SCIENCE CHINA
Life Sciences

SPECIAL ISSUE: Non-coding RNAs
• REVIEW •

October 2013 Vol.56 No.10: 953–959
doi: 10.1007/s11427-013-4556-3

Large-scale study of long non-coding RNA functions based on structure and expression features

ZHAO Yi[2†], WANG Jian[3†], CHEN XiaoWei[1], LUO HaiTao[2], ZHAO YunJie[3], XIAO Yi[3*] & CHEN RenSheng[1*]

[1]*Laboratory of Bioinformatics and Non-coding RNA, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China;*
[2]*Bioinformatics Research Group, Advanced Computing Research Laboratory, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China;*
[3]*Department of Physics, Huazhong University of Science and Technology, Wuhan 430074, China*

Received August 14, 2013; accepted September 2, 2013

Mammals and other complex organisms can transcribe an abundance of long non-coding RNAs (lncRNAs) that fulfill a wide variety of regulatory roles in many biological processes. These roles, including as scaffolds and as guides for protein-coding genes, mainly depend on the structure and expression level of lncRNAs. In this review, we focus on the current methods for analyzing lncRNA structure and expression, which is basic but necessary information for in-depth, large-scale analysis of lncRNA functions.

non-coding RNA, structure, expression, function

Citation: Zhao Y, Wang J, Chen X W, et al. Large-scale study of long non-coding RNA functions based on structure and expression features. Sci China Life Sci, 2013, 56: 953−959, doi: 10.1007/s11427-013-4556-3

The ENCODE project, which has published 30 papers to date, including a few that extensively characterize long non-coding RNAs (lncRNAs), has revealed that 76% of the human genome is transcribed to produce a range of lncRNAs [1]. The landscape of lncRNAs in mammals was unveiled by the rapid progress of deep sequencing technology [2] and computational methods to identify lncRNA [3,4]. These lncRNAs participate in a wide variety of biological processes, such as imprinting control, cell differentiation, immune responses through regulating expression, and activity and localization of protein coding genes [5,6]. However, the function and mechanisms of most lncRNAs are still unknown. Here, we combine our work and other related work to systematically illustrate the current methods to study lncRNA function through their structures and expression profiles.

1 RNA secondary structure prediction

Secondary structures of RNAs are the basis of their tertiary structures, so we will first briefly review methods for their prediction. Such methods predict standard Watson-Crick base pairs (AU and CG) and non-standard base pairs in a RNA sequence. There are many methods for RNA secondary structure prediction, which are based on different principles. Here we focus on two types of commonly used methods: the minimum free energy method [7–11], and the multiple sequence alignment method [12–14].

The minimum free energy methods are now the most widely used methods of RNA secondary structure predic-

†Contributed equally to this work
*Corresponding author (email: yxiao@mail.hust.edu.cn; crs@sun5.ibp.ac.cn)

tion. Mfold, proposed by Zuker and Stiegler [7], was the earliest-developed minimum free energy prediction algorithm. Mfold uses a dynamic programming algorithm to produce a large number of RNA secondary structure candidates, and then calculates their global free energies by adding up those of independent candidates using the nearest neighborhood approximation. The free energies of the structural units are determined experimentally. Mfold can quickly and efficiently predict the lowest free energy secondary structures of RNAs. RNAfold, developed by Hofacker and Stadler [8], is another RNA secondary structure prediction method based on the minimum free energy principle. The method is part of the Vienna Package and its advantage is that it can give the probabilities of base-pair formation, and is able to sample suboptimal free energy structures. RNAstructure by Reuter and Mathews [9] is also an RNA secondary structure prediction method based on the minimum free energy principle, but uses alternative thermodynamic parameters. Furthermore, RNAstructure can use experimental information to improve the accuracy of secondary structure prediction. RNAshapes by Steffen et al. [10] is another commonly used method of RNA secondary structure prediction, which can predict a large number of potential secondary structures with a range of free energy above the lowest value, and then clusters them into different classes. Thus, one can use other biological information (e.g., experimental data) to identify the possible native secondary structure from these classes.

Pseudoknot prediction is a very difficult problem in RNA secondary structure prediction. Most RNA secondary structure prediction methods cannot predict pseudoknot structures [15]. pknotsRG by Reeder et al. [16] is one method for predicting the pseudoknot structures, which extends the free energy parameters to include those for pseudoknot structural elements. FlexStem by Chen et al. [17] is another method for predicting pseudoknot structures, and combines global and local minimum free energy principles by using a maximal stem strategy and a stem-adding rule.

