IJBST

DOI: http://doi.org/10.5281/zenodo.4275605

ISSN: 0974 – 3987
IJBST (2020), 13(3): 34-41

In-silico Identification of lncRNA Functionality in Cancers Related to Obesity

Shalini Das, Nushrat Jahan and Chittabrata Mal
Amity Institute of Biotechnology, Amity University Kolkata, Rajarhat, Newtown, Action Area III, North 24 PGS, 700135, INDIA
shweata7@gmail.com, nushratjahan8456@gmail.com, cmal@kol.amity.edu

ABSTRACT:
Investigating the gene expression networks that regulate cancer initiation and development is crucial but remains mostly incomplete. With the innovations in RNA-sequencing technologies and computational biology, long noncoding RNAs (lncRNAs) are being known and characterized at a speedy pace. Recent findings reveal that lncRNAs are involved in serial steps of cancer development. These lncRNAs act with DNA, RNA, protein molecules and/or their combos, acting as a vital regulator in chromatin organization, as well as transcriptional and post-transcriptional regulation. Their aberrant expression confers the neoplastic cell capacities for neoplasm initiation, growth, and metastasis. Here we emphasize their aberrant expression and performance in cancers. We found that Chromosome 11 consists of the highest number of lncRNA specific to different 14 cancers related to obesity. Among these cancers, breast cancer has the highest number of lncRNA associated with it. The interacting partners of the lncRNAs were analysed and domain specific interactions were studied. The results showed that the lncRNAs H19 and MALAT1 can act as potential biomarkers for different cancers.

Keywords: lncRNA, gene expression, cancer, obesity, tumor suppression

INTRODUCTION:

In recent studies, it has been seen that only a small fraction of the human genome consists of protein coding genes (~1.5%) and the remaining fraction do not encode for any protein. [1]. The proportion of non-coding proteins proportionally correlates to the complexity of the organisms [2]. Among this class of non-coding proteins, we have long non-coding RNAs (lncRNAs) that have a length of more than 200 nucleotides. It belongs to a diverse class of transcribed RNA molecules that do not encode proteins and instead have a regulatory, catalytic or structural role [3]. LncRNAs are transcribed from intergenic or intragenic regions or from some specific chromosomal region. Transcriptional and epigenetic factors both control the expression of lncRNAs [4].

Emerging studies have shown that lncRNAs play crucial roles in a broad range of biological processes and are associated with a number of diseases, like cancer, cardiovascular disease, and neurodegeneration diseases. Mutation or dysregulations in the expression of lncRNA led to development of various complex diseases [5]. Thus, these lncRNAs are becoming critically important for the understanding of several important aspects of life sciences. lncRNAs act as drivers of tumor suppressive and oncogenic functions in some widely occurring cancers, like breast and prostate cancer.

In this work, we aim to highlight the emerging impact of ncRNAs in cancer research, with a particular focus on the mechanisms and functions of lncRNAs. These studies on lncRNA have demonstrated the importance of the non-protein coding part of the human genome in carcinogenesis, representing the cutting edge of cancer research and focusing on cancer suppression. Studies have just begun concerning the identity, function, and dysregulation of lncRNAs in cancer, and recent data suggest that they may serve as master drivers of carcinogenesis. Increased research on these RNAs will lead to a greater understanding of cancer cell function, thereby leading to novel clinical applications in oncology.

MATERIALS AND METHODS:

1. Selection of Cancer Type

Overweight or being obese is linked to a higher risk of 14 different types of cancer, as identified in a 2016 review from a working group assembled by the International Agency for Research on Cancer. In the U.S. each year, about

IJBST

DOI: http://doi.org/10.5281/zenodo.4275605

ISSN: 0974 – 3987
IJBST (2020), 13(3): 34-41

28,000 new cancer diagnoses in men and 72,000 in women are due to overweight or obesity, according to the National Cancer Institute (NCI) (http://www.health.com/breast-cancer/obesity-overweight-cancer#01 - overweight-cancer-risk) (https://www.wcrf.org/int/cancer-facts-figures/data-specific- cancers). The list of cancers contains Breast Cancer, Colon Cancer, Endometrial Cancer, Esophageal Cancer, Gallbladder Cancer, Kidney Cancer, Liver Cancer, Meningioma, Multiple Myeloma, Pancreatic Cancer, Stomach Cancer, Thyroid Cancer, Prostate Cancer.

