Review Article

Functional annotation of lncRNA in high-throughput screening

Chi Wai Yip1, Divya M. Sivaraman1,2, Anika V. Prabhu1 and Jay W. Shin1

1RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa 230-0045, Japan; 2Department of Pathology, Sree Chitra Tirunal Institute for Medical Sciences and Technology, Kerala 695011, India

Correspondence: Jay W. Shin (jay.shin@riken.jp)

Recent efforts on the characterization of long non-coding RNAs (lncRNAs) revealed their functional roles in modulating diverse cellular processes. These include pluripotency maintenance, lineage commitment, carcinogenesis, and pathogenesis of various diseases. By interacting with DNA, RNA and protein, lncRNAs mediate multifaceted mechanisms to regulate transcription, RNA processing, RNA interference and translation. Of more than 173000 discovered lncRNAs, the majority remain functionally unknown. The cell type-specific expression and localization of the lncRNA also suggest potential distinct functions of lncRNAs across different cell types. This highlights the niche of identifying functional lncRNAs in different biological processes and diseases through high-throughput (HTP) screening. This review summarizes the current work performed and perspectives on HTP screening of functional lncRNAs where different technologies, platforms, cellular responses and the downstream analyses are discussed. We hope to provide a better picture in applying different technologies to facilitate functional annotation of lncRNA efficiently.

Introduction

The characterization of the mammalian transcriptome by the Encyclopedia of DNA Elements (ENCOD) and the Functional Annotation of the Mammalian Genome (FANTOM) projects have revealed a large collection of long non-coding RNAs (lncRNAs), which are defined as greater than 200 nucleotides and account for the majority of the transcriptome [1,2]. A study summarizing multiple transcript collections has further revealed approx. 20 000 human lncRNAs with functional insight [3]. Indeed, the single-nucleotide polymorphisms (SNPs) identified by genome-wide association studies (GWAS) that are associated with diseases or traits are mainly located in the non-coding region [4], suggesting that both DNA regulatory elements and lncRNAs play some functional roles in different diseases. Except for the canonical lncRNAs that were initially characterized such as XIST, NEAT1 and MALAT1 [5,6], the vast majority of lncRNAs were characterized only in the last decade. These lncRNAs have been shown to regulate pluripotency maintenance [7], cellular reprogramming [8], lineage commitment [9], carcinogenesis [10], and pathogenesis of various diseases [11]. While the pervasive transcription of lncRNAs implies they should acquire function over evolutionary time [12], functional role of the majority of lncRNAs remains to be characterized. Notably, lncRNAs express, localize, and function specifically across different cell types [13–15], suggesting that an lncRNA may be functional in one cell type or a particular cellular phenotype but not in the others.

The diverse roles of lncRNAs in organismal development, physiological processes, and disease pathology has been revealed by the advent of high-throughput (HTP) screenings [15–25]. Various technologies have allowed loss-gain-of-function HTP screening to become popular and affordable. This has included the application of short hairpin RNA (shRNA) pooled libraries, CRISPR-based single guide RNA (sgRNA) pooled libraries, antisense oligonucleotide (ASO), and next-generation sequencing (NGS). While these HTP screening platforms have been frequently applied to mRNA functional screening, they have not yet...
become common for lncRNAs. Unlike protein-coding genes, the novelty of lncRNAs makes it challenging to hypothesize the connection between the lncRNA hits and the tested phenotypes. Thus, this review will summarize the use of HTP platforms for lncRNA screening, and the downstream analyses that help to narrow the gap between lncRNA hits and the phenotype.

Selection and prioritization of lncRNA targets

Currently, the number of known human lncRNA transcripts are over 173,000 according to NONCODE [26] and over 268,000 according to LncBook [27]. However, this could be far lower in a particular cell type since lncRNAs are highly cell type-specific [14]. Therefore, expression level is one of the key factors to select and prioritize lncRNAs in a screen to improve the success rate. Indeed, loss-of-function screening on lncRNAs with a broad range of expression levels showed positive correlations between expression level and the cell growth phenotype [15,16].

