EMBnet.journal
Bioinformatics in Action

Barracchia et al. (2021) EMBnet.journal 26(Suppl A), e955
http://dx.doi.org/10.14806/ej.26.A.955

LP-HCLUS: a novel tool for the prediction of relationships between ncRNAs and human diseases

Emanuele Pio Barracchia[1,2]✉, Gianvito Pio[1,2], Domenica D'Elia[3], Michelangelo Ceci[1,2,4]

[1] Department of Computer Science, University of Bari Aldo Moro, Bari, Italy
[2] National Interuniversity Consortium for Informatics (CINI), Rome, Italy
[3] Institute for Biomedical Technologies, National Research Council, Bari, Italy
[4] Department of Knowledge Technologies, Jozef Stefan Institute, Ljubljana, Slovenia

Competing interests: EPB none; GP none; DD none; MC none

The discovery of a functional relationship between human diseases and non-coding RNAs (ncRNAs) is not new. In the last decade, it improved the elucidation of many diseases' mechanisms and the improvement of therapeutic approaches (Lekka and Hall, 2018; Wang *et al.*, 2016; Yang *et al.*, 2014). Nevertheless, the function of many ncRNAs is still unclear or completely unknown, and therefore, their role in human diseases is difficult, if not impossible, to be identified. We have developed a new system, called LP-HCLUS, that is able to predict previously unknown disease-ncRNA associations by exploiting multi-type hierarchical clustering techniques.

Differently from other approaches, LP-HCLUS is able to analyse and benefit from heterogeneous networks of interactions/relationships among multiple types of entities (*e.g.*, diseases, ncRNAs, target genes) and relationships between them. To this aim, the proposed method first estimates the strength of the disease-ncRNA associations, exploiting both direct and indirect relationships. It constructs a hierarchy of heterogeneous clusters based on known and estimated relationships between diseases and ncRNAs. Finally, LP-HCLUS uses the generated clusters to induce new relationships, associating each of them with a certainty score. We conducted several experiments, comparing the performances achieved by LP-HCLUS with those obtained by two different competitors: HOCCLUS2 (Pio *et al.*, 2013) and ncPred (Alaimo *et al.*, 2014). In particular, we analysed two different datasets: HMDD v3.0, which contains data about relationships between diseases and miRNAs, and a dataset constructed integrating different

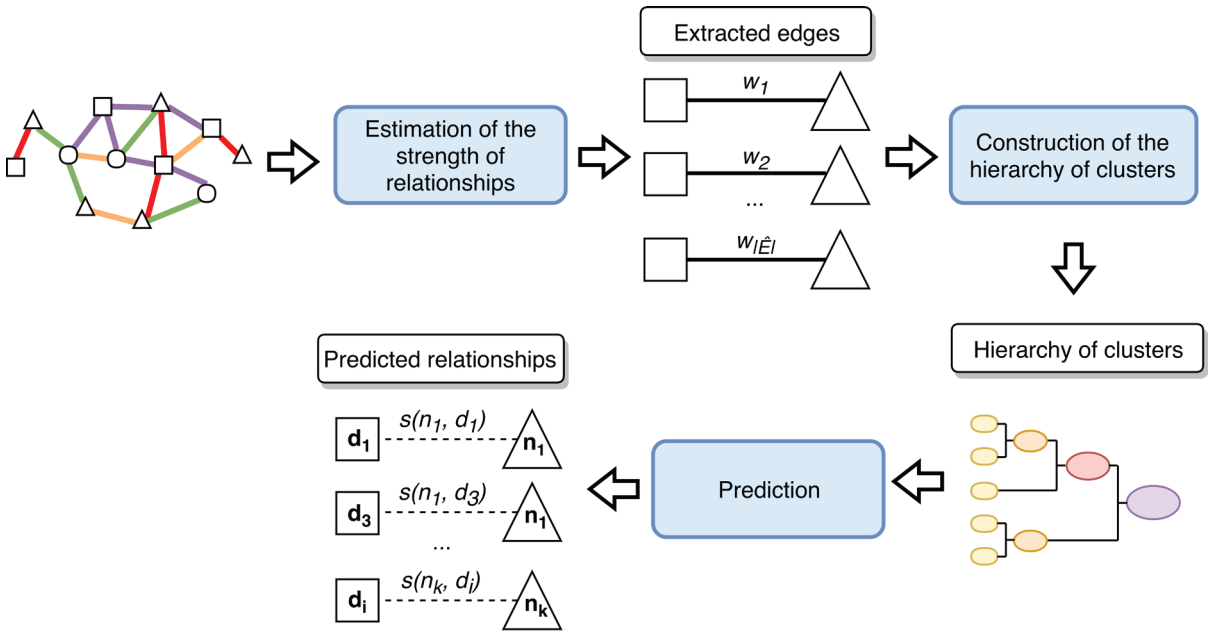

Figure 1. The workflow of the LP-HCLUS method.

