

# GWENA: Gene Co-expression Networks Analysis and Extended Modules Characterization in a Single Bioconductor Package

**Gwenaëlle Lemoine**

Département de médecine moléculaire, Faculté de médecine, Université Laval

**Marie-Pier Scott-Boyer**

Centre de recherche du Chu de Quebec-Université Laval

**Bathilde Ambroise**

L'Oréal Research and Innovation

**Olivier Perin**

L'Oréal Research and Innovation

**Amaud Droit** (✉ [amaud.droit@crchudequebec.ulaval.ca](mailto:amaud.droit@crchudequebec.ulaval.ca))

Département de médecine moléculaire, Faculté de médecine, Université Laval

---

## Research Article

**Keywords:** co-expression network, differential co-expression, R package, pipeline, aging, skeletal muscle

**Posted Date:** December 30th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-134425/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

## RESEARCH

# GWENA: gene co-expression networks analysis and extended modules characterization in a single Bioconductor package

Gwenaëlle G. Lemoine<sup>1</sup>, Marie-Pier Scott-Boyer<sup>2</sup>, Bathilde Ambroise<sup>3</sup>, Olivier Périn<sup>3</sup> and Arnaud Droit<sup>1,2\*</sup>

1. Département de médecine moléculaire, Faculté de médecine, Université Laval, 2325 rue de l'Université, G1V 0A6, Québec, Canada

2. Centre de recherche du Chu de Quebec-Université Laval, 2705 boulevard Laurier Québec, G1V 4G2, Québec, Canada

3. L'Oréal Research and Innovation, 15 rue Pierre Dreyfus, 92110, Clichy, France

## Abstract

**Background:** Network-based analysis of gene expression through co-expression networks can be used to investigate modular interactions occurring between genes toward different biological functions. An extended description of the network modules is therefore a critical step to understand the underlying processes contributing to a disease or a phenotype. Biological integration, topology study and conditions comparison (e.g. wild vs mutant) are the main methods to do so, but to date no tool combines them all into a single pipeline.

**Results:** Here we present GWENA, a new R package that integrates gene co-expression network construction and a whole characterization of the detected modules through gene set enrichment, phenotypic association, hub genes detection, topological metric computation, and differential co-expression. To demonstrate its performances, we applied GWENA on two skeletal muscle datasets from young and old patients of GTEx study. We successfully prioritized a gene whose involvement was unknown in the muscle development and growth. We also gave new insight about the variations in patterns of co-expression as the already known age-dependent loss of connectivity was found coupled to a genes interactions reorganization leading to the expression of other functions involved in aging.

**Conclusion:** GWENA is an R package available through Bioconductor (<https://bioconductor.org/packages/release/bioc/html/GWENA.html>) developed to perform extended analysis of gene co-expression networks. Thanks to biological and topological information as well as conditions comparison, it eases the understanding of genes interactions involved in diseases or phenotypes. Going beyond actual packages to perform co-expression analysis, GWENA includes new tools to fully characterize modules, such as differential co-expression, additional enrichment databases, and network visualization.

**Keywords:** co-expression network; differential co-expression; R package; pipeline; aging; skeletal muscle

## Background

The study of biological functions through discrete genes analysis methods has allowed the description of numerous pathways and the understanding of gene-disease associations [1]. The full comprehension of the complex interactions taking place in cellular processes

requires methods able to grasp the connections between the genes involved [2]. To address this issue, biological networks have been used as a framework to represent and study interactions between genes. In a gene network, a node represents a gene and an edge joining two nodes represents their interaction. Among the measures of interaction, weighted co-expression is one of the most widely used thanks to the popularity of the WGCNA R package [3] where the interactions are quantification (weight) instead of only

\* Correspondence: [Arnaud.Droit@crchudequebec.ulaval.ca](mailto:Arnaud.Droit@crchudequebec.ulaval.ca)

<sup>1</sup>Département de médecine moléculaire, Faculté de médecine, Université Laval, 2325 rue de l'Université, G1V 0A6 Québec, Canada

Full list of author information is available at the end of the article

a presence/absence information. The use of gene co-expression networks thus led to important discoveries such as the characterization of functional elements in *Arabidopsis* [4], help with the prognosis in breast cancer [5], and more generally identification and prioritization of disease candidate genes [6].