Furthermore, acquiring differential expression data by microarray or NGS-based methods between the desired phenotype and control can improve target selection. For example, lncRNAs associated with cancer subtypes and clinical prognosis were identified from microarray data of four cancer types [28], and the list was adopted in a CRISPR-based screening study [17]. Similarly, Xu and colleagues [18] selected 25 lncRNAs to screen for drug resistance by choosing the up-regulated lncRNAs after drug treatments using RNA-seq analysis. Although computationally it is possible to further prioritize to predict the most relevant lncRNAs [29–32], loss-/gain-of-function HTP screening is necessary to identify lncRNAs with a specific and relevant cellular phenotype.

Modulation of lncRNA expression

Many loss-/gain-of-function methods are available for modulating the expression of lncRNAs [33]. However, only several of them are scalable for use in HTP screening. We will only cover ASO, RNA-interference (RNAi) and CRISPR/Cas13 systems, which target the RNA directly, as well as CRISPR/Cas9 (CRISPRn) and use of the catalytically dead Cas9 (dCas9) (CRISPRi/a) which target genomic DNA (Figure 1A). For a clearer goal of the review, we
will start by comparing compatible modulation methods directly.

**ASO versus RNAi**

Both ASO and RNAi (including shRNA exogenous expression and siRNA transfection) directly target the RNA molecules for degradation. The action of RNAi depends on the RNA-induced silencing complex (RISC) to cleave the RNA target, while ASOs form a DNA–RNA duplex with the RNA target for RNase H recognition and cleavage. Typical ASOs for degrading RNA utilize LNA gapmer technology [34] and often include a 2′MOE modification for stability and nuclease resistance. Both methodologies have been used in HTP screening for functional lncRNAs and shown to knockdown lncRNAs efficiently [16,22–24].

In comparing ASO and RNAi, the key difference is the localization of the endogenous enzymes used to target the RNA molecule—RNase H is enriched in the nucleus, while RISC is enriched in the cytoplasm. Consequently, nuclear lncRNAs were more efficiently suppressed by ASOs, compared with cytoplasmic lncRNAs which responded better to RNAi [35]. However, since most lncRNAs are enriched in the nucleus where chromatin is regulated by lncRNA [36,37], targeting the nuclear lncRNAs is often more beneficial. Additionally, ASOs were shown to be equally effective in targeting introns and exons of lncRNAs [16], indicating ASOs could knockdown cytoplasmic RNA during their transcription in the nucleus. Indeed, lncRNA localization negatively affects RNAi more than ASO [35]. Therefore, ASO is recommended for modulating the expression of lncRNA for array-based screens. On the contrary, shRNA pooled library allows the RNAi modality to be scalable. Besides expression modulation, RNase H-inactive ASO has been used to block splicing of nascent RNA [38,39]. The same method has been applied to block the splicing of an lncRNA, resulting in chromatin retention and malfunction of the lncRNA [40]. This highlights the added benefits of ASO in studying lncRNA biology.

**CRISPRn versus CRISPRi/a**

CRISPR/Cas9-mediated genome editing has been commonly used in mRNA functional screening and also for lncRNA functionality. Zhu and colleagues [17] adopted paired-guide RNA to mediate deletion of various lengths from 700 lncRNA genes and identified 51 lncRNAs that affected cell growth.

In the CRISPRi system, dCas9 is fused to transcriptional repressor domains KRAB or methyl-CpG binding protein 2 (MeCP2) to achieve targeted suppression of gene expression [41–43]. Besides those for mRNA, several large-scale CRISPRi screenings have been performed to elucidate lncRNA functionality [15,21,25]. More recently, dCas9 fused with transcriptional activator VP64 and other synergistic activators have been described [44]. Unlike conventional plasmid-based overexpression, CRISPRa activates lncRNA expression from the endogenous genomic locus, which has the advantage of capturing cis-acting and nuclear lncRNA functions. Thus, CRISPRa has been utilized for the functional screening of lncRNAs [19,20].

