

GENOME-WIDE ANALYSIS OF LONG NON-CODING RNA (LNCRNA) OF AUTOIMMUNE THYROID DISEASES USING BIOINFORMATICS APPROACHES

SHOBANA SUGUMAR^{1*}, APARAJITA²

Department of Bioinformatics, School of Bioengineering, SRM University, Kattankulathur 603203, Tamilnadu, India
Email: ksemaa@gmail.com

Received: 09 Feb 2016 Revised and Accepted: 20 Apr 2016

ABSTRACT

Objective: Long non-coding RNA's (lncRNA's) have a crucial role in cancer biology. In this study, the genome sequence analysis of lncRNA expression in autoimmune thyroid disease is done to identify novel targets for further study of the disease.

Methods: All the data were collected from Disgenet and Ensemble genome browser. Gene ontology and network analysis were performed using the standard enrichment annotation method. Association of lncRNA and their targeted mRNA were analyzed by GENEMANIA.

Results: Of the all 334 lncRNA transcripts identified, only four had coding potential. lncRNA's transcripts ENST00000462973, ENST00000555326 were involved in autoimmune thyroid disease pathway which corresponds to thyroid peroxidase (TPO) and thyroid-stimulating hormone receptor (TSHR), and this could provide better insights to therapeutics.

Conclusion: Our current study on the potential link between lncRNAs and autoimmune thyroid disease presents a novel area for further investigations into the target genes of such lncRNAs, leading to therapeutic strategies for the disease.

Keywords: lncRNA, Autoimmune thyroid disease, GENEMANIA

© 2016 The Authors. Published by Innovare Academic Sciences Pvt Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

INTRODUCTION

All the genetic information related to humans is stored in the human genome. Human genome can be called as an elegant but cryptic store of information [1, 2]. With the advancement in sequencing technologies, better insight to human genome was well understood. Genomes are broadly classified into coding DNA and non-coding DNA [3, 4]. Coding DNA can be transcribed into mRNA and translated into proteins and occupy only a small fraction of the genome (<2%). Noncoding DNA doesn't encode a protein and comprises of 98% of the genome (<http://www.genome.gov/1201123>, <http://www.sanger.ac.uk/about/history/hgp/>) [5, 6]. Noncoding RNA's can be defined as any transcript or its fragment which is not used as a template in ribosomal protein synthesis [7]. They can be used for the regulation of protein-coding genes; also, they play an important role in oncogenesis and tumor prognosis. In cellular processes such as regulation of gene expression, splicing and direct chemical modification are some of the functions of noncoding RNA's [8-11].

There are a lot number of ncRNA like transcribed ultraconserved region (T-UCR's), small nucleolar RNA's (snoRNA), PIWI-interacting RNA's (piRNA), large intergenic noncoding RNA (lincRNA) and heterogeneous group of long noncoding RNA's (lncRNA) [12-14]. They are involved in a number of human diseases which include neurobehavioral and developmental disorders as well as certain forms of cancer. Certain ncRNA's encode to chromosomal regions associated with neurobehavioral disorders, including autism, bipolar, affective disorder and schizophrenia [15-17]. Long noncoding RNA's are longer than 200nt and typically expressed in a developmental-specific manner. They also have short ORF size which is <200nt [18-20]. They exhibit low sequence conservation. Five different categories of classifying lncRNA are sense/antisense, when they overlap with the exons of a different transcript on the same or on the opposite strand, intronic; originating from an intron of a different transcript, bidirectional; when the lncRNA and an adjacent transcript on the opposite strand are expressed at the same time and intergenic when the lncRNA is located in a region not affected by another coding sequence [21-23]. These RNA's are involved in various cellular processes including trans-regulation of nearby protein-coding genes, imprinting control and alternative

splicing. lncRNA's are associated with a number of diseases such as cancer, neurological disorders, heart disease and autoimmune disease. Based on the function lncRNA can be divided into three groups [24]. The first group of lncRNA's can bind and guide cellular proteins towards the target. The second group of lncRNA can bind effector molecules and initiate the formation of specific molecular complexes [25-27]. The third group of lncRNA's can bind proteins or RNA molecules and thus prevent these from exerting their function. lncRNA plays an important role in transcriptional regulation. It is a cellular process which includes transcriptional factor and polymerases to sequence-specific promoter site of genes [28]. They regulate the activity of transcriptional factor and polymerases. Some of lncRNA also act directly on transcription factors. The transcriptional factor is a means by which cell regulates the conversion of DNA to RNA [29-31].

An example of transcriptional regulation is about PcG proteins. Polycomb group proteins (PcG) silence the expression of thousands of mammalian genes. lncRNA's target PcG proteins to specific genomic locations. Ezh2 (Enhancer of zeste homolog2) a histone methyltransferase and member of polycomb repressive complex2 (PRC2) bind directly to 1.6 kb-long nc RNA known as RepA. These lncRNA's are active in post-transcriptional events also. Post transcriptional modifications include alternative splicing, editing, translation and trafficking. MALAT1 (Metastasis Associated Lung Adenocarcinoma Transcript 1) plays an important role in alternative splicing. The lncRNA HOTAIR (HOX transcript antisense intergenic RNA) is coded in the HOXC locus on chromosome 12 and regulates the HOXD genes on chromosome 2 by binding to the polycomb repressive complex PRC2 and inducing epigenetic silencing by methylation of several tumor suppressor genes on HOXD locus [32]. HOTAIR expression is deregulated in several cancers.