With the rapid growth in RNA homologous structure data, RNA secondary structure prediction methods using multiple sequence alignment have also been developed. With respect to the nucleotide sequence of RNA, base pairs are very conservative and can form conservative secondary structure elements such as helical stems. Multiple sequence analysis compares the sequence similarity of multiple RNA secondary structures to find conserved units to predict the secondary structures of unknown RNA sequences. Dynalign [12], RNAalifold [13], and ILM [14] are multiple sequence alignment prediction algorithms based on the minimum free energy principle. ILM can also predict RNA secondary structures with pseudoknots.

Although RNA secondary structure prediction methods have developed significantly, subject to the limitations of the number of available experimental structures and precise determination of thermodynamic parameters of RNA secondary structure elements, prediction accuracy is about 70% [18] and needs further improvement.

2 Three-dimensional non-coding RNA structure prediction

Non-coding RNA molecules need to form specific tertiary structures to perform their biological functions. Therefore, solving RNA tertiary structures is essential for understanding their functions. However, the number of solved tertiary structures of noncoding RNA molecules is very limited (less than 1000) due to the experimental difficulties of determining RNA structures. Therefore, many computational methods for predicting RNA tertiary structures have been proposed. However, compared with protein tertiary structure prediction, RNA tertiary structure prediction is still at a very early stage. Initially, only homology modeling methods or prediction methods for small RNA tertiary structures, like ERNA-3D [19,20] and Manip [21,22], were available. RNA tertiary structure prediction methods, in the true sense, have only begun to appear and undergo greater development in recent years [23,24].

FARNA, proposed by Das and Baker [25] at Seattle University, assembles three-dimensional (3D) fragments of RNAs into their RNA tertiary structures. The method is inspired by the protein tertiary structure prediction methods (Rosetta) of Baker et al. [26]. FARNA uses coarse-grained models to represent the base structures by taking the center of each base as a virtual atom. The experimentally determined ribosomal RNA tertiary structure is cut into 3D fragments of 3 nt and used as templates to predict the local 3D structures of RNAs. FARNA first uses the target sequence information to divide the primary structure into multiple small windows, and then replaces them with 3D fragment structures selected randomly from the 3D fragment structure library using the Monte Carlo method, and finally uses an energy function to pick out possible models of the near-native tertiary structure. For small RNA molecules with lengths less than 30 nt, prediction accuracy of about 4 Å root-mean-square-deviation (RMSD) for the main chains can be achieved. The prediction accuracy of FARNA can be further improved by considering secondary and tertiary structure information [27]. Recently, Baker et al. [28] improved FARNA into an all-atomic structure prediction method with high accuracy, FARFAR. However, FARFAR can only be used to predict tertiary structure of small RNA molecules (<20 nt).

NAST (The Nucleic Acid Simulation Tool) proposed by Jonikas et al. [29] is a method of predicting RNA tertiary structure based on molecular dynamics simulations, using a statistical potential energy function. It uses a coarse-grained model (C3′ atoms) to represent the corresponding nucleotides. The statistical potential energy function was trained on the ribosomal RNA tertiary structure. The input data of NAST includes sequence and secondary structure of

the target RNA molecules. The tertiary interactions of the target structures can also be added to guide the folding process of RNA molecules. In the case where only the sequence and secondary structure are known, the method can only predict the tertiary structures of small RNA molecules with relatively simple topology (<40 nt), and the prediction accuracy (RMSD) is about 8 Å. If tertiary interaction information is added to guide the prediction, the accuracy is greatly improved.

iFoldRNA, proposed by the Dokholyan group at the University of North Carolina, uses discrete molecular dynamics to simulate the formation of RNA tertiary structures [30,31]. The main idea of the discrete molecular dynamics is the use of square-well potentials or hard-sphere potentials instead of continuous interaction potential energy functions. Outside the potential wells, interaction forces are zero. This simplified interaction model greatly increases the computing speed. This method has been applied to simulate the folding of protein molecules since the 1980s [32]. Furthermore, the phosphate group, ribose, and base of each RNA nucleotide are simplified as three virtual atoms having certain respective size, and then the folding of this coarse-grained model is simulated using the discrete molecular dynamics to find possible tertiary structures of the corresponding RNA. iFoldRNA uses an extended chain as the initial structure of the discrete molecular dynamics simulations, then uses replica exchange to sample the conformation space, and determines the predicted tertiary structures by calculating the free energy. If only the RNA sequence is known, iFoldRNA can accurately predict the tertiary structures of small RNA molecules (<50 nt). For larger RNAs of complex topology, the prediction accuracy is severely reduced (about 23 Å RMSD). But by including secondary and tertiary structure interactions, the prediction accuracy can be greatly improved [33].