When comparing CRISPRn knockout and CRISPRi knockdown in pooled library studies targeting mRNAs, CRISPRn is more effective than CRISPRi [45,46]. However, there are several limitations when applying CRISPRn to modulate lncRNA [47]. Firstly, unlike mRNA, partial deletion of the lncRNA genes may not ablate their functions, with lncRNA gene size often too long for complete deletion [48]. While other options such as deletion of the promoter could be considered, the varied efficiency of suppression for different loci is challenging in large-scale screening. Secondly, CRISPRn could affect other proximal functional elements and their topological interactions thus confounding the mechanistic activity of lncRNAs. CRISPRn cleavage also hinders the detection of cis-regulation of lncRNAs. Finally, as many of the lncRNAs overlap with other genes, deletion by CRISPRn is less applicable. Additionally, ds-DNA damage mediated by CRISPRn is known to trigger non-specific false positives [49], which does not occur with CRISPRi [50]. Therefore, CRISPRi/a is more applicable than CRISPRn for lncRNA expression modulation.

**CRISPRi versus shRNA versus CRISPR/Cas13**

The performance of pooled libraries for RNAi, CRISPRi, and CRISPRn were compared using 46 essential and 47 non-essential mRNAs in a negative selection screen [46]. This study found that (1) CRISPRn screening performs the best in both sgRNA- and gene-based analyses, (2) shRNA screening results reflected off-target effects of individual shRNAs, which can be reduced by using multiple shRNAs, and (3) CRISPRi screening shows virtually no off-target effects, but only 50% of the sgRNA is effective, leading to a lower hit rate. As discussed earlier, CRISPRn is not the most suitable screening method for lncRNAs. For both the shRNA and CRISPRi methods, their shortcomings can be compensated by including more shRNA or sgRNA constructs. Additionally, sgRNA design can be improved with accurate transcription start sites (TSSs) positioning using FANTOM CAGE annotation [51]. Similarly, tools to improve the selection of lncRNA targets according to expression level and TSS annotation for sgRNA design are available.
Targeting the IncRNA transcripts versus IncRNA loci

In summary, targeting the transcripts of IncRNA, and not their DNA loci, is advantageous for (1) not interfering with the function of the IncRNA promoter, which may act as an enhancer for other genes, (2) targeting only individual isoforms since different isoforms could have opposing effects [57], and (3) allowing additional IncRNAs to be considered for screening, since the DNA-targeting approach must avoid affecting other genes on the same loci of intragenic, divergent, and antisense IncRNA genes. On the contrary, targeting the DNA loci could also investigate enhancer-like activities, broadening the coverage of functional non-coding regulatory elements in the genome. Furthermore, both ASO and RNAi methods exhibit independent off-target effects [16,58] whereas CRISPRi/a modulation of the DNA is less prone to off-target effects because of the narrow targeting window.

Scaling to HTP screening

HTP screening has shown promise in identifying individual IncRNAs which regulate a cellular phenotype. Moreover, as IncRNAs constitute the majority of the transcriptome and harbor many GWAS SNPs, using HTP screenings to identify the proportion of IncRNAs with the information of their genomic characteristics that participate in cellular mechanisms has become a key tool to understand the genome and its role in disease. Additionally, HTP screening can identify associations between drug compounds and their gene targets, which include both protein-coding genes [59–62] and IncRNAs [18–20,63]. For example, Bester and colleagues [19] utilized a CRISPRa-system to identify genes contributing to cytarabine resistance in an acute myeloid leukemia cell line. They revealed a group of IncRNAs driving cytarabine resistance via cis-regulation. Such studies further highlight the interconnected role IncRNAs play in various pathological mechanisms.

