Available online at **www.ijmrhs.com**

ISSN No: 2319-5886

International Journal of Medical Research & Health Sciences, 2019, 8(6): 54-63

Current Challenges of Computational Intelligent Techniques for Functional Annotation of ncRNA Genes

Qaisar Abbas*

College of Computer and Information Sciences, Al Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia

**Corresponding e-mail: qaabbas@imamu.edu.sa*

ABSTRACT

The limited understanding of functional annotation of non-coding RNAs (ncRNAs) has been largely due to the complex functionalities of ncRNAs. They perform a vital part in the operation of the cell. There are many ncRNAs available such as micro RNAs or long non-coding RNAs that play important functions in the cell. In practice, there is a strong binding of the function of RNAs that must be considered to develop computational intelligent techniques. Comprehensive modeling of the structure of an ncRNA is essential that may provide the first clue towards an understanding of its functions. Many computational techniques have been developed to predict ncRNAs structures but few of them focused on the functions of ncRNA genes. Nevertheless, the accuracy of the functional annotation of ncRNAs is still facing computational challenges and results are not satisfactory. Here, many computational intelligent methods were described in this paper to predict the functional annotation of ncRNAs. The current literature review is suggested that there is still a dire need to develop advanced computational techniques for functional annotating of ncRNA genes in terms of accuracy and computational time.

Keywords: Genes, Non-coding RNAs, Artificial intelligence, Machine learning, Functional annotation, Computational methods, Deep learning

INTRODUCTION

The discovery of non-coding RNA has made us realize that coding DNA is highly relevant in the regulation of gene expression. Due to their vital functional role in the cell operation, this type of genes has been used in many studies. Compare to coding RNAs, the recent studies showed that the non-coding RNAs (ncRNAs) play a vital role in the functional side of the cell. Nowadays, the researchers are building computational intelligent techniques for prediction of ncRNAs structure and functional annotation of ncRNAs genes [1,2]. In the case of the genome sequence, the authors utilized many ncRNA genes for classification of RNA structures and functional discovery [3-5]. As a result, the authors described in different studies that it became obvious important to predict RNA structure along with the function. However, a few studies computational intelligence techniques focused on developing intelligent system for ncRNAs genes.

MicroRNAs (miRNAs/miRs) are newly discovered that have changed the understanding of how genes are regulated. Several studies were developed based on conventional and advanced machine learning algorithms to detect the functional annotation of ncRNAs genes. But still, it is unknown for many scientists, some of the important functions of ncRNAs that must be discovered. Long non-coding RNA (lncRNAs) transcripts micro RNAs (mRNAs) genes but lacks in coding for a novel function of ncRNAs. So discovering a function of RNAs it is essential to have a good model of its structure. It is not only challenging to know the complete structure of RNA [6]. This process is also tiresome and took time as well. Few computational models for predicting RNA secondary structure have been around somewhere in the early to mid-seventies. In Figure 1, we have given a broad overview of the type of structures and resources to analyze them in many different characteristics.

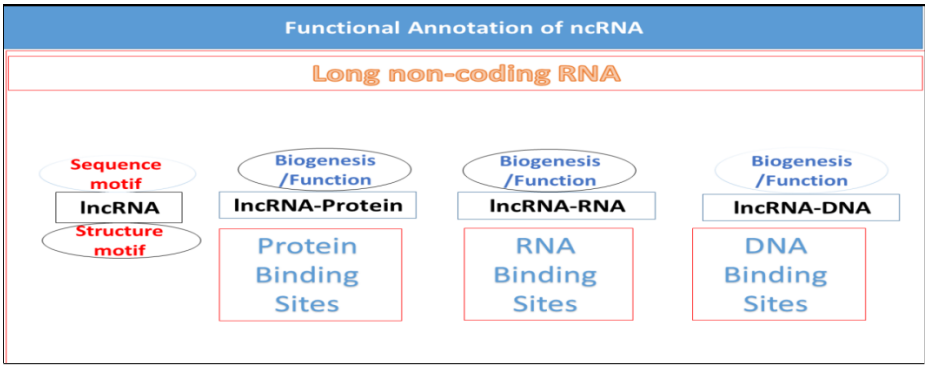

Figure 1 The branch structure of genomic variations in long ncRNAs (lncRNAs)

Good modeling of RNA structure is critical that provides some of the evidence or information towards a clear explanation of its function. Despite the advancement in high throughput computing, it is still a tedious and time-consuming task to determine the tertiary structure of RNA. In past studies, the authors have developed a number of computational algorithms to predict structure models of ncRNA genes [7]. Primary structure of RNA is a sequence-specific process that determines some functional properties, like mature miRNA and siRNA molecules base pair to their targets [8]. Modification in the primary, secondary structure and expression levels of lncRNAs, as well as their cognate RNA-binding proteins, can cause diseases ranging from neurodegeneration to cancers [9-13]. That may question the methods based on high-throughput sequencing, especially for larger structured RNAs with long-range tertiary interactions. Therefore, deriving the tertiary structure in a hierarchic way from the predicted secondary structure is not straightforward [14]. RNA tertiary structure can be model in 2 main ways [15,16].

