Application of sequence semantic and integrated cellular geography approach to study alternative biogenesis of exonic circular RNA

Rajnish Kumar[1], Rajkrishna Mondal[2], Tapobrata Lahiri[3*] and Manoj Kumar Pal[4]

*Correspondence:
tlahiri@iiita.ac.in

[1] Department of Pathology and Laboratory Medicine, Medical Center, University of Kansas, Kansas City 66160, USA
[2] Department of Biotechnology, Nagaland University, Dimapur, Nagaland 797112, India
[3] Room No. 4302, Department of Applied Sciences, Computer Centre – II, Indian Institute of Information Technology-Allahabad, Allahabad 211015, India
[4] Faculty of Engineering and Technology, United University Prayagraj, Prayagraj, UP 211012, India

Abstract

Background: Concurrent existence of lncRNA and circular RNA at both nucleus and cytosol within a cell at different proportions is well reported. Previous studies showed that circular RNAs are synthesized in nucleus followed by transportation across the nuclear membrane and the export is primarily defined by their length. lncRNAs primarily originated through inefficient splicing and seem to use NXF1 for cytoplasm export. However, it is not clear whether circularization of lncRNA happens only in nucleus or it also occurs in cytoplasm. Studies indicate that circular RNAs arise when the splicing apparatus undergoes a phenomenon of back splicing. Minor spliceosome (U12 type) mediated splicing occurs in cytoplasm and is responsible for the splicing of 0.5% of introns of human cells. Therefore, possibility of cRNA biogenesis mediated by minor spliceosome at cytoplasm cannot be ruled out. Secondly, information on genes transcribing both circular and lncRNAs along with total number of RBP binding sites for both of these RNA types is extractable from databases. This study showed how these apparently unconnected pieces of reports could be put together to build a model for exploring biogenesis of circular RNA.

Results: As a result of this study, a model was built under the premises that, sequences with special semantics were molecular precursors in biogenesis of circular RNA which occurred through catalytic role of some specific RBPs. The model outcome was further strengthened by fulfillment of three logical lemmas which were extracted and assimilated in this work using a novel data analytic approach, Integrated Cellular Geography. Result of the study was found to be in well agreement with proposed model. Furthermore this study also indicated that biogenesis of circular RNA was a post-transcriptional event.

Conclusions: Overall, this study provides a novel systems biology based model under the paradigm of Integrated Cellular Geography which can assimilate independently performed experimental results and data published by global researchers on RNA biology to provide important information on biogenesis of circular RNAs considering lncRNAs as precursor molecule. This study also suggests the possible RBP-mediated circularization of RNA in the cytoplasm through back-splicing using minor spliceosome.

Kumar *et al. BMC Bioinformatics*   (2023) 24:148

Page 2 of 21

Background

A complete physicochemical mechanism of biogenesis of circular RNAs is still under scrutiny. In some study back-splicing supported by the presence of inverted repeats in flanking Introns are considered as root cause of their biogenesis [1, 2], while in other study involvement of some trans-factors like RNA Binding Proteins (RBPs) are reported to induce the event of circularization [3–5]. Apart from this, Exon skipping events are also considered to be the influencing factor of the circularization event [6]. Recently, RBPs which are of immune factor types are reported to be involved in Circular RNAs biogenesis [7].

Interestingly, both long non-coding RNAs (lncRNAs) and Circular RNAs with length greater than 200 bp are found to share many common properties [8, 9]. Both Circular RNAs and lncRNAs are reported to work as biomarker in case of cancers and other diseases [10]. Circular RNAs differ from mRNA mainly in terms of topology, stability and translational capability and its biogenesis competes with splicing from pre-mRNA [4]. Long non-coding RNAs (lncRNAs) are mRNAs like transcripts which was first reported in Xist gene of mouse [11]. As reported in [12], in context of coding potential, lncRNAs primarily lack Open Reading Frame (ORF). Regarding differences, lncRNAs differ from mRNAs through their larger transcript sizes, longer exon lengths, low conservation of sequence, relatively low expression and more specific expression profile which are considered as features discriminating them from mRNAs [13, 14]. lncRNAs can be classified at different levels based on their function, localization and biogenesis[15].

Role of lncRNAs in post-transcriptional regulation is well reported [16]. lncRNAs are reported to be molecular address code particularly in nucleus [17]. Their associations with disease as well as cellular functions are so high that they are considered as multitasking molecules inside the cell [18]. Some of the lncRNAs undergo post-transcriptional processing resulting in alternative forms that are different than other reported lncRNAs [19, 20]. From these reports it is noticeable that field of lncRNAs is expanding. Nuclear retention or cytoplasmic enrichment of lncRNAs is controlled by some sequence features. RIDLs, a pentamer motif in BORG, and a highly structured repeat in FIRRE, which is bound by heterogeneous nuclear ribonucleoprotein U, facilitating chromatin interaction are the silent sequence features which promote the nuclear retention of lncRNAs. However, RNA Pol II pausing, inefficient splicing and m6A modification are the cytoplasmic export signals for lncRNAs [21]. Recently a new type of lncRNAs are reported that are truly exonic like mRNAs having both ends closed [22, 23]. It is however intriguing to observe that another member of RNA world, Circular RNAs (circRNAs) that resembles lncRNAs on many aspects, may provide a pack of information on roles and biogenesis of these biomolecules. Previous reports indicated that circRNAs are generated in the nucleus followed by transportation across the nuclear membrane and enriched mainly in the cytoplasm. CircRNAs that are above and less than 800 nt are efficiently exported to the cytoplasm through UAP56 and URH49 respectively [24]. However, further study is still required to explore more about spatiomolecular aspects of biogenesis, subcellular distribution, and transport mechanism of circRNA within a cell.

In the present study the process of circularization of RNA was investigated by considering results of research works so far published under an introductory data processing framework, Integrated Data Geography (IDG) in general and Integrated Cellular Geography (ICG) in particular following well known Integrated Geography (IG) approach [25]. Under this framework experimental outputs of published reports were considered as lemmas to utilize them as necessary and sufficient conditions for validation of the model of circular RNA Biogenesis proposed in this work. In the backdrop of many published results available through various investigations, introduction of this approach appeared to be the need of the hour for convergence of such apparently unconnected results towards fulfillment of a particular objective. The model however built primarily from the angles of convergence of reports on possible precursor molecules, factors leading to circularization and cellular spatiotemporal condition supporting circularization.

Towards resolving the first problem of finding precursor molecule, as for primary guess, first off, each of the exonic circular RNAs collected from 2 databases were BLASTed against the lncRNAs distributed within 3 databases. The objective was to see whether an lncRNA has the region inside their sequence from which their circular form may have originated in particular condition. As for supportive reports we find that circular RNA mostly originated from exonic regions. Other regions for their origin are also reported like, UTR and Intronic regions, lncRNA loci and antisense of known transcripts [26, 27]. Also, Recent report reveals that ciRS-7 exonic sequence is embedded in an lncRNA locus [28]. Linear-circular isoform switching (LC-switching) has been recently reported which occurs between circular transcript and its parental linear transcript [29]. Taking cue from these works and focusing this study on circular RNAs rooted from lncRNAs, a circular RNA query processing mechanism against lncRNA database was performed as discussed above. Subsequently the lncRNA data providing mapped hits and no hits were collected for further study. The target of this part of work was to find the similarity between lncRNAs and Circular RNAs at the level of sequence so that the precursor molecule (possibly lncRNA) of Circular RNA could be identified.