BARNACLE, proposed by Frellsen et al. [34], uses a random sampling method to rotate the dihedral angles to predict RNA tertiary structures. The majority of RNA tertiary structure prediction methods use fragment-assembling approaches to solve the problem of sampling. Experimentally-determined 3D structures of small fragments can be combined into a candidate of the near-native structure. However, this approach cannot perform large-scale sampling in the vicinity of the target structure, and so care must be taken with the results of the tertiary structure prediction. The Barnacle probability model, using a random sampling method, can carry out large-scale sampling in the vicinity of the local structures for the rotating dihedral angles. This may solve the bottleneck problem of sampling. Results show that, if only RNA sequence and secondary structure information are used, BARNACLE could accurately predict the tertiary structure of small RNA molecules (<50 nt, approximately 10 Å RMSD). However, with longer or more topologically-complex RNA molecules, sampling becomes very difficult due to the large number of degrees of freedom.

CG Model, established by the Gutell and Ren groups [35] at the University of Texas at Austin, also uses a coarse-grained model to predict RNA tertiary structure. The method represents each nucleotide by five simplified virtual atoms, including two virtual atoms for the main-chain atoms (a phosphate group and a sugar) and three virtual atoms for bases, to describe a stacking effect. It also uses the statistics of 688 experimental structures to fit the potential energy function of the coarse-grained model [36,37]. Molecular dynamics simulation for 15 RNAs from 12–27 nt using the CG Model showed that for 75% of the RNA molecules, at least one track can arrive at a near-native structure. By including secondary structure or tertiary interaction information, all 15 of the RNA molecules successfully folded into near-native structures using the CG Model (about 6.5 Å RMSD). The CG Model can accurately predict the tertiary structure of small RNA molecules.

RNA2D3D, proposed by Shapiro et al. [38], builds RNA tertiary structures based on the standard bases and base-pair structures. Different from most RNA tertiary structure modeling and prediction algorithms, the structural library of this method only includes the standard base structure and the standard A-form helical structures of base pairs. RNA2D3D uses secondary structures as templates and transforms them into 3D structures by replacing the corresponding RNA secondary structure with the standard A-form helical structures of the base pairs. RNA2D3D can build 3D models of RNA molecules automatically, but the generated models usually have serious steric clashes, including covalent bond cleavage, atomic overlap, or chain crossing, and therefore require further optimization and manual adjustments to generate a reasonable RNA tertiary structure [38,39]. Using RNA2D3D, Shapiro et al. [40–42] predicted the pseudoknot structure of the telomerase RNA, with a length of 48 nt, and the overall accuracy can reach 7 Å RMSD after adjustment and optimization.

The Vfold model, proposed by Chen's group at the University of Missouri, builds tertiary structures of RNA molecules based on their method of secondary structure folding kinetics [43]. Vfold is a coarse-grained model that uses the phosphorus atom (P), carbon atoms (C4), and virtual base atoms to represent each part of a nucleotide structure. This method first predicts the secondary structure of the target RNA from sequence using secondary structure folding kinetics, then uses the secondary structure information to build a coarse-grained template of the tertiary structure, and finally replaces the template structure with the experimental fragment structure, giving a final prediction. Vfold can predict the tertiary structures of small RNA molecules with a mean accuracy of 3.84 Å C4-atom RMSD.

MC-Sym, proposed by Major et al. [44,45], uses the secondary structures predicted by MC-Fold to build RNA tertiary structure. MC-Fold can predict the secondary structures of target RNAs using the minimum free energy meth-

od. MC-Sym uses sets of nucleotide cyclic motifs to build RNA structures. The advantage of this method is that both standard and non-standard base pairs can be considered, as well as experimental information like base distance, dihedral angle, rotation angle, and local 3D structure. MC-Fold/MC-Sym can predict the tertiary structures of small RNA molecules (10–30 nt) with accuracy of about 4 Å (C4 atoms) RMSD.