Since the ncRNAs are completely described the biological process in the operation of the cell for the formation of tertiary structures [17]. Those tertiary structures of ncRNA genes are enabled to interact with each other to produce proteins [18]. The authors utilized a simple folding technique to detect a transition from secondary to tertiary structure [19]. Instead of structure, the authors have also described another category of ncRNAs that is long non-coding RNA classes [20]. All those classes of ncRNAs are described in Figure 2. In fact, many kinds of ncRNA genes are used to develop non-protein coding that will ultimately generate disease in the cell [21-23].

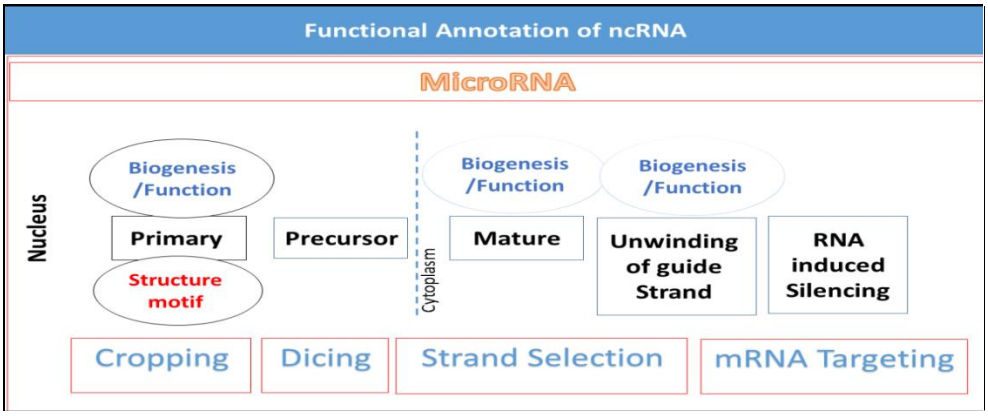

Figure 2 The branch structure of genomic variations in microRNAs

It is cleared from Figure 1 that there are many classes based on functional transcripts of ncRNAs in case of humans. All the classes are described that are difficult to cover in this review article [24]. However, the lncRNAs, miRNAs are mostly included in this study. The literature review suggested that there is also long non-coding RNA (lncRNA) class instead of just the ncRNAs annotation [25]. In practice, it was noticed by medical expert that the human ncRNAs comprise 9882 small RNAs according to the current comprehensive annotation of genes [26-29]. Accordingly, the lncRNA genes were also included in this review article to find their functional annotation in the development of the cell.

This review article focuses on the computational algorithms for predicting functional annotation of ncRNA including lncRNA class. In contrast to functional annotation of ncRNA genes, the authors have also developed some techniques to predict ncRNA structure. Those systems were not discussed in this paper such as Fold align, Pfold, Mfold, RNAfold, RNAshapes, RNAstructure, NAST, iFoldRNA, and 3dRNA [30-38]. These methods will be described in a future paper. In the coming section, state-of-the-art artificial intelligence techniques are briefly described to functional annotate the ncRNA genes [39-48].

Computational Intelligent Techniques

Functional annotation plays a vital role in identifying the genes from large sequences of ncRNAs. In recent years, the role of long ncRNAs (lncRNAs) has gained lots of interest that are not appeared to produce protein-coding [49]. Since the non-protein-coding can contain transcription noise so the experts need to identify the likely functional candidates. As a result, the functional annotation of ncRNAs or lncRNAs becomes an advanced research topic in the domain of next-generation sequencers (NGS) [50]. In regulating the gene expression and epigenetics, the lncRNAs have a crucial function such as develop many diseases. Therefore, many studies have been conducted to determine the function of lncRNAs [51]. For a typical computational method, the classification of ncRNAs depends heavily on known proteins by making it difficult to distinguish coding from non-coding transcripts. To overcome these challenges, there is a dire need to develop an advanced computational system but recent approaches have only used SVM, HMM, NN or statistical model to get 90% of accuracy for identification of protein-coding. The subsequent paragraphs briefly described these computational techniques for functional annotation of genes [52-76]. All these computational systems are presented in Table 1.

Table 1 State-of-the-art computational tools for functional annotation of non-coding RNAs

Reference	Tools	Features	Computational Methods
[49]	CM-HMM	Faster annotation of ncRNAs	[1]HMM + [2]CM
[54]	DREME	to identify and classify the lncRNA genes	[3]SVM
[55]	lncRScan-SVM	Integrate several features to identify lncRNAs	[3]SVM
[56]	DM-lncRNAs	Deep mining algorithm to identify lncRNAs	[4]DMA
[58]	PLEK	To differentiate between lncRNAs from mRNAs	k-mer scheme and [3]SVM
[59]	LncRNA-ID	Random forest algorithm with multiple features	[5]RFA
[60]	Ensembl	to interpret genomic sequence	Ensemble classifiers
[69]	DECRES	To identify cis-regularity regions	[6]DL-NN
[73]	VEP	To annotate genomic variant	Ensemble classifiers
1: Hidden Markov model; 2: Covariance model; 3: Support vector machine; 4: Deep mining algorithm; 5: Random forest algorithm; 6: Deep learning neural network			