Next part of this work was primarily devoted to identify biomolecular factors and the spatiotemporal conditions catalyzing circularization of these precursor molecules (notionally lncRNAs as indicated in this study) producing Circular RNA. The RNA-binding protein FUS regulates circRNA biogenesis by binding the introns flanking the back-splicing junctions [30]. Whereas, another RBP protein Quaking (QKI) regulates circRNA formation during Epithelial-mesenchymal transition (EMT) by binding to sites flanking circRNA-forming exons [5]. Insertion of these RBP binding sites into linear RNA can induce exon circularization [5, 30]. In this work, above RNA binding proteins (RBPs) common to both lncRNAs and Circular RNAs [31, 32] were intuitively targeted as possible important biomolecular catalysts for circularization of RNAs for their reported roles in such activities [4, 5, 7]. The cellular spatiotemporal conditions supporting circularization of RNAs were also studied on the basis of differential profile of existence (DPE) of number of lncRNAs and Circular RNAs within nucleus and cytosol of a cell in association with DPE of RBPs unique to lncRNAs and Circular RNAs. The objective of studying spatiotemporal conditions leading to formation of Circular RNAs was to ascertain whether their biogenesis is co-transcriptional or post-transcriptional.

Kumar *et al. BMC Bioinformatics*   (2023) 24:148

Page 4 of 21

Fig. 1 Chemical representation of ICG

Table 1 lncRNAs hits against circular RNAs as queries

Circular RNA database	lncRNAs in database			BLAST HITS information
	GENCODE	**NONCODE**	**LNCipedia**	
circRNADb	1211	7471	1846	Total hits obtained
	170	907	184	Hits yielding complete mapping with query circular RNA
circBase	7977	37,339	13,010	Total hits obtained
	135	1620	205	Hits yielding complete mapping with query circular RNA

Besides the major U2 spliceosome responsible for splicing of 99.5% introns, 0.5% of introns are spliced via minor spliceosome U12. Therefore, role of minor spliceosome for circulization of lncRNA within cytoplasm was also considered in this study which corroborates the phenomenon of cytoplasmic enrichment of CircRNAs.

Results

As stated in the introduction section, the objective of this work was to study molecular mechanism leading to formation of Circular RNA. To accomplice it, first possible precursor molecules were searched and thereafter, possible factors leading to conversion of these precursor molecules into Circular RNA were studied. Therefore, the results were listed in a manner so that each of the results lead to formation of a logical lemma (i.e., the interim inference drawn on the basis of contemporary results) that served as a building block to construct the proposed ICG based model of biogenesis. A simple chemical representation of this model was shown in Fig. 1 which stood as the basis for the extraction of the results.

Result to construct Lemma 1 related to identification of potential molecular precursors of Circular RNA

In this regard, methodologies as described in methodology section under subheading "Identification of possible precursor molecules of circular RNA" section was employed to check status of lncRNAs towards such function because degeneracy of transcription from both coding genes and non-coding nucleotide sequence producing them was well reported [33]. The numbers of lncRNAs sequences in GENCODE, NONCODE and LNCipedia under study were 23,898, 172,216 and 104,487 respectively. The total numbers of circular RNAs taken from circRNADb, circBase in study were 32,914 and 140,790 respectively. The total number lncRNA hits including those that were completely mapped from particular lncRNA databases against circular RNAs as queries was shown in Table 1.

Lemma 1   *Reasonable percentage of lncRNA (~5%) out of total hits obtained was found to have complete sequence similarity with circular RNA as shown in Table* 1 *indicating lncRNA as possible precursor molecule of circular RNA. Since reported abundance of circular RNA is* 1% *of the total RNA pool only* [22, 34], *a* 5% *of the hit as referred above appears to be quite an acceptable figure.*

Result to construct Lemma 2 related to identification of RBPs as molecular agent for biogenesis

The lemma related to identification of RBPs as possible molecular agent stood on the following conjectures:

1. RBPs helping circularization from precursor lncRNA to circular RNA should first attach with lncRNA and remain attached till the process of circularization.
2. For the above conjecture to be true, RBPs solely binding with lncRNA or circular RNA might not be considered for molecular agent helping circularization.
3. RBPs common to lncRNA and circular RNA might be considered as agent for the circularization of RNA.
4. For the conjectures 1–3 to be true, there must be existence of reasonable amount of common RBPs.

Furthermore, for validation of this lemma, the role of RBPs common to both lncRNA and circular RNA was explored. Towards this direction, methodology section under subheading "Molecular agents involved in circularization" section was utilized to further corroborate this result with the help of RBPs unique to Circular and lncRNAs and common to both. Results of those RBPs were obtained from star-Base, Circinteractome and POSTAR. Figure 2 showed Venn diagram generated using Venny 2.1.0 tool [35] to show binding distribution of RBPs with different RNA types. In the pool of RBPs common to both RNA types, presence of QKI and FUS was found to be important to validate the lemma since both of these RBPs were reported to be involved in circularization.

Also, for validating both Lemmas 1 and 2, steps as described in methodology section under subheading "Identification of QKI and FUS RBPs, as common to both Circular and lncRNAs", were used to extract information on number of genes producing both lncRNA and circular RNA transcripts whereas, both of these transcripts had binding sites for FUS or QKI as shown in Table 2.

At the onset of this study, the target was to zero in the RBPs common to both the lncRNA and circRNA types. Therefore, the statistical significance of the identified genes was not a matter of concern, where the result of search was utilized to prove the basis of this research, which is: RBPs common to both of these RNA types are responsible for circularization. In this study, the first objective was to find out the lncRNAs present in the POSTAR database having the binding sites for RBPs (column A of Table 2). Second objective was to retrieve circRNA genes in the circInteratome database and homologous to the selected lncRNAs (column B of Table 2). Finally,
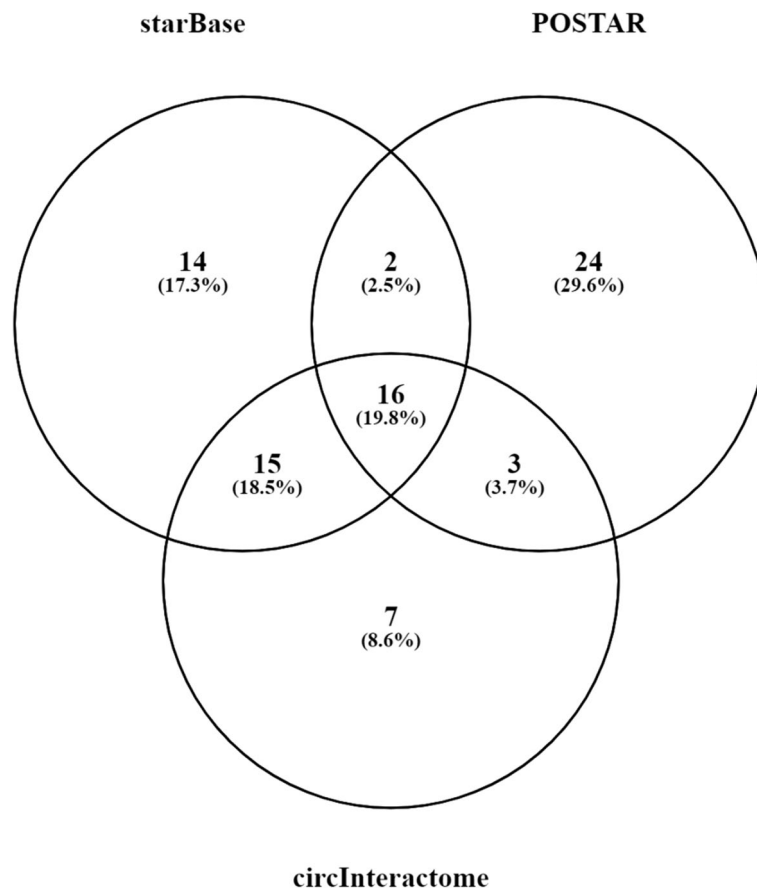
Fig. 2 Venn diagram showing percentage of RBP-types bound to both lncRNAs and Circular RNAs extracted from three different databases

Table 2 Common genes transcribing both RNA types with binding sites for FUS and QKI

RBP	lncRNA type	A	B	C
FUS	lincRNA	104	9	6
	Antisense	73	5	3
	sense_intronic	10	2	0
	sense_overlapping	9	1	0
QKI	lincRNA	47	6	1
	Antisense	24	0	0
	sense_intronic	1	1	0
	sense_overlapping	2	0	0

presence of RBP binding sites in the retrieved circRNA genes were searched manually (column C of Table 2). As binding sites for a specific RBP are not exactly identical therefore, skipping some sequences with minor variation during manual search method cannot be opted out.