When constructing gene co-expression networks, existing tools usually follow the same methodology. Using either microarray or RNA-seq gene expression, a co-expression score based on correlation is computed between each pair of genes in the samples. A clustering method is then selected to detect groups of strongly co-expressed genes called modules. Their characterization and the search for meaning in the co-expression relations classically involve the integration of biological information, as well as the study of topology [6]. Biological integration usually involves two methods, namely gene set enrichment and phenotypic association [6, 3]. A phenotypic association is based on the correlation between the eigengene (a mean gene expression profile) of the module and a phenotype measured on the samples. Despite typically having a low yet significant correlation [7] phenotypic associations are used as a surrogate to study the molecular changes related to a condition. By looking for the genes responsible for the correlation, it may be a means to find causal genes or to study the effect of the condition on the phenotype [8]. As for the gene set enrichment, the most common enrichment test is based on the over-representation analysis (ORA) of a group of genes (here our modules) compared to a reference of biological annotations such as Gene Ontology (GO) [9] or Reactome [10]. This approach, based on the guilt-by-association approach, allows the identification of new gene functions. The scale-free topology of co-expression network also allow for new analysis perspectives. The highly-connected genes also known as hub genes are often relevant for the functionality of the module, either being a regulator [11] or a gene coding for an essential function [12]. Their detection and the investigation of the neighboring gene is therefore an opportunity to understand the mechanisms at work.

Like differential expression analysis, co-expression analysis can be used in a differential way to compare conditions (e.g. wild vs. mutant). This method aims to isolate dissimilarities [13] that would not be found by solely studying the GCN of a condition of interest (e.g. disease, phenotype). Variations in gene co-expression between multiple conditions can translate into appearance/disappearance of modules, changes in gene composition of a module, or rearrangement of genes within a module potentially leading to separation into several other modules [6]. These modifications of patterns reveal insights on the biological alterations in modules

of interest and can suggest possible regulatory events linked to the studied condition (e.g. : transcription factors, miRNA). Such concept was used successfully in recent publications to namely detect specific gene modules involved in ovarian or breast cancer [14, 15] or in recovery from water stress in *Cleistanthus* [16].