Screening in an array format has the advantage of separating the individual perturbations. This allows measurement by qPCR to rule out the unsuccessful perturbation [16,22], where the degree of knockdown significantly affects the hit rate [16]. This also provides the flexibility to directly record cellular phenotype, such as through high-content imaging [24]. Oligonucleotide arrays are available for chemically synthesized siRNA, ASO, and sgRNA (Figure 1B). Previously the relatively high cost of LNA gapmer ASO limited its use in HTP screening. However, due to the lifting of the LNA patent protection, the cost is now comparable with siRNA and sgRNA (~$200 USD per oligo). Alternatively, both shRNA and sgRNA can be generated by cloning into viral vectors in an array format, however, this task can be laborious to scale-up. When the throughput of an arrayed screen is limited by cost and experimental intensiveness, pooled library screening provides another option where the throughput can be near unlimited (~$20 000 USD per library with 30 000 constructs).

Screening by oligonucleotide array

Design of the array platform

When designing the arrayed screens, at least two constructs with independent sequences showing effective knockdown and the same cellular phenotype are necessary to call a hit. From our previous study, which used 2021 ASOs to target 285 IncRNAs, 43.5% of ASOs were effective and there were 68.1% IncRNA targets with at least two effective ASOs [16]. Our results showed that higher expressing IncRNAs are more susceptible to ASO knockdown, suggesting inclusion of such targets could improve the chances of an effective ASO. For RNAi, Guttman and colleagues [22] designed five shRNAs per target for 214 intergenic IncRNAs, in which 65% had at least one effective knockdown. Designing a functional ASO sequence without off-targeting for IncRNA is challenging, since IncRNA genes harbor many repetitive elements [64] and ASOs have the potential to target the intronic sequences of nascent transcripts. For RNAi, off-target effects were partly due to the dependence on a relatively short complementary seed sequence in the 3’ end of RNAi [65]. Therefore, in order to reach enough effective oligos and deal with the off-target effects, the starting number of oligos should be sufficiently high (e.g., five or more). The phenotypic response of transfecting
### Table 1 Screenings applied to lncRNAs

<table>
<thead>
<tr>
<th>Technologies</th>
<th>Cell types</th>
<th>LncRNAs</th>
<th>Constructs</th>
<th>Phenotype (% hit)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Array screening</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>siRNA</td>
<td>HeLa</td>
<td>2231</td>
<td>4 (pooled)</td>
<td>Mitotic progression (0.1%) Chromosome segregation (0.1%) Cytokinesis (0.1%)</td>
<td>[24]</td>
</tr>
<tr>
<td>ASO</td>
<td>Human dermal fibroblast</td>
<td>285 (194), 119</td>
<td>5–15 (2–10), ≥2</td>
<td>Proliferation (7.7%), CAGE sequencing (10.9%)</td>
<td>[16]</td>
</tr>
<tr>
<td>shRNA</td>
<td>Mouse ES</td>
<td>214 (147)</td>
<td>5 (1–2)</td>
<td>Microarray (93%)</td>
<td>[22]</td>
</tr>
<tr>
<td><strong>Pooled library screening</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRISPRi</td>
<td>Epidermal keratinocyte</td>
<td>2263</td>
<td>5</td>
<td>Proliferation (0.4%)</td>
<td>[25]</td>
</tr>
<tr>
<td>Cas13</td>
<td>K562</td>
<td>25</td>
<td>10</td>
<td>Proliferation with three anti-cancer drug treatments (64%)</td>
<td>[18]</td>
</tr>
<tr>
<td>CRISPRi</td>
<td>Human glioblastoma</td>
<td>5689</td>
<td>10</td>
<td>Proliferation with fractionated radiation (8.2%)</td>
<td>[21]</td>
</tr>
<tr>
<td>CRISPRa</td>
<td>MOLM14 AML</td>
<td>14701</td>
<td>≥4</td>
<td>Proliferation with Cytarabine treatment (19.5%)</td>
<td>[19]</td>
</tr>
<tr>
<td>CRISPRa</td>
<td>Human melanoma A375</td>
<td>10504</td>
<td>~10</td>
<td>Proliferation with Vemurafenib treatment (0.2%)</td>
<td>[20]</td>
</tr>
<tr>
<td>CRISPRi</td>
<td>iPSC, MCF7, U87, K562, MDA-MB-231, HeLa, HEK293T</td>
<td>5543, 5725, 5689, 16401, 5725, 6188, 5785</td>
<td>10</td>
<td>Proliferation (5.9%), Proliferation (1%), Proliferation (1.1%), Proliferation (0.4%), Proliferation (0.5%), Proliferation (0.4%), Proliferation (0.3%)</td>
<td>[15]</td>
</tr>
<tr>
<td>CRISPR</td>
<td>Huh7.5, OC</td>
<td>671</td>
<td>~20</td>
<td>Proliferation (7.6%)</td>
<td>[17]</td>
</tr>
<tr>
<td>shRNA</td>
<td>Mouse ES</td>
<td>1280</td>
<td>≥3</td>
<td>OCT4 expression (1.6%)</td>
<td>[23]</td>
</tr>
</tbody>
</table>

