

Inference and Differential Analysis of Extended Core Networks: a way to study Anti-Sense Regulation

Marc Legeay^{*†}, Béatrice Duval^{*} and Jean-Pierre Renou[†]

^{*}LERIA, Université d'Angers 2 boulevard Lavoisier 49045 ANGERS Cedex 01, FRANCE

[†]IRHS, INRA, AGROCAMPUS-Ouest, Université d'Angers, SFR 4207 QUASAV

42 rue Georges Morel 49071 Beaucouzé cedex, FRANCE

Email: marc.legeay@univ-angers.fr, beatrice.duval@univ-angers.fr, jean-pierre.renou@inra.fr

Abstract—A key issue in bioinformatics is to decipher cell regulation mechanisms. By comparing networks observed in two different situations, differential network analysis enables to highlight differences that reveal specific cellular responses. The aim of our work is to study the role of natural anti-sense transcription on cellular regulation mechanisms. Our proposal is to build and compare networks obtained from two different sets of actors: the “usual” sense actors on one hand and the sense and anti-sense actors on the other hand. Our study only considers the most significant interactions, called an Extended Core Network; therefore our differential analysis identifies important interactions that are on the influence of anti-sense transcription. Our inference method of an Extended Core Network is inspired by C3NET, but whereas C3NET only computes one interaction per gene, we propose to consider the most significant interactions for each gene. We define the differential network analysis of two extended core networks inferred with and without anti-sense actors. This relies on *change motifs* that describe which gene-gene interactions of the extended core network are modified when we integrate anti-sense actors in the data. As our method focuses on the most significant interactions, these motifs highlight the impact of anti-sense transcription. The networks motifs obtained by our workflow are then compared with assessed biological knowledge. The study reported in this paper is realized on transcriptional data from apple fruit in a context of fruit ripening; the change motifs revealed by our analysis are matched on a protein-protein interaction network and give a small set of interesting actors that deserve further biological investigation.

I. INTRODUCTION

Gene regulation is a key issue in bioinformatics. As microarray produce large-scale expression datasets, gene network inference is a useful approach to study gene interactions, and a lot of methods have been proposed for this reverse engineering task [1]–[4]. The differential network analysis [5]–[7] proposes to study the cellular response to different situations. In medicine, the differential network analysis is used to compare healthy tissues and diseased tissues in order to reveal network rewiring induced by the disease [8]. In these approaches, the compared gene networks involve the same set of genes.

Our work has the particularity to study gene networks with sense and anti-sense transcripts. Anti-sense RNAs are endogenous RNA molecules whose partial or entire sequences

exhibit complementarity to other transcripts. Their different functional roles are not completely known but several studies suggest that they play an important role in stress response mechanisms [9]. Previous studies on *Arabidopsis Thaliana* showed that sense and anti-sense transcripts for a defense gene (RPP5) form dsRNA and generate siRNA which presumably contributes to the sense transcript degradation in the absence of pathogen infection [10].

A recent study [11] of anti-sense transcription in eight different organs (seed, flower, fruit, ...) of apple (*Malus × domestica*) shows several interesting points. Firstly it identifies anti-sense transcription for 65% of the sense transcripts expressed in at least one organ, while it is about 30% in previous *Arabidopsis Thaliana* studies. Secondly, the anti-sense transcript expression is correlated with the presence of short interfering RNAs. Thirdly, anti-sense expression levels vary depending on both organs and Gene Ontology (GO) categories. They are higher for genes belonging to the “defense” GO category and on fruits and seeds.

The work described in this paper proposes a large-scale analysis of apple transcriptomic data, with measures of anti-sense transcripts in the context of fruit ripening. The fruit ripening is a stress-related condition involving “defense” genes. To highlight the impact of anti-sense transcription, we propose to compare context-specific gene networks that involve two kinds of actors, on one hand the sense transcripts that are usually used in gene networks and on the other hand the sense and anti-sense transcripts. However gene network inference methods generally find many false positive interactions, and some authors have proposed to study the core part of a gene network [2], by only computing for each gene one significant interaction. In order to compare networks, this approach is too restrictive and we propose to compute an *Extended Core Network* where each gene is connected to its most important neighbors. We use this inference method in order to discover which interactions of the core network are modified when we integrate the anti-sense transcripts. To characterize these modifications, the notion of change motifs for the comparison graph has been defined in [12]. Here we redefine this notion to use it with our gene network inference

method. Our preliminary results on the apple datasets show that relevant informations are provided by this approach.

