RNA NRON contributes to HIV-1 latency by specifically inducing tat protein degradation. *Nat. Commun.* **2016**, *7*, 11730. [[CrossRef](#)] [[PubMed](#)]
27. Vigneau, S.; Rohrllich, P.S.; Brahic, M.; Bureau, J.F. Tmevpg1, a candidate gene for the control of Theiler's virus persistence, could be implicated in the regulation of gamma interferon. *J. Virol.* **2003**, *77*, 5632–5638. [[CrossRef](#)] [[PubMed](#)]
28. Ouyang, J.; Zhu, X.; Chen, Y.; Wei, H.; Chen, Q.; Chi, X.; Qi, B.; Zhang, L.; Zhao, Y.; Gao, G.F.; et al. NRAV, a long noncoding RNA, modulates antiviral responses through suppression of interferon-stimulated gene transcription. *Cell Host Microbe* **2014**, *16*, 616–626. [[CrossRef](#)] [[PubMed](#)]
29. Saayman, S.; Ackley, A.; Turner, A.W.; Famiglietti, M.; Bosque, A.; Clemson, M.; Planelles, V.; Morris, K.V. An HIV-encoded antisense long noncoding RNA epigenetically regulates viral transcription. *Mol. Ther.* **2014**, *22*, 1164–1175. [[CrossRef](#)] [[PubMed](#)]
30. Zhong, W.; Wang, H.; Herndier, B.; Ganem, D. Restricted expression of Kaposi sarcoma-associated herpesvirus (human herpesvirus 8) genes in Kaposi sarcoma. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 6641–6646. [[CrossRef](#)] [[PubMed](#)]
31. Merry, C.R.; Forrest, M.E.; Sabers, J.N.; Beard, L.; Gao, X.H.; Hatzoglou, M.; Jackson, M.W.; Wang, Z.; Markowitz, S.D.; Khalil, A.M. DNMT1-associated long non-coding RNAs regulate global gene expression and DNA methylation in colon cancer. *Hum. Mol. Genet.* **2015**, *24*, 6240–6253. [[CrossRef](#)] [[PubMed](#)]
32. Di Ruscio, A.; Ebralidze, A.K.; Benoukraf, T.; Amabile, G.; Goff, L.A.; Terragni, J.; Figueroa, M.E.; De Figueiredo Pontes, L.L.; Alberich-Jorda, M.; Zhang, P.; et al. DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* **2013**, *503*, 371–376. [[CrossRef](#)] [[PubMed](#)]
33. Savell, K.E.; Gallus, N.V.; Simon, R.C.; Brown, J.A.; Revanna, J.S.; Osborn, M.K.; Song, E.Y.; O'Malley, J.J.; Stackhouse, C.T.; Norvil, A.; et al. Extra-coding RNAs regulate neuronal DNA methylation dynamics. *Nat. Commun.* **2016**, *7*, 12091. [[CrossRef](#)] [[PubMed](#)]
34. Wang, L.; Zhao, Y.; Bao, X.; Zhu, X.; Kwok, Y.K.; Sun, K.; Chen, X.; Huang, Y.; Jauch, R.; Esteban, M.A.; et al. LncRNA Dum interacts with Dnmts to regulate Dppa2 expression during myogenic differentiation and muscle regeneration. *Cell Res.* **2015**, *25*, 335–350. [[CrossRef](#)] [[PubMed](#)]
35. Tsai, M.C.; Manor, O.; Wan, Y.; Mosammamaparast, N.; Wang, J.K.; Lan, F.; Shi, Y.; Segal, E.; Chang, H.Y. Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **2010**, *329*, 689–693. [[CrossRef](#)] [[PubMed](#)]

36. Li, Y.; Wang, Z.; Shi, H.; Li, H.; Li, L.; Fang, R.; Cai, X.; Liu, B.; Zhang, X.; Ye, L. HBXIP and LSD1 scaffolded by lncRNA hotair mediate transcriptional activation by c-Myc. *Cancer Res.* **2016**, *76*, 293–304. [[CrossRef](#)] [[PubMed](#)]
37. Gupta, R.A.; Shah, N.; Wang, K.C.; Kim, J.; Horlings, H.M.; Wong, D.J.; Tsai, M.C.; Hung, T.; Argani, P.; Rinn, J.L.; et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **2010**, *464*, 1071–1076. [[CrossRef](#)] [[PubMed](#)]
38. Cai, B.; Wu, Z.; Liao, K.; Zhang, S. Long noncoding RNA HOTAIR can serve as a common molecular marker for lymph node metastasis: A meta-analysis. *Tumour Biol.* **2014**, *35*, 8445–8450. [[CrossRef](#)] [[PubMed](#)]