Lemma 2 *Some of the RBPs common to both lncRNAs and circular RNA served as molecular agent for circularization of lncRNA.*

Result to construct Lemma 3 related to spatio-molecular distribution of lncRNA, exonic circular RNA and RBPs

Using the methodology section under subheading "Exploring cellular locations of lncR-NAs having sequence similarity with exonic circular RNAs" and "Exploring cellular locations of lncRNAs having no sequence similarity with exonic circular RNAs" in Methodology section, the localization profile of lncRNA was obtained and shown in Tables 3 and 4 respectively.

Taking cue from methodology section "Study on RBP types in relation to their binding with..", to obtain spatial distribution of RBPs, only those RBPs were considered, subcellular localization of which were reported in UniProtKB. As shown in Table 5, it was evident that RBPs that specifically bound to Circular RNA types were more

Table 3 Cellular spatio-molecular distribution of lncRNAs having sequence similarity with exonic circular RNA

Query circular RNA database	Distribution in cellular space (in %)				Reference lncRNA database
	Cytoplasm	Nucleus	Ribosome	Exosome	
CircRNADB	63.529	23.529	05.294	07.647	GENCODE
	59.782	25.543	08.152	06.521	Lncipedia
	58.875	24.807	06.945	09.261	NONCODE
CircBase	58.846	23.589	10.256	07.307	GENCODE
	57.900	24.900	11.700	05.500	Lncipedia
	55.700	23.100	10.800	08.300	NONCODE
Average	59.105 ± 2.577	24.245 ± 0.968	8.858 ± 2.476	7.423 ± 1.321	

Table 4 Cellular spatio-molecular distribution of lncRNAs which do not have any notable sequence similarity with exonic circular RNA

Query circular RNA database	Distribution in cellular space (in %)				Reference lncRNA database
	Cytoplasm	Nucleus	Ribosome	Exosome	
CircRNADB	52.700	24.000	12.800	10.500	GENCODE
	53.000	25.000	13.800	08.200	Lncipedia
	54.200	23.800	10.800	11.200	NONCODE
CircBase	51.500	25.800	13.700	09.000	GENCODE
	51.000	25.900	13.000	10.100	Lncipedia
	54.400	25.400	10.000	10.200	NONCODE
Average	52.8 ± 1.378	24.983 ± 0.899	12.35 ± 1.580	9.867 ± 1.084	

Table 5 The spatial distribution profile of RBPs unique for RNA types Circular and lncRNA

RNA type with which RBPs bound	Databases providing Information of RBPs related to RNA types	Sub cellular distribution of RBPs (in %) obtained from UniProtKB database against RBPs as queries			
		Cytosol only	nucleus only	Both nucleus and cytosol	Secreted
Circular RNA only	CircInteractome	26.90 ± 3.10	19.29 ± 9.29	53.81 ± 6.19	0 ± 0
lncRNA only	starBase & POSTAR	10 ± 10	51.88 ± 8.13	28.85 ± 8.85	9.38 ± 9.38
Both lncRNA and circular RNA	CircInteractome, starBase & POSTAR	18.60 ± 3.48	22.41 ± 9.36	58.99 ± 5.88	0 ± 0

localized within cytosol rather than nucleus of a cell and the contrary was true in case of RBPs bound to lncRNA types.

Facts obtained from Tables 3, 4 and 5 and publications report as building blocks of lemma 3 were as follows:

1. In general, lncRNAs were more localized within cytosol rather than nucleus of a cell.
2. Percentage of lncRNAs within cytosol that had considerable sequence similarity with circular RNA, was more than that having no notable sequence similarity with circular RNAs.
3. Exonic circular RNAs were more localized within cytosol [22, 27, 36] which is also corroborated with the outcome of Table 3.
4. Percentage of RBPs which bound with circular RNA only and localized both in nucleus and cytosol was much more than that localized in cytosol or nucleus only. However, since as per our hypothesis this RBPs have no role in circularization, these RBPs are not considered further.
5. Percentage of RBPs which bound with lncRNA only and localized in nucleus was much more than that localized in cytosol or both cytosol and nucleus. This fact together with facts 1, 2 and 3 indicated that these RBPs had hardly any role in circularization.
6. Percentage of RBPs which were common both for lncRNA and circular RNA and localized in both nucleus and cytosol was much more than that localized in cytosol or nucleus only. Furthermore, focusing on localization of this type of RBPs in cytosol only or nucleus only, their frequency was found to be reasonably greater in nucleus than cytosol. This facts together with facts 1, 2, 3 and 4 indicated that these RBPs played more significant role in circularization of RNA in comparison to other RBP types.
7. QKI was localized both in cytosol and nucleus whereas FUS was localized in nucleus.

Lemma 3   *There exists a probability that some fraction of lncRNAs were exported from nucleus to cytosol. Therefore it cannot be ruled out that RBPs common to both lncRNA and circular RNA binds with those lncRNAs within cytosol for final circularization into exonic circular RNA form along with some probability that the these RBP types lead to formation of exonic circular RNAs within nucleus.*

Also, in addition to Lemma 3 *the final lemma* (*Lemma* 4) *corroborates the availability of the machinery required for the circularization of the lncRNA in the cytoplasm.*

Lemma 4   *Minor spliceosome and circularization of lncRNA.*

Partially spliced pre-mRNAs containing minor-class introns undergo nuclear export followed by splicing via U12 spliciosome in the cytoplasm [37–39]. Splicing of pre-tRNAs in the cytoplasm also previously reported [40]. Therefore, lncRNA originated through inefficient splicing might undergo circulization though backsplicing by minor spliciosome at cytoplasm.

Kumar *et al. BMC Bioinformatics*   (2023) 24:148

Page 9 of 21

Integrated cellular geography approach to elucidate biogenesis of Circular RNA

As stated earlier in introduction, methodology and initial part of result section, ICG approach considered to build a theorem through assimilation of lemmas supporting it. Following this approach we first built the proposition in the form of a model that there should be existence of a precursor molecule of circular RNA which upon the involvement of some molecular agents would yield exonic circular RNA as shown in Fig. 4.

Towards this direction, the lemmas (Lemmas 1, 2 and 3) as a result of outcomes extracted in this work from various published experiments (as shown in Tables 1, 2, 3, 4, 5) led to formation of the theorem as shown in the Table 6 and Fig. 3.

