Bioinformatics, 32(10), 2016, 1595–1597
doi: 10.1093/bioinformatics/btw017
Advance Access Publication Date: 18 January 2016
Applications Note

OXFORD

Databases and ontologies

lncCaNet: pan-cancer co-expression network for human lncRNA and cancer genes

Yining Liu and Min Zhao*

Faculty of Science, Health, Education and Engineering, University of the Sunshine Coast, Maroochydore DC, QLD 4558, Australia

*To whom correspondence should be addressed.

Associate Editor: Ivo Hofacker

Received on 10 September 2015; revised on 20 December 2015; accepted on 12 January 2016

Abstract

Summary: Thousands of human long non-coding RNAs (lncRNAs) have been identified in cancers and played important roles in a wide range of tumorigenesis. However, the functions of vast majority of human lncRNAs are still elusive. Emerging studies revealed that the expression level of majority lncRNAs shows discordant expression pattern with their protein-coding gene neighbors in various model organisms. Therefore, it may be useful to infer lncRNAs' potential biological function in cancer development by more comprehensive functional views of co-expressed cancer genes beyond mere physical proximity of genes. To this aim, we performed thorough searches and analyses of the interactions between lncRNA and non-neighboring cancer genes and provide a comprehensive co-expression data resource, LnCaNet. In current version, LnCaNet contains the pre-computed 8 494 907 significant co-expression pairs of 9641 lncRNAs and 2544 well-classified cancer genes in 2922 matched TCGA samples. In detail, we integrated 10 cancer gene lists from public database and calculate the co-expression with all the lncRNAs in 11 TCGA cancer types separately. Based on the resulted 110 co-expression networks, we identified 17 common regulatory pairs related to extracellular space shared in 11 cancers. We expect LnCaNet will enable researcher to explore lncRNA expression pattern, their affected cancer genes and pathways, biological significance in the context of specific cancer types and other useful annotation related to particular kind of lncRNA-cancer gene interaction.

Availability and implementation: http://lncanet.bioinfo-minzhao.org/

Contact: m.zhao@uq.edu.au

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Long non-coding RNA (lncRNA) refers as non-coding transcripts longer than 200 bases and are abundant in various normal and cancer tissues (Kung *et al.*, 2013). Rather than being regarded as 'transcriptional noise', there is emerging evidence that lncRNAs play important roles in regulation of gene expression through influencing chromatin modification, transcriptional complexes targeting, mRNA splicing, protein translation (Rinn and Chang, 2012; Roberts *et al.*, 2014). However, most lncRNAs are poorly annotated, and their functions including their roles in complex diseases have not been widely studies.

2 Methods

To bring a more functional information of human lncRNAs in cancer, we first collected 10 cancer gene lists such as experimentally characterized tumor suppressors (Zhao *et al.*, 2013).

To provide the co-expression pattern of cancer genes with lncRNAs, we downloaded all the lncRNAs expression data from Mitranscriptome database (Iyer *et al.*, 2015). We estimated the expression correlation among all the 2544 cancer genes and 17 250 lncRNA transcripts from Mitranscriptome using the Spearman's correlation based on matched TCGA cancer samples. Based on the R language package (version 2.14.0), the expression correlation scores and corresponding *P*-values

were calculate. For all the *P*-values, a false discovery rate (FDR) was applied to correct the statistical significance of multiple testing. Finally, 8 494 907 pairs with 9641 lncRNAs can satisfy our criteria that the absolute expression correlation scores are greater than 0.3 and the FDR adjusted *P*-values are less than 0.01. In addition to gene-lncRNA co-expression, we also computed lncRNA-lncRNA and gene-gene co-expression pairs based on same TCGA cancer samples for broader data mining about collected cancer genes and lncRNAs in pan-cancer level.

3 Results

As shown in Figure 1A and B, a typical cancer gene entry contains seven categories information by clicking the label 'General information,' 'Expression,' 'lncRNA,' 'Regulation,' 'Mutation,' 'Homolog,' and 'Interaction' on the top. For the pre-computed co-expression pattern between cancer genes and lncRNAs, user can obtain the correlation coefficients and the corrected statistical *P*-values (Fig. 1C). In addition, the expression plots of the corresponding lncRNAs were also integrated from MiTranscriptome database. All the 110 co-expression networks are available to browse and download separately for advanced integrative study.

Based on the pre-computed results of co-expression pairs in 11 cancers, we found thousands of co-expression pairs are shared in cancers (Fig. 1E). For example, 17 pairs associated with 17 lncRNAs and 16 cancer genes are common in all 11 cancers. The gene CD24 has been co-expressed with 2 lncRNAs across 11 cancers. Based on the functional analysis, we found 6 cancer genes are mainly located extracellular space related to cell adhesion (Corrected enrichment $P = 0.0335$, Supplementary Table S1), and 4 of them are clusters of differentiation

