Predicting RNA:DNA Triplex Structures from Sequence Features Using Deep Learning Architecture

Joseph L. Tsenum[1], Maria Poptsova[2]

[1]Phystech School of Biological and Medical Physics, Moscow Institute of Physics and Technology (MIPT), Moscow Region, Russia

[2]International Laboratory of Bioinformatics, Higher School of Economics (HSE University), Moscow, Russia

*Corresponding Author: Joseph L. Tsenum; Phystech School of Biological and Medical Physics, Moscow Institute of Physics and Technology (MIPT); tsenumjosephluper@phystech.edu

Abstract

Long non-coding RNAs (lncRNAs) can perform their regulatory roles by forming triple helices through RNA–DNA interactions. Although this has been verified by a few in vivo and in vitro methods, robust *in silico* approaches that predict the potential of lncRNAs and DNA sites to form triplex structures are still required. Tools such as Triplexator have predicted vast numbers of lncRNAs and DNA sites with triplex forming potential, yet there remains a pressing need for advanced computational methods that can refine and extend these predictions. In this study, we developed ten (10) deep neural network models that predict the potential of lncRNAs and DNA sites to form triple helices on a genome-wide scale. To prepare our dataset, we first used Triplexator to screen out lncRNAs and DNA sites with low triplex-forming potential. We then trained different deep learning architectures, including two-layer convolutional neural networks (CNN), residual neural networks (ResNN), long short-term memory recurrent neural networks (LSTM-RNN), and multilayer perceptron (MLP). Among these architectures, our lncRNA_CNN and LSTM3-RNN both achieved a mean AUC of 0.99 for lncRNA features at a kernel size of 32 and a learning rate of 1e-3. For DNA site features, our DNA_CNN achieved the best performance with a mean AUC of 0.98 under the same conditions. In conclusion, we demonstrate that deep neural network architectures can effectively learn sequence features of lncRNAs and DNA to accurately predict RNA:DNA triplex formation potential, providing a scalable *in silico* framework for studying genome-wide triplex biology.

Keywords: *Long* non-coding RNAs (lncRNAs), DNA sites, triplex structures, deep learning architectures

Introduction

Advances in genome sequencing technologies in the past decade have shown the complexity of genome organization in both eukaryotic and prokaryotic organisms (Kapranov *et al.*, 2007; ENCODE Project Consortium, 2004, 2012). These genome sequencing technologies has revealed large amounts of non- coding RNAs (ncRNAs) that were previously coined as 'junk DNA' and a small fraction of protein-coding mRNA in the mammalian DNA.

Efforts have been made in the past years in sequencing ncRNAs and quantifying their various regulatory roles (Quinn and Chang, 2016; Bonasio and Shiekhat-tar, 2014; Batista and Chang, 2013; Lee, 2012). Non coding RNAs (ncRNAs) consists of a broad class of regulatory RNAs; these include small ncRNAs such as small interferingRNAs (siRNAs), PIWI-interacting RNAs, microRNAs (miRNAs), and long noncoding RNAs (lncRNAs),which are made up of more than 200 nucleotide in length (Cech and Steitz, 2014; Kowalczyk *et al.*, 2012). Majority of the mammalian non-coding RNAs are highly conserved, suggesting that they perform significant functions in biological processes (Struhl, 2007).

Long non- coding RNAs (lncRNAs) interacts with epigenetic factors to structure the results of important biological roles such as gene transcription regulation, modulation of nuclear organization, pre-mRNA splicing and X inactivation in females. This has improved our knowledge of the interaction between lncRNAs and epigenetic factors of human diseases as biomarkers of genetic diseases and their therapeutic targets (David J. H *et al.*, 2018).

Long non-coding RNAs (lncRNAs) are capable of forming RNA-DNA hybrid duplexes and RNA-DNA triplexes structures at specific DNA sequences (Wang and Chang, 2011; Li *et al.*, 2016). These triplex-forming lncRNAs are important in biological processes. Both *in-vivo* and *in-vitro* methods have verified the potential of some lncRNAs in forming triplex structures with DNA sequences. Various mathematical, statistical and computational methods have also been applied in predicting lncRNAs that have the potential of forming triplex structures with specific DNA regions. Computationally, classical machine learning and deep neural networks have shown promise in predicting these triplex-forming lncRNAs and DNA sites.

Materials and methods

Data Collection & Preprocessing for our Triplex Forming Prediction Model (TriplexFPM) for Sequence-Based Feature

531 positive dataset for triplex lncRNAs, *36021* negative dataset for lncRNA triplex; *2542* DNA Site triplex positive dataset and *12,735* DNA Site triplex negative data were collected as follow;

Triplex lncRNA Prediction Positive Dataset for TriplexFPM

We obtained *531 positive dataset for triplex lncRNAs* were collected from the work of Zhang *et al.*, 2020. They collected this positive data for triplex lncRNA prediction in two ways; They extracted lncRNAs according to TriplexRNA regions (DNA:RNA triplex forming peaks in RNA) contained in the work of Sentürk *et al*. by considering both Solid Phase Reversible Immobilization-based paramagnetic bead size selection and immunopurification with anti-DNA antibody RNA separation in Hela S3 cell. They then used GENCODE release 33 lncRNA annotation to extract the lncRNAs that covered the TriplexRNA regions and *476* unique samples were gotten. These lncRNAs were named as triplexlncRNA.

They also obtained a total number of *159* reported triplex lncRNAs that were verified by either *in vivo* or *in vitro* assays to form triplexes with DNA. They obtained these from peer-reviewed papers. These lncRNAs were named reported triplex lncRNA, which included MEG3 (Mondal *et al*., 2015), PARTICL (O'Leary *et al*., 2015), MIR100HG (Wang *et al*., 2018), FENDRR (Navarro *et al*., 2016), and HOTAIR (Kalwa *et al*., 2016). They also took into account all the variants of the reported triplex lncRNA. This brought the number of the collected positive data for triplex lncRNA samples to *635*.Triplex Domain Finder-TDF (Kuo *et al*., 2019) was then used to filter the data since the task is to predict the most likely triplex-forming lncRNAs. In the course of evaluating the triplex forming potential of the *635* lncRNAs with the whole gene promoters (except for chromosome Y and M) with TDF (default parameters), they found out that *104* of them did not contain DNA Binding Domains (DBDs)with powerful Triplex Forming Oligonucleotide (TFO) support.

The challenge of low triplex forming potential gotten from TDF in the collected positive data may be because; for reported triplex lncRNA, each lncRNA gene may have many transcripts with splice variants, but perhaps not all of them have the potential of forming triple helix with DNA, and secondly, for triplexlncRNA, they could not confirm that the overlap regions between R-Loop and TriplexRNA regions as having triplex forming potentials or not (Engreitz *et al*., 2013). To have confidence in the positive dataset, they removed the *104* lncRNAs having low triplex forming potentials and finally, *531* positive triplexlncRNAs dataset were left and we used same in our work.

Triplex lncRNA Prediction Negative Dataset for TriplexFPM

We obtained *36021 negative datasets for lncRNA triplex* from the work of Zhang *et al.*, 2020. They collected the negative lncRNA triplex samples the same way they did for the positive dataset. After removing the lncRNAs in the original positive dataset from GENCODE annotation, they then evaluated the triplex forming potential of all remaining lncRNAs with the whole gene promoters (except for chrY and chrM) using Triplex Domain Finder (TDF). The lncRNAs with at least one powerful Triplex forming oligonucleotides (TFO) supported DBD and at least *123* DNA Binding Sites (DBSs) were kept (the smallest number of DBSs was *123* in positive data). Again, from the screened lncRNAs, they removed *one* lncRNA with letter 'N' (variants of lncRNA MALAT1) which was previously confirmed to form RNA-RNA triplex (Ageeli *et al*., 2019). At last, they got *36,021* negative dataset for the lncRNAs that had the comparable potentials of forming triplehelix as their lncRNA negative data and we used same in our task.

Triplex DNA Sites Prediction Positive Dataset for TriplexFPM

We collected *2542 DNA triplex positive dataset* from the work of Zhang *et al*., 2020.

To predict triplex-forming sites in DNA based on the experimental data, they adopted TriplexDNA regions (DNA:RNA triplex forming peaks in DNA) from the work of Sentürk *et a*l., 2019 as their positive data and we did same. They enriched the RNA-associated DNAs by an unbiased approach. After thy removed *5* samples containing the letter 'N' in their sequences, they finally got *2542* DNApositive dataset and we used same in our work.

Triplex DNA Sites Prediction Negative Dataset for TriplexFPM

They obtained *12,735 DNA triplex negative data* as random regions in promoters. They downloaded all ensembl annotated promoters at Hunt *et al*., 2018, from which *12,735* regions (about 5 times amount of originally positive data) was obtained. They got these regions by randomly selecting chromosomes (except chrY and chrM) and DNA regions with the same lengths as TriplexDNA. They extracted the sequence data for both positive and negative from the DNA minus strand.

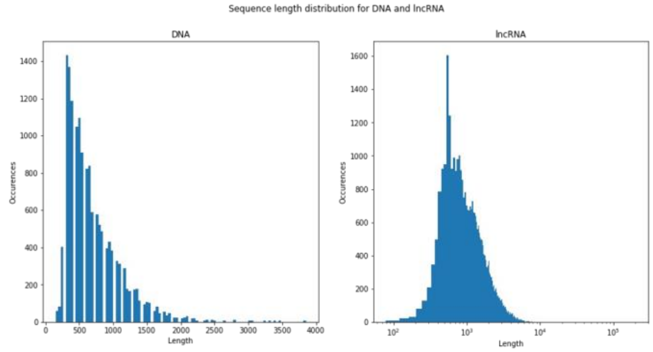


Fig 1.1: Triplex lncRNA and DNA site length distributions for TriplexFPM

Model Construction for TriplexFPM

We imported the libraries we needed from python (mainly Tensorflow and Keras), biopython (mainly SeqIO) and others. We then denoted some supplementary functions that were needed to extract sequence features used for training/testing. We balanced our data by enriching our positive data by oversampling and depleting our negative data by undersampling. We fixed the data size problem for lncRNA triplex as lncRNA_samples = 1500 (3 times bigger than the size of the positive samples), while the data size for DNA triplex as DNA_samples = 6000 (2.4 times biggerthan the size of the positive samples).

