Article

HOTAIR: Can we make sense of this anti-sense RNA?

Geremy Ralston

Department of Biochemistry, University of Otago, PO Box 56, Dunedin 9054
(email: *ralge795@student.otago.ac.nz*)

Keywords: *long non-coding RNA, cancer prognosis, modular functionality*

Introduction

The essential information flow from DNA to RNA to protein is fundamental to biochemistry. However, the function of RNA is not restricted to the task of providing messenger RNA for protein synthesis. Indeed, many further functions are provided by RNA that are not translated. Examples of these non-coding RNAs (ncRNA) are: transfer RNA, which presents the ribosome with each amino acid to be included in the construction of proteins; ribosomal RNA, which contributes significantly to the structure and function of the ribosome; and other ncRNA such as silencing RNA (siRNA) and micro RNA (miRNA), which can regulate the translation of messenger RNA. Apart from ribosomal RNA, these examples of ncRNA have a length of less than 200 nucleotides.

Long non-coding RNA (lncRNA) can be defined as non-coding RNA greater than 200 nucleotides[1]. Examples of lncRNA are HOTAIR, p21, H19 and BANCR, involved respectively in the following: cancer outcomes,[2] translation inhibition,[3] inhibition of apoptosis in gastric cancer[4] and melanoma cell migration.[5] A brief summary of some types of RNA and their respective function is presented in Fig. 1.

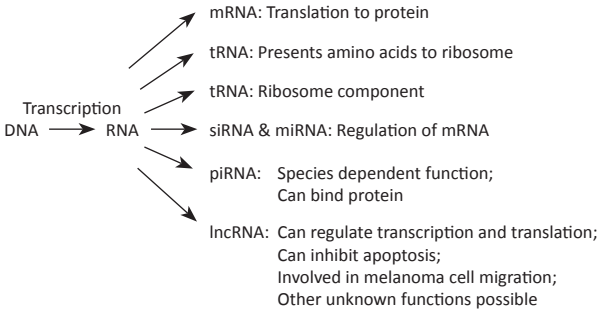

Fig. 1. Some RNA types and a brief description of their function.

lncRNA such as HOTAIR have been identified and associated with multiple types of cancer. Studies of HOTAIR provide valuable insight of the significance of lncRNA, and demonstrate the activity of functional domains that provide protein binding capacity resulting in altered transcriptional activity. This importance encourages the consideration of HOTAIR and other lncRNA as candidates for potential biomarkers and drug targets.

Identification of HOTAIR

HOTAIR (*HOX* antisense intergenic RNA) is a 2158 base pair lncRNA,[2] discovered in a study published in 2007.[6] This included determining whether differential transcription of *HOX* genes is caused by ncRNA forming distinct chromatin domains. HOTAIR was selected as a prospective causal factor as it is located on the boundary between chromatin domains which produce differential expression of genes either side of the boundary. Depletion of HOTAIR with siRNA caused much higher transcription of *HOXD* on chromosome 2, whilst having little effect on the region of origin of HOTAIR, *HOXC* from chromosome 12.

Importance of HOTAIR

The importance of HOTAIR is shown by associations between multiple cancers and expression of HOTAIR. Breast cancer metastatic samples showed increased HOTAIR by factors of hundreds to almost two thousand in comparison to non-cancerous samples.[7] Primary breast tumours also showed high HOTAIR expression, as almost 33% showed a 125 fold increase in HOTAIR.[7] Correlations between the liver cancer hepatocellular carcinoma (HCC) and HOTAIR have also been studied, with results indicating that HOTAIR levels could be used to contribute a prognostic factor in this disease. Of particular interest were samples taken from patients who had received a liver transplant, as transplant recipients with recurrence of HCC also had high levels of HOTAIR, whilst those without recurrence had low levels of HOTAIR.[8] Colorectal cancer (CRC) has also been studied with respect to HOTAIR: HOTAIR levels were found to be higher in CRC tissues than tissues without cancer.[9] Poor prognosis was also associated with high levels of HOTAIR in CRC tissues.[9] In concordance with these correlations, HOTAIR has also been found to be expressed at higher levels in human pancreatic tumours than non-tumour tissue, is a negative prognostic factor for pancreatic cancer and has been identified as a possible target for intervention.[2] These associations of HOTAIR with different types of cancer are summarised in Table 1.

