Determination of secondary and tertiary structures of cervical cancer lncRNA diagnostic and siRNA therapeutic biomarkers

Arli Aditya Parikesit[1],*, **Nihayatul Karimah[1]**, and **Didik Huswo Utomo[1]**

[1]Department of Bioinformatics, School of Life Sciences, Indonesia International Institute for Life Sciences, Jalan Pulomas Barat Kav. 88, Jakarta 13210, Indonesia
*Corresponding author: arli.parikesit@i3l.ac.id

ABSTRACT Cervical cancer is one of the primary causes of mortality in women due to human papilloma virus (HPV) infection. The fingerprint of an HPV infection could be detected using a long non-coding RNA (lncRNA) biomarker, enabling it to be utilized in molecular diagnostics. The primary structure or sequences of RNA should be annotated within conventional bioinformatics tools. Therefore, this study aimed to determine the fine-grained 2D and 3D structures of lncRNA PVT1 and its respective siRNA inhibitors. lncRNA PVT1 sequences from *Homo sapiens*, *Mus musculus*, and *Rattus norvegicus* were retrieved from Genbank (NCBI). Prediction of the 2D structure and analysis of the interactions of the lncRNA and siRNA were performed using the Vienna RNA package. The 3D structure of the RNA was computed using the SimRNA and ModeRNA software programs. The results showed that lncRNA PVT1 from *H. sapiens* and *M. musculus* had a conserved region. However, the lncRNA from both *H. sapiens* and *M. musculus* showed a low conserved region, and the 2D structure could not be determined; thus, the annotation and 2D model focused only on *H. sapiens*. Both of their lncRNA PVT1 also had a short half-life in the cell. Based on the 3D modeling pipeline, the 3D model of lncRNA PVT1 showed the stability and possible function as molecules, while the PVT1 siRNA-lncRNA interaction analysis revealed that the molecules could bind well. Based on these findings, the structures of both lncRNA PVT1 and its siRNA have the potential to be utilized as biomarkers.

KEYWORDS biomarker; cervical cancer; HPV; lncRNA; transcriptomics

1. Introduction

Cancer has long been considered as one of the main focus of World Health Organization (WHO) for developing feasible strategies for its cure, care, and prevention (WHO 2007, 2017). Approximately 8.8 million people died worldwide because of cancer(WHO 2017). Although there are many causes available, HPV is known as the primal cause of cervical cancer (zur Hausen 2009a,b), current main focus on the development of vaccine and drugs for all types of HPV-based cervical cancer is on the proteomics-based agents (Tambunan and Parikesit 2012). In order to strengthen the availability and feasibility of cervical cancer drugs, new development beyond the scope of proteomics studies should be devised. Thus, exploring the much-uncharted area of transcriptomics, as a means of focusing on RNA structures, function, and mechanisms, is a viable alternative (Dong and Chen 2013).

Transcriptomics is focusing mainly on the pattern annotation of two RNA types that available in the cell, namely messenger (m)RNA and non-coding (nc)RNA (Wang et al. 2009). ncRNA role is largely sidelined in many years as largely 'junk DNA' annotation, and has

gained its momentum now for more focus in transcriptomics research (Dong and Chen 2013). mRNAs and Micro (mi)RNAs are already common biomarkers for HPV. HPV biomarkers were already captured as miRNA in the cancer cell, and already utilized as diagnostics biomarker (Qian 2013; Qian et al. 2013), specific long non-coding (lnc)RNA as cervical cancer biomarker for diagnostic has only recently been studied.

PVT1 gene, as lncRNA diagnostic biomarker of HPV infection, is already well known (Iden et al. 2016). It is already proven that silencing PVT1 gene will eventually decrease cervical cancer cell migration, proliferation, and invasion. One of the feasible strategies for silencing PVT1 is by using silencing (si)RNA. Thus, the alternative splicing event of PVT1 gene should be carefully observed as transcript isoforms are available (Chen and Manley 2009; Wang et al. 2015). The fine-grained resolution of PVT1 biomarker can only be understood by determining its 2D and 3D structural properties. Learning the functionality and the 2/3D structure PVT1 lncRNA biomarker and its siRNA is an important step toward effective blueprint of Cervical Cancer drug and vaccine development. The objective of this study is to determine the

fine-grained 2D and 3D structure of PVT1 lncRNA and its respective siRNA inhibitor as diagnostic and potential therapeutic biomarkers.