2. Collection of lncRNA

A list of lncRNA specific to particular cancer was collected from LncRNA Disease Database 2017 (http://www.cuilab.cn/lncrnadisease). The experimentally supported lncRNA-disease association data; (Genome builds used for re-mapping in this version: hg38). The total database consisted of 888 lncRNAs. The specific 14 cancers consisted of 269 lncRNAs.

3. Chromosomal location analysis of lncRNAs

The genomic locations on the chromosome were obtained from the database which specified location of different lncRNAs in different chromosomes. Number of LncRNA present in single chromosomes was noted.

4. Identification of lncRNA interacting proteins involved in different cancers

Interacting protein partners of lncRNAs associated with different cancers are collected from RAIN Database [10] and the experimentally supported lncRNA interaction data (2015 version) from LncRNA Disease Database [6 - 8]. The proteins involved in different lncRNA were identified using keywords, "lncRNA name" from online databases, such as RAIN, lncRNA interaction data, and literature survey. Relevant published articles about the candidate genes were collected using the PubMed database (http://www.ncbi.nlm.nih.gov/pubmed). The genes selected were primarily based on scores. From the different proteins available in RAIN; genes were grouped based on the score and genes of interest. The names of proteins were deduced from UniProt through GeneName.

5. Domain analysis of lncRNA interacting proteins involved in different cancers

Domain analysis was done in order to check if there was any overlapping domain for interacting protein partners specific to particular lncRNA, using Pfam [11]. After that, the pairwise score was calculated between each domain of the interacting proteins, using EMBOSS pairwise alignment tool (https://www.ebi.ac.uk/Tools/psa/).

6. Analysis of RNA expression level of selected interacting proteins

Gene expression profiling and in situ hybridization studies have revealed that lncRNA expression is developmentally regulated, can be tissue- and cell-type specific, and can vary temporally, or in response to stimuli. Many lncRNAs are expressed in a more tissue-specific fashion and with greater variation between tissues compared to protein-coding genes. RNA Expression level is analyzed from The Human Protein atlas-which aims to map the human proteins in cells, tissues and organs using integration of various omics technologies [9]. The Human Protein Atlas consists of three sub-atlases: The Tissue Atlas, The Cell Atlas, The Pathology Atlas.

RESULTS AND DISCUSSION:

Some lncRNAs have been found to be associated with many types of cancers. For example, by analyzing 14 cancer tissues, we found that the small nucleolar RNA host gene 16 (*SNHG16*) has high expression in bladder, lung cancer. Two lncRNAs named *MALAT1* and *H19* were found to be oncogenes in lung cancer, breast cancer and other cancer types. Both of them need to interact with microRNAs to execute functions. For example, *miR-138*, which inhibits the expression of *HMGA2*, was targeted by *H19*. H19 and MALAT1 were found in maximum cancer types compared to other lncRNAs. Thus, H19 and MALAT1 have been used for further analysis.

