An exploration on possible explanation of the peculiar behaviour of lncRNA uc.339

Introduction

The genomic ultraconserved regions(UCRs) is a topic full of interest. A lot of its function remains to be elucidated. A large part of this region is known to be transcribed into long non-coding RNA(lncRNA). Studies from a number of sources have demonstrated the correlation of lncRNA expression in the carcinogenesis process across different type of cancer, and its interaction with micro RNA (miRNA). The “classic” mechanism of miRNA action is based on the dogma that the 19 to 24 nt miRNAs, generated in the cytoplasm from the longer 70 to 110 nt precursor, and eventually down-regulate gene expression in a sequence specific fashion. A preliminary study indicate that a specific UCR, designated uc.339, which is found to be up-regulated in colorectal cancer is transcribed as a lncRNA, that is significantly up-regulated as well in primary non small cell lung cancer (NSCLC) samples, compared to the adjacent lung tissue and strongly correlates with the overall survival of NSCLC patients. It is observed later that the uc.339 transcript harbors miRNA binding elements (MBE) for miR-339-3p, miR-663b-3p and miR-95-5p (from now on, referred to as miR-339, -663b, and -95, respectively), and that up-regulation of uc.339 leads to up-regulation of CCNE2, MCL1, and CDK6. All three of these predicted target genes of miR-339, -663b, and -95 have well documented pro-tumoral activity in lung carcinogenesis14-16. It is also observed that while up-regulation of uc.339 reduces the levels of miR-339, -663b and -95, over-expression of the three miRNAs does not alter the cellular level of uc.339 transcript. This finding is suggestive of an unusual “trapping” effect on multiple miRNAs by uc.339. The researchers coined the term “Non-Degrading Trapping” (NDT) to describe this unusual mechanism. Further study shows that besides complementary sequence for the miRNAs, which is often referred to as miRNA binding elements (MBE), sequences within the uc.339 transcript without miRNA complementarity are also needed for NDT to occur, these sequences are termed trapping response elements (TRE). Though the actual function of the TRE in uc.339 remains uncovered, it is suspected that this might reflect a more widespread and previously unknown type of interaction between short miRNAs and any long target RNA (coding or noncoding) and is not a phenomenon unique to uc.339. Furthermore, long RNAs could display both MBE and TRE elements, therefore the specific number of sequences per each element category and their location in the sequence of the RNA will also likely play an important role in determining the final outcome of the interaction. Due to the proposed mechanism, we will henceforth refer to the uc.339 lncRNA as “ENTRAPER”, as this transcript captures and sequesters the shorter miRNAs. The universality potential of this TRE-mediated NDP is a very intriguing topic. To have a better understanding, its biochemical mechanism need to be uncovered. I am here to explore the possibility that the deletion of 5 base cause a decline in ΔG which in turn cause the RNA more reluctant to break its secondary structure and exposes its complementary sequence to the miRNA.

