CANDIDATE GENES DISCOVERY THROUGH AUGMENTED GENE CO-EXPRESSION ANALYSIS - A CASE STUDY IN SKIN AGING

Gwenaëlle Lemoine*, Marie-Pier Scott-Boyer, Arnaud Droit

* gwenaelle.lemoine@protonmail.com
Genomics Center, Centre Hospitalier Universitaire de Québec – Université Laval Research Center, Québec, QC, Canada

AD Lab
compbio.ulaval.ca

BACKGROUND

The rapid increase in size of transcriptomic datasets requires the development of new methods able to manage them[1]. However, methods such as differential expression analysis only focuses on few genes.

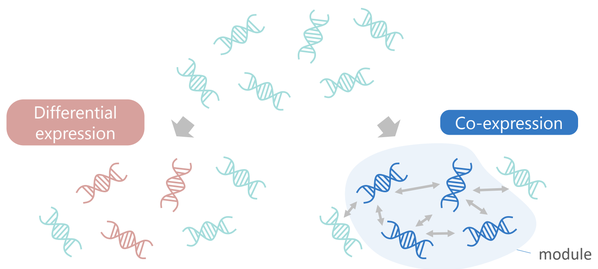

Figure 1 : Comparison of differential expression and gene co-expression. Most strongly interacting genes are grouped into clusters (modules) and assumed to be involved in a same biological function (guilt-by-association paradigm[2]).

Network-based methods such as **gene co-expression** unveil a new dimension in the study of complex molecular processes[3] (Figure 1) and multiple studies have been successfully using them[4–6].

The R package WGCNA[7] is a widely used implementation of this method and has proven its effectiveness. However, it require the users to have certain experience of co-expression subtleties to work with it and don't implement comparison of condition along with help to interpretation.

Here we present **Gene Whole co-Expression Network Analysis (GWENA)**, an R package improving information which can be retrieved from co-expression analysis applied to transcriptomic data. We applied GWENA to **skin aging** data and retrieved potentially useful information about processes involved.

CASE STUDY IN SKIN AGING

Aim : Comparing two aging ranges in order to find new potential candidate gene able to explain biological processes involved in human skin aging.