In a study conducted by Zasha, the heuristics based hidden-Markov-model (HMM) filter is designed to search the covariance models (CMs) for finding new members of the ncRNA gene family in a large genome sequence. This tool is known as CM-HMM. The authors show that the HMM model is better than the covariance models (CMs) to find the function of ncRNAs. They compared this method with other 2 systems such as tRNAs and tRNAscan-SE. The authors achieved more significant results compared to these two on hundreds of ncRNA families. Whereas in, a miRDeepFinder tool was developed to functionally analyze the plant microRNAs (miRNAs) and their targets from small RNA datasets obtained from deep sequencing [53]. In this miRDeepFinder tool, the authors identified 13 novel miRNA candidates and 12 are experimentally validated.

The authors used a support vector machine (SVM) to identify and classify the lncRNA genes [54]. They have tested this system on cDNA sequences. In that study, the authors developed a computational-based system to predict the function of lncRNAs to regulate the expression. They concluded that the ncRNA genes are an essential element of epigenetic gene regulation in plants, but fairly a minute. Similarly, the functional annotation of lncRNAs is identified using the SVM machine learning method [55]. The dataset of GENCODE annotations was used to test the proposed lncRScan-SVM system for classification of PCTs and LNCTs. The achieved show that the lncRScan-SVM is outperformed compared to state-of-the-art systems due to integrating of several features that are derived from gene structure, transcript sequence, and conservation steps.

Instead of using SVM or NN machine learning algorithms, the authors developed a deep mining algorithm to interpret

lncRNA functionality [56]. In that study, the authors used this deep mining algorithm to classify lncRNAs known as DM-lncRNAs into 4 different groups. Whereas in, the authors presented a novel method to the functional annotation of ncRNAs by using some features and SVM machine learning algorithm [57]. After applying the ten-fold cross-validation test, the coding RNAs from ncRNAs are distinguished at about 97% specificity and 98% sensitivity.

An alignment-free predictor tool was developed known as PLEK based on k-mer scheme and support vector machine (SVM) computational methods [58]. This tool was developed to differentiate between lncRNAs from mRNAs in the absence of genomic sequence or annotations. Random forest algorithm with multiple features was developed to annotate long non-coding RNAs (lncRNAs) for coding of transcript sequences [59]. This tool is known as lncRNA-ID for annotating the long coding RNAs. In that tool, the authors showed that it can compete with existing state-of-the-art methods for identification of lncRNAs. The authors have uploaded the data and programs to the GitHub website. Ensemble classifiers approach was developed to interpret the genomic sequence that provided the latest annotations methods [60]. They updated annotation models in terms of the gene, comparative genomics, regulatory regions and variations on the new human assembly, GRCh38. A TANRIC (The Atlas of Noncoding RNAs in Cancer) web-based software approach is developed to explore lncRNAs [61]. In explore of profile expression of lncRNAs, the 20 cancer types are characterized that includes TCGA sequences. It has provided the capability for researchers to analyze lncRNAs of interests (annotate lncRNAs). By using this TRANIC web-interface, the researchers can identify a lot of lncRNAs with biomedical significance results. This software is provided strong correlations to therapeutic targets and biomarkers with drug sensitivity.

They used the hypergeometric test to functionally annotate lncRNAs with significantly co-expressed with protein-coding genes [62]. From 12 different databases, the functional terms from Gene Ontology and 4,380 human biological pathways are collected to develop a web-based interface named as lncRNA2Function. This lncRNA2Function allows the researchers to browse lncRNAs and to final functional specific terms. Also, the model of KATZ that measures the lncRNA-Disease Association prediction (KATZLDA) was developed [63]. KATZLDA obtained a reliable area under the curve AUCs of 7175, significantly improving previous classical methods. In a study, a machine learning algorithm known as PredcircRNA was developed to discover circular RNA (circular RNA) is an important type of lncRNAs [64]. The PredcircRNA was used to differentiate between circular RNA and lncRNAs using multiple kernel techniques. In this approach, different features are extracted and then multiple kernel approach was used to fuse those heterogeneous features. The performance of PredcircRNA was measured by using 5-fold cross-validation test and accuracy of 0.778, the sensitivity of 0.781, specificity of 0.770, precision of 0.784 were obtained to differentiate between these two RNAs. This PredcircRNA tool is available for download.

A computational intelligence system was developed to predict diseases associated with lncRNAs [65]. In this predictor, they combined lncRNA expression profiles, gene expression profiles, and human disease-associated gene data. The comparison of non-tissue-specific linc RNAs results shows that the area under curve AUC of our algorithm is 0.7645, and the prediction accuracy is about 89%. Effective computation methods were proposed to differentiate between coding and long non-coding RNAs [66]. In the first step, they implemented a predictor to recognize lncRNAs by fusion of many features and then use the deep learning algorithm to distinguish between these two categories. They achieved 97.1% of prediction accuracy that is higher than other state-of-the-art predictors. The lncRNA-MFDL software package is freely available for academic users. A model of HyperGeometric distribution for lncRNAs was established to predict 19 associations for breast cancer, lung cancer, and colorectal cancer [67,68]. They utilized the area under the receiver operating curves (AUC) with leave-one-out cross-validation test and AUC of 0.7621 was obtained.