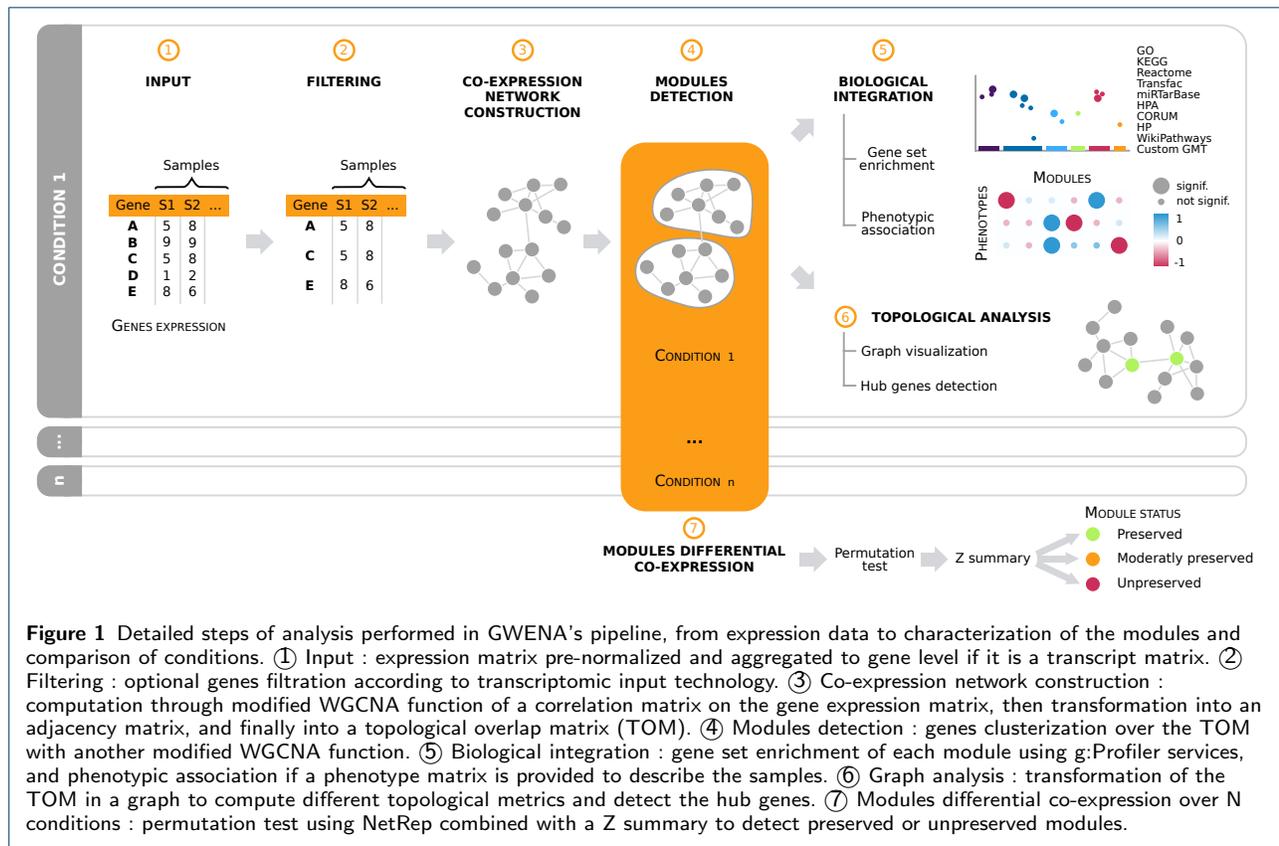
To date, multiple tools exist to perform one or some of the functionalities described previously but none combine them all into a single pipeline. Moreover, no available tool includes differential co-expression, exploit the potential of other topological metrics such as connectivity, or enables analysis to be carried out with other R packages or software as easily. In a desire to meet all these needs, we developed an R package for Gene Whole co-Expression Network Analysis (GWENA) available on Bioconductor (<https://bioconductor.org/packages/release/bioc/html/GWENA.html>). Based on a modified version of WGCNA for the network construction and modules detection, GWENA is a modular pipeline that provides ORA enrichment on 9 biological sources, phenotypic association, hub genes detection, and differential co-expression between multiple conditions. These come with a set of descriptive visualizations to help the user to understand and interpret the results.

In order to demonstrate the capabilities of our tool, we applied it to investigate skeletal muscle aging using publicly available gene expression data from donors spanning different age ranges from the GTEx database (ref). Skeletal muscle aging is indeed a major source of mobility loss in the elderly, resulting in a high fall ratio, depression, and therefore an increased mortality [17]. This decrease in the regenerative capacity of skeletal muscles and their progressive atrophy (sarcopenia) [18] gradually leads to a reduction of the contractile force and thus a loss of autonomy of the individuals [19]. Recent studies have made progress in finding factors associated to progression of sarcopenia [20, 17], such as body weight [21], but the understanding of their intricate molecular mechanisms is still lacking.

In this article, we will therefore care to detail the implementation of our new R package GWENA. A presentation of its application will be done with the study gene co-expression in young muscle, and then in the context of skeletal muscle aging by comparing samples from younger and older donors. Finally, a qualitative comparison will be made with other existing tools.

## Implementation

Designed as an R Bioconductor package, GWENA is a modular pipeline intended to ease the construction, interpretation and comparison of GCN. It reproduces a classical GCN analysis reinforced by complementary tools (Figure 1).



