Proc. Assoc. Advmt. Breed. Genet. 21: 485-488

POTENTIAL ROLE OF lncRNA CYP2C91-PROTEIN INTERACTIONS ON IMMUNE DISEASES AND OBESITY

P. Suravajhala[1], L.J.A. Kogelman[1], G. Mazzoni[1] and H.N. Kadarmideen[1]

[1] Department of Veterinary Clinical and Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Grønnegårdsvej 7, 1870 Frederiksberg, Denmark

SUMMARY

With unprecedented increase in next generation sequencing (NGS) technologies, there has been a persistent interest on transcript profiles of long noncoding RNAs (lncRNAs) and protein-coding genes forming an interaction network. Apart from protein-protein interaction (PPI), gene interaction models such as Weighted Gene Co-expression Network Analysis are used to functionally annotate lncRNAs in identifying their potential disease associations. To address this, studies have led to characterizing transcript structures and understanding expression profiles mediating regulatory roles. In the current analysis, we show how a lncRNA- cyp2c91 contributes to the transcriptional regulation localized to cytoplasm thereby making refractory environment for transcription. By applying co-expression network methods and pathway analyses on genes related to a disease such as obesity from F2 pig model, we show that we can gain deeper insight in biological processes such as the perturbances in immune system, and get a better understanding of the systems biology of diseases. We believe this study has implications for finding prognostic and diagnostic markers for obesity and immune related diseases.

INTRODUCTION

With unprecedented increase in next generation sequencing (NGS) technologies, there has been a persistent interest on transcript profiles of long noncoding RNAs (lncRNAs) and protein-coding genes forming an interaction network. Apart from protein-protein interaction (PPI), gene interaction models such as Weighted Gene Co-expression Network Analysis (WGCNA; Xue *et al.,* 2013) are used to functionally annotate lncRNAs in identifying their potential disease associations (Cogill and Wang, 2014). To address this, studies have led to characterizing transcript structures and understanding expression profiles mediating regulatory roles and comparing them with the ENCODE project (The ENCODE project. 2014). Recent reports show how lncRNAs contribute towards regulatory interactions with their non-coding peers like miRNAs (Jalali *et al*., 2013). Whether or not lncRNA-protein networks restrain interactions is little known and not detailed. How such regulatory interactions between classes of lncRNAs and proteins would have a significant influence on the organism remains a challenge.

Earlier, we have shown three regulatory genes, viz. CCR1, MSR1 and SPI1 associated with diseases like obesity and osteoporosis using gene network algorithms WGCNA and Lemon-Tree (Kogelman *et al.,* 2014a) applied to NGS-based RNAseq datasets from porcine model for obesity. These clusters of highly co-expressed genes were ranked as highly significant based on their association with obesity-related phenotypes in a F2 pig model (Kogelman *et al* 2014b). With a wide range of biological processes effectively used as regulatory molecules, we anticipate (a) if the coexpressed genes have interacting partners with any long noncoding RNAs (lncRNA), (b) if so, whether or not they affect the coexpression, consequently further changing the networks and influencing the organismal phenotype or disease outcomes, or (c) if not, what would be the outcome of such lncRNA-dependent transcription. From a putative interaction network, we have established functional classes based on several different methods, explicitly focusing on the edge-betweeness, pearson coefficient of overlapping genes, two nearest non-overlapping genes on either side, presence of subcellular location signals (not shown). These resilient methods would distinguish probability of lncRNA to show association/disassociation paradigm, RNA binding

protein-lncRNA interactivity and importantly disease association, if any.

MATERIALS AND METHODS

In the current study, we made a human concordant network from our previous WGCNA result from an animal model (*see Figure 1;* F2 pigs, Kogelman *et al.,* 2014b) and found that 340 of 540 porcine genes have orthologue peers in humans (Figure 2, panel a). The absence of orthologs in human is in agreement with the homology data available from the Pig Analysis Database (PAD) which specifies that about 73% of the sequences are covered by the both genomes (See PAD web reference). From the networks and GenBank annotation, we observed that cyp2c91, a lncRNA interacts with a host of regulatory genes. The betweenness centrality of cyp2c91 with the three regulator genes linked to obesity (CCR1, MSR1 and SPI1) was found to be consistent with the association pattern (Figure 2, panel b). With the hypothesis that lncRNA-protein interactions play an important role in regulating post-transcriptional changes and subsequent localization of the transcript, we used RNA-protein interaction predictor (RPI-pred) to predict whether or not the proteins encoded by these genes and the RNA form interaction pairs (Suresh *et al.* 2015). Considering the fact that these small molecules enter the nucleus without regulation, we asked if any gene products are localized extracellular to nucleus.