The authors utilized supervised deep learning computational algorithm for identification of active cis-regularity regions in the human genome [69]. In that research study, they suggested that this computational algorithm has significant advances in the knowledge of genomic locations of cis-regularity regions. They performed many functional annotation experiments on 300,000 candidates based on DECRES (Deep learning for identifying Cis-Regulatory Elements) computational method. In that paper, they obtained a higher value of sensitivity and specificity compared to other techniques.

TopHat and Cufflinks software were used to annotate lncRNAs [70]. However, in a study, the authors utilized an effective method called BiCliques Merging (BCM) to annotate miRNAs based on bicliques merging technique [71].

In that research, the maximal bicliques detected in the network are statistically evaluated and filtered and afterward, a greedy-based strategy was used to iteratively merge the remaining bicliques. After analyzing experiments, they showed that the proposed technique is functionally enriched compared to state-of-the-art annotation techniques by using TCGA sequence from 2 cancer datasets. They have also constructed a weighted miRNA regularity network for module discovery on Breast cancer.

Even after having intensive research has been focused on ncRNAs recently. The authors reported that the properties characteristics and signals of ncRNAs are quiet unknown [72]. As a result, many authors are trying to developed advanced computational techniques to predict the likely functional influence of ncRNA genes in the cell. The annotation of genomic variants approach was developed in by using the Ensembl variant effect predictor (VEP) in ncRNAs regions. The VEP is open source and free to use [73]. In computational intelligence, an algorithm was developed to identify pseudogenes in lncRNAs regulation in the human transcriptome [74]. In that algorithm, they identified all lncRNAs transcripts that overlap genomic spans. The experiments were performed on a retrieve sequence from the UCSC Genome Database using the UCSC Table Browser.

Likewise, an Improved LNCRNA functional similarity calculation model (ILNCSIM) was developed based on the hypothesis that lncRNAs with similar biological functions tend to be involved in similar diseases [75]. The authors presented a hierarchical structure approach and basic information to derive disease-directed acyclic graph. The combination approach is developed in by detection of lncRNA and known miRNA to functionally annotate a large number of lncRNAs [76]. In that review article, the new characteristic features are derived and technical issues posed by the diversity of lncRANs were also discussed that can be used in different bioinformatics applications. The authors suggested that the limitations of current lncRNAs-related methods that can be used to assist in the development of new computational tools. In Table 2, all the online web resources are presented for major computational methods developed in the past systems to functionally annotate ncRNAs and lncRNAs.

Table 2 Comparisons table of major State-of-the-art computational tools for functional annotation of non-coding RNAs

Reference	Name	[1]URL
[49]	CM-HMM	http://bio.cs.washington.edu/software
[54]	DREME	http://meme-suite.org/
[55]	lncRScan-SVM	http://sourceforge.net/projects/lncrscansvm/?source=directory
[56]	DM-lncRNAs	Unknown
[58]	PLEK	https://sourceforge.net/projects/plek/files/
[59]	LncRNA-ID	https://github.com/zhangy72/LncRNA-ID
[60]	Ensembl	https://github.com/Ensembl
[62]	lncRNA2Function	http://mlg.hit.edu.cn/lncrna2function
[65]	lincRNAs	http://asdcd.amss.ac.cn/MingXiLiu/lncRNA-disease.html
[69]	DECRES	https://github.com/yifeng-li/DECRES
[70]	TopHat-lncRNAs	http://www.animalgenome.org/repository/pub/MTSU2015.1014/
[73]	VEP	http://www.ensembl.org/vep

[1] Universal resource locator

Databases

To test and compare the functional annotation of RNA genes, the authors have developed some online databases. An example of online databases is presented in Table 3. As shown in this table, there are few online databases available to test the proposed computational system. However, in many examples of RNA genes, the authors have focused more on lncRNAs data sets. Therefore, Table 3 is focused on functional annotation of lncRNA genes research. To do research in this field, the databases played important roles and delivered specific annotations.

Table 3 Online databases for functional annotation

Reference	Database Name	[1]URL
[77]	NONCODE v 4.0	http://www.bioinfo.org/noncode/
[78]	lncRNAdb	http://www.lncrnadb.org/
[79]	LNCipedia	http://www.lncipedia.org
[80]	lncRNome	http://genome.igib.res.in/lncRNome/

[81]	fRNAdb	http://www.ncrna.org/frnadb
[82]	lncRNAtor	http://lncrnator.ewha.ac.kr/
[83]	lncRNAMap	http://lncrnamap.mbc.nctu.edu.tw/php/
[84]	PLncDB	http://chualab.rockefeller.edu/gbrowse2/homepage.html
[1] Universal resource locator		

Challenges for Functional Annotations

Computational intelligent techniques have many limitations that developed in the past systems for functional annotation of non-coding RNAs (ncRNAs). It was mentioned before that the ncRNAs genes are a very important function for the operation of the cell. Therefore, the authors focused more on developing methodological techniques for understanding the different functions of ncRNAs. The authors applied many computational intelligent models to determine functional divergence and resemblance ncRNAs genes in the cell along with limited data sources. There are many varieties of ncRNAs such as micro RNAs or long non-coding RNAs. Therefore to get accurate accuracy for determining the functions of ncRNAs is still a challenging task for computational intelligent methods. Compared to the functional annotation of ncRNAs, the authors have also developed structure prediction algorithms that obtained higher accuracy. Currently, many authors proposed computational methods to analyze the structural and functional annotation of ncRNAs [30-48,55-74].