In addition to associations with cancer, and as a lncRNA which has functions proven by knockdown,[7,8] HOTAIR now provides a platform for further understanding the structure-based characteristics which provide these functions.

Structural Considerations

It has been suggested that as the function of HOTAIR includes an interaction with PRC2, it may be beneficial to consider whether this structural feature is also present in other lncRNA which may be achieved through computational comparative genomics.[10]

However, this identification of structures in new ncRNAs may be difficult if a comparison is made by sequence alone, as other research has shown that lncRNA function can be conserved within a domain that does not conserve primary sequence.[11] In addition to this consideration of sequence identity between HOTAIR and other lncRNA,

Table 1. Summary of HOTAIR associations with a selection of cancer types.

Disease Analysed	Sample Type	Negative Control	Outcome of Association with HOTAIR	Reference
Breast Cancer Metastases	Breast Cancer Metastases	Non-Cancerous Samples	Cancerous samples increased HOTAIR by factors of 100x to almost 2000x	7
Primary Breast Cancer Tumour	Primary Brest Cancer Tumour	Non-Cancerous Samples	Almost 33% showed increase of 125x	7
Liver Cancer Hepatocellular Carcinoma	Transplant Recipients Tissue with Recurrence	Transplant Recipients Tissue without Recurrence	High levels of HOTAIR in recurring patients, low levels without recurrence	8
Colorectal Cancer	Colorectal Cancer Tissue	Colorectal Non-Cancerous Tissue	Higher levels of HOTAIR in cancerous tissues. Poor prognosis in Colorectal Cancer tissue with high levels of HOTAIR	9
Pancreatic Cancer	Pancreatic Cancer Tissue	Pancreatic Non-Cancerous Tissue	Higher levels of HOTAIR in cancerous tissues, HOTAIR is possible prognostic factor	2

the sequence conservation of HOTAIR can also be considered between species. An example is exon 1 (shown in Fig. 2) which has a similar RNA secondary structural prediction in humans and cows.

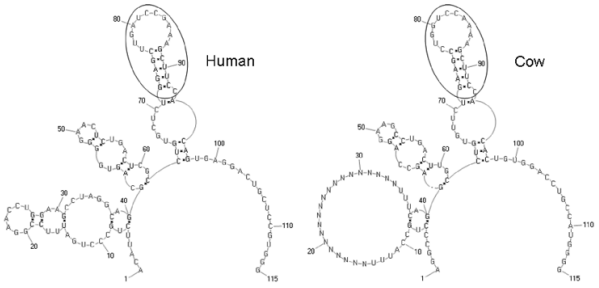

Fig. 2. PMmulti structural predictions of the 5' exon 1 of HOTAIR in human and cow. Circled are regions that have similar structural predictions with Mfold.[10]

At least two functional domains have been found in HOTAIR. At the 5' end exon 1 contains a sequence that binds PRC2 (at the Suz12 subunit of PRC2). At the 3' end domain B of exon 6 contains a sequence that binds LSD1.[10] Remarkably, these regions have been found to be highly conserved in sequence and structure within ten mammals.[10] Other regions of HOTAIR showed low conservation of sequence. This indicates the importance of those conserved regions as their consistent sequence provides the necessary structural basis for its function.