In section II, we present the Extended Core Network Inference method. In section III, we present our workflow for comparison of two extended core networks built on different sets of actors and we define the change motifs that highlight significant differences between the interactions. We provide the results obtained for the apple data and assess their biological relevance by the integration of a protein-protein network.

II. EXTENDED CORE NETWORK INFERENCE METHOD

A. Motivations

Several methods have been proposed to infer gene networks from transcriptomic data [1], [3], [4]. Some inference methods reconstruct pairwise gene interaction networks by measuring with a statistical criterion whether two genes are co-expressed or co-regulated. This statistical measure can be Spearman or Pearson correlation [13], or mutual information [2], [14], [15], that enables the detection of non-linear relationships. One major drawback of these methods is that many of the predicted interactions are false positives. To avoid false positives, the Conservative Causal Core Network (C3NET) [2] proposes to compute the core of a gene network, by selecting for each gene a unique interaction defined by the maximal mutual information value. Our aim is to compare two inferred networks to identify significant changes in the interactions when we take into account the anti-sense actors. Therefore only considering the maximal interaction for each gene is too restrictive, since several mutual information values may be very close to the maximum, and a strict comparison of the maximal values in two situations is not relevant to compare two networks. So we propose a gene network inference method, based on C3NET and named *Extended Core Network (ECN)*, where for each gene the most significant interactions are put in the inferred network.

B. Inference of an extended core network

The microarray data we use are intensity values of genes and antisense transcripts in biological samples. After a quantile normalization [16] and a copula-transformation [17], Extended Core Network uses a classical estimator of mutual information in order to estimate the connections between each pair of genes. As C3NET or ARACNE [14], we test the statistical significance of pairwise mutual information values by resampling methods and all non-significant values are set to 0 before applying the inference algorithm.

The network we obtain is represented by an adjacency matrix. We first initialize the matrix by considering that there is no connection at all. Then, for each gene, we will look for his neighbours. The neighbours of a gene g are the ones with the best mutual information. We use an accepting rate r in order to identify the threshold value that determine if a gene g' has one of the best mutual information with g . The threshold of g is fixed by the maximal mutual information of g with other genes, and the accepting rate r . The accepting rate must be between 0 and 1; 0 means that only the best

neighbour will be selected and 1 means that all significant interactions will be selected. When the accepting rate is 0, it is almost the same as in C3NET: if two interactions share the best mutual information, both of them will be selected in ECN whereas only one will be selected in C3NET. When the accepting rate is 1, the interactions of the output network are all the significant mutual information values. We have tested our method on several simulated datasets thanks to SynTREN [18]. We used precision and rappel to evaluate the performance of our inference methods with different accepting rates and we observed that an accepting rate between 5 and 10% is a good choice.

Mutual information is a symmetrical measure, that is why the final step of C3NET transforms the asymmetrical adjacency matrix into a symmetrical one: the result is thus an undirected graph. Because we want to compare two networks to decide whether the most important interactions on a gene g are modified when we integrate anti-sense actors in the algorithm, our inference method provides a directed network so that we can identify which significant changes occur in the connexions of a gene g .

III. DIFFERENTIAL NETWORK ANALYSIS

We now describe the differential analysis that we perform on two extended core networks to explore the role of anti-sense transcription. We first describe the sense and anti-sense data obtained in experiments about apple ripening. Then we define the change motifs revealed by the network comparison and we report the results. To complete this analysis, we compare the significant interaction changes with assessed knowledge about protein-protein interactions.

A. Biological material

In order to study the impact of anti-sense transcripts, we use data of apple fruit during fruit ripening. The fruit ripening is a stress-related condition involving “defense” genes. We analyse RNA extracted from apple fruits thanks to the chip AryANE v1.0 containing 63011 predicted sense genes and 63011 complementary anti-sense sequences. We study the fruit ripening process described by two conditions: harvest (H) and 60 days after harvest (60DAH), and for each condition, 22 samples of apple fruit have been analysed. We first identify transcripts displaying significant differences between the two conditions ($p\text{-val} < 1\%$). With a further threshold of 1 log change between the two conditions, we found 931 sense (S) and 694 anti-sense transcripts (AS) differentially expressed, with among them, 200 transcripts ($S \cap AS$) for which both sense and anti-sense fulfil the condition.