39. Hajjari, M.; Khoshnevisan, A.; Shin, Y.K. Molecular function and regulation of long non-coding RNAs: Paradigms with potential roles in cancer. *Tumour Biol.* **2014**, *35*, 10645–10663. [[CrossRef](#)] [[PubMed](#)]
40. Schmitz, S.U.; Grote, P.; Herrmann, B.G. Mechanisms of long noncoding RNA function in development and disease. *Cell. Mol. Life Sci.* **2016**, *73*, 2491–2509. [[CrossRef](#)] [[PubMed](#)]
41. Rossetto, C.C.; Tarrant-Elorza, M.; Pari, G.S. Cis and trans acting factors involved in human cytomegalovirus experimental and natural latent infection of CD14 (+) monocytes and CD34 (+) cells. *PLoS Pathog.* **2013**, *9*, e1003366. [[CrossRef](#)] [[PubMed](#)]
42. Zapata, J.C.; Campilongo, F.; Barclay, R.A.; DeMarino, C.; Iglesias-Ussel, M.D.; Kashanchi, F.; Romerio, F. The Human Immunodeficiency Virus 1 ASP RNA promotes viral latency by recruiting the Polycomb Repressor Complex 2 and promoting nucleosome assembly. *Virology* **2017**, *506*, 34–44. [[CrossRef](#)] [[PubMed](#)]
43. Kim, E.D.; Sung, S. Long noncoding RNA: Unveiling hidden layer of gene regulatory networks. *Trends Plant Sci.* **2012**, *17*, 16–21. [[CrossRef](#)] [[PubMed](#)]
44. Bond, C.S.; Fox, A.H. Paraspeckles: Nuclear bodies built on long noncoding RNA. *J. Cell Biol.* **2009**, *186*, 637–644. [[CrossRef](#)] [[PubMed](#)]
45. Chen, L.L. Linking long noncoding RNA localization and function. *Trends Biochem. Sci.* **2016**, *41*, 761–772. [[CrossRef](#)] [[PubMed](#)]
46. Nakagawa, S.; Hirose, T. Paraspeckle nuclear bodies—Useful uselessness? *Cell. Mol. Life Sci.* **2012**, *69*, 3027–3036. [[CrossRef](#)] [[PubMed](#)]
47. Wilusz, J.E. Long noncoding RNAs: Re-writing dogmas of RNA processing and stability. *Biochim. Biophys. Acta* **2016**, *1859*, 128–138. [[CrossRef](#)] [[PubMed](#)]
48. Dai, Q.; Li, J.; Zhou, K.; Liang, T. Competing endogenous RNA: A novel posttranscriptional regulatory dimension associated with the progression of cancer. *Oncol. Lett.* **2015**, *10*, 2683–2690. [[CrossRef](#)] [[PubMed](#)]
49. Thomson, D.W.; Dinger, M.E. Endogenous microRNA sponges: Evidence and controversy. *Nat. Rev. Genet.* **2016**, *17*, 272–283. [[CrossRef](#)] [[PubMed](#)]
50. Peng, X.; Thierry-Mieg, J.; Thierry-Mieg, D.; Nishida, A.; Pipes, L.; Bozinoski, M.; Thomas, M.J.; Kelly, S.; Weiss, J.M.; Raveendran, M.; et al. Tissue-specific transcriptome sequencing analysis expands the non-human primate reference transcriptome resource (NHPRTR). *Nucleic Acids Res.* **2015**, *43*, D737–D742. [[CrossRef](#)] [[PubMed](#)]
51. Shalem, O.; Sanjana, N.E.; Hartenian, E.; Shi, X.; Scott, D.A.; Mikkelsen, T.; Heckl, D.; Ebert, B.L.; Root, D.E.; Doench, J.G.; et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **2014**, *343*, 84–87. [[CrossRef](#)] [[PubMed](#)]
52. Cheng, D.L.; Xiang, Y.Y.; Ji, L.J.; Lu, X.J. Competing endogenous RNA interplay in cancer: Mechanism, methodology, and perspectives. *Tumour Biol.* **2015**, *36*, 479–488. [[CrossRef](#)] [[PubMed](#)]
53. Sanchez-Mejias, A.; Tay, Y. Competing endogenous RNA networks: Tying the essential knots for cancer biology and therapeutics. *J. Hematol. Oncol.* **2015**, *8*, 30. [[CrossRef](#)] [[PubMed](#)]
54. Cazalla, D.; Yario, T.; Steitz, J.A. Down-regulation of a host microRNA by a Herpesvirus saimiri noncoding RNA. *Science* **2010**, *328*, 1563–1566. [[CrossRef](#)] [[PubMed](#)]