Functions of HOTAIR

When the 5' exon 1 of HOTAIR binds PRC2, and the 3' exon 6 binds LSD1 the functional effect of these interactions involves the in *trans* methylation and demethylation of histones, namely H3K27 methylation by PRC2 and demethylation of H3K4me2 by LSD1.[12] Although the finding of HOTAIR was unique in terms of in *trans* activity, within the functional context of lncRNA, 20% of lncRNA have also been found to have a chromatin modifying function.[13] As HOTAIR binds both PRC2 and LSD1 this provides a link, hence acting as a molecular scaffold.[12] The resulting function of this scaffold interaction is the epigenetic modification of chromatin, which consequently regulates gene expression by silencing *HOXD* genes.[6] Identified regulated genes include the transcription factors *HOXD8-11*.[6] Expression of *HOXD* genes has been found to be associated with the anatomical location of cells in which they are expressed as well as differentiation, hence matching differentiation of a cell with the site it is located.[14] *HOXD8*, for example, is expressed only in the trunk and proximal regions of the leg.[14] HOTAIR is expressed preferentially at different locations, such as at the foreskin but not the lung.[6] Hence, HOTAIR epigenetically regulates cell differentiation at specific locations.

As described previously, HOTAIR has a strong association with cancer. Therefore, it is possible that HOTAIR could be targeted as either a biomarker or therapeutic for cancer. Because of this importance, furthering an understanding of the relationship between structure and function of HOTAIR could provide useful information.

Modular Functions

In addition to the functionality of HOTAIR shown thus far, further interactive capacity is possible. Modular regulatory principles, as described with respect to lncRNA,[15] suggest that modular functional domains of lncRNA could perform activities such as binding to RNA or DNA, as well as the PRC2 and LSD1 proteins found to bind HOTAIR. Consequently, the function of HOTAIR may extend beyond the previously identified interactions. The activities of these domains may occur independently of each other. These modular regulatory principles appear to apply to HOTAIR, as the two 5' and 3' domains respectively bind PRC2 and LSD1 independently. This was shown by deletion mutants in which nucleotides 1 – 300 were shown to bind PRC2, and nucleotides 1500 – 2146 bound to LSD1.[12] Further HOTAIR domains may be possible, such as an interaction with histones, which could consequently enable the H3K27 or H3K4me2 methylation and demethylation respectively.

This modular approach to understanding HOTAIR could have further utility by comparing secondary and tertiary structures of identified HOTAIR functional domains with unknown domains of other lncRNAs. This would also

place HOTAIR information in the context of other functional lncRNA domains. Modular functionality could then be more accurately understood, with possible benefits such as increasing the specificity of therapeutic drugs that target domains of HOTAIR. Another possible advantage arising from considering functional domains of HOTAIR could be knowledge of a domain that potentially forms the basis for designing a HOTAIR activity assay as a biomarker. Establishing a library of modular domains may also enable analysis of conserved nucleotides necessary for a specific function. This could allow diagnosis of diseases, and also hopefully form the basis for a successful therapeutic response.

If an active site necessary for the function of HOTAIR is identified within a modular domain, it may be beneficial to target this active site with an antagonist. A possible oversight of the modular approach could mistakenly be made if the functions of the lncRNA were assumed to be limited to individual domains. For example, two domains may each have an individual function, but these domains could also have an additional combined function. Another possibility would be if there was a switch mechanism in the lncRNA. Then a loss of function mutation in a domain other than the switch that is involved in the switch activity may appear to assign the function solely to this non-switch domain. Actually the switch region would also be a functional domain that could remain unidentified by loss of function mutations. Hence, when considering modular domains, a possible pitfall could be a false perception of accepting a simplified version of an actually more complex lncRNA function. A eukaryotic example of ncRNA that has a switch region binding to the coenzyme thiamine pyrophosphate (TPP) is the TPP riboswitch.[16,17]