The expression of HOTAIR is elevated in breast cancer which has been correlated with metastatic capacity and poor prognosis. BRCA1-binding region in the polycomb protein EZH2 overlaps with the noncoding RNA binding domain, and BRCA1 expression inhibits the binding of EZH2 to the HOTAIR. HOTAIR expression and metastasis have an adverse outcome in different types of cancer including esophageal, liver, pancreas and colorectal cancers [33]. Maternally expressed 3 (MEG3) is an lncRNA that is expressed in

normal tissues. It can activate p53 and inhibit tumor genesis and progression of various types of cancers. MEG3 gene expression is down-regulated or lost in a variety of primary human tumors and tumor cell lines, and re-expression of MEG3 has been shown to inhibit *in vitro* tumor cell proliferation [34-36]. Autoimmune diseases are caused by an immune response against constituents of the body's own tissues [37]. They occur predominantly in women. Thyroid diseases are endocrine related problems. They emerge as the age increases. This occurs due to the dysfunction of the thyroid gland. Different thyroid diseases include Hashimoto's thyroiditis, hyperthyroidism and hypothyroidism. Imbalance in the production of thyroid hormones arises from dysfunction of the thyroid gland itself. The pituitary gland produces thyroid-stimulating hormone (TSH) or the hypothalamus which regulates the pituitary gland via thyrotropin-releasing hormone. The concentration of TSH increases with age. Most common cause of hypothyroidism is when the body makes antibodies that destroy parts of the thyroid gland. Also, due to malfunctioning of pituitary problems, hypothalamus problems, and iodine deficiency this disease happens. Most common symptoms of hypothyroidism include coarse and dry hair, confusion/forgetfulness, constipation, depression, dry and scaly skin, fatigue, hair loss, increased menstrual flow, intolerance to cold temperature, irritability, muscle cramps, slower heart rate, weakness, weight gain. If proper treatment is not given on time, it can lead to a diseased state called Myxedema coma where heart failure can happen [38-41]. This is characterized by normal free thyroxin (FT4) and elevated thyrotropin (TSH) levels increase with aging and from 3 to 16%. Symptoms associated with hyperthyroidism are increased heart rate, high blood pressure, increased body temperature, increased sweating, clamminess, feeling nervous, increased appetite accompanied by weight loss, interrupted sleep. In the present study, all lncRNA's were collected from the genes associated with autoimmune thyroid disease. Those lncRNA's which were not present in NONCOD database were filtered out and their gene enrichment studies were conducted. Novel functional lncRNA's were identified for autoimmune thyroid disease. Our current study on the potential link between lncRNAs and autoimmune thyroid disease presents a novel area for further investigations into the target genes of such lncRNAs, leading to therapeutic strategies for the disease.

MATERIALS AND METHODS

Dataset

Disgenet [42] is an open access database which is a collection of human disease association studies and genes associated. A key feature of this database is to obtain information and allows the user to go back to the original source of information, i.e., to explore data in its original context. All the genes associated with autoimmune thyroid disease were searched in Disgenet database. The Ensemble data (www.ensembl.org) was used to provide genes and other annotation such as regulatory regions, conserved base pairs across species, and sequence variations. The Ensemble gene set is based on protein and mRNA evidence in UniProtKB and NCBI RefSeq databases, along with manual annotation from the VEGA/Havana group. Respective transcripts were obtained from targeted genes by searching the gene name in ensemble database. Only lncRNA or processed transcripts whose length is greater than 200 nucleotides were sorted out and used for further analysis.

Finding the functionally important lncRNA

Further, a database designed to store information about non-coding RNA's (excluding tRNA's and rRNA's) was used. Information related to lncRNA owing to similar alternative splicing pattern to mRNA was stored. There are about 210,831 number of lncRNA in NONCODE version 4. All the transcripts retrieved from ensemble database were cross-checked with the NONCODE database to find the lncRNA having functional importance.

Classification of lncRNA-based on coding potential

A Support Vector Machine-based classifier, named Coding Potential Calculator (CPC), is used to assess the protein-coding potential of a transcript based on six biologically meaningful sequence features.

10-fold cross-validation on the training dataset and independent testing on three large standalone datasets showed that CPC can discriminate coding from noncoding transcripts with high accuracy. All the transcripts which were not in NONCOD database were checked for the non-coding efficiency. The data will be further classified as coding potential, weekly and non-coding potential transcripts. coding potential transcripts have values greater than zero while weekly coding will have values near to zero and non-coding with negative values.

Exploring the coding domains of lncRNA

HMMER is used for searching sequence databases for homologs of protein sequences, and for making protein sequence alignments. It implements methods using probabilistic models called profile is hidden Markov models (profile HMMs). The advantage of using HMMs is that HMMs have a formal probabilistic basis. The FASTA sequences were obtained from the ensemble and they were converted to protein sequences and explored for any similar coding domains against pfam database. All the three open reading frames were checked.

Enrichment analysis of lncRNA

Gene ontology tool was used to perform enrichment analysis, in which biological meaning was assigned to a group of genes and this tool helps a researcher to investigate on a group of genes rather than a single gene. All the genes were added in the tool. Those GO terms which had p-value<0.05 were filtered and analyzed for biological, cellular and molecular ontologies.

DAVID functional annotation cluster analysis

The expression data was analyzed using the Functional Annotation Cluster (FAC) tool contained in the Database for Annotation, Visualization and Integrated Discovery (DAVID). As a next step, we used DAVID tool which helps to extract biological features/meaning associated with large gene lists. All the four lncRNA targeted genes were given to the DAVID tool. Default values were checked and other inputs like genetic association database disease, OMIM disease, PIR seq feature, SwissProt comment type, SwissProt PIR keyword, UniProt seq feature, GO terms, PANTHER BP pie chart (to categorize genes by biological processes), KEGG pathway, REACTOM pathway, protein domain block, InterPro, Pfam, PIR superfamily, PROSITE, SMART were also checked to generate related results. As a final step, we used GeneMANIA, a flexible, user-friendly web interface for generating hypotheses about gene function, analyzing gene lists and prioritizing genes for functional assays. GeneMANIA will output genes that likely are involved in the same process.

RESULTS

Autoimmune thyroid disease including Grave's disease and Hashimoto's thyroiditis arises due to complex interactions between environmental and genetic factors. Candidate gene analysis, whole-genome linkage screening, genome-wide association studies and whole genome sequencing are the major technologies that have advanced this field. As autoimmune thyroid disease is a complex disorder, association studies are a major tool for identifying genes conferring susceptibility. Association studies can be best studied by DISGENET database (<http://www.disgenet.org/web/DisGeNET/menu/home>). All the genes and diseases associated with autoimmune thyroid disease (Disgenet name-autoimmune thyroid disease, disease id-umls: C0178468) were collected. There were 87 genes associated with autoimmune thyroid disease and the diseases that share genes were 18,997. To obtain a statistically meaningful disease data, the 87 genes were narrowed to 25 genes based on their scores (table 1). The functionality of these identified lncRNA's was checked from NONCOD database (table 2). Genes CTLA, TG, TNFRSF25, FCRL3, RBM45, ACP1, SLC26A4, THPO, HLA-DPB1, and CD4 have a functional lncRNA, but TSHR, PTPN22, FCRL3, FLNB, TRIP13, THRAP3 genes corresponding lncRNA's are not functional. From the coding potential results, it was found that transcripts ENST0000055326, ENST00000508456, ENST00000462973, ENST00000478853 corresponding to TSHR, TRIP13, TPO, THRAP3 genes have a very low coding score and they are classified as non-coding transcripts. 32 transcripts were classified as weekly coding (table 3).