55. Guo, Y.E.; Oei, T.; Steitz, J.A. Herpesvirus saimiri microRNAs preferentially target host cell cycle regulators. *J. Virol.* **2015**, *89*, 10901–10911. [[CrossRef](#)] [[PubMed](#)]
56. Tavanez, J.P.; Quina, A.S.; Cunha, C. Virus and noncoding RNAs: Stars in the host-virus interaction game. *Future Virol.* **2014**, *9*, 1077–1087. [[CrossRef](#)]
57. Hu, S.; Wang, X.; Shan, G. Insertion of an Alu element in a lncRNA leads to primate-specific modulation of alternative splicing. *Nat. Struct. Mol. Biol.* **2016**, *23*, 1011–1019. [[CrossRef](#)] [[PubMed](#)]
58. Gutschner, T.; Hammerle, M.; Diederichs, S. MALAT1—A paradigm for long noncoding RNA function in cancer. *J. Mol. Med. (Berl.)* **2013**, *91*, 791–801. [[CrossRef](#)] [[PubMed](#)]

59. Wu, H.; Zheng, J.; Deng, J.; Zhang, L.; Li, N.; Li, W.; Li, F.; Lu, J.; Zhou, Y. LincRNA-uc002yug.2 involves in alternative splicing of RUNX1 and serves as a predictor for esophageal cancer and prognosis. *Oncogene* **2015**, *34*, 4723–4734. [[CrossRef](#)] [[PubMed](#)]
60. Barry, G.; Briggs, J.A.; Vanichkina, D.P.; Poth, E.M.; Beveridge, N.J.; Ratnu, V.S.; Nayler, S.P.; Nones, K.; Hu, J.; Bredy, T.W.; et al. The long non-coding RNA Gomafu is acutely regulated in response to neuronal activation and involved in schizophrenia-associated alternative splicing. *Mol. Psychiatry* **2014**, *19*, 486–494. [[CrossRef](#)] [[PubMed](#)]
61. Fang, Y.; Fullwood, M.J. Roles, functions, and mechanisms of long non-coding RNAs in cancer. *Genom. Proteom. Bioinform.* **2016**, *14*, 42–54. [[CrossRef](#)] [[PubMed](#)]
62. Wang, Z.; Gerstein, M.; Snyder, M. RNA-seq: A revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **2009**, *10*, 57–63. [[CrossRef](#)] [[PubMed](#)]
63. Josset, L.; Tchitchek, N.; Gralinski, L.E.; Ferris, M.T.; Einfeld, A.J.; Green, R.R.; Thomas, M.J.; Tisoncik-Go, J.; Schroth, G.P.; Kawaoka, Y.; et al. Annotation of long non-coding RNAs expressed in collaborative cross founder mice in response to respiratory virus infection reveals a new class of interferon-stimulated transcripts. *RNA Biol.* **2014**, *11*, 875–890. [[CrossRef](#)] [[PubMed](#)]
64. Mercer, T.R.; Clark, M.B.; Crawford, J.; Brunck, M.E.; Gerhardt, D.J.; Taft, R.J.; Nielsen, L.K.; Dinger, M.E.; Mattick, J.S. Targeted sequencing for gene discovery and quantification using RNA CaptureSeq. *Nat. Protoc.* **2014**, *9*, 989–1009. [[CrossRef](#)] [[PubMed](#)]
65. Trapnell, C.; Pachter, L.; Salzberg, S.L. Tophat: Discovering splice junctions with RNA-seq. *Bioinformatics (Oxf. Engl.)* **2009**, *25*, 1105–1111. [[CrossRef](#)] [[PubMed](#)]
66. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. Star: Ultrafast universal RNA-seq aligner. *Bioinformatics (Oxf. Engl.)* **2013**, *29*, 15–21. [[CrossRef](#)] [[PubMed](#)]
67. Trapnell, C.; Williams, B.A.; Pertea, G.; Mortazavi, A.; Kwan, G.; van Baren, M.J.; Salzberg, S.L.; Wold, B.J.; Pachter, L. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **2010**, *28*, 511–515. [[CrossRef](#)] [[PubMed](#)]
68. Guttman, M.; Garber, M.; Levin, J.Z.; Donaghey, J.; Robinson, J.; Adiconis, X.; Fan, L.; Koziol, M.J.; Gnirke, A.; Nusbaum, C.; et al. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat. Biotechnol.* **2010**, *28*, 503–510. [[CrossRef](#)] [[PubMed](#)]