Discussion

In this work, mechanism of biogenesis of exonic circular RNA was studied from the angle of sequence semantics and assimilation of apparently unconnected published data under the paradigm of ICG as described earlier. Sequence semantics helped in identification of precursor molecule as lncRNA that was supposed to convert into circular RNA form. The biogenesis mechanism was studied considering the spatio-molecular aspect of lncRNAs, circular RNAs and RBPs within cellular space. Overall, ICG was employed to decipher the mechanism of biogenesis of circular RNA under the theoretical formalism as described in methodology section under subheading Integrated Cellular Geography (ICG) formalism to study biogenesis of circular RNA". The concept of ICG was pressed in for such task as a generalization of naturally evolved solution instead of being forced upon into this study. The reason was a plethora of already published reports on experiments done on circular RNA and allied biomolecules which were although unconnected apparently in their present form, yet could be studied further in their assimilated form for confirming a target objective, i.e., a theorem. Although methodology for assimilation of published data, like Meta Analysis existed, it was found to be applicable only for data generated using same experimental protocol. The essence of ICG was actually that of Integrated Geography at microscopic, i.e., cellular scale which was further generalized in this work in a logical lemma based formalism as especially shown in the methodology section under subheading "Integrated Cellular Geography (ICG) formalism to study biogenesis of circular RNA" and the result section.

Circular RNAs are known as biomarkers of quite a number of diseases including cancers. This work was motivated from this fact to partially contribute to the knowledge on events leading to biogenesis of Circular RNA assuming it to be central to unearth pathobiochemical mechanism of these diseases. In this regard, the problem was divided into two segments, first to search for molecular precursors being converted into circular form and identification of factors leading to its circularization to produce Circular RNA. Secondly, the objective was to extract mechanism of biogenesis of circular RNA from the spatio-molecular distribution of lncRNA, circular RNA and RBPs. As described in the result section, lemmas were extracted as supports for solving both of these problems. These lemmas were constructed using existing databases and results of published works as research materials and through judicious assimilation of those under ICG.

To solve the first problem of identifying molecular precursor for Circular RNAs, lncRNAs were considered as primary guess material through such indication from the

Table 6 Assimilation of four lemmas to form the alternative theorem of biogenesis of exonic circular RNA

| Lemma serial no.	1	2	3	4
Extracts from Lemmas	lncRNA is possible precursor of circular RNA	Some of the RBPs common to both lncRNAs (A) and circular RNAs served as catalyst for circularization of lncRNA	Circularization of lncRNA is mediated by RBP type A both at nucleus and cytosol	Splicing machinery is abundant in cytoplasm and tRNA circularization happens in cytoplasm
Theorem	Biogenesis of exonic circular RNA is mostly post-transcriptional event which starts with the binding of RBPs common to both lncRNAs and circular RNAs with precursor molecule lncRNAs with some fraction within nucleus and also some within cytosol with exceptions following spatio-molecular arrangements of lncRNA, circular RNA and all types of RBPs			
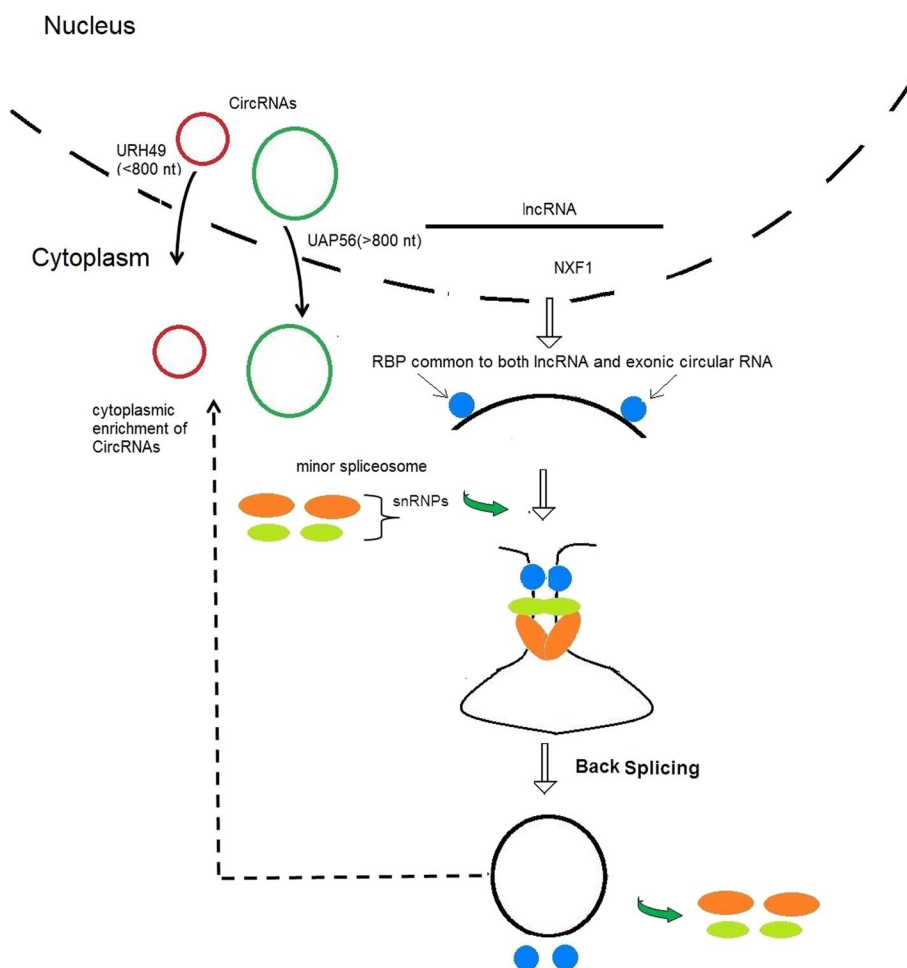
Fig. 3 Model describing different stages of circularization of lncRNA (black line and curves) with the help of RBPs common to both Circular (black circles) and lncRNAs

published research works [15, 26]. Since there was no solid evidence or proof so far for lncRNAs being circularized to form Circular RNAs, semantic similarity in sequence order between lncRNAs and Circular RNAs was first explored to establish physico-spatial relationship between these two entities. Moreover, the lncRNAs following the rules set in this study were found to have resemblance with antisense properties, which was one more attribute common for Circular RNA [22, 26, 34]. A good proportion (∼ 32%) of antisense transcripts in human was reported as lncRNAs [14]. These two findings indicated fulfillment of necessary condition for lncRNAs to be considered as molecular precursor of Circular RNA. In an apparent contradiction, although lncRNAs were sometimes found to be transcribed from coding genes, it was intriguing to find its ultimate form as whole or partial natural antisense (i.e., non-coding) transcripts (NAT) as reported by Milligan and Lipovich [41]. To check authenticity of this primary guess apart from fulfilling the requisite necessary conditions, methodology section described in subheading "Identification of possible precursor molecules of circular RNA" section was employed in this work which searched for existence of completely mapped lncRNA sequence-hits against circular RNA sequence as query. Table 1 of result section under

subheading "Result to construct Lemma 1 related to identification of potential molecular precursors of Circular RNA" section provided the outcome of this exercise from which the lemma 1 was extracted. According to lemma 1, "reasonable percentage of lncRNA (~5%) out of total hits obtained was found to have complete sequence similarity with circular RNA as shown in Table 1 indicating lncRNA as possible precursor molecule of circular RNA". Since reported abundance of circular RNA is 1% of the total RNA pool only, a 5% of the hit as referred above appears to be quite an acceptable figure.