ASSEMBLE, by Westhof's group at the University of Strasbourg, is a RNA tertiary structure modeling method that uses adjustment and optimization based on human-computer interaction [46]. This method can use multi-sequence alignment algorithms to find a large amount of secondary and tertiary structure information from homologous RNA molecules to help build RNA tertiary structure models. ASSEMBLE can employ human-computer interaction to adjust tertiary structure characteristics of base pairing, base distance, rotation angle and dihedral angle. A major advantage of the method is that, with the help of homologous information and operators, ASSEMBLE can greatly reduce the computation time. However, this method can be difficult for inexperienced operators or with less homologous information.

Liang and Schlick [47,48] systematically assessed existing RNA tertiary structure prediction methods and found that the accuracy is greater than 6.0 Å for most RNAs (50–130 nt), with a mean RMSD of 20 Å. Also, existing RNA tertiary structure prediction methods are mostly not automated, requiring human interaction for further manual adjustments to optimize the generated structures. Despite much effort on modeling and predicting noncoding RNA tertiary structures, two major problems remain. First, the highly-accurate prediction is only achievable for small or simple RNA molecules. Second, the results need manual adjustment in most cases. Therefore, automated prediction of tertiary structure of long non-coding RNA molecules remains a challenge.

Recently, we introduced a novel method for automated building of RNA tertiary structures based on RNA sequence and secondary structure [49]: 3dRNA (//122.205.6.127/3dRNA/3dRNA.html). As the tertiary structures of RNAs are restrained to a large extent by their secondary structures, we used a hierarchical approach. First, we divided a RNA structure into basic secondary structures (hairpin, double-helix, multi-way junctions, and pseudoknots), and then further divided them into the smallest structural elements (base pair, hairpin loop, inner loop, bulge loop and pseudoknot loop, junction, etc.). The smallest structural elements have greater conformation spaces to search their 3D templates, as we found that the 3D conformations of the backbones of the smallest structural elements of the same length and the same type are similar, even if their sequences are different. This can improve the accuracy of RNA structure prediction methods significantly. We selected the appropriate 3D templates for the smallest structural elements from our library extracted from experimental structures. The selected 3D templates included an additional base pair at their 5′ ends, which enabled us to use the experimental structure information to build loop structures, and this can increase the prediction accuracy. Guided by the secondary structure, the templates are first assembled into the secondary structural units, which are then assembled into a complete tertiary RNA structure.

3dRNA has been tested on a database of 300 non-redundant RNA monomers from 12–101 nt in length (including hairpins, double helices, pseudoknots, and multi-way junction structures). The results show the mean accuracy (heavy-atom RMSD) is 3.74 Å: 1.93 Å for double helices, 3.6 Å for hairpins, and 5.7 Å for complex molecular structures. Figure 1 gives two examples of structures predicted by 3dRNA. One structure is a 27-nt pseudoknot with a prediction accuracy of 3.46 Å RMSD, while the other is an L-type tRNA structure with a prediction accuracy of 3.80 Å RMSD. We also compared 3dRNA with existing RNA tertiary structure prediction methods (FARNA, RNA2D3D, iFoldRNA, V-Fold, and MC-Sym) [49]. As the FARNA and V-fold servers were unavailable, the prediction results from the papers describing FARNA and V-fold were used for comparison (Table 1) [49]. The prediction accuracy of other methods (RNA2D3D, iFoldRNA, and MC-Sym) were calculated for a database of 185 RNAs built by us (Table 2). The results show that the prediction accuracy of 3dRNA is significantly higher than these methods, especially for larger or more topologically-complex RNAs. As mentioned above, Liang and Schlick [47–48] found that the mean prediction accuracy (RMSD) of existing prediction methods is 20 Å. For example, the RMSD of MC-Sym on our database of 185 RNAs was 16.9 Å, while that of 3dRNA is 7.52 Å. This indicates that the accuracy of existing prediction methods decreases more rapidly with increasing RNA length than 3dRNA does. For longer RNAs, choosing correct conformations for various loops is critical to the highly-precise