69. Pertea, M.; Pertea, G.M.; Antonescu, C.M.; Chang, T.C.; Mendell, J.T.; Salzberg, S.L. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **2015**, *33*, 290–295. [[CrossRef](#)] [[PubMed](#)]
70. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644–652. [[CrossRef](#)] [[PubMed](#)]
71. Zhao, B.; Lu, M.; Wang, D.; Li, H.; He, X. Genome-wide identification of long noncoding RNAs in human intervertebral disc degeneration by RNA sequencing. *BioMed Res. Int.* **2016**, *2016*, 3684875. [[CrossRef](#)] [[PubMed](#)]
72. Lin, M.F.; Jungreis, I.; Kellis, M. PhyloCSF: A comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics (Oxf. Engl.)* **2011**, *27*, i275–i282. [[CrossRef](#)] [[PubMed](#)]
73. Kong, L.; Zhang, Y.; Ye, Z.Q.; Liu, X.Q.; Zhao, S.Q.; Wei, L.; Gao, G. CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* **2007**, *35*, W345–W349. [[CrossRef](#)] [[PubMed](#)]
74. Wang, L.; Park, H.J.; Dasari, S.; Wang, S.; Kocher, J.P.; Li, W. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res.* **2013**, *41*, e74. [[CrossRef](#)] [[PubMed](#)]
75. Pian, C.; Zhang, G.; Chen, Z.; Chen, Y.; Zhang, J.; Yang, T.; Zhang, L. LncRNApred: Classification of long non-coding RNAs and protein-coding transcripts by the ensemble algorithm with a new hybrid feature. *PLoS ONE* **2016**, *11*, e0154567. [[CrossRef](#)] [[PubMed](#)]
76. Ounzain, S.; Burdet, F.; Ibberson, M.; Pedrazzini, T. Discovery and functional characterization of cardiovascular long noncoding RNAs. *J. Mol. Cell. Cardiol.* **2015**, *89*, 17–26. [[CrossRef](#)] [[PubMed](#)]
77. Finn, R.D.; Clements, J.; Eddy, S.R. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **2011**, *39*, W29–W37. [[CrossRef](#)] [[PubMed](#)]

78. Robertson, G.; Schein, J.; Chiu, R.; Corbett, R.; Field, M.; Jackman, S.D.; Mungall, K.; Lee, S.; Okada, H.M.; Qian, J.Q.; et al. De novo assembly and analysis of RNA-seq data. *Nat. Methods* **2010**, *7*, 909–912. [[CrossRef](#)] [[PubMed](#)]
79. Li, B.; Dewey, C.N. RSEM: Accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinform.* **2011**, *12*, 323. [[CrossRef](#)] [[PubMed](#)]
80. Patro, R.; Duggal, G.; Love, M.I.; Irizarry, R.A.; Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **2017**, *14*, 417–419. [[CrossRef](#)] [[PubMed](#)]
81. Shapiro, E.; Biezuner, T.; Linnarsson, S. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* **2013**, *14*, 618–630. [[CrossRef](#)] [[PubMed](#)]
82. Saliba, A.E.; Westermann, A.J.; Gorski, S.A.; Vogel, J. Single-cell RNA-seq: Advances and future challenges. *Nucleic Acids Res.* **2014**, *42*, 8845–8860. [[CrossRef](#)] [[PubMed](#)]
83. Wang, Y.; Waters, J.; Leung, M.L.; Unruh, A.; Roh, W.; Shi, X.; Chen, K.; Scheet, P.; Vattathil, S.; Liang, H.; et al. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* **2014**, *512*, 155–160. [[CrossRef](#)] [[PubMed](#)]
84. Liu, S.; Trapnell, C. Single-cell transcriptome sequencing: Recent advances and remaining challenges. *F1000Research* **2016**, *5*, 182. [[CrossRef](#)] [[PubMed](#)]
85. Dinger, M.E.; Amaral, P.P.; Mercer, T.R.; Pang, K.C.; Bruce, S.J.; Gardiner, B.B.; Askarian-Amiri, M.E.; Ru, K.; Solda, G.; Simons, C.; et al. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.* **2008**, *18*, 1433–1445. [[CrossRef](#)] [[PubMed](#)]