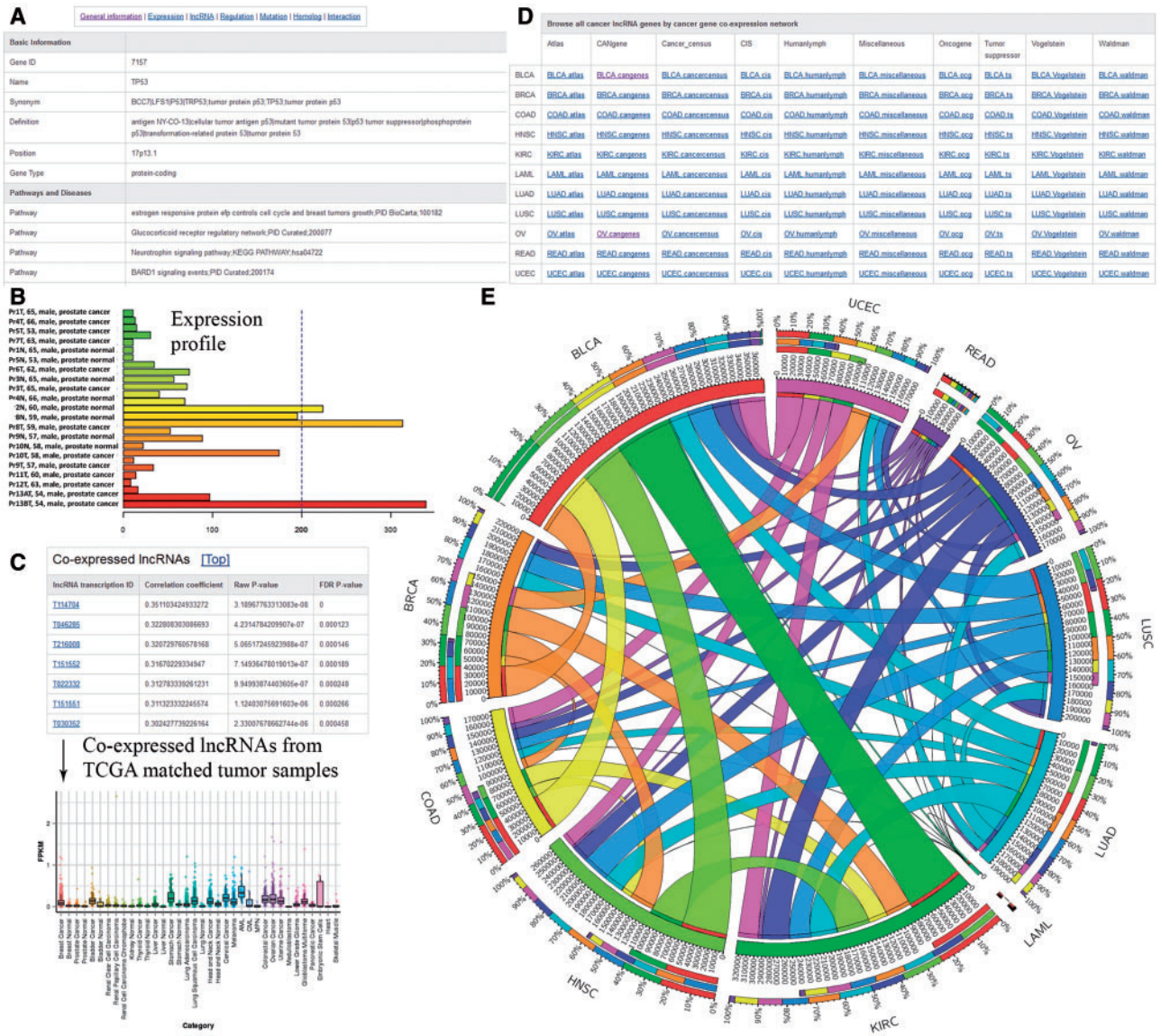

Fig. 1. (**A**–**C**) The gene information of InCaNet, including the basic information, gene expression, and pre-computed lncRNA co-expression result. (**D**) The web browser for all the 110 co-expression network. (**E**) The shared cancer genes-lncRNA pairs across 11 cancer types. The length of circularly arranged segments is proportional to the total genes in each data source. The ribbons connecting different segments represent the number of shared genes between data sources. The three outer rings are stacked bar plots that represent relative contribution of other data source to the data source's totals. BLCA (bladder urothelial carcinoma carcinoma), BRCA (breast invasive carcinoma), COAD (colon adenocarcinoma), HNSC (head and neck squamous cell carcinoma), KIRC (kidney renal clear cell carcinoma), LUAD (lung adenocarcinoma), LAML (acute myeloid leukemia), LUSC (lung squamous cell carcinoma), OV (high-grade serous ovarian cancer), READ (rectum adenocar-cinoma) and UCEC (uterine corpus endometrial carcinoma)

(CDs) which deserve the further experimental validation on the relationship of CD molecules and lncRNA.

4 Conclusion

The pan-cancer analysis of the co-expression revealed that the lncRNAs are frequently co-expressed with cancer genes in multiple cancers. To provide a reusable resource for cancer community, we published all the pre-computed networks about lncRNA and cancer genes at http://lncanet.bioinfo-minzhao.org.

Conflict of Interest: none declared.

References

Iyer,M.K. *et al*. (2015) The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet*., **47**, 199–208.

Kung,J.T. *et al*. (2013) Long noncoding RNAs: past, present, and future. *Genetics*, **193**, 651–669.

Rinn,J.L. and Chang,H.Y. (2012) Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem*., **81**, 145–166.

Roberts,T.C. *et al*. (2014) Perspectives on the mechanism of transcriptional regulation by long non-coding RNAs. *Epigenetics*, **9**, 13–20.

Zhao,M. *et al*. (2013) TSGene: a web resource for tumor suppressor genes. *Nucleic Acids Res*., **41**(Database issue), D970–D976.