oligos to cells is transient. For a longer phenotypic assay, such as differentiation, lentivirus-mediated shRNA array can be considered, as done by Gutmann and colleagues [22] to identify lncRNAs that are important in mouse ES cells.

### Phenotypes tested in array platform

When adopting an array method of screening, the cellular phenotypes of interest are abundant as compared with pooled screening. Numerous quantitative cellular assays have been utilized, such as those measuring growth, differentiation [66], infection [67], and endocytosis [68]. Among them, the inclusion of imaging is a major advantage for array screening. For instance, 50 lncRNAs were shown to affect cell morphology in human dermal fibroblasts by real-time imaging after ASO knockdown [16]. By applying high-content imaging to RNAi array screening, Stojic and colleagues [24] identified six lncRNAs for regulating mitotic progression, chromosome segregation, and cytokinesis.

### Screening by pooled library

#### Design of the pooled library platform

When designing a pooled library screening, several factors determine the scale of the study. These include the total number of lncRNA targets, how many sgRNA/shRNA constructs per lncRNA, the additional non-target scramble sequences, and the size of the coverage. A higher number of constructs could compensate for the off-target effects of shRNA and increase the number of effective sgRNAs. Typically, the number of constructs is in a range of 5–20 shRNA/sgRNA per target (Table 1). From an analytical point of view, the number of constructs should be at least 4, while a higher number allows for statistical analysis that can incorporate technical and biological variability to improve power [69]. The number of non-target controls reflects the variation of the screen, which is complicated with the randomness of construct distribution and cell-to-cell variation. Therefore, it is necessary to include a large pool of non-target controls, which usually ranges from ~100 to ~1000 [19,21,25]. Assuming a library of 10 000 constructs, 10 constructs per lncRNA with 1000 non-target controls, the throughput of the library will be 900 lncRNAs. The infection coverage represents the number of cells uptaking the same construct, where each cell contains only one construct by
restricting the multiplicity of infection (MOI) (usually ≤0.3). As the genomic integration event is random, a sufficient size of infection coverage can normalize the variability. The common infection coverage for each construct is approx. 300–500× [15,19,25] while sequencing coverage is approx. 1000×. Additional independent experimental replications of the same library are also necessary. Therefore, for a single replicate, the number of infected cells needed for 300× coverage is 3 million, and the starting number of cells is 10 million at an MOI of 0.3.