Table 1: Genes associated with autoimmune thyroid disease from disgenet

Gene	Symbol	Uniprot	Gene name	Pathway	Score
1493	CTLA4	P16410	cytotoxic T-lymphocyte-associated protein 4	Immune System	0.020
7253	TSHR	P16473	thyroid stimulating hormone receptor	Signal Transduction	0.011
7038	TG	P01266	Thyroglobulin		0.009
7173	TPO	P07202	thyroid peroxidase	Metabolism	0.009
9319	TRIP13	Q15645	thyroid hormone receptor interactor 13		0.008
140805	HT		Hashimoto thyroiditis		0.008
3126	HLA-DRB4		major histocompatibility complex, class II, DR beta 4	Immune System	0.005
9967	THRAP3	Q9Y2W1	thyroid hormone receptor associated protein 3	Immune System	0.005
3133	HLA-E	P13747	major histocompatibility complex, class I, E	Immune System	0.004
3123	HLA-DRB1	P01912;P01911	major histocompatibility complex, class II, DR beta 1	Immune System	0.003
26191	PTPN22	Q29974	protein tyrosine phosphatase, non-receptor type 22 (lymphoid)		0.003
8718	TNFRSF25	Q93038	tumor necrosis factor receptor superfamily, member 25		0.002
3630	INS	P01308	Insulin	Developmental Biology; Disease; Metabolism; Metabolism of proteins; Signal Transduction	0.002
6528	SLC5A5	Q92911	solute carrier family 5 (sodium/iodide cotransporter), member 5	Metabolism; Transmembrane transport of small molecules	0.002
115352	FCRL3	Q96P31	Fc receptor-like 3		0.002
8797	TNFRSF10A	O00220	tumor necrosis factor receptor superfamily, member 10a		0.002
129831	RBM45	Q8IUH3	RNA binding motif protein 45		0.002
348120	LINC01193		long intergenic non-protein coding RNA 1193		0.002
2317	FLNB	O75369	filamin B, beta	Immune System	0.001
3559	IL2RA	P01589	interleukin 2 receptor, alpha	Immune System	0.001
52	ACP1	P24666	acid phosphatase 1, soluble		0.001
5172	SLC26A4	O43511	solute carrier family 26 (anion exchanger), member 4	Transmembrane transport of small molecules	0.001
920	CD4	P01730	CD4 molecule	Disease; Immune System	0.001
7066	THPO	P40225	Thrombopoietin	Hemostasis	0.001
3115	HLA-DPB1	P04440	major histocompatibility complex, class II, DP beta 1	Immune System	0.001

Table 2: Functional lncRNA from NONCODE

Gene	ID	Transcript	Noncod
CTLA4	CTLA4-003	ENST00000487393	NONHSAT076500
TSHR	TSHR-008	ENST00000555326	
TG	TG-016	ENST00000522523	
	TG-014	ENST00000524151	NONHSAT129203
	TG-011	ENST00000520197	
	TG-012	ENST00000519294	
TPO	TPO-011	ENST00000497517	
	TPO-009	ENST00000425083	
	TPO-013	ENST00000462973	
	TPO-008	ENST00000479902	
PTPN22	PTPN22-009	ENST00000534519	
TNFRSF25	TNFRSF25-011	ENST00000475730	NONHSAT000647
FCRL3	FCRL3-009	ENST00000473231	
	FCRL3-007	ENST00000480682	
	FCRL3-008	ENST00000494724	
	FCRL3-010	ENST00000468507	
	FCRL3-013	ENST00000457799	NONHSAT006945
	FCRL3-012	ENST00000478179	NONHSAT006948
RBM45	RBM45-005	ENST00000464647	NONHSAT075737
FLNB	FLNB-014	ENST00000484981	
ACP1	ACP1-010	ENST00000484464	
	ACP1-012	ENST00000484125	NONHSAT068501
SLC26A4	SLC26A4-005	ENST00000480841	
	SLC26A4-007	ENST00000492030	
	SLC26A4-006	ENST00000460748	NONHSAT122703
	SLC26A4-004	ENST00000497446	
	SLC26A4-003	ENST00000477350	

THPO	THPO-004	ENST00000477594	NONHSAT093713
HLA-DPB1	HLA-DPB1-006	ENST00000471184	NONHSAT108964
	HLA-DPB1-009	ENST00000478189	NONHSAT108966
	HLA-DPB1-007	ENST00000498038	
CD4	HLA-DPB1-005	ENST00000488575	NONHSAT108965
	CD4-004	ENST00000538827	NONHSAT026148
	CD4-012	ENST00000536610	NONHSAT026146
	CD4-011	ENST00000536563	
	CD4-008	ENST00000535466	NONHSAT026149
	CD4-013	ENST00000535707	
	CD4-009	ENST00000536590	
TRIP13	TRIP13-003	ENST00000510412	
	TRIP13-007	ENST00000508456	
	TRIP13-005	ENST00000508430	
	TRIP13-004	ENST00000509210	
THRAP3	THRAP3-005	ENST00000466743	

All the lncRNA's translated information was obtained in all three reading frames. It was sent to HMMER tool to see if these lncRNA's do not have a common protein-coding region (table 4). To obtain the functional aspects of genes encoded by lncRNA, gene enrichment analysis was done. Gene ontology and DAVID (Database for Annotation, Visualization, and Integrated Discovery) were used. From the DAVID analysis, we got two functional annotation clusters having respective enrichment score 0.35 and 0.13. The greater enrichment value more refined the results are. For further analysis of the genes, we went for annotation cluster 1. Out of four genes, three genes THRAP3, TPO, TSHR share alternative splicing, splice variants, alternative products as the functional category (table 5, 6). While lncRNA's have the property of alternative splicing and thus these three genes may have the potential to regulate transcription. Also, the fold enrichment values were also less than 2. From the functional annotation table, terms having p-value<0.05 showed genes related to thyroid and related terms.