In continuation with first problem, identification of factors leading to circularization of RNA was tried to be resolved using methodology section under subheading "Molecular agents involved in circularization" section the result of which was shown in result section under subheading "Result to construct lemma 2 related to identification of RBPs as molecular agent for biogenesis" section. In this regard, the findings of result section under subheading "Result to construct lemma 2 related to identification of RBPs as molecular agent for biogenesis" section were used to build hypothesis about the possible class of RBPs responsible for biogenesis of circular RNA through its circularization. In this regard, Fig. 2 showed the binding distribution of RBPs with different RNA types indicating presence of quite high amount of RBPs (~40.32% on an average) common to both lncRNAs and circular RNAs. Authenticity of this result was further validated through the study on the RBPs, FUS [30] and QKI [5] which were known as responsible for biogenesis of circular RNA and also common to both lncRNAs and circular RNAs as described in methodology section under subheading "Identification of QKI and FUS RBPs, as common to both Circular and lncRNAs" and outcome shown in Table 2. All these results together provided the basis of the construction of lemma 2 which was: "Some of the RBPs common to both lncRNAs and circular RNAs served as molecular agent for circularization of lncRNA". It was intriguing to find that lemma 2 stood as a support for lemma 1 also.

Regarding solution of the second problem, in this work it appeared to be interesting to extract spatio-molecular root of Circular RNA via its relationship with RBPs and earlier molecular form i.e., as indicated from this study, lncRNAs. That said, problem reduced to find out overall spatial distribution of lncRNAs, Circular RNAs and RBPs within nucleus and cytosol of cells. Regarding this, the results shown in Tables 3, 4 and 5 using methodology sections, under subheading "Exploring cellular locations of lncRNAs having sequence similarity with exonic circular RNAs", subheading "Exploring cellular locations of lncRNAs having no sequence similarity with exonic circular RNAs" and subheading "Study on RBP types in relation to their binding with lncRNA and circular RNA along with differential profile of their existence within nucleus and cytosol following ICG" together with the facts generated in result under subheading "Result to construct lemma 3 related to spatio-molecular distribution of lncRNA, exonic circular RNA and RBPs" section provided the basis for formulation of lemma 3 from which the lemma 4 followed subsequently.

In the context of utilization of existing databases and apparently unconnected published reports, as stated in the beginning of this discussion and methodology section under subheading "Integrated Cellular Geography (ICG) formalism to study biogenesis of circular RNA", it appears to be quite justified to employ the formalism of integrated cellular geography (ICG) following the concept of general Integrated Geography (IG)

[25]. It grossly described spatial nature of relationship between objects, the environment holding them and the world containing both of them, which found their equivalence with RBPs, RNA types and whole cellular space comprising of cytosol and nucleus respectively. The main purpose of employing this approach was to find the origin and historical dynamism of an object in relation to the spatio-molecular environment holding it. Interestingly the lemmas extracted out of this approach appeared to be complementarily supportive to each other to lead to the final theorem (please see Table 6) on biogenesis of exonic circular RNA which was: "Biogenesis of exonic circular RNA is mostly post-transcriptional event which starts with the binding of RBPs common to both lncRNAs and circular RNAs with precursor molecule lncRNAs with some fraction within nucleus and also some within cytosol with exceptions following spatio-molecular arrangements of lncRNA, circular RNA and all types of RBPs". The salient feature of this approach is that, similar to lemma based study of a mathematical theorem, complexity in reasoning leading to an inference from apparently unconnected biological data and reports can be reduced substantially where lemmas served as filtered knowledge.

The interesting part of this work was to consider studying the role of RBPs in circularization of RNAs. In this regard, it appeared to be quite obvious that RBPs specific for Circular RNAs might not have any role for their biogenesis since they would bind with circular form only, which was not possible before circularization. The inference stemmed from the same reasoning led to the fact that RBPs common to both of these RNA types should be prime molecular factors that could be expected to help circularization of lncRNAs and thus would have to remain attached throughout transition of lncRNA to Circular RNA. This fact was further supported by the presence of such RBPs, QKI and FUS with their established role in circularization [5, 30]. However, in this work the role of QKI and FUS as a common RBP of both Circular and lncRNA types could also be found from analyses described in methodology section under subheading "Study on RBP types in relation to their binding with lncRNA and circular RNA along with differential profile of their existence within nucleus and cytosol following ICG" and Table 5 of result section under subheading "Result to construct lemma 3 related to spatio-molecular distribution of lncRNA, exonic circular RNA and RBPs" section. Therefore, role of other common RBPs remained to be an interesting area of research in relation to their role as molecular factor for circularization. Figure 3 further clarified this model. Furthermore, as reported by Barrett et al. (2017), a Circular RNA, ciRS-7 was also found to be embedded in LINC00632 which was an lncRNA. This fact was also found to be in support of our theorem [28].

However, for RBPs specific for lncRNAs the chance of their having role in the transition from linear to circular form of RNA being less following the above-mentioned reasoning, discussion on them was avoided although possibility of their involvement might not be overruled.

We have already discussed the features: sequence semantics, RBPs, and localization due to which lncRNAs have been considered as template molecules in our study. There are various types of circular RNAs, e.g., exonic, intronic etc., indicating possible existence of different precursor molecules responsible for their biogenesis. This work is actually intended to establish one among few more such relationships to see whether the present circular RNA in question is rooted from lncRNA. However, we

can proceed with other RNAs molecules adopting the similar methodology. dsRNA has been indicated as interconvertible RNA with circular RNAs [42], and can be the potential template for future study.

Conclusion

Overall, this study showed a possible model of biogenesis of exonic circular RNAs from lncRNAs as precursor molecule on the foundation of judicious fusion of results and data of already performed experiments under the paradigm of integrated cellular geography approach as analytical basis. Sequence analysis of both RNA types revealed significant similarity making the foundation of a logical lemma based study related to possibility of biogenesis of Circular RNAs from lncRNAs. Additionally, information extracted in this work from distribution of these RNA types within cellular spaces, cytosol and nucleus, in relation to their binding with specific types of RBPs further helped in building of other two lemmas in support of spatio-molecular mechanism of biogenesis of exonic circular RNA. The theorem built from these four complimentarily supportive lemmas strongly supported model of biogenesis of Circular RNA as provided in this work. According to this model: Biogenesis of exonic circular RNA is mostly post-transcriptional event which starts with the binding of RBPs common to both lncRNAs and circular RNAs with precursor molecule lncRNAs with some fraction within nucleus and also some within cytosol with exceptions following spatio-molecular arrangements of lncRNA, circular RNA and all types of RBPs. Therefore, this model indicated that the biogenesis of exonic circular RNA was mostly a post-transcriptional event. The concept of ICG, as first time introduced in this work, appeared to have prospective of a very useful methodological ingredient of systems biology for holistic study of events of cellular system.

Materials and methods

Integrated Cellular Geography (ICG) formalism to study biogenesis of circular RNA

Integrated Geography (IG) in its definition is the branch of geography that describes and explains the spatial aspects of interactions between human individuals or societies and their natural environment, called coupled human–environment systems [25]. Since biological cell can be thought of as a micro-geographic space, Integrated Cellular Geography (ICG) was employed as an alternative Systems Engineering approach to get more in-depth information about the system through pulling in apparently unconnected published results and putting them together under a framework of Data Science [43]. In this study, the equivalence of IG with ICG was drawn by substituting human by concerned biomolecules and geographic environment by cellular environment. A mathematical framework of this approach is given below.

For a pool of logical lemma, $L = \{l_i\}_{i=1}^{N}$, if these lemmas.