2. Materials and methods

The procedure was partially adapted from the existing 2D RNA prediction pipeline (Parikesit and Anurogo 2016; Parikesit et al. 2016; Parikesit and Nurdiansyah 2018). The sequences of *Homo sapiens*, *Mus musculus*, and *Rattus norvegicus* lncRNA will be downloaded in FASTA format and saved with Komodo Edit Text Editor for dynamics programming languages (https://www.activestate.com/komodo-edit). Vienna RNA Package was utilized for predicting the 2D structure and interactions of the lncRNA and siRNA (Gruber et al. 2015). The online version of Vienna RNA was used in here: http://rna.tbi.univie.ac.at/ with providing sequences in FASTA format as the inputs (TBI 2016). Only default parameters applied in running the programs. The following modules and their respective function were utilized directly in that website (Table 1).

Those respective tools were applied in order to ensure the thermodynamics and kinetics properties of PVT1 lncRNA. The 2D structure elucidation of PVT1 biomarkers was crucial to determine their exact functionality in the biochemical reactions. The output of this protocol was the thermodynamics, kinetics, and Vienna 'dot-bracket' format data files (Smit et al. 2008). Those data were forwarded to simRNA and modeRNA software for predicting

the complete 3D structure of the RNA (Rother et al. 2011; Magnus et al. 2016). The output of the 3D computational process was visualized using Chimera software.

3. Results and discussion

PVT1 genes from three species were available in Genbank (Table 2). The results showed that PVT1 of *R. norvegicus* is a composition of many transcripts, not the gene sequences itself. Only PVT1 of *H. sapiens* and *M. musculus* that could be aligned, as for *R. norvegicus* has no representative gene sequences. The pairwise alignment of PVT1 of *H. sapiens* and *M. musculus* is providing unfeasible conserved alignment. In this respect, only some parts of their sequences that are heavily conserved.

RNAfold was utilized to determine the 2D Secondary structure of lncRNA PVT1 of *H. sapiens* and *M. musculus* were further analyzed to identify the RNA stability. This computational process was conducted using RNAfold to simulate the secondary structure and folding mechanism of lncRNA PVT1 (Figure 1). It was found that both structures differ significantly and have a different structural backbone. These kinds of complex structures would definitely act in transition states of a biochemical reaction and have a very short half-time. Thus, it would be not feasible to isolate them in the wet laboratory.

Evaluation of consensus structures of conserved lnRNA PVT1 between *H. sapiens* and *M. musculus* could reveal its function (Bernhart et al. 2008). RNAalifold was utilized to determine the conserved 2D structure of

TABLE 1 The modules of Vienna RNA Package.

No.	Vienna RNA modules	Weblink	Functionality (TBI 2016)	References
1	RNA fold	http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi	Predicts minimum free energy structures and base pair probabilities from single RNA or DNA sequences.	Gruber et al. (2008); Lorenz et al. (2011)
2	RNAalifold	http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAalifold.cgi	Predicts consensus secondary structures from an alignment of several related RNA or DNA sequences. You need to upload an alignment.	Hofacker et al. (2002); Bernhart et al. (2008); Gruber et al. (2008)
3	RNAup	http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAup.cgi	Allows to predict the accessibility of a target region.	Muckstein et al. (2006); Gruber et al. (2008)
4	Barriers	http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/barriers.cgi	Allows to get insights into RNA folding kinetics.	Flamm et al. (2002); Wolfinger et al. (2004)
5	RNAxs	http://rna.tbi.univie.ac.at/cgi-bin/RNAxs/RNAxs.cgi	Assists in siRNA design.	Tafer et al. (2008)

TABLE 2 The PVT1 genes in the Genbank/NCBI database.