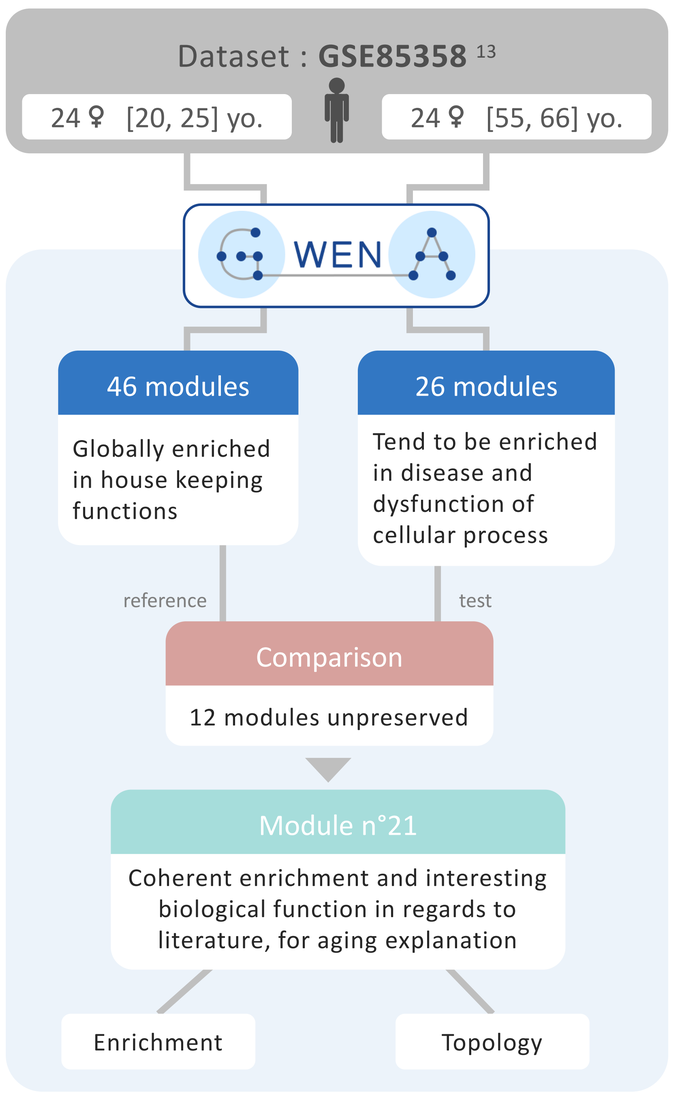

Figure 3 : Application of GWENA on GSE85358 data set and results summary.

OBJECTIVES

1. Providing a complete pipeline for co-expression analysis, from transcriptomic data to extended characterization of genes modules.
2. Bringing novel method allowing module comparison between conditions.
3. Discovering new genes of interest into skin aging.

Further exploration of module n°21

- 144 genes, with only the 69 most strongly interacting, plotted in Figure 4.
- Multiple enrichment linked to immunity processes and cellular stress.
- Proteins coded by genes into this module are mostly localized in membranes of different vacuoles, as trans-membranous proteins. In case of cell membrane, important amount of various cytokine receptor = signaling. Coherent with the immune aspect known to be related to skin aging.
- Topology:
  - 3 genes detected as hubs in young: HLA-DRB1, HLA-DMA, HLA-DM.
  - Density is significantly higher in young module compared to the old. Meaning some interactions, therefore biological function, disappear.

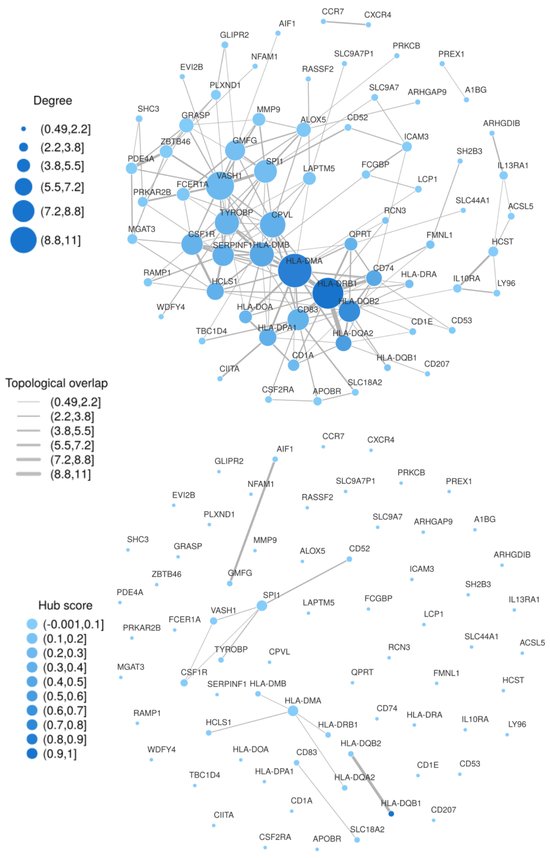

Figure 4 : Module 21's network with intensity of interactions as edge thickness. Upper module is from young population and lower one is the same genes in the old population. Highly connected genes show an important hub score.