Feature Extraction

K-mer is a commonly used approach to transform a sequence into a vector. It usually counts the frequencies of single or multiple nucleotide compositions in a sequence and represents the sequence into a 4 k dimensional vector while kmerscore is an overall measure of the k-mer nucleotide composition bias in a sequence, it is extracted from k-mer features (Zhang *et al.*, 2020). While for our CNN and ResNN, in order to extract our features, we used *K-mer and kmerscore from 1-6,* for our LSTM RNN, we used one-hot encode.

After preprocessing our data, we merged the positive and negative data, created label arrays, shuffled our data and saved our sequence data, labels and features as *processed data* for model construction. We then defined our constant hyperparameters by splitting our data into *train, validation and test* datasets with *80%, 10% and 10%* respectively as shown below:

Data size lncRNA = *3000*, Data size DNA = *12000*

We reloaded the processed_data files and scaled features for both datasets using standard scaler to enable the model to calculate distances between data for better model performance. Standardization transforms the data to have zero mean and a variance of 1, they make our data unitless (Baijayanta, 2020). Taking lncRNA and DNA sites as our features, we adopted a 2-layerConvolutional Neural Network (CNN) to construct our CNN models, which effectively learned high-level features from our data. We also used LSTM-RNN and ResNN to train our models. We first trained *5 models* predict the most likely triplex-forming lncRNAs in practice from the lncRNAs triplex forming potentials data as estimated by computational methods. We also trained *5* other models to predict the potential of DNA sites in forming triplex structures from the experimentally verified data as reported by Zhang *et al.*, 2020.

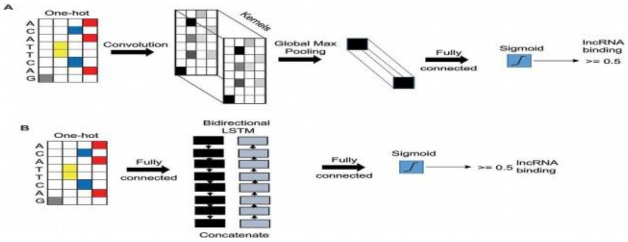


Fig. 1.2: Ideas for Model construction for TriplexFPM (Wang *et al.*, 2020)

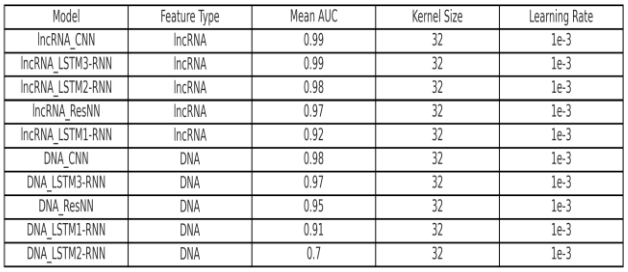
Hyperparameter Tuning for TriplexFPM

We defined our internal and external hyperparameters for our CNN, and ResNN as follows: epochs (50), input shape of (90, 1), learning rate at 1e-3 for 32 Kernel size (which performed better) and we set the verbose at 1 just to see how the training progress for each epoch, ReLU as our hidden layer activation function and Sigmoid as our output activation function. For the LSTM RNN models, our internal and external hyperparameters were as follow: epochs (80), input shape of (none, 4), learning rate at 1e-3 for 32 Kernel size, ReLU as our hidden layer activation function and Sigmoid as our output activation function. We used Adam optimization algorithm as our optimizer and since it is a binary classification task, we used binary cross entropy as our loss function. We used six different metrices to assess the performance of our models in all the architectures, including: binary accuracy, precision, recall and AUC, F1- score, MCC. We employed early stopping criteria on the optimization procedure when the values of the loss function on the validation set stopped decreasing at a particular consecutive time.

Results and discussion

We trained a total of ten (10) deep neural network models based on lncRNA and DNA site features. These models constitute our Triplex Forming Prediction Model (TriplexFPM) and include:

- **lncRNA-based models:** lncRNA_CNN, lncRNA_ResNN, lncRNA_LSTM1-RNN, lncRNA_LSTM2-RNN, lncRNA_LSTM3-RNN.
- **DNA-based models:** DNA_CNN, DNA_ResNN, DNA_LSTM1-RNN, DNA_LSTM2-RNN, DNA_LSTM3-RNN.



Model	Feature Type	Mean AUC	Kernel Size	Learning Rate
lncRNA_CNN	lncRNA	0.99	32	1e-3
lncRNA_LSTM3-RNN	lncRNA	0.99	32	1e-3
lncRNA_LSTM2-RNN	lncRNA	0.98	32	1e-3
lncRNA_ResNN	lncRNA	0.97	32	1e-3
lncRNA_LSTM1-RNN	lncRNA	0.92	32	1e-3
DNA_CNN	DNA	0.98	32	1e-3
DNA_LSTM3-RNN	DNA	0.97	32	1e-3
DNA_ResNN	DNA	0.95	32	1e-3
DNA_LSTM1-RNN	DNA	0.91	32	1e-3
DNA_LSTM2-RNN	DNA	0.7	32	1e-3

Table 1.1: TriplexFPM (lncRNA/DNA feature-based models)

Performance of TriplexFPM (lncRNA/DNA Models)

Our feature-based models trained on lncRNA and DNA sequences (TriplexFPM) demonstrated consistently high predictive performance across multiple architectures (Table 1.1). Among the lncRNA-based models, **lncRNA_CNN** and **lncRNA_LSTM3-RNN** achieved the highest performance with a mean AUC of **0.99** at a kernel size of 32 and a learning rate of 1e-3, indicating that convolutional and recurrent architectures are both capable of effectively capturing lncRNA sequence patterns relevant to triplex formation. This was closely followed by **lncRNA_LSTM2-RNN** (AUC = 0.98) and **lncRNA_ResNN** (AUC = 0.97), while **lncRNA_LSTM1-RNN** performed lowest (AUC =

0.92).

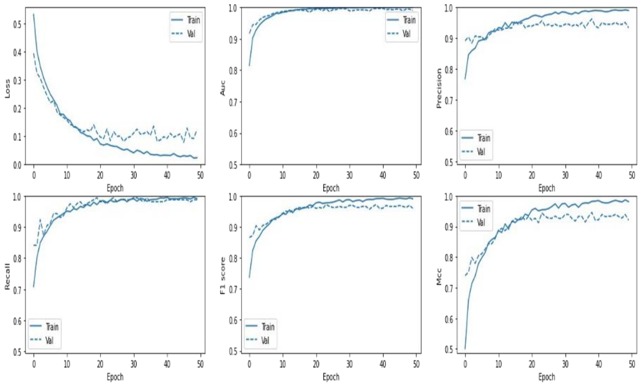


Fig. 2. 1: CNN Performance Plot for lncRNA TriplexFPM

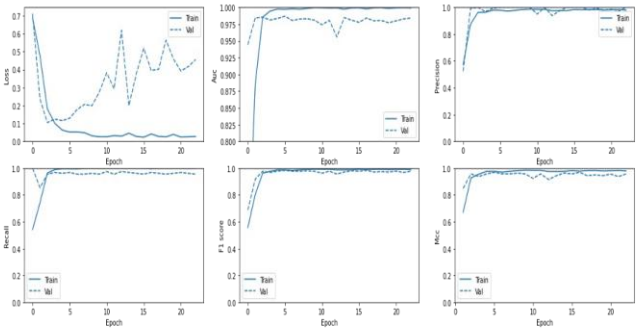


Fig 2.2: LSTM3-RNN performance plot for lncRNA TriplexFPM

For DNA feature-based models, **DNA_CNN** exhibited the best performance with a mean AUC of **0.98**, followed by **DNA_LSTM3-RNN** (AUC = 0.97) and **DNA_ResNN** (AUC = 0.95). The relatively weaker performance of **DNA_LSTM1-RNN** (AUC = 0.91) and **DNA_LSTM2-RNN** (AUC = 0.70) suggests that certain recurrent architectures may be less effective at capturing DNA sequence signals compared to convolutional and residual networks.