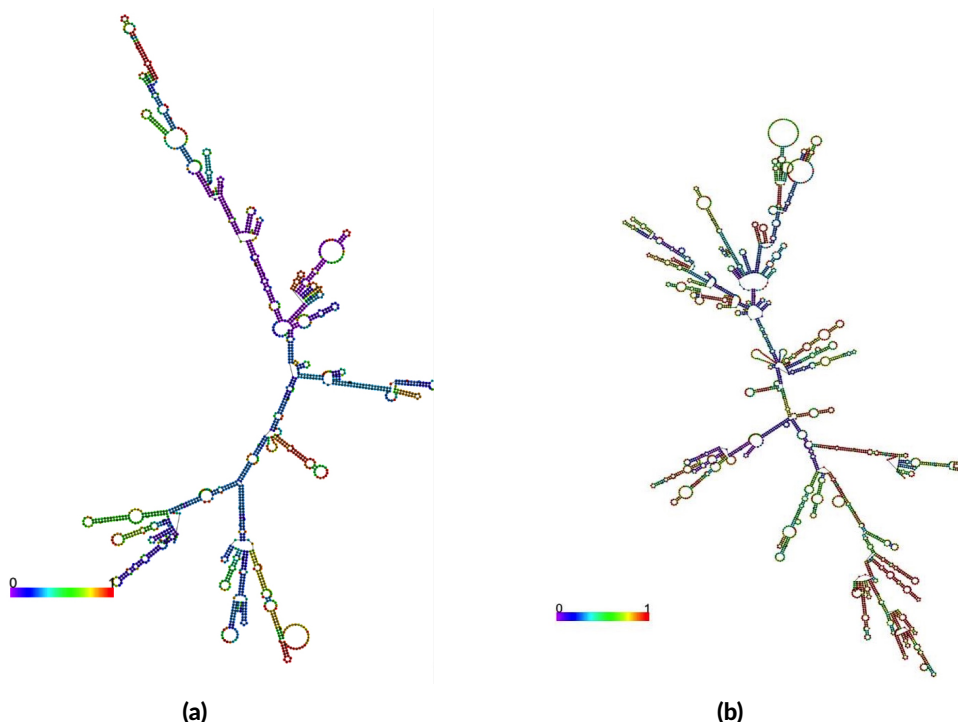| No.	Gene ID	Description	Location
1	5820	Pvt1 oncogene (non-protein coding) [*Homo sapiens* (human)]	Chromosome 8, NC_000008.11 (127794533..128101253)
2	19296	plasmacytoma variant translocation 1 [*Mus musculus* (house mouse)]	Chromosome 15, NC_000081.6 (62037986..62260215)
3	100910797	Pvt1 oncogene [*Rattus norvegicus* (Norway rat)]	Chromosome 7, NC_005106.4 (102648394..102871316)

FIGURE 1 The RNAfold computational results of: (a) lncRNA PVT1 Human; (b) lncRNA PVT1 *Mus musculus*. Both are annotated as MFE secondary structure, the one with the most minimal free energy gibbs. The scale in the left side refers to the base pair probability, from 0 (blue) to 1 (red).
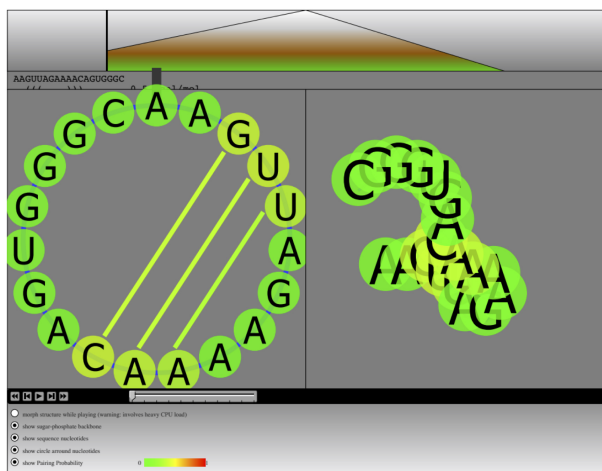

FIGURE 2 The Barriers server visualization of structural transition of siRNA lncRNA PVT1 Human. The scale of the probability of the formation of the structure can be seen in the bottom of the picture, if the green is closer to zero, and the red one is close to one. The picture above shows the probability of forming the structure according to the animation sequence. The left circle graph is the graph theory representation of each DNA sequence. The right-hand structure shows the formation prediction structure, with its probability at each sequence.