86. Guttman, M.; Amit, I.; Garber, M.; French, C.; Lin, M.F.; Feldser, D.; Huarte, M.; Zuk, O.; Carey, B.W.; Cassady, J.P.; et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **2009**, *458*, 223–227. [[CrossRef](#)] [[PubMed](#)]
87. Cabili, M.N.; Trapnell, C.; Goff, L.; Koziol, M.; Tazon-Vega, B.; Regev, A.; Rinn, J.L. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **2011**, *25*, 1915–1927. [[CrossRef](#)] [[PubMed](#)]
88. Rinn, J.L.; Chang, H.Y. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **2012**, *81*, 145–166. [[CrossRef](#)] [[PubMed](#)]
89. Pauli, A.; Valen, E.; Lin, M.F.; Garber, M.; Vastenhouw, N.L.; Levin, J.Z.; Fan, L.; Sandelin, A.; Rinn, J.L.; Regev, A.; et al. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res.* **2012**, *22*, 577–591. [[CrossRef](#)] [[PubMed](#)]
90. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550. [[CrossRef](#)] [[PubMed](#)]
91. Signal, B.; Gloss, B.S.; Dinger, M.E. Computational approaches for functional prediction and characterisation of long noncoding RNAs. *Trends Genet. TIG* **2016**, *32*, 620–637. [[CrossRef](#)] [[PubMed](#)]
92. Ramos, A.D.; Diaz, A.; Nellore, A.; Delgado, R.N.; Park, K.Y.; Gonzales-Roybal, G.; Oldham, M.C.; Song, J.S.; Lim, D.A. Integration of genome-wide approaches identifies lncRNAs of adult neural stem cells and their progeny in vivo. *Cell Stem Cell* **2013**, *12*, 616–628. [[CrossRef](#)] [[PubMed](#)]
93. Kim, D.H.; Marinov, G.K.; Pepke, S.; Singer, Z.S.; He, P.; Williams, B.; Schroth, G.P.; Elowitz, M.B.; Wold, B.J. Single-cell transcriptome analysis reveals dynamic changes in lncRNA expression during reprogramming. *Cell Stem Cell* **2015**, *16*, 88–101. [[CrossRef](#)] [[PubMed](#)]
94. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **2008**, *9*, 559. [[CrossRef](#)] [[PubMed](#)]
95. Ponjavic, J.; Oliver, P.L.; Lunter, G.; Ponting, C.P. Genomic and transcriptional co-localization of protein-coding and long non-coding RNA pairs in the developing brain. *PLoS Genet.* **2009**, *5*, e1000617. [[CrossRef](#)] [[PubMed](#)]
96. Ning, Q.; Li, Y.; Wang, Z.; Zhou, S.; Sun, H.; Yu, G. The evolution and expression pattern of human overlapping lncRNA and protein-coding gene pairs. *Sci. Rep.* **2017**, *7*, 42775. [[CrossRef](#)] [[PubMed](#)]
97. Chodroff, R.A.; Goodstadt, L.; Sirey, T.M.; Oliver, P.L.; Davies, K.E.; Green, E.D.; Molnar, Z.; Ponting, C.P. Long noncoding RNA genes: Conservation of sequence and brain expression among diverse amniotes. *Genome Biol.* **2010**, *11*, R72. [[CrossRef](#)] [[PubMed](#)]

98. Guo, X.; Gao, L.; Wang, Y.; Chiu, D.K.; Wang, T.; Deng, Y. Advances in long noncoding RNAs: Identification, structure prediction and function annotation. *Brief. Funct. Genom.* **2016**, *15*, 38–46. [[CrossRef](#)] [[PubMed](#)]
99. Mohammadin, S.; Edger, P.P.; Pires, J.C.; Schranz, M.E. Positionally-conserved but sequence-diverged: Identification of long non-coding RNAs in the Brassicaceae and Cleomaceae. *BMC Plant Biol.* **2015**, *15*, 217. [[CrossRef](#)] [[PubMed](#)]
100. Chen, J.; Shishkin, A.A.; Zhu, X.; Kadri, S.; Maza, I.; Guttman, M.; Hanna, J.H.; Regev, A.; Garber, M. Evolutionary analysis across mammals reveals distinct classes of long non-coding RNAs. *Genome Biol.* **2016**, *17*, 19. [[CrossRef](#)] [[PubMed](#)]