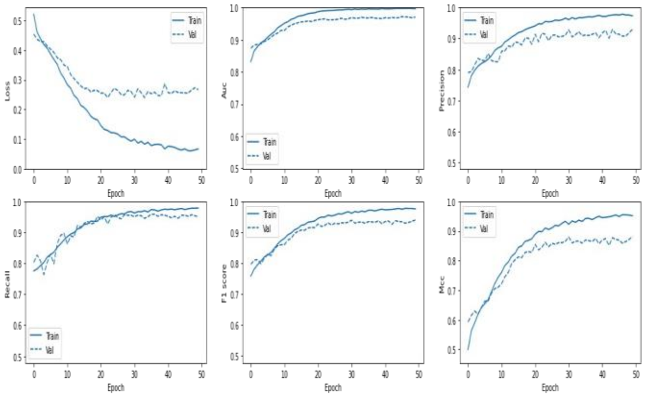


Fig 2.3: CNN performance plot for DNA site TriplexFPM

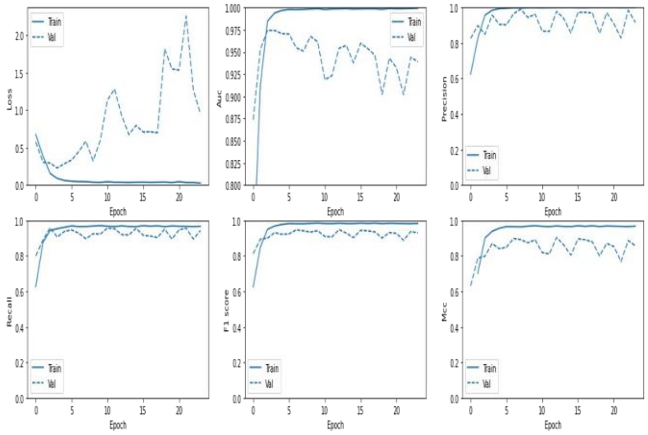


Fig 2.4: LSTM3 RNN performance plot for DNA site TriplexFP

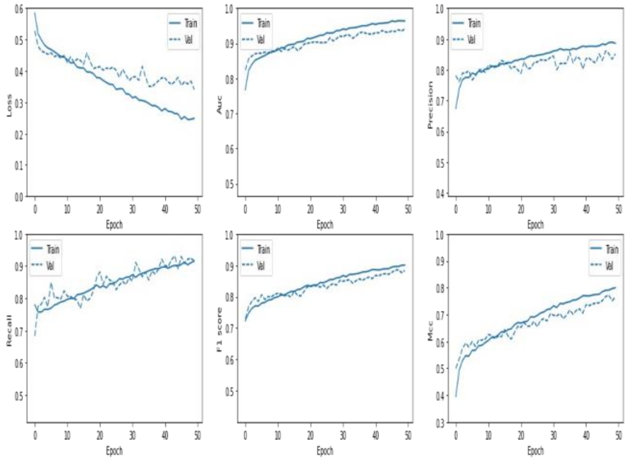


Fig 2.5: ResNN performance plot for DNA site TriplexFPM

Overall, the high area under the ROC curve (AUC) values across most architectures indicate that the lncRNAs and DNA sites identified by Triplexator as having triplex-forming potential indeed contain strong sequence signals that can be captured by deep neural networks. The superior performance of convolutional models highlights the importance of localized motif detection in triplex prediction, while the variability among recurrent architectures suggests the need for careful tuning of sequence-dependent models.

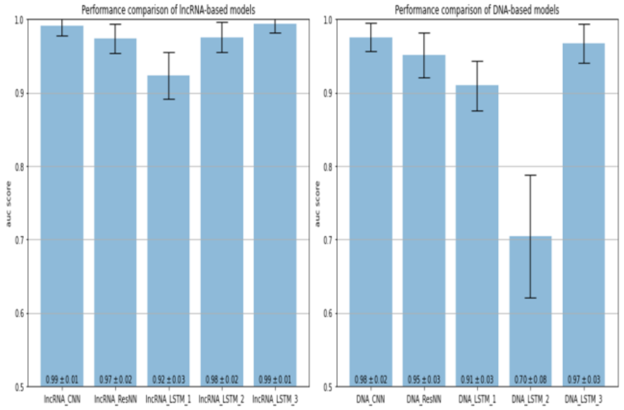


Fig 2.6: Performance comparison of different lncRNA & DNA site feature-based models on mean AUC

Limitations.

One of the challenges encountered in this study was the limited availability of lncRNA positive datasets with confirmed triplex-forming potential. Deep learning models generally benefit from large and diverse datasets, and the relatively small positive dataset used here may have constrained the performance of some models across specific metrics. In addition, since our dataset was collected from previously published studies and public repositories rather than generated experimentally, discrepancies in data quality and annotation could have contributed to inconsistencies in performance and reduced confidence in some results.

Future Directions

To address these challenges, future work should expand training beyond promoter regions to include additional genomic contexts, incorporating histone modification marks as features to further refine triplex prediction. Training different DNA regions separately and later integrating them may also improve performance. Beyond deep learning, classical machine learning models such as support vector machines, random forests, and decision trees should be benchmarked against our models to evaluate their comparative effectiveness. From a computational standpoint, implementing models in PyTorch instead of Keras-TensorFlow may provide greater flexibility and optimization opportunities. Importantly, optimizing kernel sizes and learning rates for each performance metric rather than applying uniform hyperparameters may yield further improvements. Finally, experimental validation of the predicted triplex-forming lncRNAs and DNA sites using in vitro and in vivo assays remains essential. Such efforts could establish biomarkers for complex human diseases and open new avenues for therapeutic intervention.

Declarations

Ethics Approval and Consent to Participate

Not applicable

Consent for Publication

Not applicable

Availability of Data and Materials

The datasets and deep learning models used in this study are openly available on GitHub at https://github.com/Joseph-Luper-Tsenum/Predicting-RNA-DNA-Triplex-Structures-from-Sequence-Features-Using-Deep-Learning-Architectures/tree/main

Conflicts of interest

The authors declare no conflict of interests for this article.

Funding

This research was funded by the federal budget within the quota of the Russian Government, 2019–2021.

Authors' Contribution

JLT designed the study, performed the analyses, and wrote the manuscript. MP supervised the study and contributed valuable feedback.

Acknowledgments

The authors thank the Open Doors Olympiad for providing the scholarship, the Russian Government for funding, Phystech School of Biological and Medical Physics (FBMP) at the Moscow Institute of Physics and Technology (MIPT) for their support, and the members of the International Laboratory of Bioinformatics at HSE University for their valuable assistance and guidance.

Abbreviations

- CNN = Convolutional neural network
- RNN = Recurrent neural network
- ResNN = Residual neural network
- LSTM = Long short-term memory
- MLP = Multilayer perceptron
- RNA = Ribonucleic acid
- DNA = Deoxyribonucleic acid
- lncRNA = long non-coding RNA
- TriplexFPM = Triplex Forming Prediction Model
- AUC = Area under the Receiver operating characteristic (ROC) Curve
- ReLU = Rectified Linear Unit
- lr = Learning rate
- MCC = Matthews correlation coefficient
- k-mer = A k-mer is a subsequence of length k derived from a longer nucleotide or amino acid sequence.

Supplementary Figures

Additional results of the **TriplexFPM** model are provided in the supplementary figures and tables. These include extended performance metrics, comparative analyses, and visualizations that complement the main text.