Table 3: Noncoding RNA's from coding potential calculator score

Name	Transcript ID	CPC Score	CPC Task ID
TSHR-008	ENST00000555326	-0.719541	ACCECF50-C342-11E4-8340-80687D09C235
TRIP13-007	ENST00000508456	-1.08334	73B0ED90-BC2E-11E4-8340-EC199BD12F24
TPO-013	ENST00000462973	-1.16605	236CEE90-BC4E-11E4-8340-B2BE43EA36D5
THRAP3-003	ENST00000478853	-1.11326	13741A70-BCEB-11E4-8340-A4563B8E59E1

Table 4: Functional annotation cluster1 from DAVID

Annotation cluster 1				Enrichment score 0.35					
Category	Term	Count	%	P-value	Genes	List Total	Fold Enrichment	Bonferroni	Benjamini
SP_PIR_KEYWORDS	alternative splicing	3	75	0.3366	TPO, TSHR, TRIP13	4	1.926582	0.999999	0.999727
UP_SEQ_FEATURE	splice variant	3	75	0.3379	TPO, TSHR, TRIP13	4	1.922066	0.999993	0.999993
SP_COMMENT_TYPE	alternative products	3	75	0.3433	TPO, TSHR, TRIP13	4	1.903523	0.998181	0.957356
SP_COMMENT_TYPE	similarity	3	75	0.9809	TPO, TSHR, TRIP13	4	0.817050	1	0.999949

Table 5: Functional annotation cluster2 from DAVID

Annotation cluster 2				Enrichment score-0.13					
Category	Term	Count	%	P-value	Genes	List Total	Fold Enrichment	Bonferroni	Benjamini
sp_pir_keywords	polymorphism	3	75	0.648681	THRAP3, TPO, TSHR	4	1.249025	1	0.999999
up_seq_feature	sequence variant	3	75	0.687010	THRAP3, TPO, TSHR	4	1.195359	1	0.999999952
sp_comment_type	function	3	75	0.783241	THRAP3, TPO, TSHR	4	1.072366	1	0.996764727
sp_comment_type	subcellular location	3	75	0.832350	THRAP3, TPO, TSHR	4	1.013532816	1	0.995288005

Heat map results showed thyroid peroxidase, thyroid stimulating hormone receptor, thyroid hormone receptor interactor 13 taking part in alternative splicing events (fig. 1). From the functional annotation table, KEGG pathway results

showed that lncRNA targeted genes TPO and TSHR were there in the pathway represented. The marked red genes in the pathway (shown in below fig. 2) are involved in a pathway associated with thyroid disease.

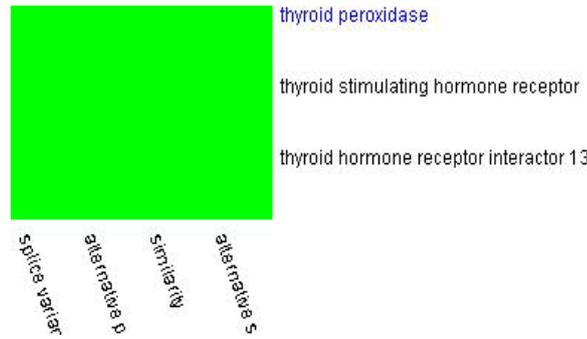


Fig. 1: Heat map generated for cluster1

Table 6: Functional annotation table from DAVID

Category	Term	P-value	Genes	Bonferroni	Benjamini
sp_pir_keywords	thyroid gland	0.0006	TPO, TSHR	0.0246	0.0246
sp_pir_keywords	congenital hypothyroidism	0.0009	TPO, TSHR	0.0367	0.0185
genetic_association_db_disease	Hypothyroidism	0.0033	TPO, TSHR	0.0640	0.0640
kegg_pathway	hsa05320:Autoimmune thyroid disease	0.0100	TPO, TSHR	0.0586	0.0586
sp_comment	alternative products: Additional isoforms seem to exist	0.0369	TPO, TSHR	0.7606	0.76065
goterm_bp_fat	GO: 0006366~transcription from RNA polymerase II promoter	0.05100	THRAP3, TRIP13	0.9990	0.9990
sp_comment_type	sequence caution	0.0570	THRAP3, TSHR, TRIP13	0.5858	0.5858
goterm_bp_fat	GO: 0006351~transcription, DNA-dependent	0.0633	THRAP3, TRIP13	0.9998	0.9871
goterm_bp_fat	GO: 0032774~RNA biosynthetic process	0.0642	THRAP3, TRIP13	0.9998	0.9472
goterm_mf_fat	GO: 0003712~transcription cofactor activity	0.0815	THRAP3, TRIP13	0.9570	0.9570
sp_pir_keywords	transmembrane protein	0.0968	TPO, TSHR	0.9829	0.7428

