

RNA Structure in Esophageal Squamous Cell Carcinoma

Lu He*

Department of Pathology, Medical Science, Sweden

Abstract

Tumorigenesis is a multi-stage, dynamic biological process that involves several genetic and epigenetic changes, aberrant non-coding RNA expression, and modifications to the expression profiles of coding genes. We refer to this group of genome-space alterations as the "cancer initiome." In the genome, long non-coding RNAs are widely expressed and have important regulatory roles in chromatin remodelling and gene regulation. In both normal development and pathological conditions, such as cancer, spatial and temporal heterogeneity in lncRNA expression has been noted. Even though several dysregulated lncRNAs have been examined in malignancies, it is still unclear how lncRNAs contribute to the development of cancer, particularly in the case of esophageal squamous cell carcinoma. From ESCC and matched nearby non-cancerous normal tissues, we performed a genome-wide screen to determine the expression of lncRNAs and coding RNAs. In comparison to matched normal tissue equivalents, we discovered which lncRNAs and coding RNAs were differently expressed in ESCC. Using polymerase chain reaction analysis, we confirmed the conclusion. Additionally, we discovered lncRNAs that are differentially expressed in ESCC and that are co-localized and expressed with differentially expressed coding RNAs. These findings suggest a possible interaction between lncRNAs and nearby coding genes that affects either lipid metabolism and that this interaction may help to develop ESCC. These findings give strong support for a potential new genetic biomarker of esophageal squamous cell carcinoma.

Minimizing free energy, which is NP-hard, is frequently used to predict RNA secondary structures with pseudoknots. During transcription from DNA into RNA, the majority of RNAs fold in a hierarchical manner where secondary structures emerge before tertiary structures. Because of kinetics, local optimization is frequently used in real RNA secondary structures rather than global optimization. By taking dynamic and hierarchical folding mechanisms into account, the accuracy of RNA structure prediction may be increased. Based on a statistical examination of the actual RNA secondary structures of all 480 sequences from RNA STRAND, which are verified by NMR or X-ray, this study presents a fresh report on RNA folding that is consistent with the golden mean feature. With L standing for the sequence length, the length ratios of the domains in these sequences are roughly $0.382L$, $0.5L$, $0.618L$, and L . The key golden sections of the sequence are just these locations. This feature allows for the building of an algorithm that simulates RNA folding by dynamically folding RNA structures in accordance with the aforementioned golden section points while also predicting RNA hierarchical structures. The Mfold, HotKnots, McQfold, ProbKnot, and Lhw-Zhu algorithms cannot match our algorithm's sensitivity and quantity of predicted pseudoknots. As a result of a novel perspective that is near to natural folding, experimental results follow the RNA folding regulations.

Keywords: Carcinoma; Tumorigenesis

Introduction

One of the most typical cancers is esophageal squamous cell carcinoma, which is also one of the leading causes of cancer-related fatalities globally. In some parts of China, there is a notable regional difference as well as an extraordinarily high prevalence [1]. Even though ESCC treatment has become more multidisciplinary, the 5-year survival rate is still low. The beginning of ESCC is a dynamic, complex biological process that occurs in the genome and may involve multiple steps of genetic and epigenetic alterations, aberrations in noncoding RNA expression, and modifications to the expression profile of coding genes. Prior to now, significant signalling pathways implicated in cancer were identified by the expression profiling of coding genes. An even more complex picture of cancer is emerging from the most recent research on actively transcribed long noncoding RNAs from high-throughput sequencing [2]. Endogenous cellular RNA transcripts called lncRNAs lack a significant-length open reading frame and have lengths of 200–100,000 nucleotides. lncRNAs exhibit more tissue- and cell-specific expression patterns than protein-coding genes, but they are typically expressed at lower levels. lncRNAs were once thought to be transcriptional noise, but current research indicates that they play crucial roles in disease progression and proliferation, including cancer, as well as in the formation and differentiation of distinct cell types. Modifying chromatin architecture and controlling gene expression in a cis or trans manner are two ways that transcribed lncRNAs are