101. Nelson, A.D.L.; Devisetty, U.K.; Palos, K.; Haug-Baltzell, A.K.; Lyons, E.; Beilstein, M.A. Evolinc: A tool for the identification and evolutionary comparison of long intergenic non-coding RNAs. *Front. Genet.* **2017**, *8*, 52. [[CrossRef](#)] [[PubMed](#)]
102. Yan, K.; Arfat, Y.; Li, D.; Zhao, F.; Chen, Z.; Yin, C.; Sun, Y.; Hu, L.; Yang, T.; Qian, A. Structure prediction: New insights into decrypting long noncoding RNAs. *Int. J. Mol. Sci.* **2016**, *17*, 132. [[CrossRef](#)] [[PubMed](#)]
103. Knudsen, B.; Hein, J. Pfold: RNA secondary structure prediction using stochastic context-free grammars. *Nucleic Acids Res.* **2003**, *31*, 3423–3428. [[CrossRef](#)] [[PubMed](#)]
104. Sundfeld, D.; Havgaard, J.H.; de Melo, A.C.; Gorodkin, J. Foldalign 2.5: Multithreaded implementation for pairwise structural RNA alignment. *Bioinformatics (Oxf. Engl.)* **2016**, *32*, 1238–1240. [[CrossRef](#)] [[PubMed](#)]
105. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **2003**, *31*, 3406–3415. [[CrossRef](#)] [[PubMed](#)]
106. Hofacker, I.L. Vienna RNA secondary structure server. *Nucleic Acids Res.* **2003**, *31*, 3429–3431. [[CrossRef](#)] [[PubMed](#)]
107. Yao, Z.; Weinberg, Z.; Ruzzo, W.L. Cmfnder—A covariance model based RNA motif finding algorithm. *Bioinformatics (Oxf. Engl.)* **2006**, *22*, 445–452. [[CrossRef](#)] [[PubMed](#)]
108. Seemann, S.E.; Mirza, A.H.; Hansen, C.; Bang-Berthelsen, C.H.; Garde, C.; Christensen-Dalsgaard, M.; Torarinnsson, E.; Yao, Z.; Workman, C.T.; Pociot, F.; et al. The identification and functional annotation of RNA structures conserved in vertebrates. *Genome Res.* **2017**, *27*, 1371–1383. [[CrossRef](#)] [[PubMed](#)]
109. Konig, R.; Chiang, C.Y.; Tu, B.P.; Yan, S.F.; DeJesus, P.D.; Romero, A.; Bergauer, T.; Orth, A.; Krueger, U.; Zhou, Y.; et al. A probability-based approach for the analysis of large-scale RNAi screens. *Nat. Methods* **2007**, *4*, 847–849. [[CrossRef](#)] [[PubMed](#)]
110. Brass, A.L.; Dykxhoorn, D.M.; Benita, Y.; Yan, N.; Engelman, A.; Xavier, R.J.; Lieberman, J.; Elledge, S.J. Identification of host proteins required for HIV infection through a functional genomic screen. *Science* **2008**, *319*, 921–926. [[CrossRef](#)] [[PubMed](#)]
111. Zhou, H.; Xu, M.; Huang, Q.; Gates, A.T.; Zhang, X.D.; Castle, J.C.; Stec, E.; Ferrer, M.; Strulovici, B.; Hazuda, D.J.; et al. Genome-scale RNAi screen for host factors required for HIV replication. *Cell Host Microbe* **2008**, *4*, 495–504. [[CrossRef](#)] [[PubMed](#)]
112. Guttman, M.; Donaghey, J.; Carey, B.W.; Garber, M.; Grenier, J.K.; Munson, G.; Young, G.; Lucas, A.B.; Ach, R.; Bruhn, L.; et al. LincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **2011**, *477*, 295–300. [[CrossRef](#)] [[PubMed](#)]
113. Lennox, K.A.; Behlke, M.A. Cellular localization of long non-coding RNAs affects silencing by RNAi more than by antisense oligonucleotides. *Nucleic Acids Res.* **2016**, *44*, 863–877. [[CrossRef](#)] [[PubMed](#)]
114. Autuoro, J.M.; Pirnie, S.P.; Carmichael, G.G. Long noncoding RNAs in imprinting and X chromosome inactivation. *Biomolecules* **2014**, *4*, 76–100. [[CrossRef](#)] [[PubMed](#)]
115. Sigoillot, F.D.; Lyman, S.; Huckins, J.F.; Adamson, B.; Chung, E.; Quattrochi, B.; King, R.W. A bioinformatics method identifies prominent off-targeted transcripts in RNAi screens. *Nat. Methods* **2012**, *9*, 363–366. [[CrossRef](#)] [[PubMed](#)]
