IPCARF

IPCARF: improving lncRNA-disease association prediction using incremental principal component analysis feature selection and a random forest classifier

Related work:

1. Evolution and Functions of Long Noncoding RNAs

(https://www.cell.com/cell/fulltext/S0092-8674(09)00142-1?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0092867409001421%3Fshowall%3Dtrue)

- This paper published in: **2009**
- RNA is not only a messenger operating between DNA and protein. Transcription of essentially the entire eukaryotic genome generates a myriad of non-protein-coding RNA species that show complex overlapping patterns of expression and regulation. Although long noncoding RNAs (lncRNAs) are among the least well-understood of these transcript species, they cannot all be dismissed as merely transcriptional "noise." Here, we review the evolution of lncRNAs and their roles in transcriptional regulation, epigenetic gene regulation, and disease.

2. NONCODE v3.0: integrative annotation of long noncoding RNAs

(https://academic.oup.com/nar/article/40/D1/D210/2903318)

- This paper published in **: 30 November 2011**
- In this study, we have updated the NONCODE database (http://www.noncode.org) to version 3.0 to include the first integrated collection of expression and functional lncRNA data obtained from re-annotated microarray studies in a single database. NONCODE has a user-friendly interface with a variety of search or browse options, a local Genome Browser for visualization and a BLAST server for sequence-alignment search. In addition, NONCODE provides a platform for the ongoing collation of lncRNAs reported

in the literature. All data in NONCODE are open to users, and can be downloaded through the website or obtained through the SOAP API and DAS services.

3. **Novel human lncRNA–disease association inference based on lncRNA expression profiles**

   (https://academic.oup.com/bioinformatics/article/29/20/2617/276977)

   - This paper published in: **15 October 2013**

   - In this study, we proposed the assumption that similar diseases tend to be associated with functionally similar lncRNAs. Then, we further developed the method of Laplacian Regularized Least Squares for LncRNA–Disease Association (LRLSLDA) in the semi supervised learning framework. Although known disease–lncRNA associations in the database are rare, LRLSLDA still obtained an AUC of 0.7760 in the leave-one-out cross validation, significantly improving the performance of previous methods. We also illustrated the performance of LRLSLDA is not sensitive (even robust) to the parameters selection and it can obtain a reliable performance in all the test classes. Plenty of potential disease–lncRNA associations were publicly released and some of them have been confirmed by recent results in biological experiments. It is anticipated that LRLSLDA could be an effective and important biological tool for biomedical research.

4. **Inferring novel lncRNA–disease associations based on a random walk model of a lncRNA functional similarity network**
   (https://pubs.rsc.org/en/content/articlelanding/2014/MB/C3MB70608G#!divAbstract)

   - This paper published in: **07 Apr 2014**

   - In this study, we proposed a global network-based computational framework, RWRlncD, to infer potential human lncRNA–disease associations by implementing the random walk with restart method on a lncRNA functional similarity network. The performance of RWRlncD was evaluated by experimentally verified lncRNA–disease associations, based on leave-one-out cross-validation. We achieved an area under the ROC curve of 0.822, demonstrating the excellent performance of RWRlncD. Significantly, the performance of RWRlncD is robust to different parameter

selections. Predictively highly-ranked lncRNA–disease associations in case studies of prostate cancer and Alzheimer's disease were manually confirmed by literature mining, providing evidence of the good performance and potential value of the RWRlncD method in predicting lncRNA–disease associations.

5. **Predicting lncRNA-disease associations and constructing lncRNA functional similarity network based on the information of miRNA (https://www.nature.com/articles/srep13186)**

- This paper Published in : **17 August 2015**
- In this study, a novel model of HyperGeometric distribution for LncRNA-Disease Association inference (HGLDA) was developed to predict lncRNA-disease associations by integrating miRNA-disease associations and lncRNA-miRNA interactions. Although HGLDA didn't rely on any known disease-lncRNA associations, it still obtained an AUC of 0.7621 in the leave-one-out cross validation. Furthermore, 19 predicted associations for breast cancer, lung cancer and colorectal cancer were verified by biological experimental studies. Furthermore, the model of LncRNA Functional Similarity Calculation based on the information of MiRNA (LFSCM) was developed to calculate lncRNA functional similarity on a large scale by integrating disease semantic similarity, miRNA-disease associations and miRNA-lncRNA interactions. It is anticipated that HGLDA and LFSCM could be effective biological tools for biomedical research.