said to work. For instance, in the same chromosomal location, H19 lncRNA controls the expression of the IGF2 gene, and HOTAIR lncRNA is transcribed on Chr 12 and controls the HoxD gene on Chr 2. A "locus control" process, which mediates the localization of genes within nuclear regions to favour their transcription through the formation of domains of histone modification and intra- or interchromosomal loops, has also been reported to be used by lncRNAs to coordinate the regulation of nearby coding genes [3]. With diverse screening techniques, dysregulated lncRNAs in various cancer types have been discovered. For instance, the cancer-related lncRNA metastasis-associated lung adenocarcinoma transcript 1 (MALAT-1) was discovered by subtractive hybridization during screening for early non-small cell lung cancer with metastasis. In early stage lung cancer, overexpression of MALAT-1 is a strong indicator of a bad prognosis and shorter survival time.

***Corresponding author:** Lu He, Department of Pathology, Medical Science, Sweden, E-mail: lu@edu.pathologyhe.in

Received: 02-Sep-2022, Manuscript No: jbc22-73565, **Editor assigned:** 07-Sep-2022, PreQC No: jbc22-73565 (PQ), **Reviewed:** 21-Sep-2022, QC No: jbc22-73565, **Revised:** 23-Sep-2022, Manuscript No: jbc22-73565 (R), **Published:** 30-Sep-2022, DOI: 10.4172/jbc2.1000161

Citation: He L (2022) RNA Structure in Esophageal Squamous Cell Carcinoma. J Biochem Cell Biol, 5: 161.

Copyright: © 2022 He L. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

A small number of dysregulated lncRNAs have been found in a variety of malignancies, which raises the possibility that they are an elusive element of the entire transcriptome that plays a role in tumorigenesis, invasion, and metastasis [4]. The "lncRNAome" of various cancers is being investigated using cutting-edge high-throughput RNA sequencing methods, and dynamic variations in lncRNA expression have been seen in cancer cells at various phases of the disease as well as after treatment. But there is still much to learn about lncRNAs' function in cancer biology, and there isn't yet a comprehensive list of lncRNAs' biological properties that can be used to predict cancer outcomes. The relationships between lncRNA and coding genes can therefore be thoroughly searched for and analysed to derive putative biological functions.

Molecularly, RNAs are adaptable. Ribosomal RNAs, transfer RNAs, and other non-coding RNAs also have crucial structural, regulatory, and catalytic roles in cells, in addition to messenger RNAs' roles as carriers of genetic information and the link between DNA and proteins [5]. We must first comprehend RNAs' structural makeup in order to fully comprehend their varied biological activities. The order of nucleotides in RNA's single-stranded polymer constitutes its fundamental structure. The sequences in question are not just lengthy strands of nucleotides, though. Three hydrogen bonds can develop between the complementary bases of the guanine and cytosine pair in RNA, two hydrogen bonds can form between the complementary bases of the adenine and uracil pair, and two hydrogen bonds can form between the complementary bases of the guanine and uracil pair. RNA forms a three-dimensional structure thanks to hydrogen bonds. Hydrogen bonds between base and backbone as well as noncanonical pairing help to keep folding stable. The secondary structure is the collection of base pairs found in the folded RNA molecule's tertiary structure, which is a 3D arrangement of atoms. Because RNA tertiary structures cannot be determined experimentally because it would be too costly and time-consuming, computational biology has turned RNA structure prediction by computers into a fundamental technique and problem [6].