RESULTS AND DISCUSSION

We observed that among the three regulator genes, CCR1 was found to be localized in cytoplasm (Figure 2, panel c). Encouraged by the outcome that the three have a plausible role of interaction with cyp2c91, we made a reliable interaction network with the mean disassociation based on the betweeness centrality (Figure 2, panel d). We found that MSR1 and CCR1 are found to be interacting with each other while SPI1 was a lone gene without an interaction pair. Nonetheless, the lncRNA-protein interactions were extended with the CCR1-cyp2c91 association mapped from network genes. The study suggests two ways forward. First, the fold change (log 2 expression) can be attributed to lncRNA-dependent transcription. Second, CCR1-cyp2c91 association is significant when compared to MSR1-cyp2c91 and SPI1-cyp2c91 (indicative of p-values, not shown) where the genes are regulatory in nature forming diseased network. The three regulatory genes are associated with obesity and immune system, possibly linking them to Lupus. This is evident by the fact that several of the genes present in the WGCNA modules of Kogelman et al., 2014 (TNIP1, GPSM3, TFEC, TES, KCP, IRF5, TNPO3, ELF1, ITGAM and TNXB, KLF6, AKR1E2) are related to immune system and systemic lupus erythematosus (SLE). This might allow us to use this network as a model for immune response or obesity.

The genome is lengthily transcribed in eukaryotes and it has been known that many transcripts have larger proportion of noncoding components. Although about 66-73% of the porcine genome (including ESTs, genes etc.) is conserved across humans, a considerable set of genes regulate interactions with lncRNAs. Further, a range of transcribed regions might tend to be regulatory and indicative of enhancing non-functional activity. Moving to a broader spectrum of calling them as junk, we asked for evidences on their regulatory potential based on their association with protein-coding genes. Consistent with the interaction networks from porcine model for obesity, subcellular localization of the products of the three protein-coding genes revealed that two are nuclear while one, CCR1 was found to be in cytoplasm. This is again in agreement with the fact that the subcellular fractions of lncRNA differ significantly from each other, with a majority enriched in the nucleus, cytoplasm and ribosomes. These results show that lncRNA-protein interactions are self-regulating and yet they are dependent on organellar specificity. Our exploratory analysis suggests that CCR1-cyp2c91 contributes to the transcriptional regulation localized to cytoplasm thereby making refractory environment for transcription. By applying co-expression network methods and pathway analyses on genes related to a disease such as obesity and systemic lupus

erythamatosis, we show that we can gain deeper insight in biological processes such as the perturbances in immune system, and get a better understanding of the systems biology of diseases. This stresses the possible need of finding genes linked to lncRNA-protein networks and further use them as potential diagnostic markers in animal and human diseases.