A. lncRNA-based models: lncRNA_ResNN, lncRNA_LSTM1-RNN, lncRNA_LSTM2-RNN

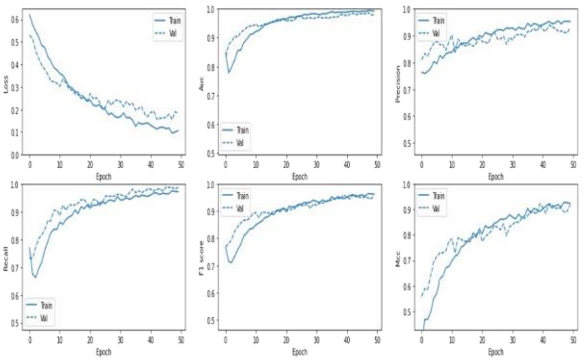


Fig 3. 1: ResNN performance plot for lncRNA TriplexFPM

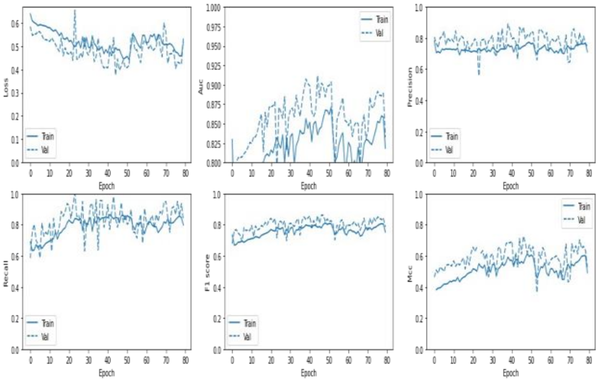


Fig 3.2: LSTM1-RNN performance plot for lncRNA TriplexFPM

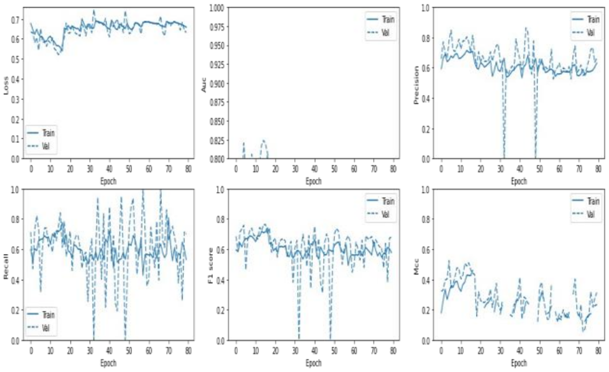


Fig 3.3: LSTM2-RNN performance plot for lncRNA TriplexFPM

B. DNA-based models: DNA_ResNN, DNA_LSTM1-RNN, DNA_LSTM2-RNN

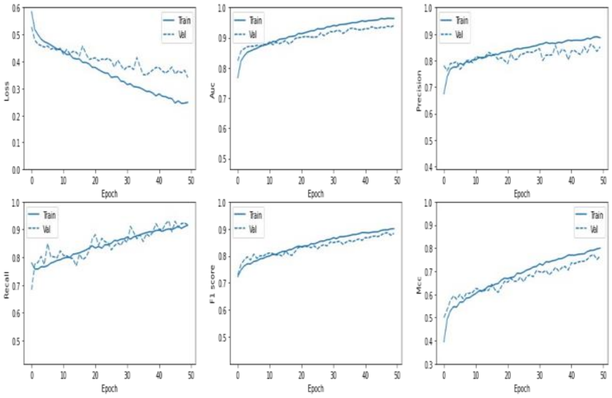


Fig 4.1: ResNN performance plot for DNA site TriplexFPM

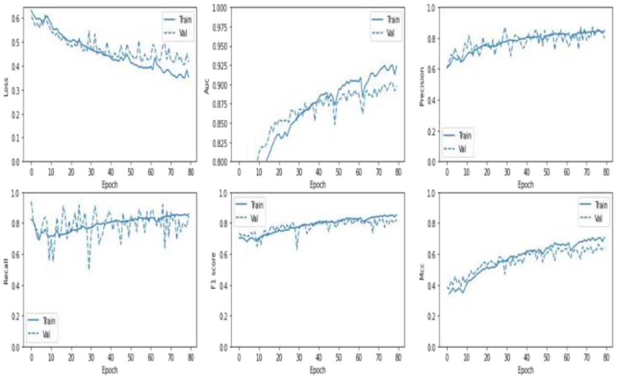


Fig 4.2: LSTM1-RNN performance plot for DNA site TriplexFPM

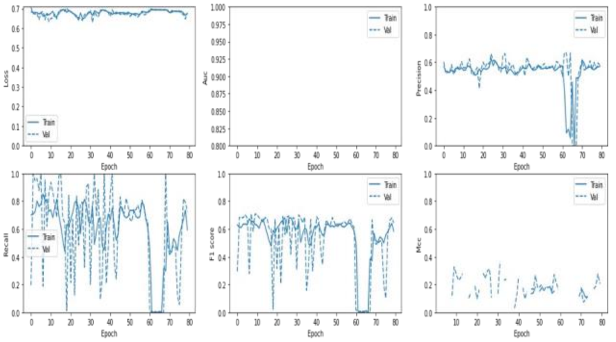


Fig 4.3: LSTM2-RNN performance plot for DNA site TriplexFPM

C. Performance comparison of different lncRNA & DNA site feature-based models

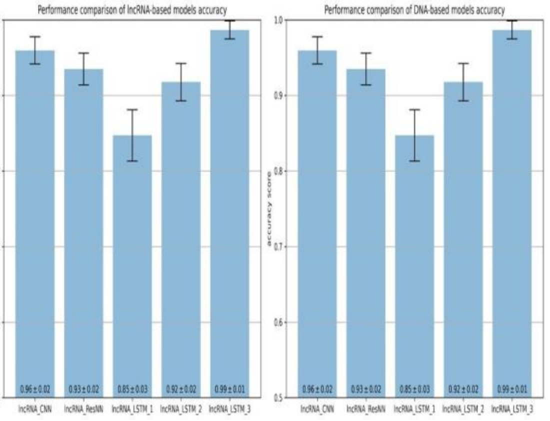


Fig 5.1: Performance comparison of different lncRNA & DNA site feature-based models on mean accuracy

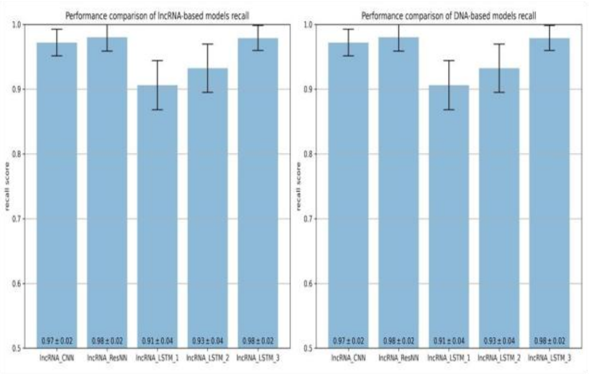


Fig 5.2: Performance comparison of different lncRNA & DNA site feature-based models on mean recall

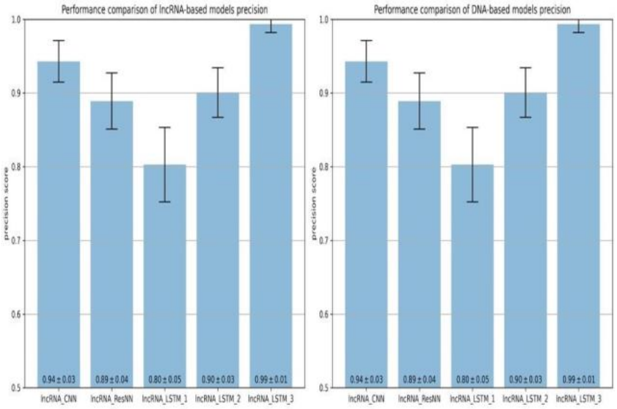


Fig 5.3: Performance comparison of different lncRNA & DNA site feature-based models on mean precision

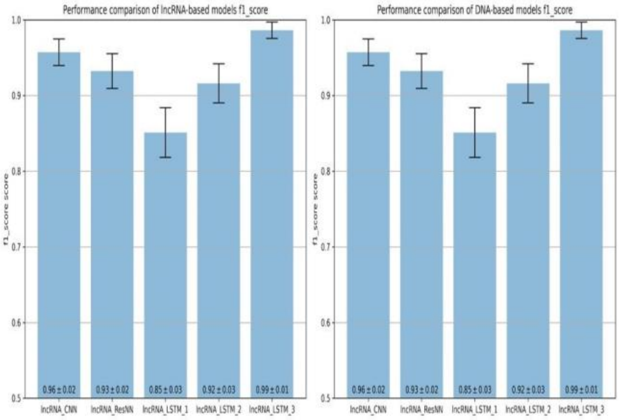


Fig 5. 4: Performance comparison of different lncRNA & DNA site feature-based models on mean F1-score

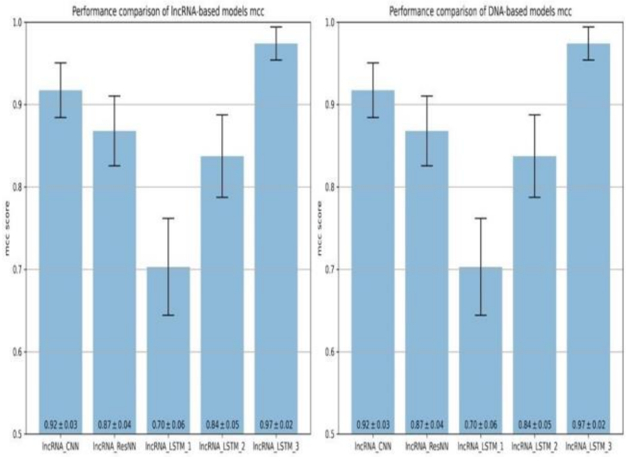


Fig 5.5: Performance comparison of different lncRNA & DNA site feature-based models on mean MCC

References

Ageeli AA, McGovern-Gooch KR, Kaminska MM, Baird NJ. Finely tuned conformationaldynamics regulate the protective function of the lncRNA MALAT1 triple helix. Nucl Acids Res. 2019;47(3):1468–81.

Aguilo F., Zhou M.M., Walsh M.J. Long noncoding RNA, polycomb, and the ghosts hauntingINK4b-ARF-INK4a expression. *Cancer Res.* 2011;71:5365–5369.

Altman N, Krzywinski M. The curse(s) of dimensionality. Nat Methods. 2018;15:399–400.

Amaral PP, Clark MB, Gascoigne DK, Dinger ME, Mattick JS. lncRNAdb: A reference database for long noncoding RNAs. Nucleic Acids Res. 2011 Jan; 39(Database issue):D146-51.

Anderson, D.M.; Anderson, K.M.; Chang, C.L.; Makarewich, C.A.; Nelson, B.R.; McAnally, J.R.; Kasaragod, P.; Shelton, J.M.; Liou, J.; Bassel-Duby, R.; et al. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. Cell 2015, 160, 595–606. [CrossRef]

Angrand, P.-O.; Vennin, C.; Le Bourhis, X.; Adriaenssens, E. The role of long non-coding RNAsin genome formatting and expression. Front. Genet. 2015, 6,
165. [CrossRef] [PubMed] 2

Antonov,I.V., Mazurov,E., Borodovsky,M. and Medvedeva,Y.A. (2018) Prediction of lncRNAs and their interactions with nucleic acids: benchmarking bioinformatics tools. Brief. Bioinformatics, 2018, 1–14.

Azhikina TL, Ignatov DV, Salina EG, Fursov MV, Kaprelyants AS (2015) Role of Small NoncodingRNAs in Bacterial Metabolism. Biochemistry 80: 1633-1646.

Bacolla A, Wang G, Vasquez KM (2015) New Perspectives on DNA and RNA Triplexes As Effectors of Biological Activity. PLoS Genet 11(12): e1005696.

Batista, P., and Chang, H. (2013). Long noncoding RNAs: cellular address codes in development and disease. Cell 152, 1298–1307.

Beltran M, Puig I, Peña C, García JM, Alvarez AB, Peña R, et al. A natural antisensetranscript regulates Zeb2/Sip1 gene expression during Snail1- induced epithelial-mesenchymal transition. *Genes Dev.* 2008;22:756–69. doi: 10.1101/gad.455708. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

Ben-Hur A, Ong CS, Sonnenburg S, Schölkopf B, Rätsch G (2008) Support Vector Machinesand Kernels for Computational Biology. PLoS Comput Biol 4(10): e1000173. https://doi.org/10.1371/journal.pcbi.1000173

Bernstein HD, Zopf D, Freymann DM, Walter P. Functional substitution of the signal recognition particle 54-kDa subunit by its Escherichia coli homolog. Proc Natl Acad Sci U S A. 1993 Jun 1; 90(11):5229-33.
Bartolomei MS, Zemel S, Tilghman SM. Parental imprinting of the mouse H19 gene. *Nature.* 1991;351:153–5. doi: 10.1038/351153a0. [PubMed] [CrossRef] [Google Scholar]
Birney E, Stamatoyannopoulos JA, Dutta A, *et al*; ENCODE Project Consortium; NISCComparative Sequencing Program; Baylor College of Medicine Human Genome Sequencing Center; Washington University Genome Sequencing Center; Broad Institute; Children's Hospital Oakland Research Institute; Identification and analysis of functional elements in 1% of the human genome by the encode pilot project. Nature. 2007;447:799–816.