The scaffold of the tertiary structures is one of RNA's secondary structures. To predict RNA tertiary structures from RNA sequences, one must first predict RNA secondary structures. Thermodynamic, comparative, and hybrid approaches are the three categories of computational methods for predicting RNA secondary structures. On the basis of a set of energy parameters that have been found through experiment, thermodynamic techniques use dynamic programming to determine the best secondary structure for a single RNA sequence with the lowest possible global free energy [7]. For relatively short RNAs, these techniques have been effective. When a large number of homologous sequences are available, manually comparing methods are more trustworthy than thermodynamic methods. To determine the structures of recognised RNA families, manual comparison methods have been employed. In statistics and mutual information, quantitative measures of covariance have been used. These methods of explicitly taking sequence phylogeny into account produced fruitful outcomes. Recently developed hybrid approaches incorporate the best features of both thermodynamic and comparative methods. On as few as three homologous sequences, hybrid techniques show promise since they take into account both sequence covariance and thermodynamic stability. Other approaches are not included in any of these three categories. Some of these techniques aim to fold and align homologous sequences simultaneously, using stochastic context-free grammars to align homologous sequences iteratively and discover a consensus structure. Based on the statistical evaluation of actual RNA secondary structures, the current study offers a fresh conclusion that RNA folding follows

the golden mean feature [8]. The terms "golden section" and "golden ratio" are also used to refer to the golden mean. In the way that plant stems are arranged with their branches and the veins in their leaves, Adolf Zeising discovered the golden ratio to be expressed. In addition to studying animal skeletons, veins, and nerves, he also conducted research on crystal geometry and chemical compound proportions, as well as the application of proportion in artistic pursuits. He discovered that the golden ratio is a fundamental law governing these phenomena [9]. The strategy is put into practise using thermodynamic data, and its effectiveness is evaluated using the PKNOTS and TT2NE data sets. The preprocessing of the GM approach improves the sensitivity and PPV of the PKNOTS data set's Lhw-Zhu (LZ) and LZ algorithms by 2% to 3%. With regard to the TT2NE data set, the GM approach shows promising results in terms of predicting secondary and pseudoknotted structures. The experimental outcomes show the RNA folding regulations from a fresh perspective that is similar to natural folding. We provide a pilot analysis of the profiles of differentially expressed lncRNAs and coding RNAs from tumour and adjacent normal tissue of certain ESCC patients in order to comprehend the role of lncRNAs in this disease [10].

Material and Methods

Specimens

Before surgery, patients provided their written informed consent, and the Zhengzhou Hospital Institutional Review Board authorised the study protocol for the use of human subjects. In May 2012, patients with ESCC who received surgical therapy at Linxian Hospital provided primary tumours and nearby nonneoplastic tissues. Instantly following surgical resection, all tissues were immediately frozen in liquid nitrogen. None of the patients had ever undergone radiotherapy or chemotherapy, and none of them had any other life-threatening illnesses. At least two senior pathologists who work independently made histopathological diagnoses of all ESCC specimens.

Microarray Hybridization

The manufacturer's instructions were followed while extracting total RNAs using Trizol reagent. A 2100 Bioanalyzer was used to determine the quality of the RNAs. According to the Agilent One-Color Microarray-Based Gene Expression Analysis Low for Input Quick Amp Labeling kit, 100 ng of total RNA was used as the input to produce Cyanine-3 labelled cRNA. Agilent SurePrint G3 Human GE K Microarray was used for sample hybridization. Data was processed using Agilent Feature Extraction 11.0.1.1 after arrays were scanned using the Agilent DNA Microarray Scanner at a 3 m scan resolution.

Characteristic of Golden Section

We contrast the secondary and pseudoknotted structures, which are verified by NMR or X-ray, with the test set's 480 sequences (nonfragment and nonredundant). The results of statistical analysis on these actual secondary structures indicate the number of domains, the 3'-end of the group, the ratio of Group 1 to Group 2, and the ratio of Group 2 to Group 3. The number of sequences is indicated on the y-axis, and the length ratio of the 3'-end of the domain to the sequence is shown on the x-axis. In the finished structure, there are not enough complimentary bases to form a helix at one point.

Dynamic Algorithm

The helices and loops that make up secondary structures for RNAs are quickly produced along a hierarchical pathway, and the subsequent gradual folding of 3D tertiary structures would consolidate the

secondary structures. RNAs also fold when DNA is converted into RNA during transcription. As a result, we compute the secondary structure first before predicting pseudoknotted structures. As DNA is converted into RNA, we fold secondary RNA structures. Only trustworthy helices are recognised, and the length of RNA sequences is steadily expanded in accordance with the aforementioned golden rules.