IMPLEMENTATION

WEN — **G**ene **W**hole co-**E**xpression **N**etwork **A**nalysis

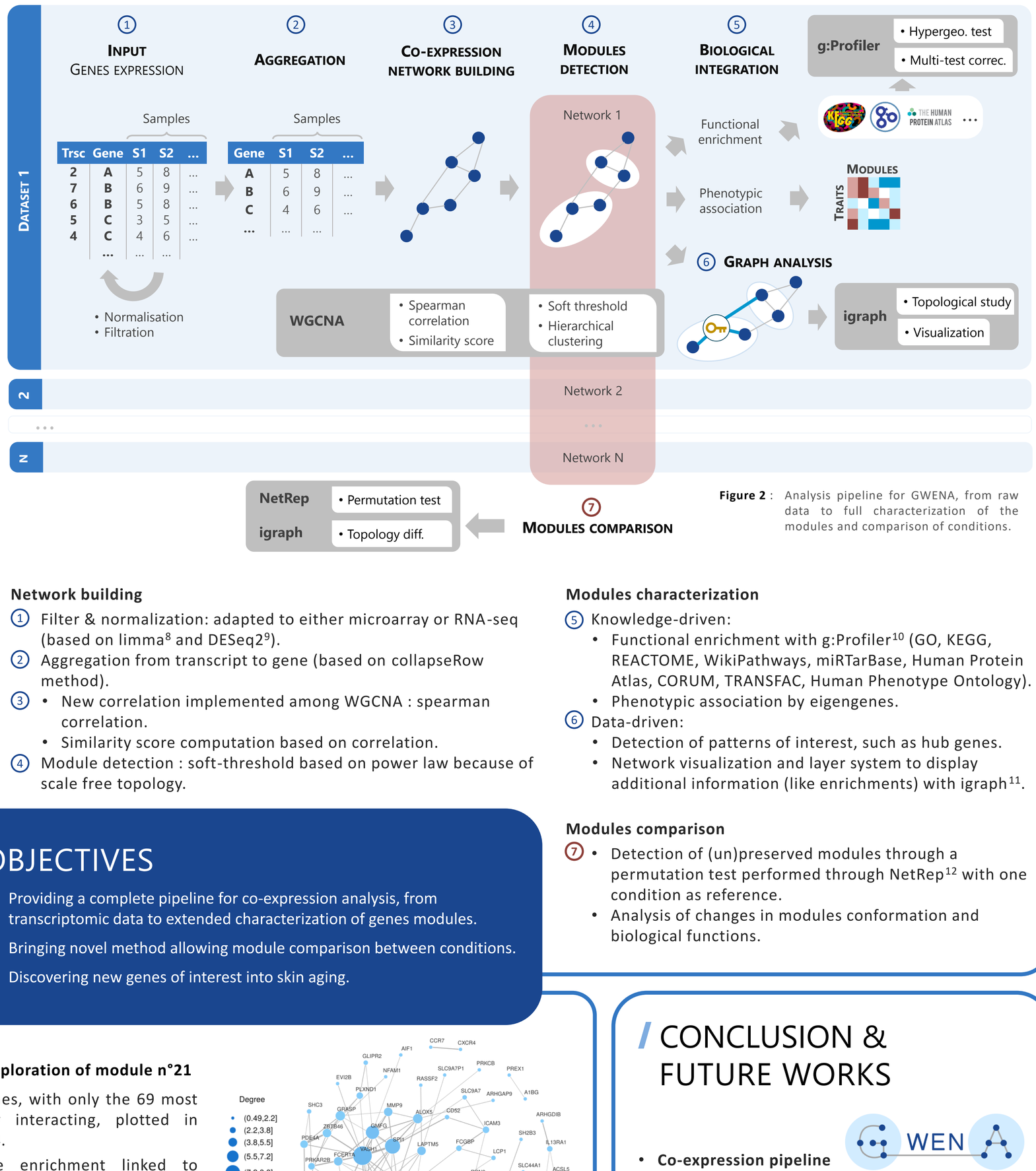

Figure 2 : Analysis pipeline for GWENA, from raw data to full characterization of the modules and comparison of conditions.

Network building

1. Filter & normalization: adapted to either microarray or RNA-seq (based on limma[8] and DESeq2[9]).
2. Aggregation from transcript to gene (based on collapseRow method).
3. • New correlation implemented among WGCNA : spearman correlation.
   • Similarity score computation based on correlation.
4. Module detection : soft-threshold based on power law because of scale free topology.