FIGURE 3 5S ribosomal RNA (PDB ID: 1VQ9, RFAM ID: RF00001) as homologous model of lncRNA PVT1 Human.

lncRNA PVT1. The conservation between two species was very low, as the conserved structure is illogical and not in accordance to the thermodynamics law due to the burdensome steric effects. As such, the RNA annotation is focused on solely human data. RNAxs was employed to determine the siRNA that knock down the lncRNA PVT1
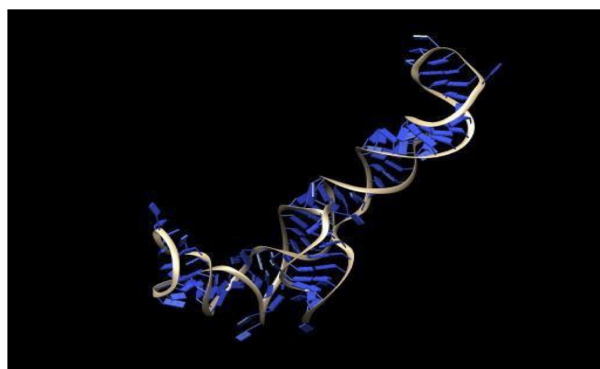
of *H. sapiens*. In this end, the siRNA was tested with RNAup to determine the feasibility of siRNA and *H. sapiens* lncRNA PVT1 interaction, and thermodynamically it interacted well. In order to test the integrity of the predicted structures, Barriers was employed to observe the transition states of siRNA-lncRNA PVT1 Human (Figure 2). An ideal siRNA structure will always be iterated for the most optimal conformation. Barriers server could predict that siRNA-lncRNA PVT1 Human has 15 structures in the energy range of 40 kcal/mol above the minimum energy. siRNA is the one that could knock off the lncRNA. Wide range siRNA for silencing viral genome already tested in the lab (Jiang and Milner 2002; Yoshinouchi et al. 2003; Chang et al. 2010).
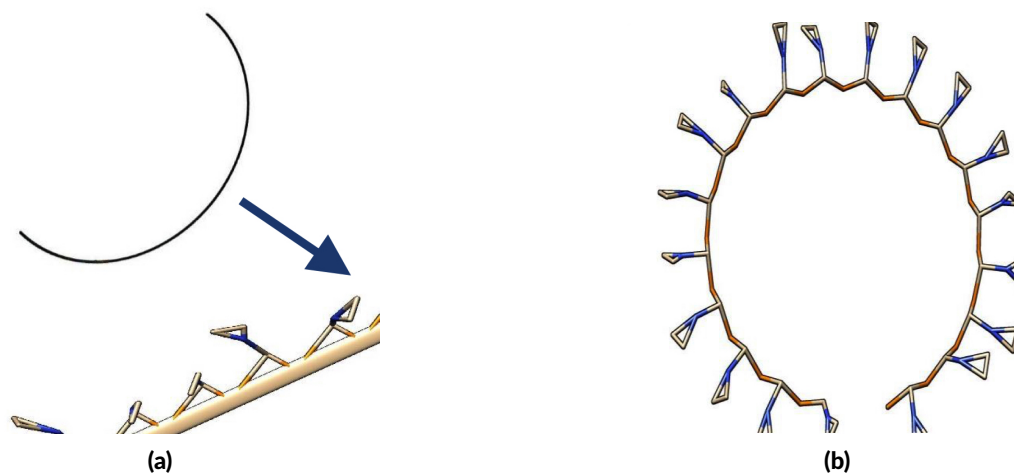
FIGURE 4 (a) 3D structure of lncRNA PVT1 Human, the residue was magnified in the right corner; (b) 3D structure of siRNA lncRNA PVT1 Human.

In respect of 3D modeling pipeline, modeRNA was employed to determine the homology model of lncRNA PVT1 Human. The structure of lncRNA PVT1 Human was homologous with the tRNA from a ribosomal unit of *Haloarcula marismortui* with the PDB ID of 1VQ9. However, further elucidation of the modeling procedure only produces the clustering of lncRNA PVT1 Human as part of RF00001 family. It acts as 5S ribosomal RNA that binds to transfer RNA (Figure 3). Fine-grained resolution of homology model is attainable due to the complete annotation of RNA database in modeRNA.

Thus, simRNA was utilized to determine the de novo computation for lncRNA PVT1 Human and its siRNA (Figure 4). RNAfold data for both structures on Figure 4a and 4b shows that no hydrogen bond information available. This also reconfirms the finding in simRNA output. It means that the internal structure of the functional groups did not support hydrogen bonding in both thermodynamics and kinetics indicator.