Results

Transcriptomic Landscape of ESCC

To look for potential lncRNA correlations with ESCC, we performed genome-wide gene expression profiling of both coding genes and lncRNAs from ESCC and nearby nonneoplastic tissue. We first checked to see if these 7,419 noncoding and 27,958 coding RNA transcripts are grouped together and distinct from the samples of normal tissue. Next, we looked at the landscapes of the entire transcriptome as well as the total transcriptomic pattern (lncRNAs + coding RNAs) from each sample. The general transitions from a particular normal to cancer state were also observed separately as a difference in the expression profile of either the lncRNA or the coding RNA. The landscape of the entire transcriptome may be changing due to a putative dynamic interaction between lncRNAs and coding RNAs, according to these observations.

Expression of lncRNAs in ESCC

Despite being widely produced throughout the genome, lncRNAs are a novel family of noncoding RNAs about which little is known about their functional properties. With the exception of a recent study of the overexpressed lncRNA AFAP1-AS1 in esophageal adenocarcinoma, high-throughput screening of lncRNAs from ESCC has received little research. 7,419 intergenic lncRNAs and other transcripts with unknown coding potential were evaluated in total, and 410 DE-lncRNAs were found in ESCC in comparison to nearby normal esophageal tissues. The ESCC Associated Long noncoding RNAs are the names we gave to the anonymous lncRNAs. Numerous malignancies have elevated HOTAIR lncRNA expression, and our investigation of ESCC shows that it is also markedly elevated. The differential expression of two more upregulated lncRNAs in ESCC was also validated by our research. The tests are split into two sections: one for pseudoknotted sequences and the other for mixed data of pseudoknot-free and pseudoknotted sequences to show the impact of our technique. Two data sets are picked. The first is the TT2NE data set for evaluating pseudoknotted structures. 47 pseudoknotted sequences taken from PDB and PseudoBase make up this data set. The PKNOTS data set is an additional one for testing secondary and fictitious knotted structures. The 116 sequences in this data set comprise pseudoknots for the HIV-1-RT ligand RNA, viral RNAs, and 25 tRNA sequences that were randomly chosen from the Sprinzl tRNA database.

Both sensitivity and PPV are used to gauge how accurate an algorithm is. Assume that the genuine RNA structure contains RP (real pair) base pairs.

Discussion

We test two models to determine the impact of the GM approach on various models, then we compare the differences between the two. The PKNOTS data set is first run via the PKONTS and LZ algorithms, and the output of the results is then obtained. We next carry out the first stage of the GM method to dynamically fold sequences at the golden points to create the frame of secondary structures, choosing the stable helix with the least amount of energy at each fold. Then, using the same energy model and settings as before, we feed the partially

folded sequences through the PKONTS and LZ algorithms. To get the statistics, we fold every sequence in the test set. Compared to the original LZ and PKNOTS algorithms, the updated LZ and PKNOTS algorithms both offer sensitivity gains of 2% to 3%. Compared to the PKNOTS algorithm, the enhanced LZ algorithm performs better (4.9%). In order to increase prediction sensitivity and decrease predicted redundant base pairs, the GM approach may also be used to other RNA structure prediction algorithms, according to the testing of enhanced PKNOTS and improved LZ.

Numerous sequences (such as Bioton, DF0660, DG7740, DI1140, DP1780, DV3200, and DY4840) have higher accuracy thanks to the updated LZ and PKNOTS algorithms, and this makes it possible for us to assess the impact of the golden mean feature. The TT2NE algorithm predicted more redundancy genera than other techniques. As an illustration, TT2NE predicts two pseudoknots even though GLV IRES, R2 retro PK, and 1y0q sequences only have one native pseudoknot. In contrast to the three anticipated by TT2NE, Bs glmS only naturally has two pseudoknots. Each algorithm's PPV and sensitivity percentages are displayed. The number of redundant genera is shown in the column Gen, and it exceeds the number of native genera. GR is the redundancy number in the anticipated genus. Average 1 is calculated as the sum of all base pairs successfully predicted over the entire database divided by the equivalent sum of native base pairs (average sensitivity) and the sum of base pairs correctly predicted.