Modules characterization

5. Knowledge-driven:
   • Functional enrichment with g:Profiler[10] (GO, KEGG, REACTOME, WikiPathways, miRTarBase, Human Protein Atlas, CORUM, TRANSFAC, Human Phenotype Ontology).
   • Phenotypic association by eigengenes.
6. Data-driven:
   • Detection of patterns of interest, such as hub genes.
   • Network visualization and layer system to display additional information (like enrichments) with igraph[11].

Modules comparison

7. • Detection of (un)preserved modules through a permutation test performed through NetRep[12] with one condition as reference.
   • Analysis of changes in modules conformation and biological functions.

CONCLUSION & FUTURE WORKS

- **Co-expression pipeline**
  - Our tool allow **modules comparison** and biological discrimination between conditions.
  - This functionality is integrated in a **complete pipeline facilitating access to co-expression analysis** and interpretation while providing advanced customization for researchers with advanced bioinformatics skills.
  - We aim to develop a graphical interface based on the same package to go further in the user-friendly aspect.

- **Skin aging**
  - Our method appears to be appropriate for highlighting potential candidate genes in skin aging processes exploration.
  - Skin aging is known to be related with immunity and inflammation. Next step will be to determine which pathways or biological function are implied and how they contribute to skin degradation.

REFERENCES

1. van Dam, S., Võsa, U., van der Graaf, A., Franke, L. & de Magalhães, J. P. Gene co-expression analysis for functional classification and gene–disease predictions. *Briefing in Bioinformatics* (2017).
2. Barabási, A. L. & Oltvai, Z. N. Network biology: Understanding the cell's functional organization. *Nature Review Genetics* (2004).
3. López-Kleine, L., Leal, L. & López, C. Biostatistical approaches for the reconstruction of gene co-expression networks based on transcriptomic data. *Briefing in Functional Genomics* (2013).
4. Li, J. *et al.* Application of Weighted Gene Co-expression Network Analysis for Data from Paired Design. *Scientific Report* (2018).
5. MacLennan, N. K. *et al.* Weighted gene co-expression network analysis identifies biomarkers in glycerol kinase deficient mice. *Molecular Genetics and Metabolism* (2009).
6. Kogelman, L. J. A. *et al.* Identification of co-expression gene networks, regulatory genes and pathways for obesity based on adipose tissue RNA Sequencing in a porcine model. *BMC Medical Genomics*, (2014).
7. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* (2008).
8. Ritchie, ME, Phipson, B, Wu, D, Hu, Y, Law, CW, Shi, W, and Smyth, GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* (2015).
9. M. I. Love, W. Huber, S. Anders: Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *BMC Genome Biology* (2014).
10. Reimand J et *al*. g:Profiler–a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res* (2007).
11. Csardi, Gabor, and Tamas Nepusz. The igraph software package for complex network research. InterJournal, Complex Systems (2006)
12. Ritchie, Scott C., et al. A scalable permutation approach reveals replication and preservation patterns of network modules in large datasets. Cell systems (2016)
13. Kuehne, Andreas, et al. An integrative metabolomics and transcriptomics study to identify metabolic alterations in aged skin of humans in vivo. *BMC Genomics* (2017)

ACKNOWLEDGMENT

Thank you to L'Oréal – Recherche et Innovation for financing my thesis subject.
Also many thanks to Audrey Lemaçon, Régis Ongaro-Carcy, and Julien Prunier for their help along this project and beyond.

UNIVERSITÉ LAVAL
CHAIRE DE RECHERCHE ET D'INNOVATION L'ORÉAL EN BIOLOGIE NUMÉRIQUE
AFFILIÉ À UNIVERSITÉ LAVAL
CHU de Québec

Designed by Gwenaëlle Lemoine