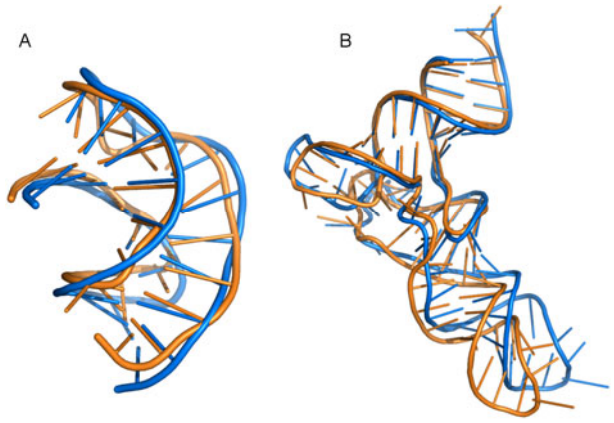

Figure 1 Predicted tertiary structures of typical RNA molecules. A, Pseudoknot RNA (PDB ID: 1KPZ). B, tRNA (PDB ID: 1J1U). The predicted structures (blue) are superimposed on their respective experimental structures (gold).

prediction of their global structures.

3 Using lncRNA expression profiles derived from microarrays to predict lncRNA functions

Researchers have identified tens of thousands of lncRNAs from mammalian genomes, and have systematically studied the function of many of them [6,50,51]. The functions of some lncRNAs, such as HOTAIR [52], AIR [53], and Kcnq1ot1 [54], have been characterized experimentally, and computational methods for large-scale prediction of lncRNA function have also been established [55,56]. It is important to note that determining lncRNA expression levels is the most important factor for studying their functions. To effectively and conveniently detect the expression level of lncRNAs, we developed a combined microarray containing probes that targeted both mRNAs and lncRNAs. Version 3 of the combined microarray will be released soon.

We collected lncRNAs from 15 data sources (Table 3). GENCODE released a comprehensive set of 22444 lncRNAs in March 2013 [57]. Most of the other data sources focused on specific lncRNAs, such as lncRNAs from HOX gene loci and ultraconserved regions. Some lncRNAs in one data source were also included in other data sources, usually with different IDs. To obtain a non-redundant set of lncRNAs, we based it on lncRNAs released by GENCODE, then added lncRNAs from other data sources that were not identified by GENCODE. Version 3 of the combined microarray contained approximately 37000 non-redundant lncRNAs.

One probe was designed for each lncRNA, based on its sequence. We tried to design probes that would specifically bind to their targets without cross-hybridization and would hybridize under similar conditions. The mRNA probes were provided by CapitalBio Corporation (Beijing, China).

The combined microarray of lncRNAs and mRNAs will enable studies of lncRNA function, especially in diseases. The combined microarray will also provide clues for molecular mechanisms of a large number of unknown lncRNAs.

Based on expression profiles of both lncRNAs and mRNAs, we have developed a computational pipeline for large-scale annotation of lncRNA functions [55,56,66]. First, gene expression values for several datasets were used to construct a two-color coding/non-coding gene co-expression network. Then, multiple methods, including hub-based, module-based, and global-based analysis, can be used to predict lncRNA functions based on the co-expression network. The hub-based method assigns functions to central

Table 1 Comparison of the mean C4-atom RMSD values of predictions of 13 RNA tertiary structures by 3dRNA, FARNA, V-fold, and MC-Sym

Method	3dRNA	FARNA	V-fold	MC-Sym
Mean RMSD (Å)	3.18	4.37	3.84	3.58

Table 2 Comparison of predictions (the mean heavy-atom RMSD values) of 185 RNA tertiary structures by 3dRNA, iFoldRNA, RNA2D3D, and MC-Sym

Method	3dRNA	iFoldRNA	RNA2D3D	MC-Sym
Mean RMSD(Å)	3.97	6.87	6.37	5.87

Table 3 The number of lncRNAs from various data sources with all three versions of the combined microarray