Blume, S.W.; Meng, Z.; Shrestha, K.; Snyder, R.C.; Emanuel, P.D. The 5 0 - untranslated RNA of the human dhfr minor transcript alters transcription pre- initiation complex assembly at themajor (core) promoter. J. Cell. Biochem. 2003, 88, 165–180. [CrossRef] [PubMed]
Bo Chang, Lili Meng, Eldad Haber, Frederick Tung, and David Begert. Multi- level Residual Networks From Dynamical Systems View. ICLR, pp. 1–14, 2018.
Bonasio, R., and Shiekhattar, R. (2014). Regulation of transcription by long noncoding RNAs. Annu. Rev. Genet. 48, 1–23.
Breiman L. Random Forests, *Mach Learn*, 2001, vol. 45 (pg.5-32) Google Scholar Crossref

Brockdorff N, Ashworth A, Kay GF, Cooper P, Smith S, McCabe VM, et al. Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. *Nature.*1991;351:329–31. doi: 10.1038/351329a0. [PubMed] [CrossRef] [Google Scholar]

Brown CJ, Hendrich BD, Rupert JL, Lafrenière RG, Xing Y, Lawrence J, Willard HF. The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and ishighly localized within the nucleus. Cell. 1992 Oct 30; 71(3):527-42.

Brown CJ, Ballabio A, Rupert JL, Lafreniere RG, Grompe M, Tonlorenzi R, et al. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature.* 1991;349:38–44. doi: 10.1038/349038a0. [PubMed] [CrossRef] [Google Scholar]

Buske FA, Bauer DC, Mattick JS, Bailey TL. Triplex-Inspector: an analysis tool for triplex-mediated targeting of genomic loci. *Bioinformatics*. 2013;29(15):1895-1897.
Buske,F.A., Bauer,D.C., Mattick,J.S. and Bailey,T.L. (2012) Triplexator: detecting nucleic acid triple helices in genomic and transcriptomic data. Genome Res., 22, 1372–1381.

Campalans A, Kondorosi A, Crespi M. Enod40, a short open reading frame- containing mRNA, induces cytoplasmic localization of a nuclear RNA binding protein in Medicago truncatula. Plant Cell. 2004 Apr;16(4):1047-59. doi: 10.1105/tpc.019406. Epub 2004 Mar 22. Erratum in: Plant Cell. 2017 Apr;29(4):912. PMID: 15037734; PMCID: PMC412876.

Carl Kingsford, Steven L Salzberg. Nat Biotechnol. Author manuscript; available in PMC 2009 Jun 24. Published in final edited form as: Nat Biotechnol. 2008 Sep; 26(9): 1011–1013. doi:10.1038/nbt0908-1011. PMCID: PMC2701298

Cech, T., and Steitz, J. (2014). The noncoding RNA revolution—trashing old rules to forgenew ones. Cell 157, 77–94.
Chang Z, Li G, Liu J, Zhang Y, Ashby C, Liu D, Cramer CL, Huang X. Bridger: a new framework for de novo transcriptome assembly using RNA-seq data. Genome Biol. 2015;16:30. doi: 10.1186/s13059-015-0596-2.

Charles Richard J.L., Ogawa Y. (2015) Understanding the Complex Circuitry of lncRNAs atthe X-inactivation Center and Its Implications in Disease Conditions. In: Morris K. (eds) Long Non-coding RNAs in Human Disease. Current Topics in Microbiology and Immunology, vol 394. Springer, Cham. https://doi.org/10.1007/82_2015_443.

Chu C, Quinn J, Chang HY. Chromatin isolation by RNA purification (ChIRP). J Vis Exp.2012;61. pii: 3912. doi: 10.3791/3912.

Chu C, Qu K, Zhong FL, Artandi SE, Chang HY. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol Cell. 2011;44:667–678. doi: 10.1016/j.molcel.2011.08.027.

Chu C, Zhang QC, da Rocha ST, Flynn RA, Bharadwaj M, Calabrese JM, Magnuson T, Heard E, Chang HY.

Systematic discovery of Xist RNA binding proteins. Cell. 2015;161:404–416. doi: 10.1016/j.cell.2015.03.025.10. Costa FF. Non-coding RNAs: new players in eukaryotic biology.Gene. 2005;357:83–94. doi: 10.1016/j.gene.2005.06.019.

Chugunova, A.; Navalayeu, T.; Dontsova, O.; Sergiev, P. Mining for small translated ORFs. J. Proteome Res. 2018, 17, 1–11. [CrossRef]

Clark MB, Mattick JS. Long noncoding RNAs in cell biology. *Semin Cell Dev Biol.* 2011;22:366–76. doi: 10.1016/j.semcdb.2011.01.001. [PubMed] [CrossRef] [Google Scholar]

Clemson CM, McNeil JA, Willard HF, Lawrence JB. XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. J Cell Biol. 1996 Feb; 132(3):259-75.

Deng, Z.; Norseen, J.; Wiedmer, A.; Riethman, H.; Lieberman, P.M. TERRA RNA binding toTRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. Mol. Cell. 2009, 35, 403–413. [CrossRef] [PubMed]

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, MerkelA, 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 Sep; 22(9):1775- 89.

Diederik P. Kingma and Jimmy Ba Adam: A Method for Stochastic Optimization. 3rd International Conference for Learning Representations, San Diego. (2015). https://arxiv.org/abs/1412.6980

Eldad Haber and Lars Ruthotto. Stable Architectures for Deep Neural Networks. ArXiv, 2017.

ENCODE Project Consortium. (2004). The ENCODE (ENCyclopedia of DNA elements) project. Science 306, 636–640.

ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74.

Engreitz JM, Pandya-Jones A, McDonel P, Shishkin A, Sirokman K, Surka C, Kadri S, Xing J, Goren A, Lander ES, Plath K. The Xist lncRNA exploits three- dimensional genome architecture to spread across the X chromosome. Science. 2013;341(6147):1237973.

Espinoza CA, Allen TA, Hieb AR, Kugel JF, Goodrich JA. B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. *Nat Struct Mol Biol.* 2004;11:822–9. doi:10.1038/nsmb812. [PubMed] [CrossRef] [Google Scholar]

Fang S., Zhang L., Guo J., Niu Y., Wu Y., Li H., Zhao L., Li X., Teng X., Sun X., et al.NONCODEV5: A comprehensive annotation database for long non-coding RNAs. *Nucleic AcidsRes.* 2018;46:D308–D314. doi: 10.1093/nar/gkx1107.

Fan Wang, Pranik Chainani, Tommy White, Jin Yang, Yu Liu & Benjamin Soibam (2018)Deep learning identifies genome-wide DNA binding sites of long noncoding RNAs, RNA Biology, 15:12, 1468-1476, DOI: 10.1080/15476286.2018.1551704.

Kopp F (2019) Molecular functions and biological roles of long non-coding RNAs in humanphysiology and disease. J Gene Med 21(8):e3104

Fullwood MJ, Wei CL, Liu ET, Ruan Y. Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. Genome Res. 2009;19:521–532. doi: 10.1101/gr.074906.107.

Gaddis,S.S., Wu,Q., Thames,H.D., DiGiovanni,J., Walborg,E.F., MacLeod,M.C. and Vasquez,K.M. (2006) A web-based search engine for triplex-forming oligonucleotide target sequences. Oligonucleotides, 16, 196–201

Gardini A. Global run-on sequencing (GRO-Seq). Methods Mol Biol. 2017;1468:111–120. doi: 10.1007/978-1-4939-4035-6_9.

Garzia, A.; Meyer, C.; Morozov, P.; Sajek, M.; Tuschl, T. Optimization of PAR- CLIP for transcriptome-wide identification of binding sites of RNA-binding proteins. Methods 2017, 118–119, 24–40. [CrossRef]

Ginger MR, Shore AN, Contreras A, Rijnkels M, Miller J, Gonzalez-Rimbau MF, et al. A noncoding RNA is a potential marker of cell fate during mammary gland development. *Proc Natl Acad Sci U S A.* 2006;103:5781–6. doi: 10.1073/pnas.0600745103. [PMC free article] [PubMed][CrossRef] [Google Scholar]

Goñi,J.R., de la Cruz,X. and Orozco,M. (2004) Triplex-forming oligonucleotide target sequences in the human genome. Nucleic Acids Res., 32, 354–360.

Grabherr MG, Haas BJ, Yassour M, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29:644– 652. doi: 10.1038/nbt.1883.

Grote, P.; Herrmann, B.G. The long non-coding RNA Fendrr links epigenetic control mechanisms to gene regulatory networks in mammalian embryogenesis. RNA Boil. 2013, 10, 1579–1585. [CrossRef] [PubMed]

Grote P., Wittler L., Hendrix D., Koch F., Wahrisch S., Beisaw A. The tissue- specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Dev Cell*.2013;24:206–214.

Guil, S., and Esteller, M. (2012). Cis-acting noncoding RNAs: friends and foes. Nat. Struct. Mol. Biol. 19, 1068–1075.

Guttman, M., and Rinn, J. (2012). Modular regulatory principles of large non- coding RNAs. Nature 482, 339–346.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O.,Carey, B.W., Cassady, J.P., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature 458, 223–227.

Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, et al. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell.* 2010;142:409–19. doi: 10.1016/j.cell.2010.06.040. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

He,S., Zhang,H., Liu,H. and Zhu,H. (2014) LongTarget: a tool to predict lncRNA DNA-binding motifs and binding sites via Hoogsteen base-pairing analysis. Bioinformatics, 31, 178–186.

Hirose T, Virnicchi G, Tanigawa A, Naganuma T, Li R, Kimura H, et al. NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies. Mol Biol Cell 2014;25:169–83.