## 1 - Input

Both microarray and RNA-seq normalized expression can be used as input. The choice of normalization method is left to the user as it is highly dependent on the technology used to produce the raw data and the experimental design. Data must be stored in a table with genes as columns and samples as rows, or in a SummarizedExperiment object [22]. For an optimal analysis, the minimal number of samples recommended is about 20 samples [23].

Transcriptome level data (probes or transcript) need to be aggregated to the gene level for the next steps. Its execution is left to the user as the transcriptomic technology impacts the aggregation method to choose. However, it is recommended to use the highest mean probes expression for microarray data, and the counts sum for RNA-seq [24]. This can be achieved with the collapsing R function as described by Miller et al. [24].

## 2 - Filtering

Genes are not always informative for modules detection as genes not always vary and their expression can be linked to technical biases. An additional filtering step can thus be applied to avoid noise and speed up the pipeline analysis. This operation must be carried out with caution as it may impact the network

construction. Over-filtering may result in loss of informative signal and changing the data distribution could break the scale-free topology [25, 23]. In addition, co-expression network analysis is a method designed to handle larger amount of data than differential expression analyses and capture the slightest significant genes expression variations.

Two filters meeting these criteria are available in GWENA:

- 1 Low count filter : removes genes having a lower count than a pre-defined threshold (default is 5). It prevents confusing the true expression of a gene with an expression due to technical background noise.
- 2 Low variation filter : removes genes which expression is too similar across samples. As co-expression modules detection relies on the discrimination of similarity between gene expression profiles across samples, genes that do not vary sufficiently across samples may be randomly clustered in the same (or in different) modules which would not reflect the biological reality.

## 3 - Co-expression network construction

The well-known R package WGCNA [3] was modified in order to be integrated it in our modular pipeline

: the co-expression network construction step which computes the genes pairwise co-expression score has been isolated in its own R function. The first step of the co-expression score computation is the calculation of a correlation matrix based on the gene expression matrix. The Spearman correlation was added to the automated version of network construction in WGCNA as it ensures a better representation of genes monotonic relationships [26]. A power law distribution is then fitted on the correlation matrix and the estimated power is used to elevate the correlation matrix, resulting in an adjacency matrix [27]. According to the hierarchical organisation of gene co-expression networks [28], a topological overlap matrix (TOM) [27] is then computed using the adjacency matrix which represents the final gene co-expression score matrix. Finally, the function return this matrix along with metadata information regarding the computation to ensure a good tracking of the performed operations.

#### 4 - Modules detection

The modules detection part from WGCNA was isolated in a new R function using the previously calculated gene co-expression score matrix as input. A hierarchical clustering is performed on the matrix which is then cut according to a defined cut-off value in order to define the modules and the genes they contain. The first component of the principal component analysis of each module is used as a representative of their respective gene expression profile and is called an eigengene. In addition to its summarizing function, the eigengene is used to merge the highly-correlated modules. The gene co-expression profile of each module is visible using a dedicated function, with the eigengene highlighted. The function finally returns a detailed object with the detected modules as lists of genes identifiers, the dendrogram of the clustering, and the modules before merge.

#### 5 - Biological integration

Biological integration consists of two different analyses, namely gene set enrichment and phenotypic association.

The gene set enrichment (or functional enrichment) analysis is performed using *g:Profiler* [29] through their *profiler2* R package. Their enrichment function covers 9 biological functional databases: Gene Ontology (GO) [9], Kyoto Encyclopedia of Genes and Genomes (KEGG) [30], Reactome [10], Transfac [31], miRTarBase [32], Human Protein Atlas (HPA) [33], CORUM [34], Human Phenotype ontology (HP) [35], WikiPathways [36]. Realizing a custom enrichment file through a Gene Matrix Transposed (GMT) format in *gprofiler2* requires the use of additional functions. Also, *gprofiler2* does not provide a merging function between the

output of classical and custom enrichment to return all the enrichments in a single output. GWENA therefore provides a wrapper of these functions to have a all-in-one function.

The phenotypic association uses the eigengene returned in the output of the module detection function to perform a correlation test on a matrix of given phenotypes. If a phenotype is qualitative instead of quantitative, the variable encoding the phenotype is transformed into a binary variable (also known as dummy variable).