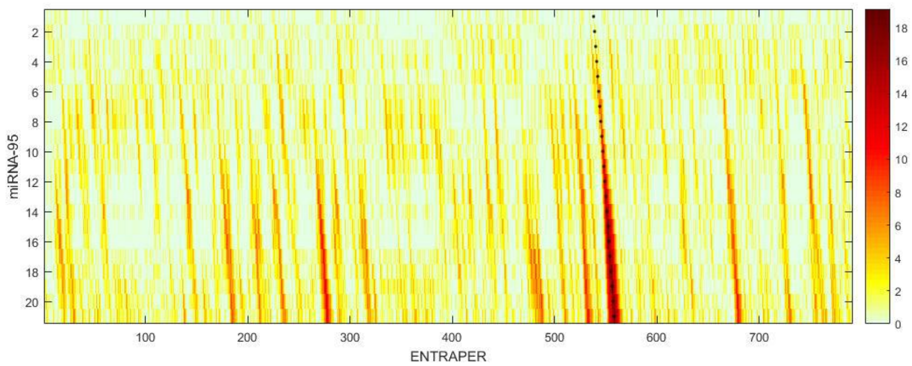
Methods

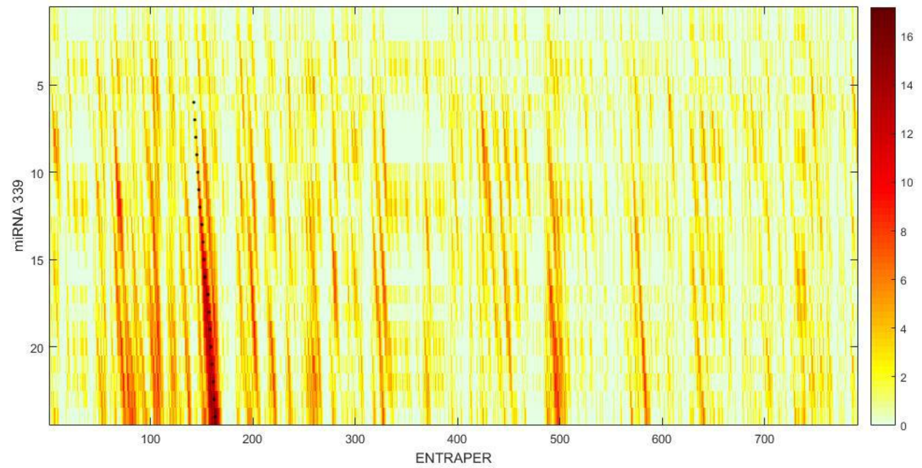
The sequence for ENTRAPER is acquired through inputting the partial sequence in the lab through blastn and retrieve the first hit for *Homo. Sapiens*. From here, the sequence of miRNA and ENTRAPER is aligned by Smith-Waterman algorithm to exclude the complementary sequence for TRE searching. Subsequently, all possible 5 base and 4 base deletion is created from the original whole sequence in Matlab. And said sequence is saved in fasta format for secondary structure prediction.

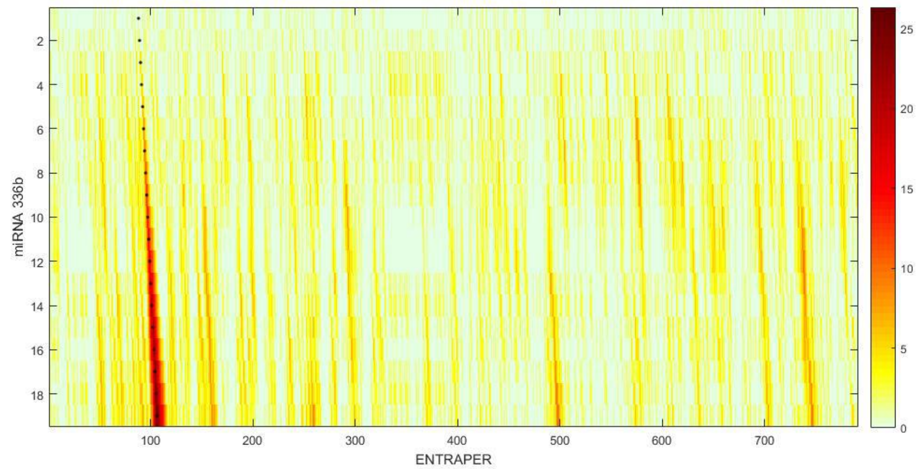
Secondary structure is predicted by a online source called Quikfold. Choosing this program is because this is the only program I can find to fold multiple sequence at the same time. Considering it have a maximum inputting limit of 150,000 nucleotides at a time, the data generated were saved in 140 sequences per file knowing the length of ENTRAPER is 791 nucleotides (the exact length of the molecule in vivo is unknown, this is the approximate length I used for this project). ΔG of all mutants is retrieved manually into an excel file (I tried to find some more clever way with no prevail), and subsequently plotted against the start point for all deletions in Matlab.