Sources	CBC lncRNA V1	CBC lncRNA V2	CBC lncRNA V3
GENCODE/ENSEMBL [57,58]	–	12754	22444
Human lincRNA catalog [51]	–	8195	14353
RefSeq [59]	4765	4765	4814
UCSC [60]	13521	13521	5596
NRED [61]	1289	1289	13701
H-InvDB [62]	17203	17203	1038
Enhancer-like lncRNA [63]	2975	2975	3019
RNAdb [64]	–	–	1599
Antisense ncRNA pipeline	1053	1053	1053
UCRs	481	481	962
CombinedLit	529	529	529
Hox ncRNAs	–	407	407
snoRNA	389	389	389
lncRNAdb [65]	78	78	104
ncRNAs from Chen lab	848	848	848
Total	42283	63639	70856
Unique lncRNAs	30622	35024	37491

unannotated lncRNAs (hub) according to the functional enrichment of their neighboring protein-coding genes. Moreover, previous studies have demonstrated that genes within a co-expressed module usually possess similar functions [56]. Thus, lncRNA functions could also be deduced from the protein-coding genes within the same modules. The above two methods were exclusively based on a local strategy that can only annotate a limited number of lncRNAs. We have recently developed a long non-coding RNA global function predictor (lnc-GFP) for large-scale prediction of lncRNAs functions (http://www.bioinfo.org/ncfans/), based on the same co-expression network [55].

4 Future directions

Although the abundance and functional importance of lncRNAs has been verified, exploring their mechanisms is a developing but difficult field. A deep understanding of their structure, expression profile, and other information, such as chromatin signature, will be required to characterize the functional mechanisms of lncRNAs, using advances in high-throughput approaches including RNA-seq [2], chromatin immunoprecipitation sequencing (ChIP-seq) [67], RNA immunoprecipitation sequencing (RIP-seq) [68], and chromatin isolation by RNA purification (CHIRP-seq) [69]. Many questions remain to be addressed in this rapidly expanding field. Specifically, detailed knowledge of structure-function relationships in lncRNAs is still limited, and the functional significance of tissue- or cell line-specific expression of lncRNAs is also uncharacterized, and these issues should be addressed in future studies.

This work was supported by the National High Technology Research and Development Program of China (2012AA020402) and National Natural Science Foundation of China (11074084, 30970558).