116. Puschnik, A.S.; Majzoub, K.; Ooi, Y.S.; Carette, J.E. A CRISPR toolbox to study virus-host interactions. *Nat. Rev. Microbiol.* **2017**, *15*, 351–364. [[CrossRef](#)] [[PubMed](#)]
117. Zhu, S.; Li, W.; Liu, J.; Chen, C.H.; Liao, Q.; Xu, P.; Xu, H.; Xiao, T.; Cao, Z.; Peng, J.; et al. Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR-Cas9 library. *Nat. Biotechnol.* **2016**, *34*, 1279–1286. [[CrossRef](#)] [[PubMed](#)]
118. Qi, L.S.; Larson, M.H.; Gilbert, L.A.; Doudna, J.A.; Weissman, J.S.; Arkin, A.P.; Lim, W.A. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **2013**, *152*, 1173–1183. [[CrossRef](#)] [[PubMed](#)]

119. Gilbert, L.A.; Larson, M.H.; Morsut, L.; Liu, Z.; Brar, G.A.; Torres, S.E.; Stern-Ginossar, N.; Brandman, O.; Whitehead, E.H.; Doudna, J.A.; et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **2013**, *154*, 442–451. [[CrossRef](#)] [[PubMed](#)]
120. Kampmann, M.; Bassik, M.C.; Weissman, J.S. Functional genomics platform for pooled screening and generation of mammalian genetic interaction maps. *Nat. Protoc.* **2014**, *9*, 1825–1847. [[CrossRef](#)] [[PubMed](#)]
121. Gilbert, L.A.; Horlbeck, M.A.; Adamson, B.; Villalta, J.E.; Chen, Y.; Whitehead, E.H.; Guimaraes, C.; Panning, B.; Ploegh, H.L.; Bassik, M.C.; et al. Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **2014**, *159*, 647–661. [[CrossRef](#)] [[PubMed](#)]
122. Liu, S.J.; Horlbeck, M.A.; Cho, S.W.; Birk, H.S.; Malatesta, M.; He, D.; Attenello, F.J.; Villalta, J.E.; Cho, M.Y.; Chen, Y.; et al. CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science* **2017**, *355*. [[CrossRef](#)] [[PubMed](#)]
123. Tanenbaum, M.E.; Gilbert, L.A.; Qi, L.S.; Weissman, J.S.; Vale, R.D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **2014**, *159*, 635–646. [[CrossRef](#)] [[PubMed](#)]
124. Chavez, A.; Scheiman, J.; Vora, S.; Pruitt, B.W.; Tuttle, M.; Iyer, E.P.R.; Lin, S.; Kiani, S.; Guzman, C.D.; Wiegand, D.J.; et al. Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* **2015**, *12*, 326–328. [[CrossRef](#)] [[PubMed](#)]
125. Konermann, S.; Brigham, M.D.; Trevino, A.E.; Joung, J.; Abudayyeh, O.O.; Barcena, C.; Hsu, P.D.; Habib, N.; Gootenberg, J.S.; Nishimasu, H.; et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **2015**, *517*, 583–588. [[CrossRef](#)] [[PubMed](#)]
126. Koirala, P.; Huang, J.; Ho, T.T.; Wu, F.; Ding, X.; Mo, Y.Y. lncRNA AK023948 is a positive regulator of AKT. *Nat. Commun.* **2017**, *8*, 14422. [[CrossRef](#)] [[PubMed](#)]
127. Bassik, M.C.; Kampmann, M.; Lebbink, R.J.; Wang, S.; Hein, M.Y.; Poser, I.; Weibezahn, J.; Horlbeck, M.A.; Chen, S.; Mann, M.; et al. A systematic mammalian genetic interaction map reveals pathways underlying ricin susceptibility. *Cell* **2013**, *152*, 909–922. [[CrossRef](#)] [[PubMed](#)]
128. Adamson, B.; Norman, T.M.; Jost, M.; Cho, M.Y.; Nunez, J.K.; Chen, Y.; Villalta, J.E.; Gilbert, L.A.; Horlbeck, M.A.; Hein, M.Y.; et al. A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell* **2016**, *167*. [[CrossRef](#)] [[PubMed](#)]
129. Boettcher, M.; Tian, R.; Blau, J.; Markegard, E.; Wu, D.; Biton, A.; Zaitlen, N.; McCormick, F.; Kampmann, M.; McManus, M.T. Decoding directional genetic dependencies through orthogonal CRISPR/Cas screens. **2017**. [[CrossRef](#)]