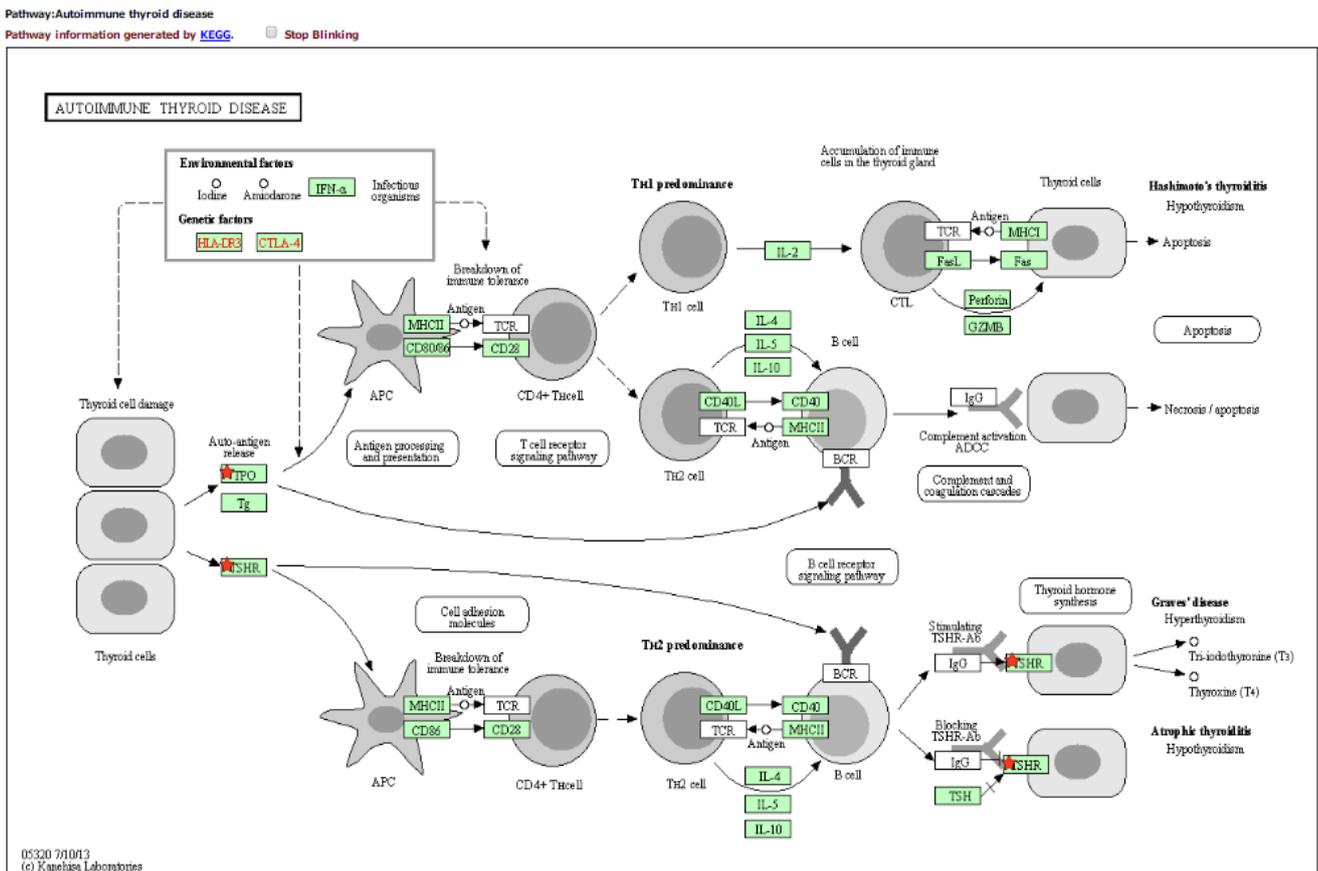


Fig. 2: Pathway of autoimmune thyroid disease from KEGG (kyoto encyclopedia of genes and genomes, <http://www.genome.jp/kegg/>)

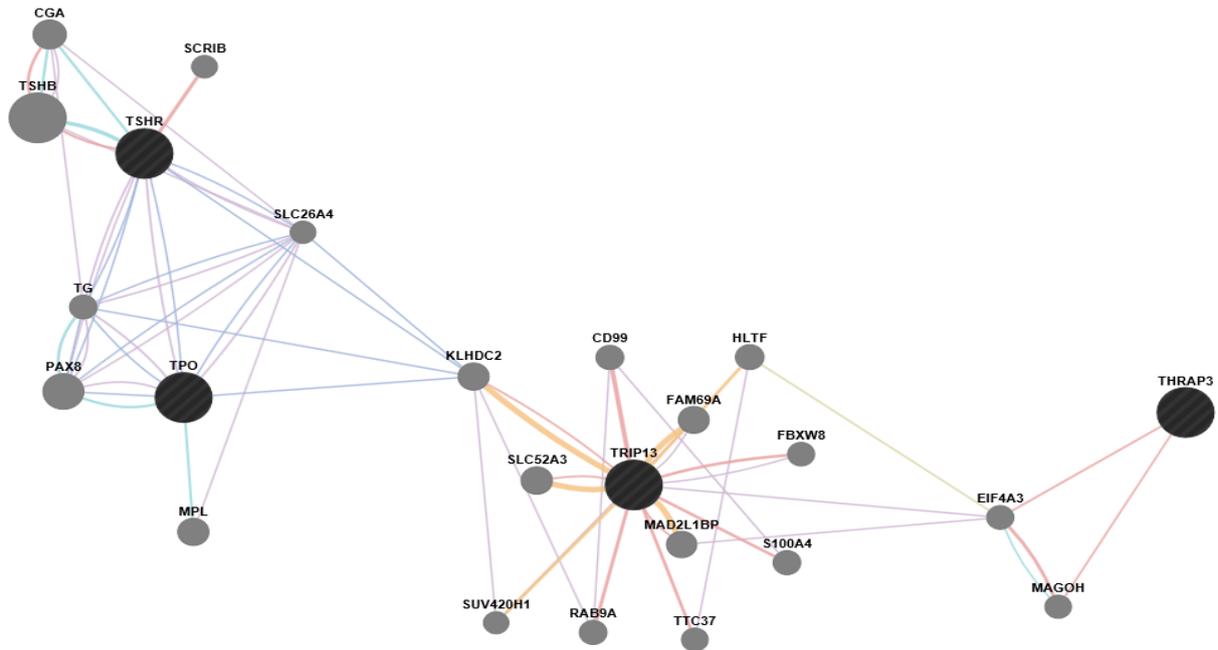


Fig. 3: GeneMANIA interaction diagram autoimmune thyroid disease (<http://www.genemania.org/>)

Four differently transcribed mRNA's regulated by lncRNA, including TSHR, TPO, THRAP3, TRIP13 were subjected to GENEMANIA analysis (fig. 3). Three TSHR, TPO and TRIP13 were found to be in a functional network in terms of co-expression. Also, TSHR, TRIP13 and THRAP3 are involved in physical interactions.

GO analyses predicted that lncRNAs targeted mRNA were associated with the metabolic process (ontology: biological process), cell (ontology: cellular component) and binding (ontology: molecular function).