B. Change motifs

Our differential analysis compares the network inferred from the sense data (S) with the network inferred from the sense and anti-sense data ($S \cup AS$, noted SAS). We compare these two networks thanks to *change motifs*: change motifs allow to highlight the part of the core network where anti-sense transcripts have an impact. The notion of change motif

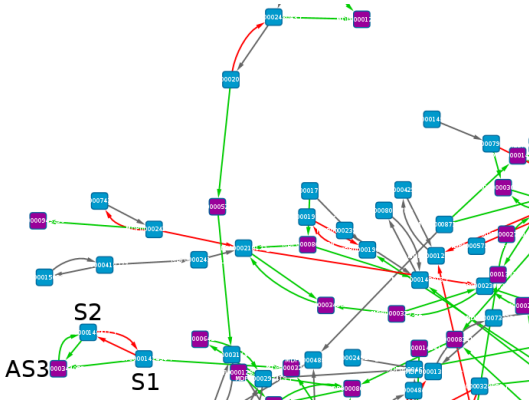


Fig. 1. Part of the graph of change motifs from the 60DAH experiment. A blue node is a sense node, a purple node is an anti-sense node. A red link is an interaction only in the S network, a green link is an interaction only in the SAS network, and a grey link is an interaction in both S and SAS network. Here we can see that S1 is linked by S2 with a red outer link, and by AS3 with a green outer link meaning those three genes form a change motif.

has been used in [12] in a restricted case where the networks are computed with C3NET. When comparing Extended Core Networks, we define a change motif as a sense actor s that is linked to one or many other sense actors in the S network, but these interactions are no longer present in the SAS network. This change of interactions occurs because in the SAS network, the most significant interactions (greater values of the mutual information criterion) for this gene s are interactions with anti-sense actors. So change motifs allow us to identify interactions that are “modified” by adding anti-sense actors into the network inference. The information provided by these change motifs would be omitted by a classical gene network inference relying only on sense data. In Figure 1 we have an illustration of what we define as change motifs. The S and SAS network are both drawn in the same graph. A sense node is blue and an anti-sense node is purple. A red link is a connection present only in S network, a green link is a connection present only in SAS network and a gray link is a connection present in both networks. A change motif is a blue node with no gray outer link and at least one green outer link.

The accepting rate of Extended Core Network determines the number of neighbours a node has, so it changes the number of change motifs in a network. Due to the experimentations on simulated data, we decided to use the accepting rate of 5% and with this rate, we identified 300 change motifs in the H experiment and 308 motifs in the 60DAH experiment. The number of change motifs shows that anti-sense transcripts have a real impact because about 30% of sense actors are involved into a change motif. Figure 1 shows the graph of change motifs from the 60DAH experiment. The graph of change motifs is obtained by combining the S and SAS network inferred with ECN 5%. By using directed graph we can thus identify more modifications of interactions. The graph of change motifs allows biologists to explore easily the modifications that occurred in their experiment. Change motifs

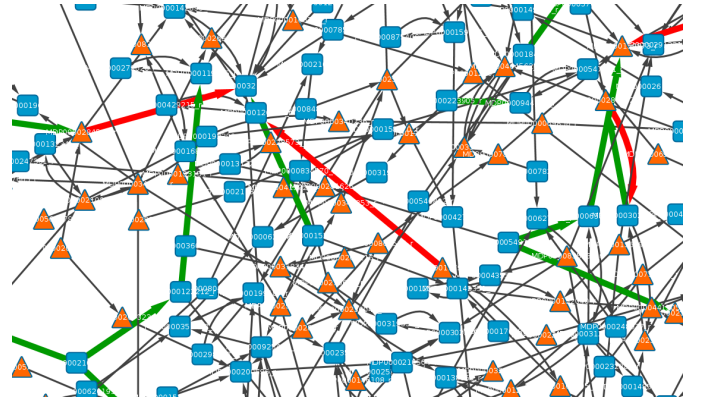


Fig. 2. Part of the ECN 5% graph for sense-only 60DAH experiment. Orange triangle-shaped nodes represent nodes involved in a change motif. A bold green link represent a connection present in the interactome. A bold red link represent a connection present in the interactome and involved in a change motif.

allow us to identify areas in the sense core network where anti-sense transcription have a deep impact, and we can identify anti-sense actors that modify the core network connections. The genes involved into a change motif will be studied to analyse the impact of anti-sense transcription.