**Unique molecular identifiers**

Since cell-to-cell variability has posed challenges in interpreting phenotypes, strategies such as incorporating unique molecular identifiers (UMIs) in sgRNA libraries have been established [70]. The UMIs have allowed for the screening of clonally expanded and individually tagged cells, resulting in an increased sensitivity and robustness compared with conventional analyses. The statistical methods, including using the UMIs as internal replicates and in lineage dropout analyses, increase both the precision and the accuracy of the screen, as well as reducing the infection coverage needed to reach the same statistical power [71].

**Phenotypes tested in pooled library platform**

The cellular phenotypes assessed after genetic perturbation are diverse (Figure 1C), including survival advantage for robust cell growth [15,17,25], after drug treatments [18–20] or with fractionated radiation [21]. By combining with cell sorting, a wide variety of phenotypes can be measured, such as pluripotency maintenance [15,23,72], differentiation [73], protein transport [74], oxidative stress [75], and many more.

Almost all the pooled library screenings for IncRNAs thus far have relied on survival advantage as the phenotype of interest. For example, Liu and colleagues [21] identified 434 and 33 IncRNAs, that respectively support and reduce cell growth, in human glioblastoma cells in the presence of clinically relevant doses of fractionated radiation. Additionally, Cai and colleagues [25] identified 9 IncRNAs that support robust cell proliferation of epidermal keratinocyte cells. The lower hit rate of this study may be due to a lower number of sgRNAs designed per target and thus lower statistical power. Another reason is that including treatments that place the cells under stress, as in the Liu and colleagues’ study [21], can provoke the expression or function of IncRNAs, which constitute a significant fraction of the genes differentially expressed in response to cell stress [76]. As summarized for all IncRNA screenings in Table 1, the hit rate is generally higher if drug treatment is included.

When combined with cell sorting, cell loss from the staining and washing steps of fluorescence-activated cell sorting (FACS) is a factor to consider, although CRISPRi screen combined with FACS has been reported [15]. A larger population of transduced cells are needed to compensate for this cell loss and reach the final coverage. Formaldehyde fixation can reduce the degree of cell loss while de-crosslinking is required to rescue the genomic DNA for PCR library construction [15]. Indeed, many screens rely on expression of exogenous genes carrying fluorescent signals [73,74] or fluorescent probe live trackers [75]. For example, Liu and colleagues [73] performed a CRISPRa screen in mouse ES cells, where the cell surface marker hCD8 was inserted downstream of Tubb3. Differentiated neurons were then separated by magnetic-activated cell sorting (MACS), which combined with use of cell surface markers, can minimize cell loss.

**Analytical efforts**

Pooled library screening requires rigorous bioinformatics analyses to interpret the results, with detection of false positives remaining a critical issue. Nevertheless, advanced analysis strategies for CRISPR applications are currently available [77], with several distinct algorithms established for evaluating the results of pooled library screening. Briefly, redundant siRNA activity [78] and HiTSelect [79] are designed for RNai screening. Redundant siRNA activity ranks the targeted genes by log fold change and generates P-values from the ranking against a uniform distribution, while HiTSelect ranks target genes by considering both the effect on the phenotype and the number of active constructs using a random-effects model. The MAGeCK robust ranking algorithm [80] is commonly used in CRISPR-based screens. It uses the raw sgRNA read counts and adopts a negative binomial model to generate sgRNA P-values, which are combined to gene level by a modified robust ranking algorithm. CRISPhieRmix [81] uses the log fold change value of each sgRNA generated from standard count software such as DESeq2 [82], and provides empirical FDR for the target genes using a hierarchical mixture distribution. BAGEL [83] uses data from prior screens to build null distribution and positive effects to rank target genes. Besides, some algorithms such as MAGeCK maximum likelihood estimation [84] and JACKS [85] are designed to compare and pool multiple screens. CERES [86] is specifically designed to correct the side effects mediated by DNA damage from CRISPR cuts for cancer cells which exhibit large copy number variation. Bodapati and colleagues [69] compared these algorithms with CRISPR pooled screening data and suggested...
using MAGeCK robust ranking algorithm in most cases for its robustness and performance, while CRISPhieRmix was touted as the only algorithm taking the various sgRNA efficiencies into account for CRISPRi/a screenings. If UMIs are included, there is an advantage to incorporating additional statistical methods, such as internal replicate analysis and lineage dropout analysis [71]. Because the hits identified by pooled library screening are statistical outcomes from a population of perturbations with considerable false positives [87], validation by reproducing the phenotypic result with individual perturbation is sometimes necessary.