DISCUSSION

In recent years, there were a lot of evidence supporting lncRNA's associated with cancer. However, very few studies have been conducted on the potential role of lncRNA's in autoimmune diseases. Investigations into the molecular mechanism of the autoimmune disease especially thyroid diseases have focused only on protein coding part. Therefore, our understanding of lncRNA function in autoimmune thyroid disease is poor. For this reason, current study focusses on non-coding part (lncRNA) for further investigation of the therapeutic potential of autoimmune thyroid disease. Since long non-coding RNA's fall under the transcriptome study, it was necessary to find the corresponding transcripts from the gene associated. For lncRNA transcriptome study, more emphasis was given on the reference database used. In this study, ensemble project and related annotation from the Biomart/Havana group at Sanger Institute provide effective identification, classification, and counting of differentially expressed non-coding transcriptome associated with autoimmune thyroid disease.

All the associated transcripts of autoimmune thyroid disease from genes were retrieved from the ensemble gene browser (<http://www.ensembl.org/index.html>). There were 1711 transcripts identified from 25 genes which included both protein-coding and non-coding transcripts. As our interest was on lncRNA which comes under non-coding part, the results were filtered, and 334 transcripts were identified. The non-coding part contains processed not only transcripts/lncRNA but also 23 retained introns, 1 pseudogene, 18 processed transcripts, 15 nonsense-mediated decay, 209 known protein-coding transcripts, 18 novel protein-coding transcripts and 25 putative protein coding transcripts. retained introns are alternatively spliced transcript that is believed to contain intronic sequence relative to other coding transcripts in a given locus, nonsense-mediated decay is a process which detects nonsense mutations and prevents the expression of truncated proteins, novel protein coding transcripts are having a sequence matched outside Ensemble for an alternate species and known protein-coding transcripts have a sequence match in a sequence repository external to Ensemble for same species.

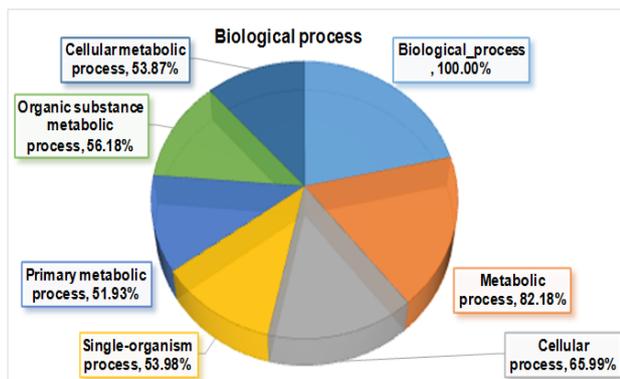


Fig. 4: Gene ontology for biological process (<http://geneontology.org>)

All the genes were showing biological significance and 56.18% showing the metabolic process. There are seven ontologies associated with the biological process (fig. 4). 100% of genes corresponds biological process, then metabolic process and finally cellular process. Other ontologies associated are a single-organism process, primary metabolic process, organic substance metabolic process and cellular metabolic process. The function carried out by the genesis tetrapyrrole binding. This refers to molecular ontology (fig. 5).

Molecular ontology illustrates seven top ten ontologies which have tetrapyrrole binding with 36%, heme binding 33%, antioxidant activity 13%, oxidoreductase active acting on peroxide as receptor 8%, and peroxidase activity 7%. There are nine ontology terms associated with the cellular process, and 21% of genes are involved in cellular processes (fig. 6), 14% in the cell, and 13% in cell part, intracellular part, an integral component of the membrane and intrinsic component of cell membrane. From the cellular process, it was found that most of the gene product functions are related to cell.

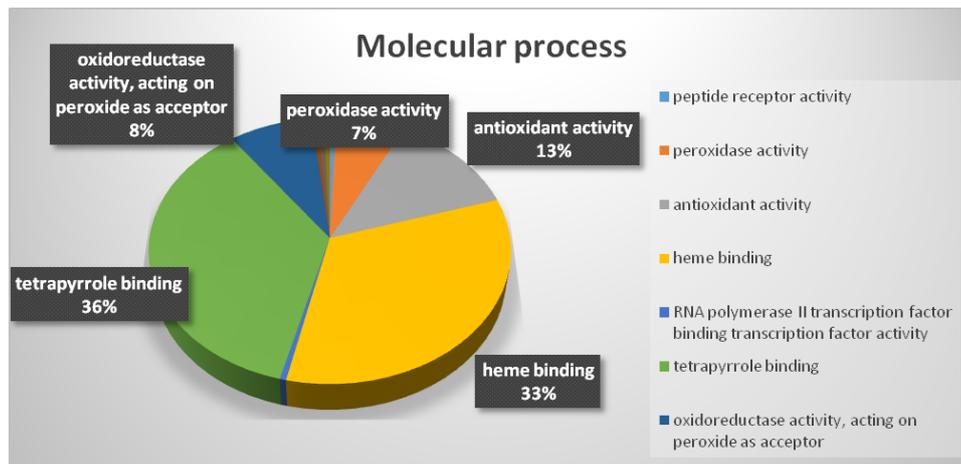


Fig. 5: Gene ontology for molecular process (<http://geneontology.org>)

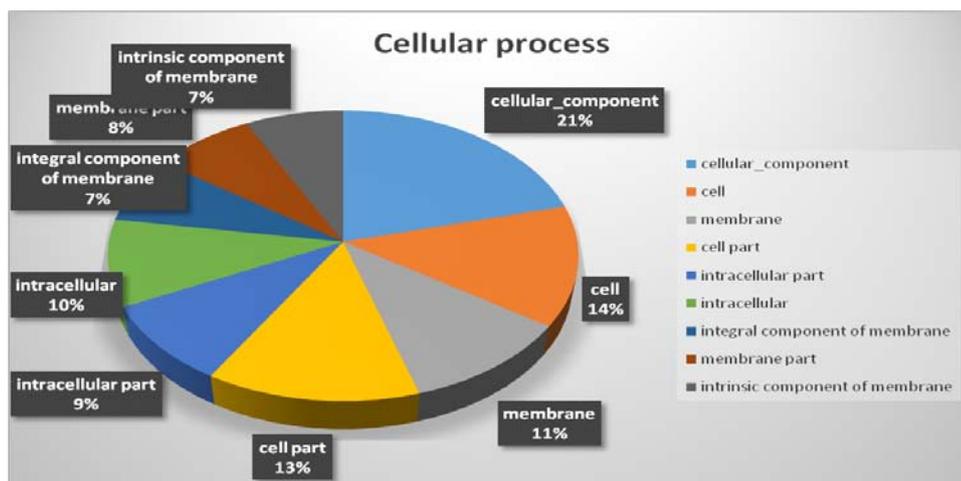


Fig. 6: Gene ontology for cellular process (<http://geneontology.org>)

In this study, we identified 1711 protein transcripts from 25 genes related to autoimmune thyroid disease. Of these 334 transcripts contained lncRNA's which were greater than 200nt. Unlike protein-coding gene or miRNA's the function of lncRNA's cannot be currently inferred from sequence/structure. Therefore, to date, most of the studies have predicted function via, a genomic association of lncRNA's with protein-coding genes because lncRNA's often regulate the expression of their overlapping or neighboring protein-coding genes [19].