C. Biological knowledge integration

By exploring Figure 1, biologists can look for interesting interaction modifications in their experiments. The analysis we propose can be completed by the integration of biological knowledge. More precisely we can compare the inferred interactions and the change motifs with biologically assessed interactions.

Here, we propose to match gene networks obtained by transcription data with interactome knowledge. We look for every interactions of the gene networks obtained thanks to Extended Core Network into the interactome. All the interactions of a co-expression network implies an interaction in a protein-protein network. To identify the interactions that can be found at the transcriptomic and molecular level is interesting. We also identify if change motifs are involved in such interactions.

As an interactome is not available for the apple, we use *Arabidopsis* interactome [19] and orthologs in order to integrate this biological knowledge in our analysis. The gene networks are inferred with all the data, meaning the 931 transcripts for the S network and the 1625 transcripts for the SAS network, then we filter the nodes of the networks to keep only the ones which have an *Arabidopsis* ortholog. This step changes the topology of the networks, because only 791 sense transcripts have an ortholog, and 575 anti-sense transcripts have an ortholog. The S network has now 791 nodes and the SAS network has 1366 nodes.

We use Extended Core Network with an accepting rate of 5% in order to create S networks in both H and 60DAH experiments. We look if there exists a path in the interactome between two directly connected genes in the ECN network. For example, if g_1 and g_2 are directly connected in the

ECN network, we look for a path between g_1 and g_2 in the interactome. We use the AI-1 version of interactome composed by protein-protein interactions from the main and repeat version, and also from literature. It represents a graph of 4866 nodes and 11374 interactions.

We perform this differential network analysis only with the S networks because anti-sense transcripts are not supposed to interact with proteins. First, we kept only genes with *Arabidopsis* orthologs and more than two-thirds of change motifs are remaining. Second, we observe that only 3% of edges from the ECN sense-only network can be found in the interactome. However, we can note that some of these edges are also involved in a change motif. If we find an edge of the ECN sense only network in the interactome, it means that the two sense actors are involved into a same pathway that allows proteins to interact with each other. An edge of the ECN sense only network in a change motif and in the interactome means that the two sense actors are involved into a same pathway but there is at least one of these sense actors that has a greater mutual information with an anti-sense actor. This is why sense and anti-sense actors linked by these edges are interesting to study. Figure 2 shows the graphical result of the differential network analysis for the 60DAH experiment. The graph was drawn using Cytoscape [20]. We can see that sense nodes that are involved in a change motif (orange triangles) are spread all over the network. Some interactions involved in a change motif which are found in the interactome are in the center of the network and will be studied in the following of our work.

IV. CONCLUSION

In order to study on a large scale the impact of anti-sense transcription, we propose to infer gene networks, with the particularity to integrate anti-sense transcripts in the process. We extend the recent ideas introduced in the field of differential network analysis, because in our framework, the main idea is to compare a network inferred from sense data and a network inferred from sense and anti-sense data. To achieve such a comparison, we first propose the Extended Core Network inference method. Extending the idea proposed by C3NET, this method builds a co-expression gene network where only the core of the network is shown by selecting the most important interactions for each gene. Secondly we extend the differential network analysis that we performed in [12] by using the Extended Core Network method instead of C3NET. The comparison of two networks inferred on different sets of actors leads to define change motifs that highlight gene interactions that are impacted by anti-sense transcription. The visualisation of these change motifs helps the biologist to focus on interesting parts of the networks. We also propose to complete the differential analysis by integrating the revealed motifs in a biological network like a protein-protein network.

ACKNOWLEDGEMENT

The authors would like to thank the GRIOTE project, funded by the Pays-de-la-Loire Region. We thank the Institut

de Recherche en Horticulture et Semences for providing microarray data on apple with sense and anti-sense probes.