As a matter of fact, the structure prediction of ncRNAs or lncRNAs genes is still having an accuracy that is not up-to-the-mark. The author's utilized conventional machine learning algorithms instead of the advanced computational deep learning algorithm. The advantage of using the deep-neural network (DNN) is that these models can be trained on a large sequence data and the prediction response is in a very limited amount of time compared to convolutional machine learning approaches such as SVM, AdaBoost or NN classifiers. Unfortunately, at present, we find only one effective methodological approach that can predict functional of ncRNAs or long ncRNAs genes through advanced machine learning approaches such as deep learning algorithm [69].

To highlight these issues, let us consider a study in which the authors developed RNAz system to predict the structure of ncRNAs but it requires high computational cost due to support vector machine (SVM) classifiers [28]. Therefore, it is unsuitable for the next-generation sequence (NGS) or where there is a requirement for big data. Also, the authors developed an RNA-sample system to predict secondary structure but this system required high computational complexity [36]. This information is briefly described in Table 3 about structure prediction. From this table, it is noticed that the high-throughput sequences are not possible to extract from the large-sequence ncRNA genes. Therefore, these approaches must require deep-learning algorithms to effectively detect large sequence structures without focusing on domain expert knowledge about machine learning or data mining algorithms.

CONCLUSION

Advances in bioinformatics have made enormous progress towards the development of computational techniques for the structure and functions of ncRNA or lncRNA. In this review article, the advanced computational methods are focused to annotate the function of ncRNAs. This paper described the computational methods to predict the structure or to functionally annotate the ncRNAs genes. To increase the accuracy of functional annotation, there is a dire need for domain expert knowledge. In fact, the state-of-the-art computational techniques that have been used in the past required for pre-processing of raw input data, feature selection and to fine-tune parameters for getting high accuracy. However, the authors put many efforts to do pre- or post-processing on data instead of developing effective computational intelligent techniques. It was noticed that the deep learning algorithms are classifying the data without doing pre- or post-processing steps. The methods were developed using deep learning algorithms to get higher accuracy for data classification. In practice, the deep learning algorithms have outperformed compared to conventional machine learning algorithms for functional annotation of ncRNAs but still, lack of expert knowledge is required to define multilayer architecture. As reported in the past studies, the techniques developed using deep learning algorithm must be further investigated in terms of performance.

DECLARATIONS

Conflict of Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

REFERENCES

[1] Zhang, Tao Jiang Michael Q. *Current topics in computational molecular biology*. MIT Press, 2002.

[2] Xue Chenghai, et al. "Classification of real and pseudo microRNA precursors using local structure-sequence features and support vector machine." *BMC Bioinformatics*, Vol. 6, No.1, 2005, p. 310.

[3] Lowe Todd M and Sean R Eddy. "tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence." *Nucleic Acids Research,* Vol. 25, No. 5, 1997, pp. 955-64.

[4] Washietl Stefan, Ivo L Hofacker, and Peter F Stadler. "Fast and reliable prediction of noncoding RNAs." *Proceedings of the National Academy of Sciences,* Vol. 102, No. 7, 2005, pp. 2454-59.

[5] Gardner Paul P, Andreas Wilm, and Stefan Washietl. "A benchmark of multiple sequence alignment programs upon structural RNAs." *Nucleic Acids Research,* Vol. 33, No. 8, 2005, pp. 2433-39.

[6] Washietl, Stefan, et al. "Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome." *Nature Biotechnology,* Vol. 23, No. 11, 2005, p. 1383.

[7] Tinoco Ignacio, Olke C Uhlenbeck and Mark D Levine. "Estimation of secondary structure in ribonucleic acids." *Nature,* Vol. 230, No. 5293, 1971, p. 362.

[8] Carthew Richard W and Erik J Sontheimer. "Origins and mechanisms of miRNAs and siRNAs." *Cell,* Vol. 136, No. 4, 2009, pp. 642-655.

[9] Wapinski Orly and Howard Y Chang "Corrigendum: Long noncoding RNAs and human disease." *Trends in Cell Biology,* Vol. 21, No. 10, 2011, pp. 354-561

[10] Tsai Miao-Chih, Robert C Spitale and Howard Y Chang. "Long intergenic noncoding RNAs: new links in cancer progression." *Cancer Research,* Vol. 71, No. 1, 2011, pp.3-7.