1 Bernstein B E, Birney E, Dunham I, et al. An integrated encyclopedia of DNA elements in the human genome. Nature, 2012, 489: 57–74
2 Ozsolak F, Milos P M. RNA sequencing: Advances, challenges and opportunities. Nat Rev Genet, 2011, 12: 87–98
3 Sun L, Luo H, Bu D, et al. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. Nucleic Acids Res, 2013, doi: 10.1093/nar/gkt646
4 Bu D, Yu K, Sun S, et al. NONCODE v3.0: Integrative annotation of long noncoding RNAs. Nucleic Acids Res, 2012, 40: D210–D215
5 Mercer T R, Dinger M E, Mattick J S. Long non-coding RNAs: Insights into functions. Nat Rev Genet, 2009, 10: 155–159
6 Sun L, Luo H, Liao Q, et al. Systematic study of human long intergenic non-coding RNAs and their impact on cancer. Sci China Life Sci, 2013, 56: 324–334
7 Zuker M, Stiegler P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res, 1981, 9: 133–148
8 Hofacker I L, Stadler P F. Memory efficient folding algorithms for circular RNA secondary structures. Bioinformatics, 2006, 22: 1172–1176
9 Reuter J S, Mathews D H. RNAstructure: Software for RNA secondary structure prediction and analysis. BMC Bioinformatics, 2010, 11: 129
10 Steffen P, Voss B, Rehmsmeier M, et al. RNAshapes: An integrated RNA analysis package based on abstract shapes. Bioinformatics, 2006, 22: 500–503
11 Mathews D H, Turner D H. Prediction of RNA secondary structure by free energy minimization. Curr Opin Struct Biol, 2006, 16: 270–278
12 Mathews D H, Turner D H. Dynalign: An algorithm for finding the secondary structure common to two RNA sequences. J Mol Biol, 2002, 317: 191–203
13 Bernhart S H, Hofacker I L, Will S, et al. RNAalifold: Improved consensus structure prediction for RNA alignments. BMC Bioinformatics, 2008, 9: 474
14 Ruan J, Stormo G D, Zhang W. ILM: A web server for predicting RNA secondary structures with pseudoknots. Nucleic Acids Res, 2004, 32: W146–W149
15 Lyngso R B, Pedersen C N. RNA pseudoknot prediction in energy-based models. J Comput Biol, 2000, 7: 409–427
16 Reeder J, Steffen P, Giegerich R. pknotsRG: RNA pseudoknot folding including near-optimal structures and sliding windows. Nucleic Acids Res, 2007, 35: W320–W324
17 Chen X, He S M, Bu D, et al. FlexStem: Improving predictions of RNA secondary structures with pseudoknots by reducing the search space. Bioinformatics, 2008, 24: 1994–2001
18 Mathews D. Predicting the secondary structure common to two RNA sequences with Dynalign. Curr Protoc Bioinform, 2004, Chapter 12: Unit 12.4
19 Zwieb C, Muller F. Three-dimensional comparative modeling of RNA. Nucleic Acids Symp Ser, 1997, 41: 69–71
20 Burks J, Zwieb C, Muller F, et al. Comparative 3-D modeling of tmRNA. BMC Mol Biol, 2005, 6: 14
21 Massire C, Westhof E. MANIP: An interactive tool for modelling RNA. J Mol Graph Model, 1998, 16: 197–205, 255–257
22 Tsai H Y, Masquida B, Biswas R, et al. Molecular modeling of the three-dimensional structure of the bacterial RNase P holoenzyme. J Mol Biol, 2003, 325: 661–675
23 Shapiro B A, Yingling Y G, Kasprzak W, et al. Bridging the gap in RNA structure prediction. Curr Opin Struct Biol, 2007, 17: 157–165
24 Hajdin C E, Ding F, Dokholyan N V, et al. On the significance of an RNA tertiary structure prediction. RNA, 2010, 16: 1340–1349
25 Das R, Baker D. Automated *de novo* prediction of native-like RNA tertiary structures. Proc Natl Acad Sci USA, 2007, 104: 14664–14669
26 Rohl C A, Strauss C E M, Misura K M S, et al. Protein structure prediction using Rosetta. Methods Enzymol, 2004, 383: 66–93
27 Das R, Kudaravalli M, Jonikas M, et al. Structural inference of native and partially folded RNA by high-throughput contact mapping. Proc Natl Acad Sci USA, 2008, 105: 4144–4149
28 Das R, Karanicolas J, Baker D. Atomic accuracy in predicting and designing noncanonical RNA structure. Nat Methods, 2010, 7: 291–294
29 Jonikas M A, Radmer R J, Laederach A, et al. Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. RNA, 2009, 15: 189–199
30 Sharma S, Ding F, Dokholyan N V. iFoldRNA: Three-dimensional RNA structure prediction and folding. Bioinformatics, 2008, 24: 1951–1952
31 Ding F, Sharma S, Chalasani P, et al. *Ab initio* RNA folding by discrete molecular dynamics: From structure prediction to folding mechanisms. RNA, 2008, 14: 1164–1173
32 Ding F, Dokholyan N V. Emergence of protein fold families through rational design. PLoS Comput Biol, 2006, 2: e85
33 Gherghe C M, Leonard C W, Ding F, et al. Native-like RNA tertiary structures using a sequence-encoded cleavage agent and refinement by discrete molecular dynamics. J Am Chem Soc, 2009, 131: 2541–2546
34 Frellsen J, Moltke I, Thiim M, et al. A probabilistic model of RNA conformational space. PLoS Comput Biol, 2009, 5: e1000406
35 Xia Z, Gardner D P, Gutell R R, et al. Coarse-grained model for simulation of RNA three-dimensional structures. J Phys Chem B,