HOTAIR in Relation to Pancreatic Cancer

In a 2012 study,[2] some conclusions were made regarding the association of HOTAIR with pancreatic cancer. Firstly, HOTAIR was found to be a negative prognostic factor for pancreatic cancer, as comparative HOTAIR expression in normal pancreas and pancreatic tumours was found to be significantly increased in tumour samples. Patients with tumours that had spread from the pancreas to the regional lymph nodes also had increased levels of HOTAIR. Similarly, samples from patients with tumours that extended beyond the pancreas, when compared with tumours only in the pancreas, showed increased HOTAIR. A Kaplan-Meier plot showed significantly shorter survival times for patients with high HOTAIR expression compared with low HOTAIR. The combination of these results, all of which show HOTAIR as a negative prognosis factor for pancreatic cancer, is useful, as it demonstrates the importance of HOTAIR with respect to pancreatic cancer. Despite this success, the majority of these increased levels of HOTAIR within the tumour, primary lymph node and extended tumour remained within the span of the negative control results. This means that HOTAIR as a biomarker for these outcomes does not appear to provide a definitive answer.

Secondly, this study indicated that HOTAIR is pro-oncogenic. This was shown by overexpression of HOTAIR, resulting in significantly increased cell growth in one of two pancreatic cell lines. In addition, cell growth was significantly reduced upon knockdown of HOTAIR in two other pancreatic cell lines. Results from gene set enrichment analysis also indicate potential for pro-oncogenic HOTAIR activity, as it was shown to regulate genes with the function of cell proliferation and cell cycle progression. Cell invasion was also decreased with knockdown of HOTAIR, and apoptosis was increased upon knockdown of HOTAIR in two pancreatic cell lines.

Thirdly, the genes regulated by HOTAIR in pancreatic cancer cell lines were concluded to be significantly different from those regulated in breast cancer cell lines. However, this aspect of the study only considered gene expression of Panc1 and MDA-MB-231, one example each respectively of pancreatic and breast cancer cell lines. It would be interesting to repeat this with different cell lines of each cancer type, as considerable variation has been shown in the results of other experiments within this study, which considered more diverse samples of cancer cell lines.

Although the gene regulation effects of HOTAIR have been shown to be dependent on an interaction with PRC2, this study also showed that some regulated HOTAIR genes are independent of PRC2. This finding was achieved by comparing mRNA expression with each of HOTAIR knockdown and EZH2 (a subunit of PRC2) knockdown, with results of significant induction of selected tumour suppression genes and no significant induction of the same genes, respectively. As HOTAIR has been shown previously to also bind LSD1, a further experiment that could be of interest is to replace the knockdown of EZH2 with knockdown of LSD1. If the regulation of these tumor suppression genes were found to be independent of LSD1, this could suggest either direct regulation by HOTAIR, or another regulatory binding partner of HOTAIR exists, or either of PRC2 or LSD1 may be sufficient to regulate these genes. This could be checked by a double knockdown of PRC2 (via EZH2) and LSD1, to determine if this gene regulation can occur with neither present.

Discussion

The results from these studies encourage the consideration of HOTAIR as a biomarker or a targeted response to pancreatic cancer. Although HOTAIR has been shown to be overexpressed in the Panc1 and L3.6pL pancreatic cancer cell lines, it could also be of interest to determine how well these particular cell lines represent the general population with pancreatic cancer. This is because there is considerable variation in HOTAIR levels from other cell lines, so it would be useful to determine whether this variation is also present in a broader range of pancreatic cancer cell lines which would better represent pancreatic cancer patients.

In future research, it may also be of interest to attempt identification of further structural details of HOTAIR, such as possible methylation of adenosines that could alter function. These types of post-transcriptional modification are often found in regions of functionality[18] which may benefit the identification of further modular regulatory

domains within HOTAIR. Also, as HOTAIR may interact not only with the proteins PRC2 and LSD1 but possibly also with RNA, identifying these possible interactions could provide further understanding of HOTAIR interactions. As lncRNA has been described as having scaffold characteristics, it may be possible that speculative RNA interacting with HOTAIR may also interact with either or both of PRC2 and LSD1. Therefore, any RNA binding to these proteins could be tested for also having interactions with HOTAIR. As we traverse the unknown characteristics of HOTAIR, we progress in the direction of making sense of this anti-sense.