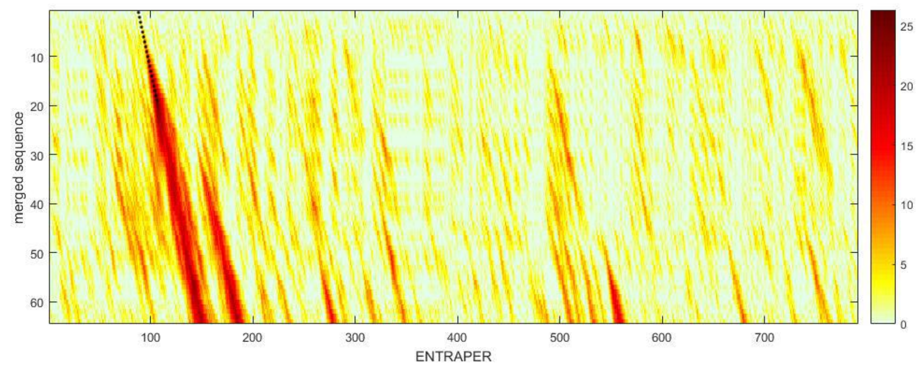
Results

The alignment was committed both separately to each miRNA and a sequence merging all three like below. The alignment for merged sequence is very different and deemed not credible.

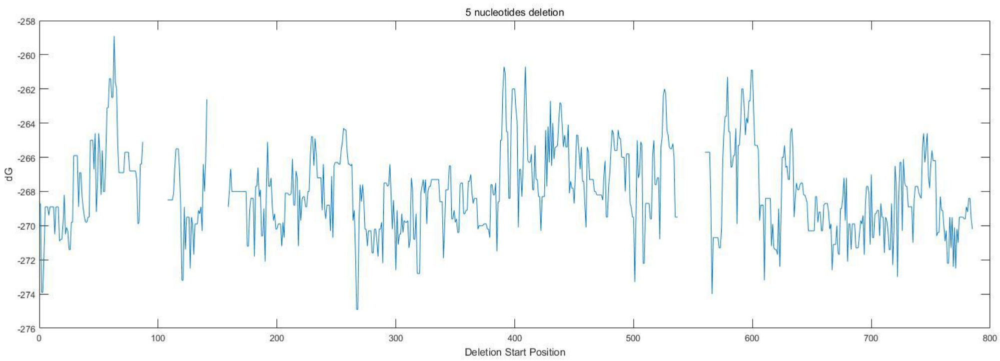


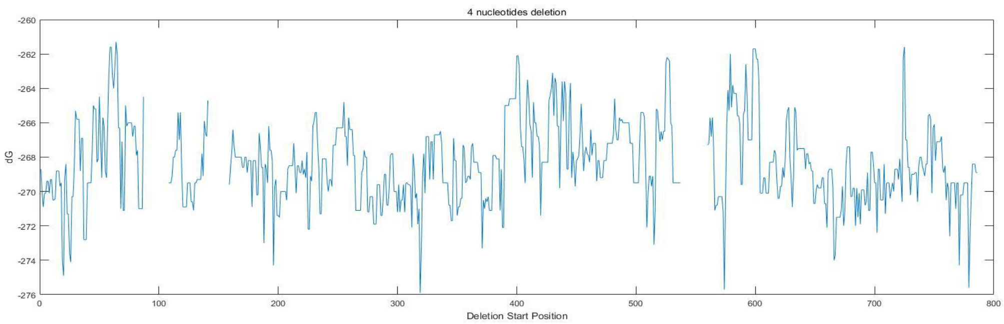






After saving the sequences in fasta format, Quikfold start to give back right amount of data. And ΔG of each sequence is plotted like below for both 5 nucleotides deletion and 4 nucleotides deletion with MBE regions skipped.