Conclusion

Based on the statistical analysis of actual RNA secondary structures, we present a novel finding in this study that RNA folding follows the golden mean feature. Nearby to 0.382L, 0.5L, 0.618L, and L are the sequence's folding 3'-end sites. These key golden sequence points are listed above. With this trait in mind, we create a GM algorithm by dynamically folding RNA secondary structures in accordance with the aforementioned golden section sites and by creating pseudoknots by crossing subsequences. Utilizing thermodynamic data, we put the technique into practise and evaluate its efficacy using the PKNOTS and TT2NE data sets. For the PKNOTS data set, we first preprocess the sequence with the first GM step before obtaining the partially folded sequence output to feed into the PKNOTS and LZ algorithms. The two algorithms then get 2% to 3% better. This is due to the half folded sequence serving as its structural framework. In other words, folding at the golden spots regulates the course of folding and therefore inhibits the development of some duplicate structures. In order to increase prediction accuracy and decrease projected duplicated base pairs, preprocessing of GM can also be used to other RNA structure prediction algorithms. In comparison to Mfold, HotKnots, McQfold, ProbKnot, and LZ for the TT2NE data set, GM has higher sensitivity and more predicted pseudoknots. Over Mfold, HotKnot, ProbKnot, and LZ, the PPV of GM is superior. According to these results, the GM approach performs well in terms of forecasting secondary and pseudoknotted structures. When compared to other RNA structure prediction algorithms, the sensitivity and PPV of the GM method are superior. The experimental outcomes show the RNA folding regulations from a fresh perspective that is similar to natural folding.

Acknowledgement

The author would like to acknowledge his Department of Pathology, Medical Science, Sweden for their support during this work.

Conflicts of Interest

The author has no known conflict of interest associated with this paper.

References

1. Fan J, Liu Z, Mao X, Tong X, Zhang T (2020) Global Trends in the Incidence and Mortality of Esophageal Cancer from 1990 to 2017. *Cancer Med* 9: 33-38.
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA (2018) Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer J Clinicians* 68(6): 394-424.
3. Wong MCS, Hamilton W, Whiteman DC, Jiang JY, Qiao Y (2018) Global Incidence and Mortality of Oesophageal Cancer and Their Correlation with Socioeconomic Indicators Temporal Patterns and Trends in 41 Countries. *Sci Rep* 8(1): 45-122.
4. Ohashi S, Miyamoto Si, Kikuchi O, Goto T, Amanuma Y (2015) Recent Advances from Basic and Clinical Studies of Esophageal Squamous Cell Carcinoma. *Gastroenterology* 149(7): 1700-1815.
5. Maher B (2012) ENCODE: The Human Encyclopaedia. *Nature* 489(7414): 46-80.
6. Lee JT (2012) Epigenetic Regulation by Long Noncoding RNAs. *Science* 338(6113): 1435-1439.
7. Yoon JH, Kim J, Gorospe M (2015) Long Noncoding RNA Turnover. *Biochimie* 117: 15-21.
8. Wang X, Sun W, Shen W, Xia M, Chen C (2016) Long Non-coding RNA DILC Regulates Liver Cancer Stem Cells via IL-6/STAT3 axis. *J Hepatol* 64(6): 1283-1294.
9. Zhou J, Yang L, Zhong T, Mueller M, Men Y (2015) H19 lncRNA Alters DNA Methylation Genome Wide by Regulating S-Adenosylhomocysteine Hydrolase. *Nat Commun* 6: 10221-10285.
10. Wu Y, Zhang L, Wang Y, Li H (2014) Long Noncoding RNA HOTAIR Involvement in Cancer. *Tumor Biol* 35(10): 9531-9538.