[11] Nussinov Ruth and Ann B Jacobson. "Fast algorithm for predicting the secondary structure of single-stranded RNA." *Proceedings of the National Academy of Science,* Vol. 77, No. 11, 1980, pp. 6309-13.

[12] Washietl, Stefan, et al. "Computational analysis of noncoding RNAs." *Wiley Interdisciplinary Reviews: RNA,* Vol. 3, No. 6, 2012, pp. 759-78.

[13] Novikova Irina, Scott Hennelly and Karissa Sanbonmatsu. "Tackling structures of long noncoding RNAs." *International Journal of Molecular Sciences,* Vol. 14, No. 12, 2013, pp. 23672-84.

[14] Tsai Miao-Chih, Robert C Spitale, and Howard Y Chang. "Long intergenic noncoding RNAs: new links in cancer progression." *Cancer Research,* Vol. 71, No. 1, 2011, pp. 3-7.

[15] Gautheret, Daniel, Francois Major, and Robert Cedergren. "Pattern searching/alignment with RNA primary and secondary structures: an effective descriptor for tRNA." *Bioinformatics,* Vol. 6, No. 4 1990, pp. 325-31.

[16] Pedersen Jakob Skou, et al. "Identification and classification of conserved RNA secondary structures in the human genome." *PLoS Computational Biology,* Vol. 2, No. 4, 2006, p. e33.

[17] Dann III Charles E, et al. "Structure and mechanism of a metal-sensing regulatory RNA." *Cell,* Vol. 130, No. 5, 2007, pp. 878-92.

[18] Wan Yue, et al. "Understanding the transcriptome through RNA structure." *Nature Reviews Genetics,* Vol. 12, No. 9, 2011, pp. 641.

[19] Brion Philippe and Eric Westhof. "Hierarchy and dynamics of RNA folding." *Annual Review of Biophysics and Biomolecular Structure,* Vol. 26, No. 1, 1997, pp. 113-37.

[20] Yan Kun, et al. "Structure prediction: new insights into decrypting long noncoding RNAs." *International Journal of Molecular Sciences,* Vol. 17, No. 1, 2016, p. 132.

[21] Zhao Yi, et al. "Large-scale study of long non-coding RNA functions based on structure and expression features." *Science China Life Science,* Vol. 56, No. 10, 2013, pp. 953-59.

[22] Guo Xingli, et al. "Advances in long noncoding RNAs: identification, structure prediction, and function annotation." *Briefings in Functional Genomics,* Vol. 15, No. 1, 2015, pp. 38-46.

[23] Harrow Jennifer, et al. "GENCODE: the reference human genome annotation for The ENCODE Project." *Genome research,* Vol. 22, No. 9, 2012, pp. 1760-74.

[24] Bhartiya, Deeksha, and Vinod Scaria. "Genomic variations in non-coding RNAs: structure, function, and regulation." *Genomics,* Vol. 107, No. 2, 2016, pp. 59-68.

[25] Derrien Thomas, et al. "The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression." *Genome Research,* Vol. 22, No. 9, 2012, pp. 1775-89.

[26] Mercer Tim R, Marcel E Dinger, and John S. Mattick. "Long non-coding RNAs: insights into functions." *Nature Reviews Genetics,* Vol. 10, No. 3, 2009, p. 155.

[27] Rivas Elena and Sean R Eddy. "Secondary structure alone is generally not statistically significant for the detection of noncoding RNAs." *Bioinformatics,* Vol. 16, No. 7, 2000, pp. 583-605.

[28] Washietl Stefan, Ivo L Hofacker, and Peter F Stadler. "Fast and reliable prediction of noncoding RNAs." *Proceedings of the National Academy of Sciences,* Vol. 102, No. 7, 2005, pp. 2454-59.

[29] Washietl Stefan, et al. "Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome." *Nature Biotechnology,* Vol. 23, No. 11, 2005, p. 1383.

[30] Havgaard Jakob Hull, et al. "Pairwise local structural alignment of RNA sequences with sequence similarity less than 40%." *Bioinformatics,* Vol. 21, No. 9, 2005, pp. 1815-1824.

[31] Sætrom Pål, et al. "Predicting non-coding RNA genes in Escherichia coli with boosted genetic programming." *Nucleic Acids Research,* Vol. 33, No. 10, 2005, pp. 3263-70.

[32] Uzilov Andrew V, Joshua M Keegan, and David H Mathews. "Detection of non-coding RNAs on the basis of predicted secondary structure formation free energy change." *BMC Bioinformatics,* Vol. 7, No. 1, 2006, p. 173.

[33] Dall Deniz, et al. "STRAL: progressive alignment of non-coding RNA using base pairing probability vectors in quadratic time." *Bioinformatics,* Vol. 22, No. 13, 2006, pp. 1593-99.

[34] Lindgreen Stinus, Paul P Gardner and Anders Krogh. "MASTR: multiple alignment and structure prediction of non-coding RNAs using simulated annealing." *Bioinformatics,* Vol. 23, No. 24, 2007, pp. 3304-11.

[35] Torarinsson Elfar, Jakob H Havgaard, and Jan Gorodkin. "Multiple structural alignment and clustering of RNA sequences." *Bioinformatics,* Vol. 23, No. 8, 2007, pp. 926-32.