The computational results were deposited in the Mendeley cloud system for further retrieval (Parikesit 2017). The pure prediction pipeline with solely computational approach would only useful for basic transcriptomics research. In this end, for applied research, fine-grained resolution of 3D RNA structure prediction could be applied when combined with experimental data, such as with utilizing hydroxyl radical probing (HRP) protocol (Ding et al. 2012).

4. Conclusions

It can be concluded that the design of 2D and 3D structure of both lncRNA PVT1 and its siRNA could be developed further as diagnostics and potential therapeutic biomarkers for cervical cancer in the wet laboratory. These 2D and 3D structural information could be utilized for predicting a finer grained biochemical mechanism with molecular

docking and dynamics method. In this end, the repertoire of the biochemical reaction and functionality could be elucidated in an exact manner.

Authors' contributions

AP designed the study. AAP, NK, DHU, carried out the laboratory work. AAP, NK, DHU analyzed the data. AAP, NK, DHU wrote the manuscript. All authors read and approved the final version of the manuscript.

Competing interests

We declare that there is no competing interest.

References

Bernhart SH, Hofacker IL, Will S, Gruber AR, Stadler PF. 2008. RNAalifold: improved consensus structure prediction for RNA alignments. BMC Bioinf. 9:474. doi:10.1186/1471-2105-9-474.

Chang JTC, Kuo TF, Chen YJ, Chiu CC, Lu YC, Li HF, Shen CR, Cheng AJ. 2010. Highly potent and specific siRNAs against E6 or E7 genes of HPV16- or HPV18-infected cervical cancers. Cancer Gene Ther. 17(12):827–36. doi:10.1038/cgt.2010.38.

Chen M, Manley JL. 2009. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat Rev Mol Cell Biol. 10(11):741–754. doi:10.1038/nrm2777.

Ding F, Lavender CA, Weeks KM, Dokholyan NV. 2012. Three-dimensional RNA structure refinement by hydroxyl radical probing. Nat Methods 9(6):603–608. doi:10.1038/nmeth.1976.

Dong Z, Chen Y. 2013. Transcriptomics: advances and approaches. Sci China: Life Sci. 56(10):960–967. doi:10.1007/s11427-013-4557-2.

Flamm C, Hofacker IL, Stadler PF, Wolfinger T, Wolfinger MT. 2002. Barrier trees of degenerate landscapes. Z Phys Chem. 216:155. doi:10.1524/zpch.2002.216.2.155.

Gruber AR, Bernhart SH, Lorenz R. 2015. The ViennaRNA web services. Methods Mol Biol. 1269:307–326. doi:10.1007/978-1-4939-2291-8_19.

Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. 2008. The Vienna RNA websuite. Nucleic Acids Res. 36(Web Server issue). doi:10.1093/nar/gkn188.

Hofacker IL, Fekete M, Stadler PF. 2002. Secondary structure prediction for aligned RNA sequences. J Mol Biol. 319(5):1059–1066. doi:10.1016/S0022-2836(02)00308-X.

Iden M, Fye S, Li K, Chowdhury T, Ramchandran R, Rader JS. 2016. The lncRNA PVT1 contributes to the cervical cancer phenotype and associates with poor patient prognosis. PLoS One 11(5):e0156274. doi:10.1371/journal.pone.0156274.

Jiang M, Milner J. 2002. Selective silencing of viral gene expression in HPV-positive human cervical carcinoma cells treated with siRNA, a primer of RNA interference. Oncogene 21(39):6041–6048. doi:10.1038/sj.onc.1205878.

Lorenz R, Bernhart SH, zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA Package 2.0. Algorithms Mol Biol. 6(1):26. doi:10.1186/1748-7188-6-26.

Magnus M, Boniecki MJ, Dawson W, Bujnicki JM. 2016. SimRNAweb: a web server for RNA 3D structure modeling with optional restraints. Nucleic Acids Res. 44(14):gkw279. doi:10.1093/nar/gkw279.

Muckstein U, Tafer H, Hackermuller J, Bernhart SH, Stadler PF, Hofacker IL. 2006. Thermodynamics of RNA-RNA binding. Bioinformatics 22(10):1177–1182. doi:10.1093/bioinformatics/btl024.

Parikesit AA, Anurogo D. 2016. Prediksi struktur 2-dimensi non-coding RNA dari biomarker kanker payudara triple-negative dengan Vienna RNA package [The 2-dimensional structure prediction of Triple-Negative Breast Cancer (TNBC) biomarker non-coding RNA with Vienna RNA Package]. Chimica et Natura Acta 4(1):27. doi:10.24198/cna.v4.n1.10445.