IJBST

DOI: http://doi.org/10.5281/zenodo.4275605

ISSN: 0974 – 3987
IJBST (2020), 13(3): 34-41

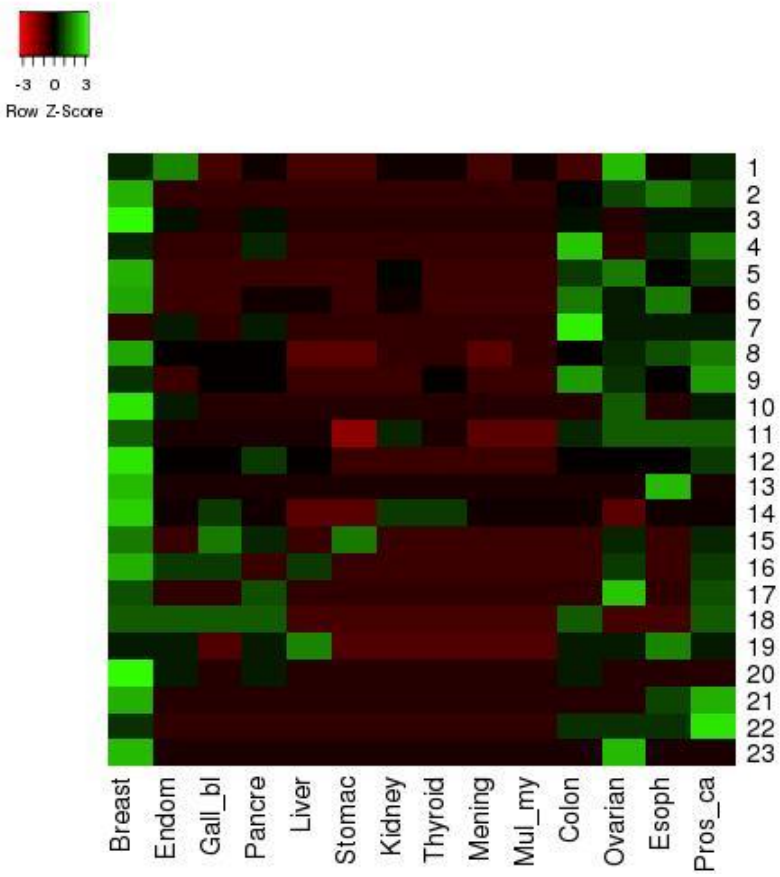

Fig. 1. Heatmap of cancers and lncRNAs originating from different chromosomes. Name of the cancers given in the Table 1.

Table 1. Total number of lncRNAs in 14 cancers

Cancer	Number of lncRNAs
Breast Cancer	63
Prostate Cancer	38
Ovarian Cancer	32
Colon Cancer	30
Esophageal Cancer	28
Pancreatic Cancer	18
Endometrial Cancer	17
Gallbladder cancer	12
Kidney Cancer	9
Thyroid Cancer	7
Liver Cancer	7
Multiple myeloma	4
Meningioma	2
Stomach Cancer	2

IJBST

DOI: http://doi.org/10.5281/zenodo.4275605

ISSN: 0974 – 3987
IJBST (2020), 13(3): 34-41

Table 2. Interacting protein partners of the lncRNAs H19 and MALAT1

LncRNAs	Interacting Proteins
H19	METTL16
	DENR
	HSPB8
	AKAP5
MALAT1	MMP8
	HCN2
	HCN4
	HCN3
	ATOH7
	PEX5L

Table 3(a). Pairwise distance scores between domains of the interacting protein partners for H19

	METTL16	HSPB8	DENR	AKAP5
METTL16	0	12	22	11
HSPB8		0	12	8
DENR			0	7
AKAP5				0

Table 3(b). Pairwise distance scores between domains of the interacting protein partners for MALAT1

	ATOH7	HCN2	HCN3	HCN4	MMP8	PEX5L
ATOH7	0	11	12	12	15.5	20
HCN2		0	1167.5	1246	12	5
HCN3			0	1228.5	16	5
HCN4				0	5	5
MMP8					0	21.5
PEX5L						0