**Annotate the roles of the IncRNA hits**

Transcriptome profiling

Except for several of the IncRNAs with structural implications [88] or those regulating translation [89], the majority of IncRNAs regulate other genes at the transcriptional [36,37] or RNA [90] levels. This signature allows us to unveil the functions of most IncRNAs by studying the transcriptomic changes. Therefore, follow-up studies restricted to IncRNA hits using Perturb-seq or CROP-seq [91–93] will be the most compatible with pooled library screening. Perturb-seq and CROP-seq are sequencing platforms designed to combine single cell RNA-seq and CRISPR-based genetic screens. To facilitate the detection of the non-polyadenylated gRNA in single cell transcriptome, Perturb-seq lentiviral vector harbors a gRNA-matched barcode upstream of the poly-A tail of the puromycin gene [91], while CROP-seq introduces an additional gRNA copy upstream of the poly-A of the puromycin gene [93]. Besides, a single cell RNA-seq platform has been applied to the shRNA screen by using a pol II-dependent shRNA backbone [94]. Single-cell transcriptomes containing the sgRNA/shRNA identity can unveil the mechanism mediated by specific IncRNAs, while a single-cell readout is advantageous if the cell population is heterogeneous (e.g., study design with differentiation).

More recently, targeted Perturb-seq [95] was developed, allowing profiling of a subset of the transcriptome (e.g., genes near the loci of IncRNA hits). Alternatively, molecular phenotyping with bulk RNA-seq or CAGE-seq after IncRNA perturbation can reveal genes or pathways modulated by the IncRNA [16]. Such transcriptomic profiling will identify the global transcriptomic changes, which can be captured by differential expression profile followed by Gene Set Enrichment Analysis (GSEA) [96] and Gene Ontology (GO) [97] analyses. Ramilowski and colleagues [16] characterized ZNF213-AS1 by GSEA in controlling migration in dermal fibroblasts and validated this in a wound closure migration assay. Besides global transcriptomic changes, transcriptomic profiling combined with computational analyses will also identify direct effector genes of the IncRNA.

**Prediction of effector genes with known IncRNA mechanisms**

Predicting direct effector genes is often necessary to connect IncRNA hits to the tested phenotype with cellular mechanisms. This will also reflect the functional mechanisms of the IncRNAs so that validation experiments can be designed. Since cis-regulation is one of the major mechanisms mediated by IncRNA [36,37], defining proximal genes either by 2D distance or 3D chromatin structure from Hi-C data will yield functional interactions between the IncRNA locus and the effector genes (Figure 2A). From certain disease phenotypes, effector genes can be identified by connecting IncRNAs to GWAS [98] and expression quantitative trait loci (eQTL) [99–101]. Additionally, co-expression analysis among the tissue-wide or cell type-wide data from various consortia [1,102] between the IncRNAs and effector genes should improve confidence in these networks.