[36] Xu Xing, Yongmei Ji, and Gary D Stormo. "RNA Sampler: a new sampling based algorithm for common RNA secondary structure prediction and structural alignment." *Bioinformatics,* Vol. 23, No. 15, 2007, pp. 1883-91.

[37] Bernhart Stephan H, et al. "RNAalifold: improved consensus structure prediction for RNA alignments." *BMC Bioinformatics,* Vol. 9, No. 1, 2008, p. 474.

[38] Friedländer Marc R, et al. "Discovering microRNAs from deep sequencing data using miRDeep." *Nature Biotechnology,* Vol. 26, No. 4, 2008, p. 407.

[39] Tran Thao T, et al. "De novo computational prediction of non-coding RNA genes in prokaryotic genomes." *Bioinformatics,* Vol. 25, No. 22, 2009, pp. 2897-2905.

[40] Hamada Michiaki, et al. "Prediction of RNA secondary structure using generalized centroid estimators." *Bioinformatics,* Vol. 25, No. 4, 2008, pp. 465-73.

[41] Hackenberg Michael, et al. "miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments." *Nucleic Acids Research,* Vol. 37, 2009, pp. W68-W76.

[42] Reuter Jessica S and David H Mathews. "RNAstructure: software for RNA secondary structure prediction and analysis." *BMC Bioinformatics,* Vol. 11, No. 1, 2010, p. 129.

[43] Kertesz Michael, et al. "Genome-wide measurement of RNA secondary structure in yeast." *Nature,* Vol. 467, No. 7311, 2010, p. 103.

[44] Mathelier Anthony and Alessandra Carbone. "MIReNA: finding microRNAs with high accuracy and no learning at genome scale and from deep sequencing data." *Bioinformatics,* Vol. 26, No. 18, 2010, pp. 2226-34.

[45] Tsai Miao-Chih, Robert C Spitale, and Howard Y. Chang. "Long intergenic noncoding RNAs: new links in cancer progression." *Cancer Research,* Vol. 71, No. 1, 2011, pp. 3-7.

[46] Sato Kengo, et al. "IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming." *Bioinformatics,* Vol. 27, No. 13, 2011, pp. 85-93.

[47] Novikova, Irina V, et al. "Rise of the RNA machines: exploring the structure of long non-coding RNAs." *Journal of Molecular Biology,* Vol. 425, No. 19, 2013, pp. 3731-46.

[48] Fan Xiao-Nan and Shao-Wu Zhang. "lncRNA-MFDL: identification of human long non-coding RNAs by fusing multiple features and using deep learning." *Molecular BioSystems,* Vol. 11, No. 3, 2015, pp. 892-97.

[49] Khachane Amit N and Paul M. Harrison. "Mining mammalian transcript data for functional long non-coding RNAs." *PLoS One*, Vol. 5, No. 4, 2010, p. e10316.

[50] Sacco Letizia Da, Antonella Baldassarre, and Andrea Masotti. "Bioinformatics tools and novel challenges in long non-coding RNAs, lncRNAs) functional analysis." *International Journal of Molecular Sciences*, Vol. 13, No. 1, 2012, pp. 97-114.

[51] Guttman Mitchell and John L. Rinn. "Modular regulatory principles of large non-coding RNAs." *Nature,* Vol. 482, No. 7385, 2012, p. 339.

[52] Weinberg Zasha and Walter L Ruzzo. "Sequence-based heuristics for faster annotation of non-coding RNA families." *Bioinformatics,* Vol. 22, No. 1, 2005, pp. 35-39.

[53] Xie Fuliang, et al. "miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs." *Plant Molecular Biology,* Vol. 80, No. 1, 2012, pp. 75-84.

[54] Boerner Susan and Karen M McGinnis. "Computational identification and functional predictions of long noncoding RNA in Zea mays." *PloS One*, Vol. 7, No. 8, 2012, p. e43047.

[55] Sun Lei, et al. "lncRScan-SVM: a tool for predicting long non-coding RNAs using support vector machine." *PloS One,* Vol. 10, No. 10, 2015, p. e0139654.

[56] Ma Lina, Vladimir B Bajic, and Zhang Zhang. "On the classification of long non-coding RNAs." *RNA Biology,* Vol. 10, No. 6, 2013, pp. 924-33.

[57] Liu Jinfeng, Julian Gough, and Burkhard Rost. "Distinguishing protein-coding from non-coding RNAs through support vector machines." *PLoS Genetics,* Vol. 2, No. 4, 2006, p. e29

[58] Li Aimin, Junying Zhang and Zhongyin Zhou. "PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme." *BMC Bioinformatics,* Vol. 15, No. 1, 2014, p. 311.

[59] Achawanantakun Rujira, et al. "LncRNA-ID: Long non-coding RNA IDentification using balanced random forests." *Bioinformatics,* Vol. 31, No. 24, 2015, pp. 3897-3905.

[60] Cunningham, Fiona, et al. "Ensembl 2015." *Nucleic Acids Research,* Vol. 43, No. D1, 2014, pp. D662-D669.