IJBST

DOI: http://doi.org/10.5281/zenodo.4275605

ISSN: 0974 – 3987
IJBST (2020), 13(3): 34-41

Table 4. RNA expression analysis of lncRNA interacting proteins

lncRNAs	Interacting Protein partners	RPKM Value of Breast Cancer RNAs	RPKM Value of Colon Cancer RNAs
H19	METTL16	4.1	3.6
	DENR	15.6	14.4
	HSPB8	19.6	12.87
	AKAP5	0.2	0.5
MALAT1	MMP8	0.1	0.1
	HCN2	0	0.4
	HCN4	0	0
	HCN3	2	4.2
	ATOH7	0.1	0.1
	PEX5L	0	0

1. Chromosomal location of the lncRNAs

By analysing the location of the lncRNAs on different chromosomes, it is predicted that chromosome 11 consists of the highest number (34) of lncRNA specific to different cancers related to obesity. Chromosome Y (Chromosome 24) has no lncRNA. The result shown in the heatmap below (Fig.1) could be further helpful for the researchers to analyze the specific functionality of lncRNA and their location.

2. Interaction of lncRNA with other proteins involved in cancer related to obesity

The selected lnRNA interacting proteins were retrieved from RAIN database and through literature search. Over 300 different proteins have been demonstrated to have direct association with the lncRNA relation to 14 cancers related to obesity. Here H19 and MALAT1 were selected as lncRNA of interest and, therefore, the proteins associated with these lncRNA were given prior importance. Table 1 lists the total number of lncRNAs for a particular cancer type. Breast cancer has the maximum number of lncRNAs whereas meningioma and stomach cancer has just 2 lncRNAs associated with it.

Two lncRNAs (H19 and MALAT1) that were found in maximum cancer types were analysed to find their interacting proteins. 10 proteins were selected on the basis of RAIN interaction. Table 2 depicts the interacting proteins selected for the study of domain and their related functions.

3. Overlapping domains between the interacting proteins

Domain region and the domain sequence of the interacting proteins were found. Then, the pairwise distance score was calculated.

From Table 3(a) we can find that out of all the interacting proteins, domains of DENR and METTL16 have the highest pairwise distance score. This suggests there might be overlapping domains between these proteins and those are interacting with the lncRNA H19.

Table 3(b) suggests that out of all the interacting proteins, domains of HCN2 and HCN4 as well as HCN3 and HCN3 have the highest pairwise distance score. Thus we can say that there might be overlapping domains between these proteins and those are interacting with the lncRNA MALAT1.

4. RNA expression of lncRNA interacting proteins

Gene expression profiling and in situ hybridization studies have revealed that lncRNA expression is developmentally

IJBST

DOI: http://doi.org/10.5281/zenodo.4275605

ISSN: 0974 – 3987
IJBST (2020), 13(3): 34-41

regulated, can be tissue- and cell-type specific, and can vary temporally, or in response to stimuli. Many lncRNAs are expressed in a more tissue-specific fashion and with greater variation between tissues compared to protein-coding genes. Considering the GTEx Dataset, RNA expression data is reported as median RPKM (reads per kilobase per million mapped reads) generated by the GTEx datasets. This dataset consisted of 214 samples for Breast Cancer and 345 for colon cancer. RNA expression data of lncRNA interacting protein partners are presented in Table 4. While the expression of some interacting partners is higher, some have lower expressions as well.

CONCLUSION:

In recent studies, lncRNA has been proved to be a powerful regulator of adipocyte differentiation and gene expression. Blnc1 (brown fat lncRNA) was identified as a conserved lncRNA regulator of brown and beige adipocyte differentiation [12]. Again, a long non-coding RNA, HOTAIR was expressed in gluteal adipose and may regulate key processes in adipocyte differentiation [13].

From the above analysis, it is concluded that the lncRNAs H19 & MALAT1 which is common in both colon cancer & breast cancer, is expressed differently in different tissues. The expression level of lncRNAs is much higher in breast cancer and comparatively lower in colon cancer. The endogenous *H19* gene is frequently abundant in Breast cancer patients, and the overexpression of *H19* plays an important role in Breast cancer development. Since that discovery, accumulating reports of dysregulated expression of lncRNAs in Breast cancer have suggested that aberrant lncRNAs are involved in all stages of Breast cancer. Relative expression of MALAT1 was seen to be increased in Breast cancer tissues and cell lines.