Hunt SE, McLaren W, Gil L, Thormann A, Schuilenburg H, Sheppard D, Parton A, Armean IM, Trevanion SJ, Flicek P, Cunningham F. Ensembl variation resources. Database. 2018;1:2018.

Hombach S., Kretz M. (2016) Non-coding RNAs: Classification, Biology and Functioning. In: Slaby O., Calin G. (eds) Non-coding RNAs in Colorectal Cancer. Advances in Experimental Medicine and Biology, vol 937. Springer, Cham. *https://doi.org/10.1007/978-3-319-42059-2_1*.

Ingolia N.T., Lareau L.F., Weissman J.S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell.* 2011;147:789–802.

Inoue F, Kircher M, Martin B, et al. A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome Res*. 2017;27(1):38-52. PMID: 27831498; PMCID: PMC5204343

Jalali, S.; Gandhi, S.; Scaria, V.S. Distinct and modular organization of protein interactingsites in long non-coding RNAs. Front. Mol. Biosci. 2018, 5, 27. [CrossRef]

Jeon Y, Lee JT. YY1 tethers Xist RNA to the inactive X nucleation center. *Cell.* 2011;**146**(1):119–133. doi: 10.1016/j.cell.2011.06.026. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

Jessica A. Brown, Max L. Valenstein, Therese A. Yario, Kazimierz T. Tycowski, Joan A. Steitz. MALAT1 and MENβ RNA 3′ ends form triple helices Proceedings of the National Academy of Sciences. (2012)109 (47) 19202-19207; DOI: 10.1073/ pnas.1217338109

Jianyin Long, … , Paul A. Overbeek, Farhad R. Danesh. The long noncoding RNA *Tug1* connects metabolic changes with kidney disease in podocytes. *J Clin Invest.* 2016;126(11):4205–4218.doi:10.1172/JCI87927.

Kaikkonen MU, Lam MT, Glass CK. Non-coding RNAs as regulators of gene expression and epigenetics. Cardiovasc Res. 2011;90:430–440. doi: 10.1093/cvr/cvr097.

Kallen A.N., Zhou X.B., Xu J., Qiao C., Ma J., Yan L. The imprinted H19 lncRNA antagonizes let-7 microRNAs. *Mol Cell.* 2013;52:101–112.

Kalwa M, Hänzelmann S, Otto S, Kuo CC, Franzen J, Joussen S, Fernandez- Rebollo E, Rath B, Koch C, Hofmann A, Lee SH. The lncRNA HOTAIR impacts on mesenchymal stem cells via triple helix formation. Nucl Acids Res. 2016;44(22):10631–43.

Kapranov, P., Willingham, A.T., and Gingeras, T.R. (2007). Genome-wide transcription and the implications for genomic organization. Nat. Rev. Genet. 8, 413–423.

Karapetyan, A.E.; Coen Buiting, C.; Kuiper, R.A.; Marcel, W.; Coolen, M.W. Regulatory rolesfor long ncRNA and mRNA. Cancers 2013, 5, 462–490. [CrossRef] [PubMed]

Katayama S, Tomaru Y, Kasukawa T, *et al*; RIKEN Genome Exploration Research Group;Genome Science Group (Genome Network Project Core Group); FANTOM Consortium. Antisense transcription in the mammalian transcriptome.Science. 2005;309:1564–1566. doi: 10.1126/science.1112009.

Kalwa M, Hänzelmann S, Otto S, et al. The lncRNA HOTAIR impacts on mesenchymal stem cells via triple helix formation. *Nucleic Acids Res*. 2016;44(22):10631-10643. doi:10.1093/nar/gkw802

Kazachenka A, Bertozzi TM, Sjoberg-Herrera MK, *et al*. Identification, Characterization, and Heritability of Murine Metastable Epialleles: Implications for Non-genetic Inheritance. *Cell*.2018;175(5):1259-1271.e13. PMID: 30454646; PMCID: PMC6242299

Kazimierczyk M, Kasprowicz MK, Kasprzyk ME, Wrzesinski J. Human Long Noncoding RNA Interactome: Detection, Characterization and Function. Int J Mol Sci.

2020 Feb 4;21(3):1027. doi:10.3390/ijms21031027. PMID: 32033158; PMCID: PMC7037361.

Ke H, Zhao L, Feng X, Xu H, Zou L, Yang Q, et al. NEAT1 is required for survival of breast cancer cells through FUS and miR-548. Gene Regul Syst Bio 2016;10:11–7.

Kevin L Howe, Premanand Achuthan, James Allen, Jamie Allen, Jorge Alvarez-Jarreta, M Ridwan Amode, Irina M Armean, Andrey G Azov, Ruth Bennett, Jyothish Bhai, Konstantinos Billis, Sanjay Boddu, Mehrnaz Charkhchi, Carla Cummins, Luca Da Rin Fioretto, Claire Davidson,Kamalkumar Dodiya, Bilal El Houdaigui, Reham Fatima, Astrid Gall, Carlos Garcia Giron, Tiago Grego, Cristina Guijarro-Clarke, Leanne Haggerty, Anmol Hemrom, Thibaut Hourlier, Osagie G Izuogu, Thomas Juettemann, Vinay Kaikala, Mike Kay, Ilias Lavidas, Tuan Le, Diana Lemos, Jose Gonzalez Martinez, José Carlos Marugán, Thomas Maurel, Aoife C McMahon, Shamika Mohanan, Benjamin Moore, Matthieu Muffato, Denye N Oheh, Dimitrios Paraschas, Anne Parker, Andrew Parton, Irina Prosovetskaia, Manoj P Sakthivel, Ahamed I Abdul Salam, Bianca M Schmitt, Helen Schuilenburg, Dan Sheppard, Emily Steed, Michal Szpak, Marek Szuba, Kieron Taylor, Anja Thormann, Glen Threadgold, Brandon Walts, Andrea Winterbottom, Marc Chakiachvili, AmeyaChaubal, Nishadi De Silva, Bethany Flint, Adam Frankish, Sarah E Hunt, Garth R IIsley, Nick Langridge, Jane E Loveland, Fergal JMartin, Jonathan M Mudge, Joanella Morales, Emily Perry, Magali Ruffier, John Tate, David Thybert, Stephen J Trevanion, Fiona Cunningham, Andrew D Yates, Daniel R Zerbino, Paul Flicek Ensembl 2021. Nucleic Acids Res. 2021, vol. 49(1):884–891. PubMed PMID: 33137190.

Kloc M, Wilk K, Vargas D, Shirato Y, Bilinski S, Etkin LD. Potential structural role of non-coding and coding RNAs in the organization of the cytoskeleton at the vegetal cortex of Xenopus oocytes. *Development.* 2005;132:3445–57. doi: 10.1242/dev.01919. [PubMed] [CrossRef] [Google Scholar]

Kotake Y., Nakagawa T., Kitagawa K., Suzuki S., Liu N., Kitagawa M. Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene.*Oncogene.* 2011;30:1956–1962.

Kowalczyk, M.S., Higgs, D.R., and Gingeras, T.R. (2012). Molecular biology: RNA discrimination. Nature 482, 310–311.

Krol, J.; Loedige, I.; Filipowicz, W. The widespread regulation of microRNA biogenesis,function and decay. Nat. Rev. Genet. 2010, 11, 597–610. [CrossRef].

Kung, J., Colognori, D., and Lee, J. (2013). Long noncoding RNAs: past, present, andfuture. Genetics 193, 651–669.

Kuo CC, Hänzelmann S, Sentürk Cetin N, Frank S, Zajzon B, Derks JP, Akhade VS, Ahuja G, Kanduri C, Grummt I, Kurian L. Detection of RNA–DNA binding sites in long noncoding RNAs. NuclAcids Res. 2019;47(6):e32.

Lars Ruthotto and Eldad Haber. Deep Neural Networks motivated by Partial Differential Equations. arXiv, pp. 1–7, 2018.

LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015 May 28;521(7553):436–44. [PubMed] [CrossRef] [Reference list]

Lee, J. (2012). Epigenetic regulation by long noncoding RNAs. Science 338, 1435–1439.

Li, J.; Ma, W.; Zeng, P.; Wang, J.; Geng, B.; Yang, J.; Cui, Q. LncTar: A tool for predicting the RNA targets of long noncoding RNAs. Brief. Bioinform. 2015. [CrossRef]

Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J. SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics. 2009;25:1966–1967. doi: 10.1093/bioinformatics/btp336.

Li, T., Mo, X., Fu, L., Xiao, B., and Guo, J. (2016). Molecular mechanisms of longnoncoding RNAs on gastric cancer. Oncotarget 7, 8601–8612.

Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, Merkurjev D, Zhang J, Ohgi K, Song X, Oh S, Kim HS, Glass CK, Rosenfeld MG. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature. 2013;498:516–520. doi: 10.1038/nature12210

Li W, Notani D, Rosenfeld MG. Enhancers as non-coding RNA transcription units: recentinsights and future perspectives. Nat Rev Genet. 2016;17:207–223. doi: 10.1038/nrg.2016.4.

Li X.L., Wu Z.Q., Fu X.B., Han W.D. LncRNAs: insights into their function and mechanics in underlying disorders. *Mutat Res Rev Mut Res.* 2014;762:1–21.

Li, Y.; Syed, J.; Sugiyama, H. RNA-DNA triplex formation by longnoncoding RNAs. Cell Chem. Biol. 2016, 23, 1325–1333. [CrossRef]

Loewer S, Cabili MN, Guttman M, Loh YH, Thomas K, Park IH, et al. Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat Genet.* 2010;42:1113–7. doi: 10.1038/ng.710. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

Maass PG, Luft FC, Bähring S. Long non-coding RNA in health and disease. J Mol Med(Berl). 2014;92:337–346. doi: 10.1007/s00109-014-1131-8.