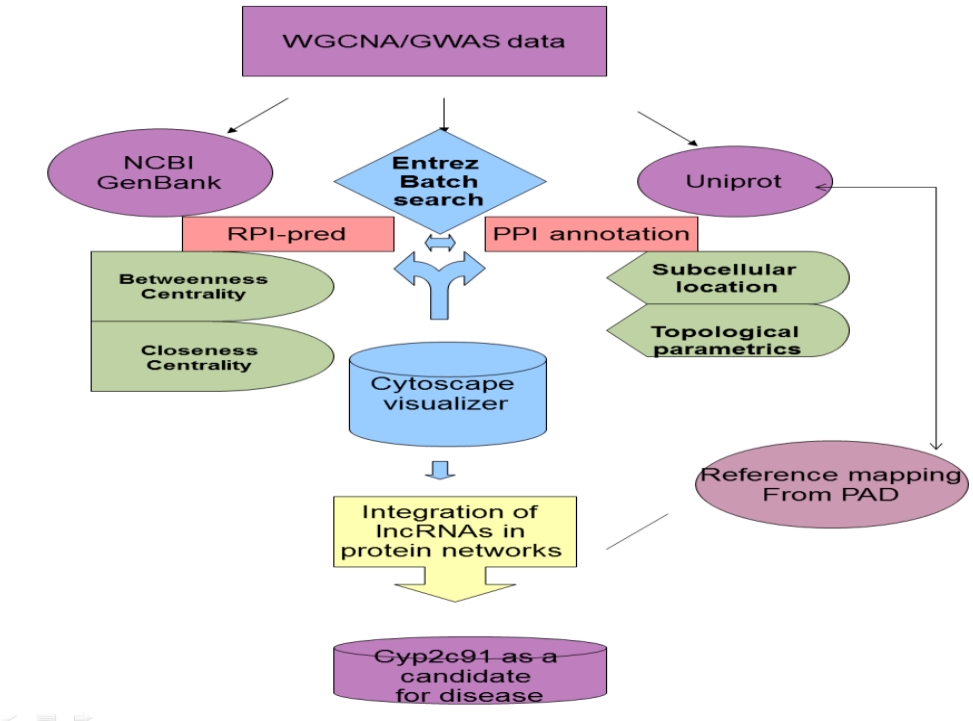

Figure 1: The GWAS data from Kogelman et al. were checked for candidacy across the GenBank. After obtaining the *bona fide* accessions in humans, the sequences were checked using RPI-Pred and protein annotation. The betweenness centrality and closeness centrality values for the nodes were then computed and visualized using Cytoscape. The centrality values are computed for those that do not contain multiple edges. They are the normalized values for each gene/node by dividing the number of pairs of nodes existing in the network. The range would be between 0 to 1 with the condensed values in exponential form as calculated by cytoscape (centrality of vertex). We considered the candidate lncRNA linked to disease after reference mapping and linkage to disease while integrating it into the protein interaction network.

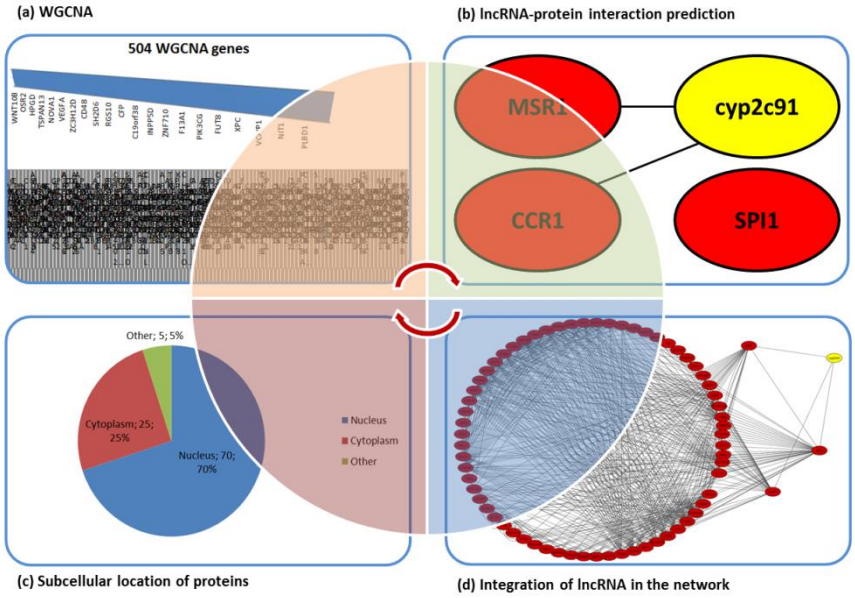

Figure 2: (a) The 504 genes from WGCNA across different modules linked to diseases not limited to obesity and immune response. (b) Representative lncRNA cyp2c91 gene (in yellow) shown to be interacting with three regulatory protein-coding genes (c) Subcellar location of the genes associated in the network and (d) the network topology showing the profiled expression across the regulatory genes associated with cyp2c91. This is indicative of global protein-RNA interaction data.

REFERENCES

Cogill S.B., Wang L. (2014). *Cancer Inform*. ***13***(Suppl 5):49-59.

Jalali S., Bhartiya D., Lalwani M.K., Sivasubbu S., Scaria V. (2013). *PLoS ONE*. ***8***(2): p. e53823.

Kogelman L.J.A., Cirera S., Zhernakova D.V., Fredholm M., Franke L, Kadarmideen H.N. (2014). *BMC Med Genomics*. ***7*:**57.

Kogelman, L.J.A., Pant S.D., Fredholm M., Kadarmideen H.N. 2014. *Frontiers in Genetics (Systems biology)*, ***5***:214.

PAD: http://www.piggenome.dk/

Suresh V., Liu L., Adjeroh D., Zhou X. (2015). *Nucleic Acids Res*. ***43***(3):1370-9.

The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science*, 2004. ***306***(5696): p. 636-40.

van Heesch S, van Iterson M, Jacobi J, Boymans S, Essers PB, de Bruijn E et al. (2014). *Genome Biol.* **15**(1):R6.

Xue Z., Huang K., Cai C., Cai L., Jiang C.Y., Feng Y. et al. (2013). *Nature*. ***29***;500(7464):593-7.