Parikesit AA, Anurogo D, Parikesit AA. 2016. Bioin-formatics approach towards transcriptomics of Filaggrin. J Agromed Med Sci. 2(3):8–16. doi:10.19184/ams.v2i3.3254.

Parikesit AA, Nurdiansyah R. 2018. Generating two-dimensional repertoire of siRNA Linc-ROR and siRNA mRNA ARF6 from the lincRNA-RoR/miR-145/ARF6 expression pathway that involved in the progression of Triple Negative Breast Cancer. IOP Conf Ser: Mater Sci Eng. 299(1):012059. doi:10.1088/1757-899X/299/1/012059.

Parikesit AAP. 2017. Research data of the "The discovery of secondary and tertiary structure of cervical cancer lncRNA biomarker". doi:10.17632/v8xdgjnkvn.1. Mendeley Data, v1.

Qian K. 2013. Bioinformatics analysis of HPV associated host microRNA functions and identification of viral microRNA. [Doctoral thesis]. [Helsinki]: University of Helsinki.

Qian K, Pietilä T, Rönty M, Michon F, Frilander MJ, Ritari J, Tarkkanen J, Paulín L, Auvinen P, Auvinen E. 2013. Identification and validation of human papillomavirus encoded microRNAs. PLoS One 8(7):e70202. doi:10.1371/journal.pone.0070202.

Rother M, Rother K, Puton T, Bujnicki JM. 2011. ModeRNA: a tool for comparative modeling of RNA 3D structure. Nucleic Acids Res. 39(10):4007–4022. doi:10.1093/nar/gkq1320.

Smit S, Rother K, Heringa J, Knight R. 2008. From knotted to nested RNA structures: A variety of computational methods for pseudoknot removal. RNA 14(3):410–416. doi:10.1261/rna.881308.

Tafer H, Ameres SL, Obernosterer G, Gebeshuber CA, Schroeder R, Martinez J, Hofacker IL. 2008. The impact of target site accessibility on the design of effective siRNAs. Nat Biotechnol. 26(5):578–583. doi:10.1038/nbt1404.

Tambunan USF, Parikesit AA. 2012. HPV bioinformatics: in silico detection, drug design and prevention agent development. In: R Rajkumar, editor. Topics on cervical cancer with an advocacy for prevention. chapter 14. Rijeka, Croatia: Intech Publishing. p. 237–252. doi:10.13140/2.1.2298.3689.

TBI. 2016. Vienna RNA Package Web Version 2.0.

Wang Y, Liu J, Huang BO, Xu YM, Li J, Huang LF, Lin J, Zhang J, Min QH, Yang WM, Wang XZ. 2015. Mechanism of alternative splicing and its regulation. Biomed Rep. 3(2):152–158. doi:10.3892/br.2014.407.

Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 10(1):57–63. doi:10.1038/nrg2484. NIHMS150003.

WHO. 2007. The World Health Organization's fight against cancer: strategies that prevent, cure and care. Geneva: World Health Organization. Available from: http://www.who.int/iris/handle/10665/43665.

WHO. 2017. The World Health Organization Can-

cer key facts. [cited 23 Apr 2017]. Available from: http://www.who.int/cancer/media/news/cancer-key-facts/en/.

Wolfinger MT, Svrcek-Seiler WA, Flamm C, Hofacker IL, Stadler PF. 2004. Efficient computation of RNA folding dynamics. J Phys A: Math Gen. 37(17):4731–4741. doi:10.1088/0305-4470/37/17/005.

Yoshinouchi M, Yamada T, Kizaki M, Fen J, Koseki T, Ikeda Y, Nishihara T, Yamato K. 2003. In vitro and in vivo growth suppression of human papillomavirus 16-positive cervical cancer cells by e6 siRNA. Mol Ther. 8(5):762–768. doi:10.1016/j.ymthe.2003.08.004.

zur Hausen H. 2009a. Papillomaviruses in the causation of human cancers - a brief historical account. Virology 384(2):260–265. doi:10.1016/j.virol.2008.11.046.

zur Hausen H. 2009b. The search for infectious causes of human cancers: where and why (Nobel lecture). Angew Chem. 48(32):5798–5808. doi:10.1002/anie.200901917.