Marco Ciccone, Marco Gallieri, Jonathan Masci, Christian Osendorfer, and Faustino Gomez. NAISNET: Stable Deep

Networks from Non-Autonomous Differential Equations. arXiv, 2018.

Martianov I, Ramadass A, Serra Barros A, Chow N, Akoulitchev A. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature.* 2007;445:666–70. doi: 10.1038/nature05519. [PubMed] [CrossRef] [Google Scholar]

Martorell-Marugán J, Tabik S, B4. enhammou Y, et al. Deep Learning in Omics Data Analysis and Precision Medicine. In: Husi H, editor. Computational Biology [Internet]. Brisbane (AU): Codon Publications; 2019 Nov 21. Chapter 3. Available from: https://www.ncbi.nlm.nih.gov/books/NBK550335/ doi: 10.15586/ computationalbiology.2019.ch3

Matsumoto, A.; Nakayama, K.I. Hidden peptides encoded by putative noncoding RNAs. Cell Struct. Funct. 2018, 43, 75–83. [CrossRef] [PubMed]

Matsumoto, A.; Pasut, A.; Matsumoto, M.; Yamashita, R.; Fung, J.; Monteleone, E.; Saghatelian, A.; Nakayama, K.I.; Clohessy, J.G.; Pandolfi, P.P. mTORC1 and muscle regenerationare regulated by the LINC00961-encoded SPAR polypeptide. Nature 2017, 541, 228–232. [CrossRef]
Mattick JS. The central role of RNA in human development and cognition. FEBS Lett. 2011;585:1600–1616. doi: 10.1016/j.febslet.2011.05.001.

McKenna SJ, Amaral T, Akbar S, Jordan L, Thompson A. Immunohistochemical analysisof breast tissue microarray images using contextual classifiers. *J Pathol Inform*.2013;4(Suppl):S13. Published 2013 Mar 30. doi:10.4103/2153-3539.109871

Mercer TR, Mattick JS. Structure and function of long noncoding RNAs in epigenetic regulation. Nat Struct Mol Biol. 2013;20(3):300.

Misawa A, Takayama KI, Inoue S. Long non-coding RNAs and prostate cancer. CancerScience. 2017 Nov;108(11):2107-2114. DOI: 10.1111/cas.13352.

Mishra K, Kanduri C. Understanding Long Noncoding RNA and Chromatin Interactions: WhatWe Know So Far. *Noncoding RNA*. 2019;5(4):54. Published 2019 Dec 3. doi:10.3390/ncrna5040054

Mondal T, Subhash S, Vaid R, Enroth S, Uday S, Reinius B, Mitra S, Mohammed A, James AR, Hoberg E, Moustakas A. MEG3 long noncoding RNA regulates the TGF-β pathway genes through formation of RNA–DNA triplex structures. Nat Commun. 2015;24(6):7743.

Moran V.A., Perera R.J., Khalil A.M. Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs. *Nucleic Acids Res.* 2012;40:6391–6400.

Morris, K., and Mattick, J. (2014). The rise of regulatory RNA. Nat. Rev. Genet. 15, 423–437.

Mourtada-Maarabouni M, Hedge VL, Kirkham L, Farzaneh F, Williams GT. Growth arrest in human T-cells is controlled by the non-coding RNA growth- arrest-specific transcript 5 (GAS5) *J Cell Sci.* 2008;121:939–46. doi: 10.1242/jcs.024646. [PubMed] [CrossRef] [Google Scholar]

Muddashetty R, Khanam T, Kondrashov A, Bundman M, Iacoangeli A, Kremerskothen J, et al. Poly(A)-binding protein is associated with neuronal BC1 and BC200 ribonucleoprotein particles. *J Mol Biol.* 2002;321:433–45. doi: 10.1016/S0022-2836(02)00655-1. [PubMed] [CrossRef] [GoogleScholar]

Navarro C, Cano C, Cuadros M, Herrera-Merchan A, Molina M, Blanco A. A mechanistic study of lncRNA Fendrr regulation of FoxF1 lung cancer tumor supressor. In: International conference on bioinformatics and biomedical engineering 2016 Apr 20 (pp. 781–789). Springer,Cham.

Nelson, B.R.; Makarewich, C.A.; Anderson, D.M.; Winders, B.R.; Troupes, C.D.; Wu, F.;Reese, A.L.; McAnally, J.R.; Chen, X.; Kavalali, E.T.; et al. A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. Science 2016, 351, 271–275. [CrossRef].

O'Leary VB, Smida J, Buske FA, Carrascosa LG, Azimzadeh O, Maugg D, Hain S, Tapio S, Heidenreich W, Kerr J, Trau M, Ovsepian SV, Atkinson MJ. PARTICLE triplexes cluster in the tumor suppressor WWOX and may extend throughout the human genome. Sci Rep. 2017 Aug 2;7(1):7163. doi: 10.1038/s41598-017-07295-5. PMID: 28769061; PMCID: PMC5541130.

O'Leary VB, Ovsepian SV, Carrascosa LG, Buske FA, Radulovic V, Niyazi M, Moertl S, Trau M, Atkinson MJ, Anastasov N. PARTICLE, a triplex-forming long ncRNA, regulates locus-specific methylation in response to low-dose irradiation. Cell Rep. 2015;11(3):474–85.

Orozco-Arias S, Piña JS, Tabares-Soto R, Castillo-Ossa LF, Guyot R, Isaza G. Measuring Performance Metrics of Machine Learning Algorithms for Detecting and Classifying Transposable Elements. *Processes*. 2020; 8(6):638. https://doi.org/10.3390/pr8060638

Ozsolak F, Platt AR, Jones DR, Reifenberger JG, Sass LE, McInerney P, Thompson JF, Bowers J, Jarosz M, Milos PM. Direct RNA sequencing. Nature. 2009;461:814–818. doi: 10.1038/nature08390.

Pandey R.R., Mondal T., Mohammad F., Enroth S., Redrup L., Komorowski J. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell.* 2008;32:232–246.

Pang KC, Frith MC, Mattick JS. Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. Trends Genet. 2006 Jan; 22(1):1- 5.

Park J, Yi D, Ji S. A Novel Learning Rate Schedule in Optimization for Neural Networks and It's Convergence. *Symmetry*. 2020; 12(4):660. https://doi.org/10.3390/sym12040660.

Philippe N, Bou Samra E, Boureux A, Mancheron A, Rufflé F, Bai Q, De Vos J, Rivals E,Commes T. Combining DGE and RNA-sequencing data to identify new polyA+ non-coding transcripts in the human genome. Nucleic Acids Res. 2014;42:2820–2832. doi:10.1093/nar/gkt1300.

Ponting, C.P., Oliver, P.L., and Reik, W. (2009). Evolution and functions of long noncoding RNAs. Cell 136, 629–641.

Postepska-Igielska A, Giwojna A, Gasri-Plotnitsky L, Schmitt N, Dold A, Ginsberg D, et al. LncRNA Khps1 regulates expression of the proto-oncogene SPHK1 via triplex-mediated changes in chromatin structure. *Mol Cell.* 2015;**60**:626–636. doi: 10.1016/j.molcel.2015.10.001. [PubMed] [CrossRef] [Google Scholar]

Pratik Chaudhari, Adam Oberman, Stanley Osher, Stefano Soatto, and Guillaume Carlier. Deep Relaxation: partial differential equations for optimizing deep neural networks. Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, PMLR, 2017. URL http://arxiv.org/abs/1704.04932.

Qianli Liao and Tomaso Poggio. Bridging the Gaps Between Residual Learning, Recurrent Neural Networks and Visual Cortex. arXiv, (047):1–16, 2016.

Quinn, J., and Chang, H. (2016). Unique features of long non-coding RNA biogenesis and function. Nat. Rev. Genet. 17, 47–62.

Quinodoz, S.; Guttman, M. Longnoncoding RNAs: An emerging link between gene regulation and nuclear organization. Cell Chem. Biol. 2014, 24, 652–663.

Reeves MB, Davies AA, McSharry BP, Wilkinson GW, Sinclair JH. Complex I binding by a virally encoded RNA regulates mitochondria-induced cell death. Science. 2007 Jun 1; 316(5829):1345-8.

Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. Annu Rev Biochem.2012;81:145–166. doi: 10.1146/annurev-biochem-051410-092902.

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell. 2007;129:1311–1323. doi: 10.1016/j.cell.2007.05.022. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

Rintala-Maki ND, Sutherland LC. Identification and characterisation of a novel antisensenon-coding RNA from the RBM5 gene locus. *Gene.* 2009;445:7–16. doi: 10.1016/j.gene.2009.06.009. [PubMed] [CrossRef] [Google Scholar]

Ritambhara Singh, Jack Lanchantin, Gabriel Robins, Yanjun Qi, DeepChrome: deep-learning for predicting gene expression from histone modifications, *Bioinformatics*, Volume 32, Issue 17, 1 September 2016, Pages i639–i648.

Romero-Barrios, N.; Legascue, M.F.; Benhamed, M.; Federico Ariel, F.; Crespi, M. Splicingregulation by long non-coding RNAs. Nucleic Acids Res. 2018, 46, 2169–2184. [CrossRef] [PubMed]

Ruiz-Orera, J.; Verdaguer-Grau, P.; Villanueva-Canas, J.L.; Messegue, X.; Alba, M.M.Functional and non-functional classes of peptides produced by long non-coding RNAs. biorXiv2016. [CrossRef]

Sallam T, Jones MC, Gilliland T, Zhang L, Wu X, Eskin A, Sandhu J, Casero D, Vallim TQ, Hong C, Katz M, Lee R, Whitelegge J, Tontonoz P. Feedback modulation of cholesterolmetabolism by the lipid-responsive non-coding RNA LeXis. Nature. 2016;534:124–128. doi:10.1038/nature17674.