One of the most abundant lncRNAs, whose expression alters during numerous cancers is MALAT1 (metastasis-associated lung adenocarcinoma transcript 1). It is a highly conserved lncRNA and it exhibits an uncommon 3'-end processing [14]. It was found that when the MALAT1 gene is knocked down in the mouse mammary carcinoma model, there was slower tumor growth and metastasis rate had reduced. Also there were alterations in splicing patterns and thus changes in expression of the genes involved in protumorigenic signaling pathways [15]. MALAT1 has also been proposed to be a potential biomarker of colorectal cancer as the MALAT1 expression was upregulated 2.26 times in the Colorectal cancer tissues as compared to the non-cancerous tissues [16]. Besides cancer, MALAT1 abnormal upregulation was seen in diabetes-induced microvascular dysfunction [17]. Thus, several studies in-vivo have been made on MALAT1 which shows that it is a potential lncRNA which can be targeted in several types of cancer.

H19 is another abundantly expressed lncRNA that has been implicated in human genetic disorders and cancer. This lncRNA is developmentally regulated and is highly expressed in fetal tissues and adult muscles [18]. H19 was also found to be a valuable target for breast cancer therapy as it is an estrogen-inducible gene and helps in the breast cancer cell survival and proliferation [19]. H19 also plays a role in gastric cancer by competing with the microRNA (MiR-141) which suppresses malignancy in human cancer [20]. A lot of research in lncRNA has shown H19 and MALAT1 to be major genes leading to aberrant gene expression and cancer. Thus, H19 and MALAT1 could serve as potential biomarkers for early diagnosis of cancers that are found to be related to obesity.

Here we have analysed the biological functions of lncRNAs in different cancers related to obesity. Newly discovered lncRNAs have been found to have a huge impact on cancer development and progression. Future studies involving the motifs, secondary structure or tertiary structure of lncRNAs as well as the gene regulatory network involving lncRNAs will help us to understand them better and establish more effective strategies for these lncRNA as a potential biomarker for early clinical diagnosis, prognosis and therapeutics in patients

IJBST

ISSN: 0974 – 3987
IJBST (2020), 13(3): 34-41

DOI: http://doi.org/10.5281/zenodo.4275605

suffering from cancer. However, a lot of clinical samples will be required for further studies to confirm how specific and sensitive these lncRNAs are as clinical biomarkers.

ACKNOWLEDGEMENTS

Authors gratefully acknowledge Amity University Kolkata for computational facilities.

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CONFLLICT OF INTEREST

The authors declare no potential conflict of interest.

REFERENCES:

[1] Bertone, P., Stolc, V., Royce, T. E., Rozowsky, J. S., Urban, A. E., Zhu, X., ... & Gerstein, M. (2004). Global identification of human transcribed sequences with genome tiling arrays. *Science*, *306*(5705), 2242-2246.

[2] Taft, R. J., Pheasant, M., & Mattick, J. S. (2007). The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays*, *29*(3), 288-299.

[3] Carninci, P., & Hayashizaki, Y. (2007). Noncoding RNA transcription beyond annotated genes. *Current opinion in genetics & development*, *17*(2), 139-144.

[4] Yang, G., Lu, X., & Yuan, L. (2014). LncRNA: a link between RNA and cancer. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, *1839*(11), 1097-1109.

[5] Chen, X., & Yan, G. Y. (2013). Novel human lncRNA–disease association inference based on lncRNA expression profiles. *Bioinformatics*, *29*(20), 2617-2624.

[6] Chen, G., Wang, Z., Wang, D., Qiu, C., Liu, M., Chen, X., ... & Cui, Q. (2012). LncRNADisease: a database for long-non-coding RNA-associated diseases. *Nucleic acids research*, *41*(D1), D983-D986.