2010, 114: 13497–13506
36 Cannone J J, Subramanian S, Schnare M N, et al. The comparative RNA web (CRW) site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. BMC Bioinformatics, 2002, 3: 2
37 Gutell R R, Lee J C, Cannone J J. The accuracy of ribosomal RNA comparative structure models. Curr Opin Struct Biol, 2002, 12: 301–310
38 Martinez H M, Maizel J V Jr., Shapiro B A. RNA2D3D: A program for generating, viewing, and comparing 3-dimensional models of RNA. J Biomol Struct Dyn, 2008, 25: 669–683
39 Zhao Y, Gong Z, Xiao Y. Improvements of the hierarchical approach for predicting RNA tertiary structure. J Biomol Struct Dyn, 2011, 28: 815–826
40 Yingling Y G, Shapiro B A. Dynamic behavior of the telomerase RNA hairpin structure and its relationship to dyskeratosis congenita. J Mol Biol, 2005, 348: 27–42
41 Yingling Y G, Shapiro B A. The prediction of the wild-type telomerase RNA pseudoknot structure and the pivotal role of the bulge in its formation. J Mol Graph Model, 2006, 25: 261–274
42 Yingling Y G, Shapiro B A. The impact of dyskeratosis congenita mutations on the structure and dynamics of the human telomerase RNA pseudoknot domain. J Biomol Struct Dyn, 2007, 24: 303–320
43 Cao S, Chen S J. Physics-based *de novo* prediction of RNA 3D structures. J Phys Chem B, 2011, 115: 4216–4226
44 Major F, Turcotte M, Gautheret D, et al. The combination of symbolic and numerical computation for three-dimensional modeling of RNA. Science, 1991, 253: 1255–1260
45 Parisien M, Major F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. Nature, 2008, 452: 51–55
46 Jossinet F, Ludwig T E, Westhof E. Assemble: An interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. Bioinformatics, 2010, 26: 2057–2059
47 Laing C, Schlick T. Computational approaches to RNA structure prediction, analysis, and design. Curr Opin Struct Biol, 2011, 21: 306–318
48 Laing C, Schlick T. Computational approaches to 3D modeling of RNA. J Phys Condens Matter, 2010, 22: 283101
49 Zhao Y, Huang Y, Gong Z, et al. Automated and fast building of three-dimensional RNA structures. Sci Rep, 2012, 2: 734
50 Mercer T R, Mattick J S. Structure and function of long noncoding RNAs in epigenetic regulation. Nat Struct Mol Biol, 2013, 20: 300–307
51 Cabili M N, Trapnell C, Goff L, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev, 2011, 25: 1915–1927
52 Rinn J L, Kertesz M, Wang J K, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell, 2007, 129: 1311–1323
53 Nagano T, Mitchell J A, Sanz L A, et al. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science, 2008, 322: 1717–1720
54 Pandey R R, Mondal T, Mohammad F, et al. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell, 2008, 32: 232–246
55 Guo X, Gao L, Liao Q, et al. Long non-coding RNAs function annotation: A global prediction method based on bi-colored networks. Nucleic Acids Res, 2013, 41: e35
56 Liao Q, Liu C, Yuan X, et al. Large-scale prediction of long non-coding RNA functions in a coding-non-coding gene co-expression network. Nucleic Acids Res, 2011, 39: 3864–3878
57 Derrien T, Johnson R, Bussotti G, et al. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. Genome Res, 2012, 22: 1775–1789
58 Flicek P, Amode M R, Barrell D, et al. Ensembl 2012. Nucleic Acids Res, 2012, 40: D84–D90
59 Pruitt K D, Tatusova T, Brown G R, et al. NCBI Reference Sequences (RefSeq): Current status, new features and genome annotation policy. Nucleic Acids Res, 2012, 40: D130–D135
60 Chan P P, Holmes A D, Smith A M, et al. The UCSC Archaeal Genome Browser: 2012 update. Nucleic Acids Res, 2012, 40: D646–D652
61 Dinger M E, Pang K C, Mercer T R, et al. NRED: A database of long noncoding RNA expression. Nucleic Acids Res, 2009, 37: D122–D126
62 Takeda J, Yamasaki C, Murakami K, et al. H-InvDB in 2013: An omics study platform for human functional gene and transcript discovery. Nucleic Acids Res, 2013, 41: D915–D919
63 Orom U A, Derrien T, Beringer M, et al. Long noncoding RNAs with enhancer-like function in human cells. Cell, 2010, 143: 46–58
64 Pang K C, Stephen S, Dinger M E, et al. RNAdb 2.0—an expanded database of mammalian non-coding RNAs. Nucleic Acids Res, 2007, 35: D178–D182
65 Amaral P P, Clark M B, Gascoigne D K, et al. lncRNAdb: A reference database for long noncoding RNAs. Nucleic Acids Res, 2011, 39: D146–D151
66 Liao Q, Xiao H, Bu D, et al. ncFANs: A web server for functional annotation of long non-coding RNAs. Nucleic Acids Res, 2011, 39: W118–W124
67 Blecher-Gonen R, Barnett-Itzhaki Z, Jaitin D, et al. High-throughput chromatin immunoprecipitation for genome-wide mapping of *in vivo* protein-DNA interactions and epigenomic states. Nat Protoc, 2013, 8: 539–554
68 Zhao J, Ohsumi T K, Kung J T, et al. Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol Cell, 2010, 40: 939–953
69 Chu C, Qu K, Zhong F L, et al. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol Cell, 2011, 44: 667–678