REFERENCES

- [1] M. Bansal, V. Belcastro, A. Ambesi-Impiombato, and D. di Bernardo, "How to infer gene networks from expression profiles," *Molecular Systems Biology*, vol. 3, no. 1, p. 78, Feb. 2007.
- [2] G. Altay and F. Emmert-Streib, "Inferring the conservative causal core of gene regulatory networks," *BMC Systems Biology*, vol. 4, no. 1, p. 132, Sep. 2010.
- [3] D. Marbach, J. C. Costello, R. Kffner, N. M. Vega, R. J. Prill, D. M. Camacho, K. R. Allison, The DREAM5 Consortium, M. Kellis, J. J. Collins, and G. Stolovitzky, "Wisdom of crowds for robust gene network inference," *Nature Methods*, vol. 9, no. 8, pp. 796–804, Aug. 2012.
- [4] S. Friedel, B. Usadel, N. von Wiren, and N. Sreenivasulu, "Reverse Engineering: A Key Component of Systems Biology to Unravel Global Abiotic Stress Cross-Talk," *Frontiers in Plant Science*, vol. 3, p. 294, Dec. 2012.
- [5] R. Sharan and T. Ideker, "Modeling cellular machinery through biological network comparison," *Nature Biotechnology*, vol. 24, no. 4, pp. 427–433, Apr. 2006.
- [6] G. Altay, M. Asim, F. Markowetz, and D. E. Neal, "Differential C3net reveals disease networks of direct physical interactions," *BMC Bioinformatics*, vol. 12, no. 1, p. 296, Jul. 2011.
- [7] T. Ideker and N. J. Krogan, "Differential network biology," *Molecular Systems Biology*, vol. 8, p. 565, Jan. 2012.
- [8] A.-L. Barabási, N. Gulbahce, and J. Loscalzo, "Network Medicine: A Network-based Approach to Human Disease," *Nature reviews. Genetics*, vol. 12, no. 1, pp. 56–68, Jan. 2011.
- [9] V. Pelechano and L. M. Steinmetz, "Gene regulation by antisense transcription," *Nature Reviews Genetics*, vol. 14, no. 12, pp. 880–893, Dec. 2013.
- [10] H. Yi and E. J. Richards, "A Cluster of Disease Resistance Genes in Arabidopsis Is Coordinately Regulated by Transcriptional Activation and RNA Silencing," *The Plant Cell*, vol. 19, no. 9, pp. 2929–2939, Jan. 2007.
- [11] J.-M. Celton, S. Gaillard, M. Bruneau, S. Pelletier, S. Aubourg, M.-L. Martin-Magniette, L. Navarro, F. Laurens, and J.-P. Renou, "Widespread anti-sense transcription in apple is correlated with siRNA production and indicates a large potential for transcriptional and/or post-transcriptional control," *New Phytologist*, vol. 203, no. 1, pp. 287–299, Mar. 2014.
- [12] M. Legeay, B. Duval, and J.-P. Renou, "Differential functional analysis and change motifs in gene networks to explore the role of anti-sense transcription," in *International Symposium on Bioinformatics Research and Applications*. Springer, 2016, pp. 117–126.
- [13] P. Langfelder and S. Horvath, "WGCNA: an R package for weighted correlation network analysis," *BMC Bioinformatics*, vol. 9, no. 1, p. 559, Dec. 2008.
- [14] A. A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R. D. Fava, and A. Califano, "ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context," *BMC Bioinformatics*, vol. 7, no. Suppl 1, p. S7, Mar. 2006.
- [15] X. Zhang, K. Liu, Z.-P. Liu, B. Duval, J.-M. Richer, X.-M. Zhao, J.-K. Hao, and L. Chen, "NARROMI: a noise and redundancy reduction technique improves accuracy of gene regulatory network inference," *Bioinformatics*, vol. 29, no. 1, pp. 106–113, Jan. 2013.
- [16] X. Qiu, H. Wu, and R. Hu, "The impact of quantile and rank normalization procedures on the testing power of gene differential expression analysis," *BMC Bioinformatics*, vol. 14, p. 124, 2013.
- [17] R. Steuer, J. Kurths, C. O. Daub, J. Weise, and J. Selbig, "The mutual information: Detecting and evaluating dependencies between variables," *Bioinformatics*, vol. 18, no. suppl 2, pp. S231–S240, Jan. 2002.
- [18] T. V. d. Bulcke, K. V. Leemput, B. Naudts, P. v. Remortel, H. Ma, A. Verschoren, B. D. Moor, and K. Marchal, "SynTReN: a generator of synthetic gene expression data for design and analysis of structure learning algorithms," *BMC Bioinformatics*, vol. 7, no. 1, p. 43, Jan. 2006.
- [19] Arabidopsis Interactome Mapping Consortium, "Evidence for Network Evolution in an Arabidopsis Interactome Map," *Science*, vol. 333, no. 6042, pp. 601–607, Jul. 2011.
- [20] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks," *Genome Research*, vol. 13, no. 11, pp. 2498–2504, Nov. 2003.