Moreover, IncRNAs are known to function by interacting with protein, DNA and RNA [103,104], while their subcellular localization can suggest their functional mechanisms [105]. Fractionation data will be useful to estimate mechanisms of the IncRNAs, but data with matched cell type are necessary as subcellular localization of IncRNAs is cell type specific [13]. For trans-regulation, experimental genome-wide RNA–DNA interaction analysis, such as GRID-seq [106] and RADICL-seq [107] and in silico RNA–DNA interaction such as triplex formation prediction [108] can be used as references (Figure 2B). Both cis and trans RNA–DNA interactions are likely to involve proteins as the executors while linking a protein (chromatin modifier or transcription factor) with RNA–DNA interaction will help characterizing the downstream effects. Many DNA-binding proteins were found to be capable of binding RNA [109,110]. RNA-binding protein (RBP) is a class of protein that cooperates with IncRNAs for post-transcriptional processes, such as splicing, cleavage and polyadenylation. ENCODE Phase III has generated RNA–RBP interaction for 356 RBPs in K562 and HepG2 [109] while ChIP-seq data are also available for 58 of these RBPs [111]. Combining these two datasets could reveal RNA–DNA interaction with RBP content (Figure 2C). Last but not least, sequestering miRNA is one of the known mechanisms of IncRNA while the differential expression of the mRNA competitor can be revealed from transcriptome profiling. A number of databases supporting the prediction of miRNA sponge interaction have been described [112] (Figure 2D).
Figure 2. Mechanisms of lncRNA in gene regulation

(A) Regulation of proximal genes. (B) Regulation of genes in trans. (C) Mediation of protein-DNA interaction to regulate gene expression. (D) Sequestering miRNA from mRNA by acting as miRNA sponge.

Conclusion

Array and pooled library screenings have been established in many studies to identify functional protein coding genes in different cellular phenotypes, but the work on lncRNAs is lagging. However, the majority of lncRNAs have yet to be characterized. Advances in HTP screening platforms present an opportunity to explore the functionality of lncRNAs. Following the genome-wide screening of lncRNAs, it will also be imperative to investigate molecular mechanisms of individual lncRNAs for determining their roles across different cell types, including in disease, to identify functional conservation and redundancy.

Summary

- Direct RNA-targeting perturbation methods present the advantage of (1) distinguishing between isoforms, (2) avoiding interference with the enhancer function of lncRNA promoters, and (3) targeting lncRNAs even when their loci overlap with other genes.

- DNA-targeting perturbation methods provide more consistent results by having fewer off-target effects and allowing functional screening of cis-regulatory elements.

- Array screening can be easily combined with different phenotypic readouts and allow the quantification of perturbation efficiency.

- Pooled library screening benefits from high throughput, but the results are preliminary and often require secondary screening and validations.
Competing Interests
The authors declare that there are no competing interests associated with the manuscript.

Funding
This work was supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) funding to RIKEN Integrative Medical Sciences (IMS); and the Japan Society for the Promotion of Science (JSPS) [grant number 19F19386 (to A.V.P.)].

Author Contribution
C.W.Y. and J.W.S. conceived the content of the paper. C.W.Y. wrote the majority of the paper. D.M.S. and A.V.V. contributed to the writing. All authors reviewed, edited and approved the final version of the paper.

Abbreviations
ASO, antisense oligonucleotide; dCas9, dead Cas9; ENCODE, Encyclopedia of DNA Elements; ES, embryonic stem; FANTOM, Functional Annotation of the Mammalian Genome; FDR, false discovery rate; GSEA, gene set enrichment analysis; GWAS, genome-wide association study; HTP, high-throughput; IncRNA, long non-coding RNA; MOI, multiplicity of infection; NGS, next-generation sequencing; pol II, RNA polymerase II; qPCR, quantitative PCR; RBP, RNA-binding protein; RISC, RNA-induced silencing complex; RNAi, RNA-interference; sgRNA, single guide RNA; shRNA, short hairpin RNA; SNP, single-nucleotide polymorphism; TSS, transcription start site; UMI, unique molecular identifier.

References

© 2021 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY-NC-ND).


61 Tian, R., Abairienos, A., Hong, J., Hashemi, S.H., Yan, R., Nalls, M.A. et al. (2021) Genome-wide CRISPRi/a screens in human neurons link lysosomal failure to ferroptosis. bioRxiv, https://doi.org/10.1101/2020.06.27.175679


© 2021 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY-NC-ND).