#### 6 - Graph analysis

To analyze the topology of the graph and allow its visualization, GWENA imports the *igraph* [37] R package. A wrapping function including integrity checks use the gene co-expression score matrix to build a graph object on which all *igraph* topological metrics can be computed (e.g. degree, connectivity, strength). Among the multiple metrics computable on a network, hub genes remain the most studied structure. As they can be defined according to different methods, the three most popular ones were implemented: highest connectivity, highest degree, and Kleinberg's score. GWENA visualization function simplifies the native plotting function of *igraph* and accommodate it to GCN to assist in their interpretation (e.g. the native implementation of an edge filter parameter, as these are complete graphs). The layout selection was also favored towards scale-free topology compatible layouts as they are a main property of GCN.

#### 7 - Modules differential co-expression

Analysis of module preservation or unpreservation can be performed between different conditions such as treatments or phenotype. To isolate modules whose topology changes between conditions, GWENA first performs a permutation test using the *NetRep* R package [38]. Seven topological metrics are computed on each module in each condition. A permutation is then applied to the selected control condition where each node label of the modules is randomly reassigned without replacement to another and the seven metrics are then recomputed on it. Using these permutations as a null distribution [39], modules are considered preserved if all seven topological metrics are significant for the alternative hypothesis (one or two sided) with the chosen alpha error.

As the unpreservation of a module cannot be assumed from the non significant modules, a second step of preservation evaluation is carried out using a Z summary score [40, 41]. The final score returned by GWENA is the combination of these two steps (Additional file 1.1).

## Results and discussion

To present GWENA's use and its capability to isolate genes groups or co-expression patterns of interest in a single condition or multiple conditions, we analyzed RNA-seq skeletal muscle data from GTEx (v8)[42] (Additional file 1.2). This data set contains 19,312 genes from 803 samples representing ages ranging from 20 to 70 years old. Low read counts and the low variation genes were discarded using the filtering function of GWENA to decrease the noise, resulting in 18 870 genes.

As GTEx data is known to be subject to multiple confounding factors (batch effect, experimental bias, read contamination, etc.) [43, 44, 25], a partial PC-correction[25] was applied to correct the data (Additional file 1.3). To investigate the aging process two subsets representing contrasting age classes were selected from the corrected data set: 73 samples between 20 and 30 years old (referred as young in this report), and 292 samples between 60 and 70 years old (referred as old in this report). Both datasets were analyzed using GWENA's pipeline with default parameters, with the exception of the correlation method parameter which was selected to be "spearman" instead of the default "pearson" as it is less sensible to outliers.

### Single condition modules analysis

To illustrate the process of analyzing a single condition with GWENA, we first focused on studying the muscle gene co-expression computed in the young sub-population. The 95 modules detected on the co-expression score matrix with GWENA were merged according to their similarity indices, which resulted in a total of 35 modules (Figure 2.A.). Each module was then tested for its association with a selected set of phenotypes related to muscle aging (i.e. age, sex, ethnicity, body weight and BMI) to isolate modules of interest. As shown in Figure 2.B, 15 of these modules were significantly associated with at least one of the phenotypes.