[61] Li Jun, et al. "TANRIC: an interactive open platform to explore the function of lncRNAs in cancer." *Cancer Research,* Vol. 75, No. 18, 2015, pp. 3728-37.

[62] Jiang Qinghua, et al. "LncRNA2Function: a comprehensive resource for functional investigation of human lncRNAs based on RNA-seq data." *BMC Genomics*, Vol. 16, No. 3, 2015.

[63] Chen Xing. "KATZLDA: KATZ measure for the lncRNA-disease association prediction." *Scientific Reports,* Vol. 5, 2015, p. 16840.

[64] Pan Xiaoyong and Kai Xiong. "PredcircRNA: computational classification of circular RNA from other long non-coding RNA using hybrid features." *Molecular Biosystems,* Vol. 11, No. 8, 2015, pp. 2219-26.

[65] Liu Ming-Xi, et al. "A computational framework to infer human disease-associated long noncoding RNAs." *PloS One,* Vol. 9, No .1, 2014, pp. e84408.

[66] Fan Xiao-Nan and Shao-Wu Zhang. "lncRNA-MFDL: identification of human long non-coding RNAs by fusing multiple features and using deep learning." *Molecular BioSystems,* Vol. 11, No. 3, 2015, pp. 892-97.

[67] Chen Xing. "Predicting lncRNA-disease associations and constructing lncRNA functional similarity network based on the information of miRNA." *Scientific Reports,* Vol. 5, 2015, p. 13186.

[68] Chen Xing, et al. "Constructing lncRNA functional similarity network based on lncRNA-disease associations and disease semantic similarity." *Scientific Reports,* Vol. 5, 2015, p. 11338.

[69] Li Yifeng, Wenqiang Shi and Wyeth W. Wasserman. "Genome-wide prediction of cis-regulatory regions using supervised deep learning methods." *BMC Bioinformatics,* Vol. 19, No. 1, 2018, p. 202.

[70] Al-Tobasei Rafet, Bam Paneru, and Mohamed Salem. "Genome-wide discovery of long non-coding RNAs in rainbow trout." *PLoS One,* Vol. 11, No. 2, 2016, p. e0148940.

[71] Liang Cheng, Yue Li, and Jiawei Luo. "A novel method to detect functional microRNA regulatory modules by bicliques merging." *IEEE/ACM Transactions on Computational Biology and Bioinformatics, TCBB),* Vol. 13, No. 3, 2016, pp. 549-56.

[72] Arruda Wosley, et al. "ncRNA-Agents: A multiagent system for non-coding RNA annotation." *Brazilian Symposium on Bioinformatics*, Springer, Cham, 2013.

[73] McLaren William, et al. "The ensemble variant effect predictor." *Genome Biology,* Vol. 17, No. 1, 2016, p. 122.

[74] Milligan Michael J, et al. "Global intersection of long non-coding RNAs with processed and unprocessed pseudogenes in the human genome." *Frontiers in Genetics,* Vol. 7, 2016, p. 26.

[75] Huang Yu-An, et al. "ILNCSIM: improved lncRNA functional similarity calculation model." *Oncotarget,* Vol. 7, No. 18, 2016, p. 25902.

[76] Yotsukura Sohiya, et al. "Computational recognition for long non-coding RNA, lncRNA, pp. software and databases." *Briefings in Bioinformatics,* Vol. 18, No. 1, 2016, pp. 9-27.

[77] Xie Chaoyong, et al. "NONCODEv4: exploring the world of long non-coding RNA genes." *Nucleic acids Research,* Vol. 42.D1, 2013, pp. D98-D103.

[78] Quek Xiu Cheng, et al. "lncRNAdb v2. 0: expanding the reference database for functional long noncoding RNAs." *Nucleic Acids Research,* Vol. 43.D1, 2014, pp. D168-D173.

[79] Volders Pieter-Jan, et al. "LNCipedia: a database for annotated human lncRNA transcript sequences and structures." *Nucleic Acids Research,* Vol. 41, No. D1, 2012, pp. D246-D251.

[80] Bhartiya Deeksha, et al. "lncRNome: a comprehensive knowledgebase of human long noncoding RNAs." *Database,* 2013.

[81] Mituyama Toutai, et al. "The Functional RNA Database 3.0: databases to support mining and annotation of functional RNAs." *Nucleic Acids Research,* Vol. 37, No. 1, 2008, pp. D89-D92.

[82] Park Charny, et al. "lncRNAtor: a comprehensive resource for functional investigation of long non-coding RNAs." *Bioinformatics,* Vol. 30, No. 17, 2014, pp. 2480-85.

[83] Chan Wen-Ling, Hsien-Da Huang and Jan-Gowth Chang. "lncRNAMap: a map of putative regulatory functions in the long non-coding transcriptome." *Computational Biology and Chemistry,* Vol. 50, 2014, pp. 41-49.

[84] Jin Jingjing, et al. "PLncDB: plant long non-coding RNA database." *Bioinformatics,* Vol. 29, No. 8, 2013, pp. 1068-71.