[7] Bao, Z., Yang, Z., Huang, Z., Zhou, Y., Cui, Q., & Dong, D. (2019). LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases. *Nucleic acids research*, *47*(D1), D1034-D1037.

[8] Jia, K., Gao, Y., Shi, J., Zhou, Y., Zhou, Y., & Cui, Q. (2020). Annotation and curation of the causality information in LncRNADisease. *Database*, *2020*.

[9] Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., ... & Wernerus, H. (2010). Towards a knowledge-based human protein atlas. *Nature biotechnology*, *28*(12), 1248-1250.

[10] Junge, A., Refsgaard, J. C., Garde, C., Pan, X., Santos, A., Alkan, F., ... & Gorodkin, J. (2017). RAIN: RNA–protein association and interaction networks. *Database*, *2017*.

[11] Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., ... & Studholme, D. J. (2004). The Pfam protein families database. *Nucleic acids research*, *32*(suppl_1), D138-D141.

[12] Mi, L., Zhao, X. Y., Li, S., Yang, G., & Lin, J. D. (2017). Conserved function of the long noncoding RNA Blnc1 in brown adipocyte differentiation. *Molecular metabolism*, *6*(1), 101-110.

[13] Divoux, A., Karastergiou, K., Xie, H., Guo, W., Perera, R. J., Fried, S. K., & Smith, S. R. (2014). Identification of a novel lncRNA in gluteal adipose tissue and evidence for its positive effect on preadipocyte differentiation. *Obesity*, *22*(8), 1781-1785.

[14] Zhang, B., Arun, G., Mao, Y. S., Lazar, Z., Hung, G., Bhattacharjee, G., ... & Spector, D. L. (2012). The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell reports*, *2*(1), 111-123.

[15] Arun, G., Diermeier, S., Akerman, M., Chang, K. C., Wilkinson, J. E., Hearn, S., ... & Brogi, E. (2016). Differentiation of mammary tumors and reduction in metastasis upon Malat1 lncRNA loss. *Genes & development*, *30*(1), 34-51.

[16] Zheng, H. T., Shi, D. B., Wang, Y. W., Li, X. X., Xu, Y., Tripathi, P., ... & Cai, S. J. (2014). High expression of lncRNA MALAT1 suggests a biomarker of poor prognosis in colorectal cancer. *International journal of clinical and experimental pathology*, *7*(6), 3174.

[17] Liu, J. Y., Yao, J., Li, X. M., Song, Y. C., Wang, X. Q., Li, Y. J., ... & Jiang, Q. (2014). Pathogenic role of lncRNA-MALAT1 in endothelial cell dysfunction in diabetes mellitus. *Cell death & disease*, *5*(10), e1506-e1506.

[18] Kallen, A. N., Zhou, X. B., Xu, J., Qiao, C., Ma, J., Yan, L., ... & Min, W. (2013). The imprinted H19 lncRNA antagonizes let-7 microRNAs. *Molecular cell*, *52*(1), 101-112.

[19] Sun, H., Wang, G., Peng, Y., Zeng, Y., Zhu, Q. N., Li, T. L., ... & Zhu, Y. S. (2015). H19 lncRNA mediates 17β-estradiol-induced cell

IJBST

DOI: http://doi.org/10.5281/zenodo.4275605

ISSN: 0974 – 3987
IJBST (2020), 13(3): 34-41

proliferation in MCF-7 breast cancer cells. *Oncology reports*, *33*(6), 3045-3052.

[20] Zhou, X., Ye, F., Yin, C., Zhuang, Y., Yue, G., & Zhang, G. (2015). The interaction between MiR-141 and lncRNA-H19 in regulating cell proliferation and migration in gastric cancer. *Cellular Physiology and Biochemistry*, *36*(4), 1440-1452.