To further define the biological processes lncRNA's may be involved in, gene ontology enrichment analysis was done with a protein coding genes associated with lncRNA's in genomic content. GO analysis predicted that lncRNA's targeted mRNA were associated with the metabolic process (ontology: biological), cell (ontology: cellular) and binding (ontology: molecular). Similar studies were done with hepatoblastoma tissues [36].

With the four lncRNA targeted genes, functional annotation clustering was done with DAVID with default GO term libraries. Parameters were set with high stringency and ease=0.1. The rank of the four impacted GO biological processes, GO molecular function, Swiss-prot (SP) and protein information resource (PIR) keywords (SP_PIR_KEYWORD) show that the lncRNA targeted genes was able to capture GO biological processes which can be easily expected in transcription from RNA polymerase II promoter which had p -value<0.05 [44].

To check the regulating mRNA's associated with lncRNA, all the genes were given to GENEMANIA software [43]. Three TSHR, TPO and TRIP13 were found to be in a functional network in terms of co-

expression. Also, TSHR, TRIP13 and THRAP3 are involved in physical interactions and it constitutes around 64% of the interactions. THRAP3 is involved with exon-exon junction complex and regulation of alternative mRNA splicing via spliceosome. TPO have the function like hormone metabolic process, thyroid hormone generation, thyroid hormone metabolic process, phenol containing metabolic process which can be more relevant to our study [20].

The data from the current study shows that expression of this altered lncRNAs could contribute to autoimmune thyroid disease therapeutics. To understand the functions of lncRNAs further, in the current study pathway analysis was used to associate these differentially expressed lncRNAs with their target genes and found that one pathway corresponded to transcripts; the most enriched network was autoimmune thyroid regulation composed of two targeted genes.

CONCLUSION

A total of 334 processed transcripts/lncRNA's were identified, only four of them ENST00000555326, ENST00000508456, ENST0000-0462973, ENST00000478853 corresponding to TSHR, TRIP13, TPO, THRAP3 had coding potential. These lncRNA transcripts ENST00000555326, ENST00000462973 were involved in autoimmune thyroid disease pathway. The data from the current study shows that expression of these lncRNAs could contribute to autoimmune thyroid disease. Our current study on the potential link between lncRNAs and autoimmune thyroid disease presents a novel area for further investigations into the target genes of such lncRNAs, leading to therapeutic strategies for the disease.

ACKNOWLEDGMENT

The authors thank the management of SRM University for providing the facilities to carry out this work.

CONFLICTS OF INTERESTS

There is no potential conflict of interest or competing interest.