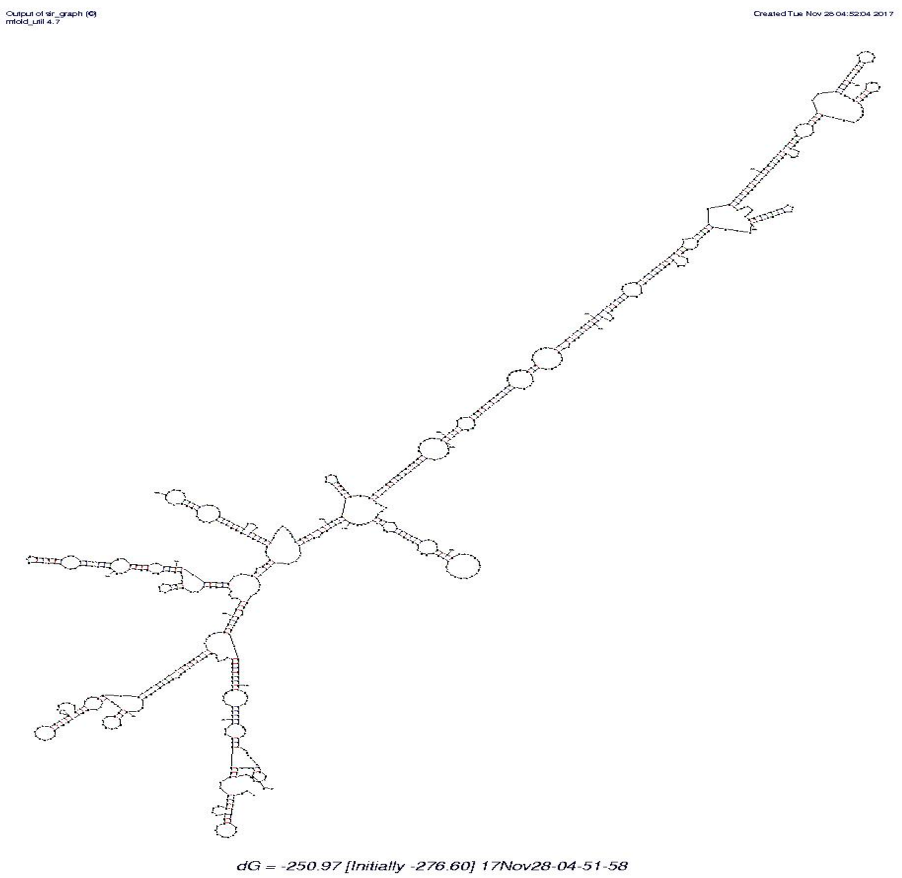




ΔG as expected can fluctuate in a reasonable range, yet the region formerly described as TRE (start position 63) stand out on the unstable side. The lowest therefore most thermodynamically stable mutants for 5 nucleotide deletion would start deletion at 264 and for 4 nucleotide deletion it will start at 319.

Discussion

This analysis do not show ΔG as a promising impact factor of the NDP mechanism, making a 5 nucleotide deletion mutant start at 264 may quickly prove this. However, this algorithm is very problematic to use. For each sequence, Quikfold will generate 5-20 different secondary structures. Using the mean and anova test may sound better for statistic analysis, however, this do not make sense in other ways. We can have an average number but an average structure is an absurd definition, thus when retrieving data, I retrieved the most stable structure of each sequence. However even doing this comparison can not give us a perfectly quantifiable secondary structure, because even when we have the most viable structure for the wild type (see below), this structure is not the actual structure almost for certain.



RNA usually do not form base pairs if they are too far away on sequence, but how much is the right amount to set up is still a question remains to be answered. And RNA have tertiary structure in real solution, our algorithm only calculate watson crick base pair and base pair stacking energy for the ΔG, which do not necessarily translate to better stability in the solution or even complexes with proteins.

It is also important to understand that ΔG is not the most likely option to cause a conformational change in RNA, there are a lot of other more likely idea, i.e this is a recognition site for miRNA-protein complex, or it can recruit other factors upon miRNA binding that also results in miRNA trapping. This open other possible directions for exploration like what kind of protein may be closely related to miRNA binding, does ENTRAPER float around cytosol on its own or with company of other molecules, if latter, what are they? LncRNA is a field with a lot more unanswered questions than answered ones.

The question of why this 5 nucleotide deletion can cause ENTRAPER to lose its function can not be answered by ΔG prediction, however it does not exclude that it can be caused by altered secondary structure. We are planning for a SHAPE sequence experiment. This experiment can give us some data on which nucleotide do not make a pair, which might be enough to answer if a certain structural change is present.

References

1.Sana J, Hankeova S, Svoboda M, Kiss I, Vyzula R, Slaby O, Expression Levels of Transcribed Ultraconserved Regions <i>uc.73</i> and <i>uc.388</i> Are Altered in Colorectal Cancer. Oncology 2012;82:114-118

2.Unpublished result from Dr. Nikonowicz and his collaborators

3.Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature. 2010;465(7301):1033-8. doi: 10.1038/nature09144.

4.Vannini I, Wise PM, Challagundla KB, et al. Transcribed ultraconserved region 339 promotes carcinogenesis by modulating tumor suppressor microRNAs. *Nature Communications*. 2017;8:1801. doi:10.1038/s41467-017-01562-9.