state-of-the-art data sources (Chen *et al.*, 2013; Helwak *et al.*, 2013; Bauer-Mehren *et al.*, 2010; Jiang *et al.*, 2009).

The results show that our system is able to outperform its competitors, and it can help biologists to conduct more focused research. Such a conclusion is also confirmed by a qualitative analysis conducted on the predicted associations that showed that many associations predicted by LP-HCLUS with a high certainty score have been subsequently validated and introduced in a more recent version of HMDD dataset (v3.2). The importance of such a development is also in its easy transfer for applications in any biological study involving heterogeneous data from different sources and types (*e.g.*, different omics data, chemicals, biochemical and structural data, *etc.*).

Availability of data and materials

The system LP-HCLUS, the adopted datasets and all the results are available at: http://www.di.uniba.it/~gianvitopio/systems/lphclus/

Acknowledgements

We would like to acknowledge the financial support of the European Commission through the project MAESTRA - Learning from Massive, Incompletely annotated, and Structured Data (Grant Number ICT-2013-612944). We also acknowledge the financial support of the Ministry of Education, Universities and Research (MIUR) through the PON projects “Big Data Analytics” (AIM1852414 - Activity 1, Line 1) and TALIsMAn - Tecnologie di Assistenza personALizzata per il Miglioramento della quAlità della vitA (Grant N. ARS01_0111), and of Italian National Research Council (CNR) through the InterOmics Flagship project.

References

1. Alaimo S, Giugno R, Pulvirenti A (2014) ncPred: ncRNA-Disease Association Prediction through Tripartite Network-Based Inference. Frontiers in Bioengineering and Biotechnology **2**. http://dx.doi.org/10.3389/fbioe.2014.00071
2. Bauer-Mehren A, Rautschka M, Sanz F, Furlong LI (2010) DisGeNET: a Cytoscape plugin to visualize, integrate, search and analyze gene-disease networks. Bioinformatics **26** (22):2924–2926. http://dx.doi.org/10.1093/bioinformatics/btq538
3. Chen G, Wang Z, Wang D, Qiu C, Liu M *et al.* (2013) LncRNADisease: a database for long-non-coding RNA-associated diseases. Nucleic Acids Research **41** (Database issue):983–986. http://dx.doi.org/10.1093/nar/gks1099
4. Helwak A, Kudla G, Dudnakova T, Tollervey D (2013) Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Cell **153** (3):654–665. http://dx.doi.org/10.1016/j.cell.2013.03.043
5. Jiang Q, Wang Y, Hao Y, Juan L, Teng M *et al.* (2009) miR2disease: a manually curated database for microRNA deregulation in human disease. Nucleic Acids Research **37** (Database issue):98-104. http://dx.doi.org/10.1093/nar/gkn714
6. Lekka E, Hall J (2018) Noncoding RNAs in disease. FEBS Letters **592** (17):2884-2900. http://dx.doi.org/10.1002/1873-3468.13182
7. Pio G, Ceci M, D'Elia D, Loglisci C, Malerba D (2013) A Novel Biclustering Algorithm for the Discovery of Meaningful Biological Correlations between microRNAs and their Target Genes. BMC Bioinformatics **14** (Suppl 7):8. http://dx.doi.org/10.1186/1471-2105-14-S7-S8
8. Wang P, Guo Q, Gao Y, Zhi H, Zhang Y *et al.* (2016) Improved method for prioritization of disease associated lncRNAs based on ceRNA theory and functional genomics data. Oncotarget **8** (3):4642–4655. http://dx.doi.org/10.18632/oncotarget.13964
9. Yang X, Gao L, Guo X, Shi X, Wu H *et al.* (2014) A Network Based Method for Analysis of lncRNA-Disease Associations and Prediction of lncRNAs Implicated in Diseases. Plos One **9** (1):e87797. http://dx.doi.org/10.1371/journal.pone.0087797