130. Jaitin, D.A.; Weiner, A.; Yofe, I.; Lara-Astiaso, D.; Keren-Shaul, H.; David, E.; Salame, T.M.; Tanay, A.; van Oudenaarden, A.; Amit, I. Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-seq. *Cell* **2016**, *167*. [[CrossRef](#)] [[PubMed](#)]
131. Xie, S.; Duan, J.; Li, B.; Zhou, P.; Hon, G.C. Multiplexed engineering and analysis of combinatorial enhancer activity in single cells. *Mol. Cell* **2017**, *66*. [[CrossRef](#)] [[PubMed](#)]
132. Datlinger, P.; Rendeiro, A.F.; Schmidl, C.; Krausgruber, T.; Traxler, P.; Klughammer, J.; Schuster, L.C.; Kuchler, A.; Alpar, D.; Bock, C. Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **2017**, *14*, 297–301. [[CrossRef](#)] [[PubMed](#)]
133. Wagner, D.E.; Klein, A.M. Genetic screening enters the single-cell era. *Nat. Methods* **2017**, *14*, 237–238. [[CrossRef](#)] [[PubMed](#)]
134. Willingham, A.T.; Orth, A.P.; Batalov, S.; Peters, E.C.; Wen, B.G.; Aza-Blanc, P.; Hogenesch, J.B.; Schultz, P.G. A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. *Science* **2005**, *309*, 1570–1573. [[CrossRef](#)] [[PubMed](#)]
135. Du, Z.; Sun, T.; Haciosuleyman, E.; Fei, T.; Wang, X.; Brown, M.; Rinn, J.L.; Lee, M.G.; Chen, Y.; Kantoff, P.W.; et al. Integrative analyses reveal a long noncoding RNA-mediated sponge regulatory network in prostate cancer. *Nat. Commun.* **2016**, *7*, 10982. [[CrossRef](#)] [[PubMed](#)]
136. Tripathi, V.; Ellis, J.D.; Shen, Z.; Song, D.Y.; Pan, Q.; Watt, A.T.; Freier, S.M.; Bennett, C.F.; Sharma, A.; Bubulya, P.A.; et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell* **2010**, *39*, 925–938. [[CrossRef](#)] [[PubMed](#)]
137. Adriaens, C.; Standaert, L.; Barra, J.; Latil, M.; Verfaillie, A.; Kalev, P.; Boeckx, B.; Wijnhoven, P.W.; Radaelli, E.; Vermi, W.; et al. P53 induces formation of NEAT1 lncRNA-containing paraspeckles that modulate replication stress response and chemosensitivity. *Nat. Med.* **2016**, *22*, 861–868. [[CrossRef](#)] [[PubMed](#)]

138. Goyal, A.; Myacheva, K.; Gross, M.; Klingenberg, M.; Duran Arque, B.; Diederichs, S. Challenges of CRISPR/Cas9 applications for long non-coding RNA genes. *Nucleic Acids Res.* **2017**, *45*, e12. [[CrossRef](#)] [[PubMed](#)]
139. Pulido-Quetglas, C.; Aparicio-Prat, E.; Arnan, C.; Polidori, T.; Hermoso, T.; Palumbo, E.; Ponomarenko, J.; Guigo, R.; Johnson, R. Scalable design of paired CRISPR guide RNAs for genomic deletion. *PLoS Comput. Biol.* **2017**, *13*, e1005341. [[CrossRef](#)] [[PubMed](#)]
140. Alvarez-Dominguez, J.R.; Bai, Z.; Xu, D.; Yuan, B.; Lo, K.A.; Yoon, M.J.; Lim, Y.C.; Knoll, M.; Slavov, N.; Chen, S.; et al. De novo reconstruction of adipose tissue transcriptomes reveals long non-coding RNA regulators of brown adipocyte development. *Cell Metab.* **2015**, *21*, 764–776. [[CrossRef](#)] [[PubMed](#)]
141. Hon, C.C.; Ramilowski, J.A.; Harshbarger, J.; Bertin, N.; Rackham, O.J.; Gough, J.; Denisenko, E.; Schmeier, S.; Poulsen, T.M.; Severin, J.; et al. An atlas of human long non-coding RNAs with accurate 5' ends. *Nature* **2017**, *543*, 199–204. [[CrossRef](#)] [[PubMed](#)]
142. Gong, Y.; Huang, H.T.; Liang, Y.; Trimarchi, T.; Aifantis, I.; Tsiganos, A. LncRNA-screen: An interactive platform for computationally screening long non-coding RNAs in large genomics datasets. *BMC Genom.* **2017**, *18*, 434. [[CrossRef](#)] [[PubMed](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).