6. **Lnc2Cancer: a manually curated database of experimentally supported lncRNAs associated with various human cancers**

(https://academic.oup.com/nar/article/44/D1/D980/2502607)

- This paper published in: **4 January 2016**
- Lnc2Cancer (http://www.bio-bigdata.net/lnc2cancer) is a manually curated database of cancer-associated long non-coding RNAs (lncRNAs) with experimental support that aims to provide a high-quality and integrated resource for exploring lncRNA deregulation in various human cancers. LncRNAs represent a large category of functional RNA molecules that play a significant role in human cancers. A curated collection and summary of deregulated lncRNAs in cancer is essential to thoroughly understand the

mechanisms and functions of lncRNAs. Here, we developed the Lnc2Cancer database, which contains 1057 manually curated associations between 531 lncRNAs and 86 human cancers. Each association includes lncRNA and cancer name, the lncRNA expression pattern, experimental techniques, a brief functional description, the original reference and additional annotation information. Lnc2Cancer provides a user-friendly interface to conveniently browse, retrieve and download data. Lnc2Cancer also offers a submission page for researchers to submit newly validated lncRNA-cancer associations. With the rapidly increasing interest in lncRNAs, Lnc2Cancer will significantly improve our understanding of lncRNA deregulation in cancer and has the potential to be a timely and valuable resource.

7. Long non-coding RNA expression in bladder cancer

(https://link.springer.com/article/10.1007/s12551-017-0379-y)

- This paper published in: **08 December 2017**

- The advent of novel high-throughput sequencing methods has facilitated identification of non-coding RNAs with fundamental roles in cellular biological and pathological conditions. A group of these consisting of at least 200 nucleotides are called long non-coding RNAs (lncRNAs). Their participation in the pathogenesis of cancer has been highlighted in recent years. Bladder cancer, one of the most prevalent cancers worldwide, exhibits altered expression levels of several lncRNAs. Several in vitro and in vivo studies have assessed the effects of silencing RNAs on cancer cell phenotypes and in vivo tumor growth. For instance, in vitro studies have shown that nuclear paraspeckle assembly transcript 1 (NEAT1), promoter of CDKN1A antisense DNA damage-activated RNA(PANDAR) and metastasis-associated lung adenocarcinoma transcript 1(MALAT1) have oncogenic effects while Maternally expressed 3 (MEG3) and BRAF activated non-coding RNA (BANCR) are tumor suppressors. Analysis of these data will help to identify a panel of lncRNAs that can be potentially used for both early detection and prognosis in bladder cancer patients. Here, we review the roles of several lncRNAs in the on cogenesis, tumor suppression, early detection, and prognosis of bladder cancer.

8. **TPGLDA: Novel prediction of associations between lncRNAs and diseases via lncRNA-disease-gene tripartite graph**

   (https://www.nature.com/articles/s41598-018-19357-3)

   - This paper published in: **18 January 2018**

   - In this study, we propose a novel prediction of lncRNA-disease associations via lncRNA-disease-gene tripartite graph (TPGLDA), which integrates gene-disease associations with lncRNA-disease associations. Compared to previous studies, TPGLDA can be used to better delineate the heterogeneity of coding-non-coding genes-disease association and can effectively identify potential lncRNA-disease associations. After implementing the leave-one-out cross validation, TPGLDA achieves an AUC value of 93.9% which demonstrates its good predictive performance. Moreover, the top 5 predicted rankings of lung cancer, hepatocellular carcinoma and ovarian cancer are manually confirmed by different relevant databases and literatures, affording convincing evidence of the good performance as well as potential value of TPGLDA in identifying potential lncRNA-disease associations.

9. **Long non-coding RNAs: Functional regulatory players in breast cancer**

   (https://www.sciencedirect.com/science/article/pii/S2468054018300842?via%3Dihub **)**

   - This paper published in: **March 2019**

   - In this study, a specific focus on lncRNAs as prognostic markers among BC patients showing molecular subtype heterogeneity was also tackled in this review. Finally, the functional and the mechanistic roles of such booming ncRNA molecules in shaping the fate of the BC progression have been highlighted.

10. **DBS: a fast and informative segmentation algorithm for DNA copy number analysis**

    (https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-018-2565-8)

    - This paper published in: **2019**

    - DBS is implemented in a platform-independent and open-source Java application (ToolSeg), including a graphical user interface and simulation data generation, as well as various segmentation methods in the native Java language.

11. **Predicting MicroRNA-Disease Associations Using Kronecker Regularized Least Squares Based on Heterogeneous Omics Data** (https://ieeexplore.ieee.org/document/7862218)
    - This paper published in: **2020**

    - In this study, we present a semi-supervised method of Kronecker regularized least squares for predicting the potential or missing miRNA-disease associations (KRLSM). KRLSM integrates different omics data to assist various diseases or miRNAs with sparsely known associations to make predictions, and combines the disease space and miRNA space into a whole miRNA-disease space by Kronecker product. Finally, the semi-supervised classifier of regularized least squares is adopted to identify disease-related miRNAs. The experiment results demonstrate that the proposed method outperforms the other state-of-the-art approaches. In addition, case studies of several common diseases further indicate the effectiveness of KRLSM to identify potential miRNA-disease associations.