Salmena, L.; Poliseno, L.; Tay, Y.; Kats, L.; Pandolfi, P.P. A ceRNA hypothesis: The Rosettastone of a hidden RNA language? Cell 2011, 146, 353–358. [CrossRef]

Sentürk Cetin N, Kuo CC, Ribarska T, Li R, Costa IG, Grummt I. Isolation and genome-wide characterization of cellular DNA: RNA triplex structures. Nucl Acids Res. 2019;47(5):2306–21.

Schulz MH, Zerbino DR, Vingron M, Birney E. Oases: robust de novo RNA- seq assembly across the dynamic range of expression levels. Bioinformatics. 2012;28:1086–1092. doi: 10.1093/bioinformatics/bts094.

Shamovsky I, Ivannikov M, Kandel ES, Gershon D, Nudler E. RNA-mediated response to heat shock in mammalian cells. *Nature.* 2006;440:556–60. doi: 10.1038/nature04518. [PubMed] [CrossRef] [Google Scholar]

Shaw, N.N.; Arya, D.P. Recognition of the unique structure of DNA: RNA hybrids. Biochimie2008, 90, 1026–1039. [CrossRef] [PubMed]

Sheik Mohamed J, Gaughwin PM, Lim B, Robson P, Lipovich L. Conserved longnoncoding RNAs transcriptionally regulated by Oct4 and Nanog modulate pluripotency in mouse embryonic stem cells. *RNA.* 2010;16:324–37. doi: 10.1261/rna.1441510. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

Shen, Z., Bao, W. & Huang, DS. Recurrent Neural Network for Predicting Transcription Factor Binding Sites. *Sci Rep* **8,** 15270 (2018). https://doi.org/10.1038/s41598-018-33321-1
Smillie, C.L.; Sirey, T.; Ponting, C.P. Complexities of post-transcriptional regulation and the modeling of ceRNA crosstalk. Crit. Rev. Biochem. Mol. Biol. 2018, 53, 231–245. [CrossRef].

St Laurent G, Wahlestedt C, Kapranov P. The landscape of long noncoding RNA classification. Trends Genet. 2015;31:239–251. doi: 10.1016/j.tig.2015.03.007.

Struhl, K. (2007). Transcriptional noise and the fidelity of initiation by RNA polymerase II. Nat. Struct. Mol. Biol. 14, 103–105.

Sunwoo H, Dinger ME, Wilusz JE, Amaral PP, Mattick JS, Spector DL. MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res.* 2009;19:347–59. doi: 10.1101/gr.087775.108. [PMC free article] [PubMed] [CrossRef] [Google Scholar].

Sun, X.; Haider Ali, M.S.S.; Moran, M. The role of interactions of long non- coding RNAsand heterogeneous nuclear ribonucleoproteins in regulating cellular functions. Biochem. J.2017, 474, 2925–2935. [CrossRef]

Swiezewski S, Liu F, Magusin A, Dean C. Cold-induced silencing by long antisensetranscripts of an Arabidopsis Polycomb target. Nature. 2009 Dec 10; 462(7274):799-802.

Tafer, H.; Hofacker, I.L. RNAplex: A fast tool for RNA-RNA interaction search. Bioinformatics 2008, 24, 2657–2663. [CrossRef]
Taft RJ, Pang KC, Mercer TR, Dinger M, Mattick JS. Non-coding RNAs: regulators ofdisease. J Pathol. 2010;220:126–39. doi: 10.1002/path.2638.

Takahashi H, Kato S, Murata M, Carninci P. CAGE (cap analysis of gene expression): aprotocol for the detection of promoter and transcriptional networks. Methods Mol Biol. 2012;786:181–200. doi: 10.1007/978-1-61779- 292-2_11.
Tano K, Akimitsu N. Long non-coding RNAs in cancer progression. *Front Genet.* 2012;3:219.doi: 10.3389/fgene.2012.00219. [PMC free article] [PubMed] [CrossRef] [Google Scholar]
Tay, Y.; Rinn, J.; Pandolfi, P.P. The multilayered complexity of ceRNA crosstalk andcompetition. Nature 2014, 505, 344–352. [CrossRef]

Tripathi, V.; Ellis, J.D.; Shen, Z.; Song, D.Y.; Pan, Q.; Watt, A.T.; Freier, S.M.; Bennett, C.F.; Sharma, A.; Bubulya, P.A.; et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol. Cell 2010, 39, 925–938. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

Wang J, Zhang J, Zheng H, Li J, Liu D, Li H, Samudrala R, Yu J, Wong GK. Mouse transcriptome: neutral evolution of 'non-coding' complementary DNAs. Nature. 2004 Oct 14;431(7010):1 p following 757; discussion following 757.
Wang, K.C., and Chang, H.Y. (2011). Molecular mechanisms of long noncod- ing RNAs. Mol. Cell 43, 904–914. [CrossRef]

Wang, Q., Ma, Y., Zhao, K. *et al.* A Comprehensive Survey of Loss Functions in Machine Learning. *Ann. Data. Sci.* (2020). https://doi.org/10.1007/s40745- 020-00253-5

Wang S, Ke H, Zhang H, Ma Y, Ao L, Zou L, Yang Q, Zhu H, Nie J, Wu C, Jiao B. LncRNA MIR100HG promotes cell proliferation in triple-negative breast cancer through triplex formation with p27 loci. Cell Death Dis. 2018;9(8):1–1. https://doi.org/10.1038/s41419-018-0869-2

Wapinski O, Chang HY. Long noncoding RNAs and human disease. *Trends Cell Biol.* 2011;21:354–61. doi: 10.1016/j.tcb.2011.04.001. [PubMed] [CrossRef] [Google Scholar]

Waseem Rawat, Zenghui Wang; Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Comput* 2017; 29 (9): 2352–2449. doi: https://doi.org/10.1162/neco_a_00990
Watanabe Y, Yamamoto MS. S. pombe mei2+ encodes an RNA-binding protein essential for premeiotic DNA synthesis and meiosis I, which cooperates with a novel RNA species meiRNA. *Cell.* 1994;78:487–98. doi: 10.1016/0092- 8674(94)90426-X. [PubMed] [CrossRef] [Google Scholar]

Wilusz, J.E., Sunwoo, H., and Spector, D.L. (2009). Long noncoding RNAs: functional surprises from the RNA world. Genes Dev. 23, 1494–1504.
Wouter G. Touw, Jumamurat R. Bayjanov, Lex Overmars, Lennart Backus, Jos Boekhorst, Michiel Wels, Sacha A. F. T. van Hijum, Data mining in the Life Sciences with Random Forest: awalk in the park or lost in the jungle?, *Briefings in Bioinformatics*, Volume 14, Issue 3, May 2013, Pages 315–326, https://doi.org/10.1093/bib/bbs034

Xu, C., Jackson, S.A. Machine learning and complex biological data.

Genome Biol **20,** 76 (2019). https://doi.org/10.1186/s13059-019-1689-0

Yang L., Froberg J.E., Lee J.T. Long noncoding RNAs: fresh perspectives into the RNA world. *Trends Biochem Sci.* 2014;39:35–43.
Yiping Lu, Aoxiao Zhong, Quanzheng Li, Massachusetts General Hospital, and Bin Dong.Beyond Finite Layer Neural Networks: Bridging Deep Architectures and Numerical Differential Equations. arxiv, pp. 1–15, 2017.

Yoon, J.H.; Gorospe, M. Cross-Linking Immunoprecipitation and qPCR (CLIP- qPCR) Analysis to map interactions between longnoncoding RNAs and RNA- binding proteins. Methods Mol. Biol. 2016, 1402, 11–17.

Yongqing Zhang, Jianrong Yan, Siyu Chen, Meiqin Gong, Dongrui Gao, Min Zhu* and Wei Gan, "Review of the Applications of Deep Learning in Bioinformatics", Current Bioinformatics 2020; 15(8) . https://doi.org/10.2174/1574893615999200711165743

Yue Li, Junetha Syed, Hiroshi Sugiyama. RNA-DNA Triplex Formation by Long NoncodingRNAs. *Cell Chemical Biology* (2016), 23 (11): 1325-1333

Zhang He, Michael J. Zeitz, Hong Wang, Beibei Niu, Shengfang Ge, Wei Li, Jiuwei Cui, Guanjun Wang, Guanxiang Qian, Michael J. Higgins, Xianqun Fan, Andrew R. Hoffman, Ji-Fan Hu. Long noncoding RNA-mediated intrachromosomal interactions promote imprinting at the *Kcnq1* locus. *J Cell Biol* (2014)204 (1): 61–75. https://doi.org/10.1083/jcb.201304152

Zhang Y, Jia C, Fullwood MJ, Kwoh CK. DeepCPP: a deep neural network based on nucleotide bias information and minimum distribution similarity feature selection for RNA coding potential prediction. Brief Bioinform. 2020.

Zhang Y, Jia C, Kwoh CK. Predicting the interaction biomolecule types for lncRNA: an ensemble deep learning approach. Brief Bioinform. 2020.

Zhang, Y., Long, Y. & Kwoh, C.K. Deep learning based DNA:RNA triplex forming potential prediction. BMC Bioinformatics 21, 522 (2020).

Zhao J, et al. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science.* 2008;**322**(5902):750–756. doi: 10.1126/science.1163045. [PMC freearticle] [PubMed] [CrossRef] [Google Scholar]

Zhu, S.; Wang, J.; He, Y.; Meng, N.; Yan, G.R. Peptides/Proteins encoded by non-coding RNA: A novel resource bank for drug targets and biomarkers. Front. Pharmacol. 2018, 9, 1295.[CrossRef]

Zitnik M, Nguyen F, Wang B, Leskovec J, Goldenberg A, Hoffman MM. Machine learning for integrating data in biology and medicine: principles, practice, and opportunities. Inf Fusion. 2019;50:71–91.