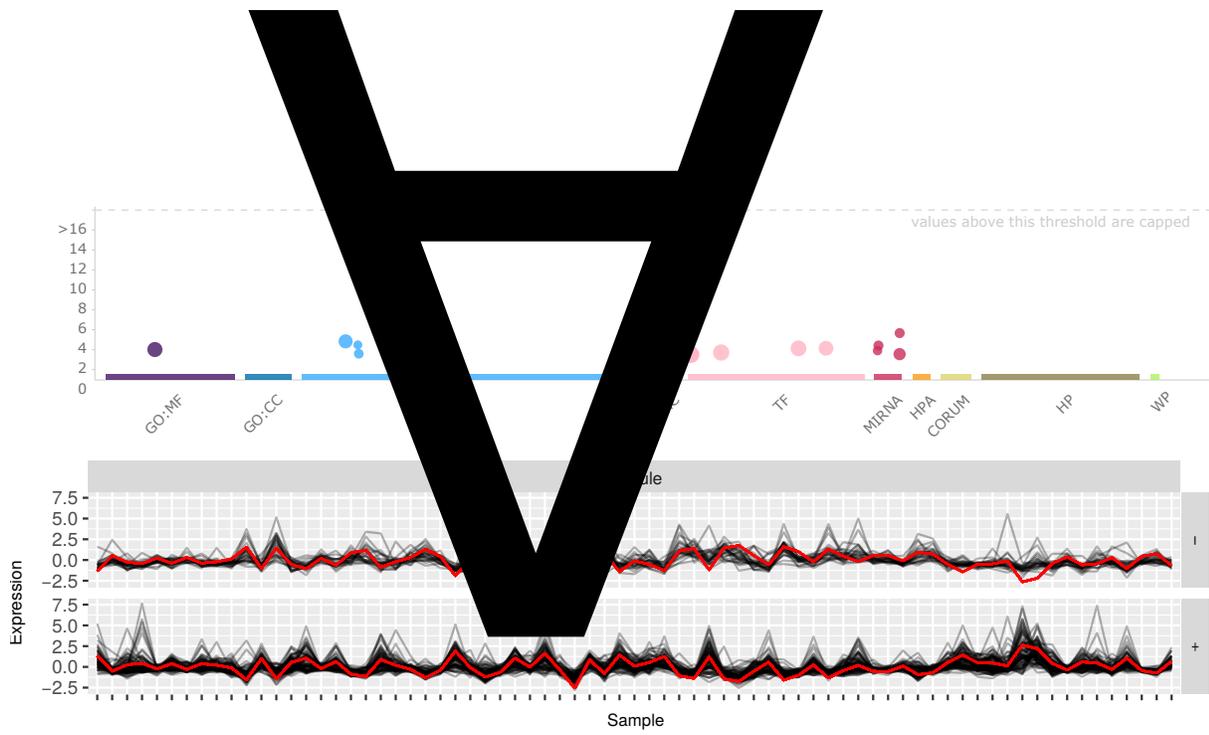
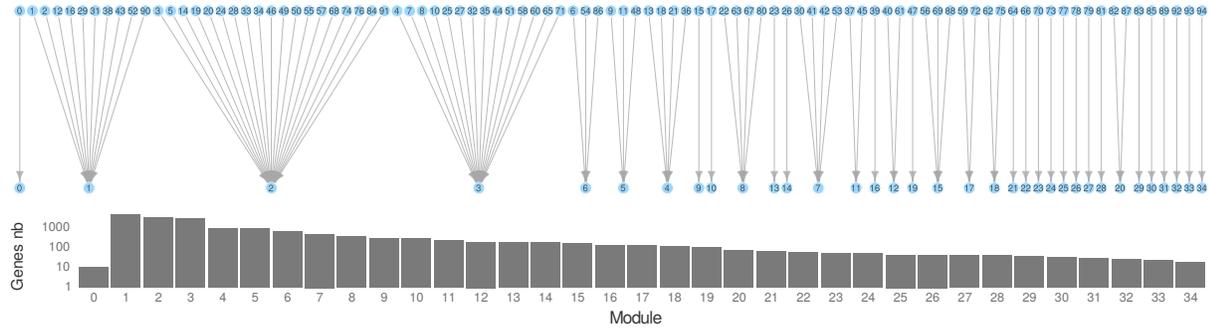
These modules were provided to GWENA enrichment analysis (p value <0.05 with g:SCS multiple testing correction) to identify their biological functions and assess their potential involvement in muscle function (Table 1). All modules were at least enriched in one term and 8 obtained enrichment terms related to muscle activity or metabolism (Table 1). Modules 19, 21 and 25 were the top 3 enriched for terms related to muscle function. However, modules 21 and 25 terms were mostly coming from Human Protein Atlas and were also related to a wide range of additional tissues such as the pancreas, the cervix, the bladder, the stomach, or the skin and were thus deemed less specific for muscle aging than module 19.