1. serve as supports to prove theorem, $T = f(l)$, for f is a function in broader sense which is derived through judiciously made association of all the members of L, and,
2. represent complimentary form of necessary and sufficient proof of T,

We may consider the analysis of the system is compatible and doable under the broad paradigm of Integrated Data Geography (IDG) and specifically, for this work, Integrated Cellular Geography (ICG).

In our work which dealt with biomechanical root of formation of circular RNA, the geographical space of our choice was a biological cell and its very important sub-cellular spaces (nucleus and cytosol) from where the referred biomolecule was generated. We analyzed the spread of this biomolecular entity outside and inside the sub-cellular space along with assimilating other reported pieces of information (e.g., about RBP molecules involved in circularization of such molecules) to come up with necessary and sufficient conditions of root of formation of such molecule.

Collection of data and Pre processing

Data collection strategy

The objective of data collection for this work was primarily to study the effectiveness of the newly proposed Integrated Cellular Geography approach in predicting biogenesis of circular RNA. Under this approach, following two types of data were considered for their utilization:

1. Data type 1 (DT1): Apparently unconnected reports of published papers, and,
2. Data type 2 (DT2): Formal data collected from databases.

Therefore for the model constructed under this approach, the intention was to look into the convergence of DT1 and DT2 to extract information on:

1. possible precursor molecules of circular RNA,
2. molecular agents involved in circularization, and,
3. cellular spatiotemporal condition supporting circularization.

Collection of data for identifying possible precursor molecules of circular RNA

Considering the specific class of circular RNA that were exonic and were intuitively formed from lncRNAs, the sequences of circular RNAs were collected from circBase [44] and circRNADb [45], while lncRNAs sequence were downloaded from GENCODE [14], NONCODE [46] and LNCipedia [47]. The current version of the dataset was downloaded like gencode.v19.lncRNA_transcripts.fa from GENCODE, high confidence lncRNAs lncipedia_5_0_hc.fasta from LNCipedia and NONCODEv5_human.fa from NONCODE. These datasets were pre-processed by writing Perl scripts to extract relevant information for further study.

Collection of data for identifying factors leading to circularization of RNA

As mentioned in the introduction section, RNA Binding Proteins were reported to be involved in the circularization of some RNAs. To get a comprehensive view and extract relevant information, circInteractome [31], starBase [32] and POSTAR [48] databases were used. starBase and POSTAR are the databases where binding of RBPs on lncRNAs are available, whereas Circinteractome contains the RBP binding information on Circular RNAs.

Collection of data for detecting cellular spatiotemporal condition supporting circularization

For this purpose, online database and web server were used by taking specific queries like, RBP-types or lncRNAs to search out their special location inside cellular spaces. UniProtKB [49] was used for getting the annotation of RBPs which included their localization information. Also web server iLoc-LncRNA [50] was used to get locations of lncRNAs within cellular spaces.

Identification of possible precursor molecules of circular RNA

Mapping the circular RNAs sequence on lncRNAs

In this phase, the objective was to search out the region of homology of circular RNAs on lncRNAs to initially check if there was any indication for lncRNAs to serve as precursor molecules of circular RNAs. For this purpose, we have taken two circular RNA databases, circBase and circRNADb in our study. Also for lncRNAs we have taken datasets from three lncRNAs database like GENCODE, LNCipedia and NONCODE. Following dataset, query-set and tools were used for the mapping:

1. Datasets for reference data of BLAST [51] tool: GENCODE, LNCipedia and NON-CODE
2. Query-set: Queries obtained from circular RNA datasets, circBase and circRNADb
3. Search tool: BLASTN of BLAST+ 2.8.0 package.

Mapping process was given in Table 7.

The process was performed for all possible combinations of query and reference databases.

Molecular agents involved in circularization

Study on RBP types in relation to their binding with lncRNA and circular RNA

along with differential profile of their existence within nucleus and cytosol following ICG

As clarified in the introduction section, in this work, role of some RBPs was studied in relation to circularization of RNA. Therefore, it was important to get account of their

Table 7 Steps for mapping of circular RNA query onto lncRNA reference dataset

Mapping steps	Executable used	Parameter	Value of parameters
Creation of BLAST database	makeblastdb	-in	lncRNA database sequences
		-input_type	'fasta'
		-dbtype	'nucl'
		-out	BLAST_generated_lncRNA_database_name
Searching homology	blastn	-query	circular RNAs database sequences
		-task	'blastn'
		-db	BLAST_generated_lncRNA_database_name
		-evalue	1e-05
		-outfmt	6
		-perc_identity	100
		-num_threads	8
		-out	Output_file name
		-max_target_seqs	1
		-max_hsps	1

spatial existence especially within nucleus and cytosol of a cell along with information on the types of RNAs they bound. Therefore, databases, circInteractome [31], starBase [32] and POSTAR [48] were utilized to extract information on RBPs unique to Circular and lncRNAs as well as common to both.

Subsequently percentage existences of RBPs for each of the RNA-types along with RBPs unique to both Circular and lncRNAs were computed in:

1. Cytosol only
2. Nucleus only
3. Both nucleus and cytosol, and,
4. Secreted from the cell

To accomplish this task, supplementary files extracted from databases, CircInteractome and starBase were utilized to extract RBPs and RNAs they bound to by writing Perl scripts to read these files. However, in case of POSTAR database following methodological steps were utilized to extract list of RBPs for lncRNAs:

Step 1 First, POSTAR.csv file was downloaded and the list of lncRNAs was extracted by parsing of this file.
Step 2 Each of the lncRNAs within the list was submitted as query to POSTAR to find out the RBPs having binding sites on it.
Step 3 The RBPs extracted in step 2 were further processed to eliminate redundant instances of them.

Furthermore, to get information on spatial location of RBPs, each one of them was inputted into UniProtKB [36] to get its spatial location from the output annotated data.

Identification of QKI and FUS RBPs, as common to both Circular and lncRNAs

Among many different types of functions mediated by QKI type of RBPs, in this work, its role for circularization was investigated from the report [5]. This report highlighted that:

1. Epithelial to mesenchymal transition (EMT) coincides with regulation of many circular RNAs,
2. Circular RNA formation was regulated by QKI during EMT,
3. Binding sites of QKI was found to be flanking circRNA-forming exons, and,
4. Circularization of linear RNA could be induced by insertion of QKI binding site into them.

Similarly the role of another RBP, FUS in regulating circularization was documented in [30]. Therefore, it was imperative in this study to investigate on RBPs common to both Circular and lncRNAs which was carried out using existing literatures and databases, POSTAR and CircInteractome. POSTAR provided the information about the post transcriptional regulatory role of RBPs while CircInteractome gave information of circular RNAs, their RBPs and microRNAs. Towards this direction, the methodology adopted comprised of two stages:

Stage1: At this stage following steps were applied to get first hand information about RNAs for a specific RBP-type:

Step 1 Specific RBP (either QKI or FUS) was submitted as query in POSTAR database.

Step 2 The database returned the names of the genes and corresponding RBP-type binding sites on RNAs transcribed from these genes.

Stage2: Since direct description of RNAs (whether mRNA or lncRNA) could not be extracted at stage 1, following steps were applied to get such information:

Step 1 Out of the output gene pools extracted through stage 1, only those transcribing for lncRNAs were screened through writing Perl script.

Step 2 Each of the screened genes was submitted as query in CircInteractome database and the outputs yielding at least 1 hit were chosen only.

Step 3 Each of the outputs obtained from step 2 was further analyzed to check manually the RNA types transcribed by this gene (say, Circular RNA, since this database checks for Circular RNA only), RBP types for these RNA types (say, QKI or FUS) and binding sites for these RBPs on these RNA types. Only those genes transcribing for Circular RNA with specific RBP-types (QKI or FUS) were retained for further analysis.