REFERENCES

- Wilson BJ, Nicholls SG. The human genome project and recent advances in personalized genomics. *J Risk Management Healthcare Policy* 2015;8:9.
- Shendure J, Mitra RD, Varma C, Church GM. Advanced sequencing technologies: methods and goals. *Nat Rev Genet* 2004;5:335-44.
- Hernandez LM. editor. *Diffusion and Use of Genomic Innovations in Health and Medicine: Workshop Summary*. National Academies Press; 2008.
- De Vargas AF. The human genome project and its importance in clinical medicine. *Int Congr Ser* 2002;1237:3-13.
- National Human Genome Research Institute, USA, est; 1989. Available from: <http://www.genome.gov/12011238>. [Last accessed on 10 Jan 2016].
- Wellcome trusts Sanger institute. The human genome project. Available from: <https://www.sanger.ac.uk/about/history/hgp/>. [Last accessed on 10 Jan 2016].
- Louro R, El-Jundi T, Nakaya HI, Reis EM, Verjovski-Almeida S. Conserved tissue expression signatures of intronic noncoding RNAs transcribed from human and mouse loci. *Genomics* 2008;92:18-25.
- Wong DT. Salivary extracellular noncoding RNA: emerging biomarkers for molecular diagnostics. *Clin Ther* 2015;37:540-51.
- Koshimizu TA, Fujiwara Y, Sakai N, Shibata K, Tsuchiya H. Oxytocin stimulates expression of a noncoding RNA tumor marker in a human neuroblastoma cell line. *Life Sci* 2010;86:455-60.
- Szymanski M, Barciszewska MZ, Erdmann VA, Barciszewski J. A new frontier for molecular medicine: noncoding RNAs. *Biochim Biophys Acta (BBA)-Rev Cancer* 2005;1756:65-75.
- Storz G, Opdyke JA, Zhang A. Controlling mRNA stability and translation with small, noncoding RNAs. *Curr Opin Microbiol* 2004;7:140-4.
- Vogel J, Wagner EG. Target identification of small noncoding RNAs in bacteria. *Curr Opin Microbiol* 2007;10:262-70.
- Li C, Li Y, Shen L, Huang J, Sun Y, Luo Y, *et al.* The role of noncoding regions of classical swine fever virus C-strain in its adaptation to the rabbit. *Virus Res* 2014;183:117-22.
- Santosh B, Varshney A, Yadava PK. Non-coding RNAs: biological functions and applications. *Cell Biochem Function* 2015;33:14-22.
- Alvarez ML, DiStefano JK. The role of non-coding RNAs in diabetic nephropathy: potential applications as biomarkers for disease development and progression. *Diabetes Res Clin Pract* 2013;99:1-11.
- Fu M, Huang G, Zhang Z, Liu J, Huang Z, Yu B, *et al.* Expression profile of long noncoding RNAs in cartilage from knee osteoarthritis patients. *Osteoarthritis Cartilage* 2015;23:423-32.
- Beltrami C, Angelini TG, Emanuelli C. Noncoding RNAs in diabetes vascular complications. *J Mol Cell Cardiol* 2015;89:42-50.
- Lv J, Liu H, Huang Z, Su J, He H, Xiu Y, *et al.* Long non-coding RNA identification over mouse brain development by integrative modeling of chromatin and genomic features. *Nucleic Acids Res* 2013;41:10044-61.
- Sun J, Lin Y, Wu J. Long non-coding RNA expression profiling of mouse testis during postnatal development. *PloS One* 2013;8:750-7.
- Luo X, Pan J, Wang L, Wang P, Zhang M, Liu M, *et al.* Epigenetic regulation of lncRNA connects ubiquitin-proteasome system with infection-inflammation in preterm births and preterm premature rupture of membranes. *BMC Pregnancy Childbirth* 2015;15:21.
- Xin M, Wang Y, Yao Y, Song N, Hu Z, Qin D, *et al.* Identification and characterization of wheat long non-protein coding RNAs responsive to powdery mildew infection and heat stress by using microarray analysis and SBS sequencing. *BMC Plant Biol* 2011;11:61.
- Hao Y, Wu W, Shi F, Dalmolin RJ, Yan M, Tian F, *et al.* Prediction of long noncoding RNA functions with co-expression network in esophageal squamous cell carcinoma. *BMC Cancer* 2015;15:1.
- Isin M, Dalay N. LncRNAs and neoplasia. *Clin Chim Acta* 2015;444:280-8.
- Parolia A, Crea F, Xue H, Wang Y, Mo F, Ramnarine VR, *et al.* The long non-coding RNA PCGEM1 is regulated by androgen receptor activity *in vivo*. *Mol Cancer* 2015;14:1.
- Gao W, Zhu M, Wang H, Zhao S, Zhao D, Yang Y, *et al.* Association of polymorphisms in long non-coding RNA H19 with coronary artery disease risk in a Chinese population. *Mutat Res* 2015;772:15-22.
- Li Z, Zhao X, Zhou Y, Liu Y, Zhou Q, Ye H, Wang Y, *et al.* The long non-coding RNA HOTTIP promotes progression and gemcitabine resistance by regulating HOXA13 in pancreatic cancer. *J Transl Med* 2015;13:84.
- Zhao W, Mu Y, Ma L, Wang C, Tang Z, Yang S, *et al.* Systematic identification and characterization of long intergenic non-coding RNAs in fetal porcine skeletal muscle development. *Sci Rep* 2015;5. Doi:10.1038/srep08957. [Article in Press]
- Gibb EA, Brown CJ, Lam WL. The functional role of long non-coding RNA in human carcinomas. *Mol Cancer* 2011;10:1-38.
- Han D, Wang M, Ma N, Xu Y, Jiang Y, Gao X. Long noncoding RNAs: novel players in colorectal cancer. *Cancer Lett* 2015;361:13-21.
- Haemmerle M, Gutschner T. Long non-coding RNAs in cancer and development: where do we go from here? *Int J Mol Sci* 2015;16:1395-405.
- Villegas VE, Zaphiropoulos PG. Neighboring gene regulation by antisense long non-coding RNAs. *Int J Mol Sci* 2015;16:3251-66.
- Kornienko AE, Guenzl PM, Barlow DP, Pauler FM. Gene regulation by the act of long non-coding RNA transcription. *BMC Biol* 2013;11:1.
- Loewen G, Jayawickramarajah J, Zhuo Y, Shan B. Functions of lncRNA HOTAIR in lung cancer. *J Hematol Oncol* 2014;7:90.
- Zhang YC, Liao JY, Li ZY, Yu Y, Zhang JP, Li QF, *et al.* Genome-wide screening and functional analysis identifies a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Genome Biol* 2014;15:512.
- Xu G, Chen J, Pan Q, Huang K, Pan J, Zhang W, *et al.* Long noncoding RNA expression profiles of lung adenocarcinoma ascertained by microarray analysis. *PloS One* 2014;9:e104044. Doi:10.1371/journal.pone.0104044. [Article in Press]
- Dong R, Jia D, Xue P, Cui X, Li K, Zheng S, *et al.* Genome-wide analysis of long noncoding RNA (lncRNA) expression in hepatoblastoma tissues. *PloS One* 2014;9:e85599. Doi:10.1371/journal.pone.0085599. [Article in Press]
- Pearce SH, Merriman TR. Genetics of type 1 diabetes and autoimmune thyroid disease. *Endocrinol Metab Clin North Am* 2009;38:289-301.
- Taylor JC, Gough SC, Hunt PJ, Brix TH, Chatterjee K, Connell JM, Franklyn JA, *et al.* A genome-wide screen in 1119 relative pairs with autoimmune thyroid disease. *J Clin Endocrinol Metab* 2006;91:646-53.
- Ban Y, Ban Y, Ban Y. Autoimmune thyroid disease genes identified in non-caucasians. *Open J Endocr Metab Dis* 2012;2:107-16.
- Shiva S, Ilkhchooyi F, Rezamand A. Thyroid autoimmunity at the onset of type 1 diabetes mellitus in children. *Open J Immunol* 2013;3:37-40.
- Gesing A, Lewiński A, Karbownik-Lewińska M. The thyroid gland and the process of aging; what is new? *Thyroid Res* 2012;5:1-5.
- Bauer-Mehren A, Bundschus M, Rautschka M, Mayer MA, Sanz F, Furlong LI. Gene-disease network analysis reveals functional modules in mendelian, complex and environmental diseases. *PloS One* 2011;6:e20284. Doi:10.1371/journal.pone.0020284. [Article in Press]
- Montoyo J, Zuberi K, Rodriguez H, Kazi F, Wright G, Donaldson SL, *et al.* GeneMANIA cytoscape plugin: fast gene function predictions on the desktop. *Bioinformatics* 2010;26:2927-8.

44. Bionaz M. Nutrigenomics approaches to fine-tune metabolism and milk production: is this the future of ruminant nutrition?

Adv Dairy Res 2014;2:e107. Doi:10.4172/2329-888X.1000e107. [Article in Press].