Acknowledgements

Thanks to Dr Chris Brown for his inspiring, thoughtful teaching of lncRNA; Dr Harold Bernhardt for an abundance of enthusiastic, insightful discussion; and everyone from the Tate lab for helping it all precipitate.

References

1. Taft, R. J.; Pang, K. C.; Mercer, T. R.; Dinger, M.; Mattick, J. *J. Pathol.* **2010**, *220*, 126-139.
2. Kim, K.; Jutooru, I.; Chadalapaka, G.; Johnson, G.; Frank, J.; Burghardt, R.; Kim, S.; Safe, S. *Oncogene* **2012**, *32*, 1616.
3. Yoon, J.; Abdelmohsen, K.; Srikantan, S.; Yang, X.; Martindale, J.; De, S.; Huarte, M.; Zhan, M.; Becker, K.; Gorospe, M. *Mol. Cell* **2012**, *47*, 1-8.
4. Yang, F.; Bi, J.; Xue, X.; Zheng, L.; Zhi, K.; Hua, J.; Fang, G. *FEBS J.* **2012**, *279*, 3159-3165.
5. Flockhart, R. J.; Webster, D. E.; Qu, K.; Mascarenhas, N.; Kovalski, J.; Kretz, M.; Khavari, P. A. *Genome Res.* **2012**, *22*, 1006-1014.
6. Rinn, J. L.; Kertesz, M.; Wang, J. K.; Squazzo, S. L.; Xu, X.; Brugmann, S. A.; Goodnough, L. H.; Helms, J. A.; Farnham, P. J.; Segal, E.; Chang, H. Y. *Cell* **2007**, *129*, 1311-1323.
7. Gupta, R. A.; Shah, N.; Wang, K.; Kim, J.; Horlings, H. M.; Wong, D. J.; Tsai, M.; Hung, T.; Argani, P.; Rinn, J. L.; Wang, Y.; Brzoska, P.; Kong, B.; Li, R.; West, R. B.; van de Vijver, M. J.; Sukumar, S.; Chang, H. Y. *Nature* **2010**, *464*, 1071-1078.
8. Yang, Z.; Zhou, L.; Wu, L.; Lai, M.; Xie, H.; Zhang, F.; Zheng, S. *Ann. Surg. Oncol.* **2011**, *18*, 1243-1250.
9. Kogo, R.; Shimamura, T.; Mimori, K. *Cancer Res.* **2011**, *71*, 6320-6326.
10. He, S.; Liu, S.; Zhu, H. *BMC Evol. Biol.* **2011**, *11*, 102-115.
11. Ulitsky, I.; Shkumatava, A.; Jan, C. H.; Sive, H.; Bartel, D. P. *Cell* **2011**, *147*, 1537-1550.
12. Tsai, M.; Manor, O.; Wan, Y.; Mosammaparast, N.; Wang, J.; Lan, F.; Shi, Y.; Segal, E.; Chang, H. *Science* **2010**, *329*, 689-693.
13. Khalil, A. M.; Guttman, M.; Huarte, M.; Garber, M.; Raj, A.; Morales, D. R.; Thomas, K.; Presser, A.; Bernstein, B. E.; van Oudenaarden, A.; Regev, A.; Lander, E. S.; Rinn, J. L. *Proc. Natl Acad. Sci. USA* **2009**, *106*, 11667-11672.
14. Rinn, J. L.; Bondre, C.; Gladstone, H. B.; Brown, P. O.; Chang, H. Y. *PLoS Genet.* **2006**, *2*, 1084-1096.
15. Guttman, M.; Rinn, J. L. *Nature* **2012**, *482,* 339-346.
16. Cheah, M. T.; Wachter, A.; Sudarsen, N.; Breaker, R. R. *Nature* **2007**, *447*, 497-501.
17. Bocobza, S.; Aharoni, A. *Cell* **2008**, *13*, 526-533.
18. Higa, S.; Maeda, N.; Kenmochi, N.; Tanaka, T. *BBRC* **2002**, *297*, 1344-1349.