Cellular spatiotemporal condition supporting circularization

Exploring cellular locations of lncRNAs having sequence similarity with exonic circular RNAs

As described in subheading under the methodology section "Identification of possible precursor molecules of circular RNA", to examine whether exonic circular RNAs can be originated from lncRNAs following steps were performed using fixed number of Circular RNAs and lncRNAs obtained from the process mentioned above.

1. Based on the total BLAST hits obtained only those hits were extracted (A) in which the circular RNAs sequences fully mapped in lncRNAs.
2. Using A, specific web-server, iLoc-LncRNA ('http://lin-group.cn/server/iLoc-LncRNA/predictor.php') was used to know probable spatial location of these lncRNAs inside the cellular space.

Exploring cellular locations of lncRNAs having no sequence similarity with exonic circular RNAs

Following steps were performed for this purpose:

1. Based on the total BLAST hits obtained only those hits were extracted (A) in which the circular RNAs sequences were not at all mapped in lncRNAs.
2. Finally cellular location of such lncRNAs was found using iLoc-LncRNA web server.

To reduce computational complexity in dealing with very large database where number of data is greater than 1000, sample datasets were used through random selection of data from the original database.

Workflow depicting the ICG model for biogenesis of circular RNA was shown in Fig. 4.
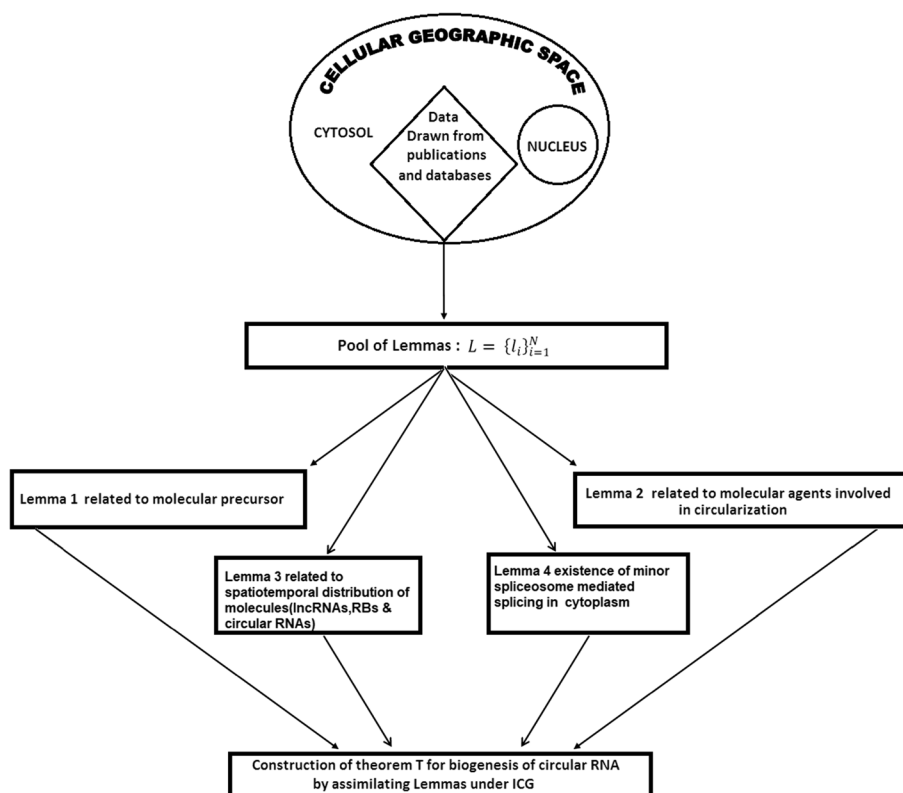
Fig. 4 Workflow based ICG model for biogenesis of circular RNA

Abbreviations

lncRNAs	Long non-coding RNAs
Circular RNAs	CircRNAs
IDG	Integrated Data Geography
ICG	Integrated Cellular Geography
IG	Integrated Geography
RBPs	RNA binding proteins
DPE	Differential profile of existence
ENCODE	ENCyclopedia Of DNA Elements
circRNADb	Circular RNA Database
ORF	Open Reading Frame
EMT	Epithelial-mesenchymal transition
FIRRE	Firre intergenic repeating RNA element
RIDLs	Repeat Insertion Domains of LncRNAs

Author contributions
RK: led the implementation of the work and supported in writing the manuscript. RM: supported implementation of the work and helped TL in preparing figures. TL: conceptualized the work and wrote the main manuscript. MP supported implementation of the work. All authors read and approved the final manuscript.

Availability of data and materials
All the data used in this work are available with the corresponding author and can be shared on such demand.

Declarations

Ethics approval and consent to participate
Not applicable.

References

1. Ivanov A, Memczak S, Wyler E, Torti F, Porath HT, Orejuela MR, Piechotta M, Levanon EY, Landthaler M, Dieterich C, Rajewsky N. Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. Cell Rep. 2015;10(2):170–7.
2. Zhang XO, Wang HB, Zhang Y, Lu X, Chen LL, Yang L. Complementary sequence-mediated exon circularization. Cell. 2014;159(1):134–47.
3. Zhang Y, Xue W, Li X, Zhang J, Chen S, Zhang JL, Yang L, Chen LL. The biogenesis of nascent circular RNAs. Cell Rep. 2016;15(3):611–24.
4. Ashwal-Fluss R, Meyer M, Pamudurti NR, Ivanov A, Bartok O, Hanan M, Evantal N, Memczak S, Rajewsky N, Kadener S. circRNA biogenesis competes with pre-mRNA splicing. Mol Cell. 2014;56(1):55–66.
5. Conn SJ, Pillman KA, Toubia J, Conn VM, Salmanidis M, Phillips CA, Roslan S, Schreiber AW, Gregory PA, Goodall GJ. The RNA binding protein quaking regulates formation of circRNAs. Cell. 2015;160(6):1125–34.
6. Kelly S, Greenman C, Cook PR, Papantonis A. Exon skipping is correlated with exon circularization. J Mol Biol. 2015;427(15):2414–7.
7. Li X, Liu CX, Xue W, Zhang Y, Jiang S, Yin QF, Wei J, Yao RW, Yang L, Chen LL. Coordinated circRNA biogenesis and function with NF90/NF110 in viral infection. Mol Cell. 2017;67(2):214–27.
8. Wu H, Yang L, Chen LL. The diversity of long noncoding rnas and their generation. Trends Genet. 2017;33(8):540–52.
9. Dang Y, Yan L, Hu B, Fan X, Ren Y, Li R, Lian Y, Yan J, Li Q, Zhang Y, Li M, Ren X, Huang J, Wu Y, Liu P, Wen L, Zhang C, Huang Y, Tang F, Qiao J. Tracing the expression of circular RNAs in human pre-implantation embryos. Genome Biol. 2016;17(1):130.
10. Xi X, Li T, Huang Y, Sun J, Zhu Y, Yang Y, Lu ZJ. RNA biomarkers: frontier of precision medicine for cancer. Noncoding RNA. 2017;3(1):9.
11. Brockdorff N, Ashworth A, Kay GF, McCabe VM, Norris DP, Cooper PJ, Swift S, Rastan S. The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell. 1992;71(3):515–26.
12. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature. 2009;458(7235):223–7.
13. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. Integrative annotation of human large inter-genic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. 2011;25(18):1915–27.
14. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, Lagarde J, Veeravalli L, Ruan X, Ruan Y, Lassmann T, Carninci P, Brown JB, Lipovich L, Gonzalez JM, Thomas M, Davis CA, Shiekhattar R, Gingeras TR, Hubbard TJ, Notredame C, Harrow J, Guigó R. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. 2012;22(9):1775–89.
15. Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function. Nat Rev Genet. 2016;17(1):47–62.
16. Yoon JH, Abdelmohsen K, Gorospe M. Posttranscriptional gene regulation by long noncoding RNA. J Mol Biol. 2013;425(19):3723–30.
17. Batista PJ, Chang HY. Long noncoding RNAs: cellular address codes in development and disease. Cell. 2013;152(6):1298–307.
18. Gomes AQ, Nolasco S, Soares H. Non-coding RNAs: multi-tasking molecules in the cell. Int J Mol Sci. 2013;14(8):16010–39.
19. Wilusz JE. Long noncoding RNAs: Re-writing dogmas of RNA processing and stability. Biochim Biophys Acta. 2016;1859(1):128–38.
20. Zhang Y, Yang L, Chen LL. Life without A tail: new formats of long noncoding RNAs. Int J Biochem Cell Biol. 2014;54:338–49.
21. Bridges MC, Daulagala AC, Kourtidis A. LNCcation: lncRNA localization and function. J Cell Biol. 2021;220(2):e202009045.
22. Salzman J, Gawad C, Wang PL, Lacayo N, Brown PO. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. PLoS ONE. 2012;7(2):e30733.
23. Chen I, Chen CY, Chuang TJ. Biogenesis, identification, and function of exonic circular RNAs. Wiley Interdiscip Rev RNA. 2015;6(5):563–79.
24. Li Z, Kearse MG, Huang C. The nuclear export of circular RNAs is primarily defined by their length. RNA Biol. 2019;16(1):1–4.
25. Castree N, Demeritt D, Liverman D. Introduction: making sense of environmental geography. In: Castree N, Demeritt D, Liverman D, Rhoads B, editors. A Companion to environmental geography. Oxford: Wiley; 2009. p. 1–15.
26. Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak SD, Gregersen LH, Munschauer M, Loewer A, Ziebold U, Landthaler M, Kocks C, le Noble F, Rajewsky N. Circular RNAs are a large class of animal RNAs with regulatory potency. Nature. 2013;495(7441):333–8.

27. Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu J, Marzluff WF, Sharpless NE. Circular RNAs are abundant, conserved, and associated with ALU repeats. RNA. 2013;19(2):141–57.

28. Barrett SP, Parker KR, Horn C, Mata M, Salzman J. ciRS-7 exonic sequence is embedded in a long non-coding RNA locus. PLoS Genet. 2017;13(12):e1007114.

29. Zhang J, Chen S, Yang J, Zhao F. Accurate quantification of circular RNAs identifies extensive circular isoform switching events. Nat Commun. 2020;11(1):90.

30. Errichelli L, Dini Modigliani S, Laneve P, Colantoni A, Legnini I, Capauto D, Rosa A, De Santis R, Scarfò R, Peruzzi G, Lu L, Caffarelli E, Shneider NA, Morlando M, Bozzoni I. FUS affects circular RNA expression in murine embryonic stem cell-derived motor neurons. Nat Commun. 2017;8:14741.

31. Dudekula DB, Panda AC, Grammatikakis I, De S, Abdelmohsen K, Gorospe M. CircInteractome: a web tool for exploring circular RNAs and their interacting proteins and microRNAs. RNA Biol. 2016;13(1):34–42.

32. Li JH, Liu S, Zheng LL, Wu J, Sun WJ, Wang ZL, Zhou H, Qu LH, Yang JH. Discovery of protein-lncRNA interactions by integrating large-scale CLIP-Seq and RNA-Seq datasets. Front Bioeng Biotechnol. 2015;2:88.

33. Mattick JS, Rinn JL. Discovery and annotation of long noncoding RNAs. Nat Struct Mol Biol. 2015;22(1):5–7.

34. Salzman J, Chen RE, Olsen MN, Wang PL, Brown PO. Cell-type specific features of circular RNA expression. PLoS Genet. 2013;9(9):e1003777. https://doi.org/10.1371/journal.pgen.1003777.

35. Oliveros JC. VENNY. An interactive tool for comparing lists with Venn Diagrams. 2007; http://bioinfogp.cnb.csic.es/tools/venny/index.html.

36. Jeck WR, Sharpless NE. Detecting and characterizing circular RNAs. Nat Biotechnol. 2014;32:453–61.

37. König H, Matter N, Bader R, Thiele W, Müller F. Splicing segregation: the minor spliceosome acts outside the nucleus and controls cell proliferation. Cell. 2007;131(4):718–29.

38. Caceres JF, Misteli T. Division of labor: minor splicing in the cytoplasm. Cell. 2007;131(4):645–7.

39. Steitz JA, Dreyfuss G, Krainer AR, Lamond AI, Matera AG, Padgett RA. Where in the cell is the minor spliceosome? Proc Natl Acad Sci USA. 2008;105(25):8485–6.

40. Yoshihisa T, Ohshima C, Yunoki-Esaki K, Endo T. Cytoplasmic splicing of tRNA *in Saccharomyces cerevisiae*. Genes Cells. 2007;12(3):285–97.

41. Milligan MJ, Lipovich L. Pseudogene-derived lncRNAs: emerging regulators of gene expression. Front Genet. 2015;5:476.

42. Chen YG, Hur S. Cellular origins of dsRNA, their recognition and consequences. Nat Rev Mol Cell Biol. 2022;23(4):286–301.

43. Hayashi C. What is data science? Fundamental concepts and a heuristic example. In: Hayashi C, Yajima K, Bock H-H, Ohsumi N, Tanaka Y, Baba Y, editors. Data Science classification and related methods. Tokyo: Springer; 1998. p. 40–51.

44. Glažar P, Papavasileiou P, Rajewsky N. circBase: a database for circular RNAs. RNA. 2014;20(11):1666–70.

45. Chen X, Han P, Zhou T, Guo X, Song X, Li Y. circRNADb: A comprehensive database for human circular RNAs with protein-coding annotations. Sci Rep. 2016;6:34985.

46. Zhao Y, Li H, Fang S, Kang Y, Wu W, Hao Y, Li Z, Bu D, Sun N, Zhang MQ, Chen R. NONCODE 2016: an informative and valuable data source of long non-coding RNAs. Nucleic Acids Res. 2016;44(D1):D203-208.

47. Volders PJ, Helsens K, Wang X, Menten B, Martens L, Gevaert K, Vandesompele J, Mestdagh P. LNCipedia: a database for annotated human lncRNA transcript sequences and structures. Nucleic Acids Res. 2013;65:D246–51.

48. Hu B, Yang YT, Huang Y, Zhu Y, Lu ZJ. POSTAR: a platform for exploring post-transcriptional regulation coordinated by RNA-binding proteins. Nucleic Acids Res. 2017;45(D1):D104–14.

49. Breuza L, Poux S, Estreicher A, Famiglietti ML, Magrane M, Tognolli M, Bridge A, Baratin D, Redaschi N; UniProt Consortium. The UniProtKB guide to the human proteome. Database (Oxford). 2016; 2016.

50. Su ZD, Huang Y, Zhang ZY, Zhao YW, Wang D, Chen W, Chou KC, Lin H. iLoc-lncRNA: predict the subcellular location of lncRNAs by incorporating octamer composition into general PseKNC. Bioinformatics. 2018;34(24):4196–204.

51. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.

Publisher's Note