module	# genes	# pheno. asso.	# enrichment	% muscle enrich.
0	11	NA	NA	NA
1	5335	1	3288	0.6
2	3661	2	1098	0.4
3	3355	0	1620	0.3
4	1001	1	1883	1.0
5	987	0	1626	0.0
6	699	0	457	0.6
7	546	0	428	1.1
8	409	0	561	0.7
9	310	1	729	0.3
10	308	0	58	5.2
11	261	0	857	0.0
12	214	0	847	15.0
13	207	0	452	1.4
14	197	1	767	0.5
15	175	2	233	0.0
16	137	1	6	0.0
17	136	0	1	0.0
18	129	2	20	0.0
19	108	2	18	3.7
20	77	2	52	0.0
21	72	1	233	2.8
22	63	0	24	0.0
23	57	0	10	0.0
24	55	0	8	0.0
25	47	2	82	8.5
26	47	0	32	0.0
27	46	0	12	0.0
28	43	0	147	4.7
29	40	0	17	0.0
30	35	1	2	0.0
31	31	1	12	0.0
32	27	0	1	0.0
33	24	0	23	0.0
34	20	1	3	0.0

**Table 1** Summary of biological integration for modules detected in young age range. The number of enrichment terms is the sum of the terms across all sources of biological databases. The ratio of terms linked to muscle is the number of terms containing a term in a defined corpus of regex over the total number of enrichment found. The corpus is composed of the following regex: "muscle", "sarco\\w+", "myo\\w+", "muscul\\w+", "actin\\w+", "myosin\\w+".

Briefly, the remaining module 19 presented 77% of genes positively correlated to its eigengene (therefore 23 negatively, Figure 2.D.), and the muscle enriched terms involved muscle adaptation and negative regulation of hypertrophy (Table 2, Figure 2.C.).

The detection of hub genes by GWENA returned 12 hub genes, some of which are known as transcription factors. Among them, ARID5B (ENSG00000150347) is a transcription factor strongly co-expressed with KLF15 (ENSG00000163884) and TRIM63 (ENSG00000158022) (Figure 2.E). These two genes are present in the GO term GO:0014888 (striated muscle adaptation) to which ARID5B is not associated. The function of ARID58 is well known in adipocytes and hepatocytes but is still rarely studied in skeletal muscle metabolism. However, the knockout of this gene in mice has shown structural defects in the sarcomere structure [45]. Coupled with the results of GWENA, this may corroborate the involvement of ARID5B in the adaptation of striated muscle in response to a stimulus. Moreover, it has recently been shown that ARID5B knockout in mice was associated with increased glucose metabolism via an increased translocation of SLC2A4 (ENSG00000181856)[46]. Since SLC2A4 is a gene that is also regulated by KLF15



E

[47, 48], this supports the idea that ARID5B has implications in skeletal muscle function and more precisely in glucose metabolism. GWENA thus allowed the identification of a gene that may give new insight in the muscle development and growth which needs to be confirmed by further experiments.

source	term name	p val.
GO:BP	response to hormone	0.0015
GO:BP	negative regulation of muscle hypertrophy	0.0033
GO:BP	muscle adaptation	0.0118
GO:BP	response to peptide hormone	0.0129
GO:BP	striated muscle adaptation	0.0255
GO:BP	platelet-derived growth factor receptor signaling pathway	0.0328
GO:BP	regulation of muscle adaptation	0.0434
GO:MF	enzyme binding	0.0097
MIRNA	hsa-miR-6882-5p	0.0002
MIRNA	hsa-miR-197-5p	0.0039
MIRNA	hsa-miR-152-5p	0.0125
MIRNA	hsa-miR-6878-5p	0.0282
REAC	Regulation of FOXO transcriptional activity by acetylation	0.0126
TF	Factor: Zbtb37; motif: NYACCGCRNTCACCGCR; match class: 1	0.0073
TF	Factor: RNF96; motif: BCCCGCRGCC; match class: 1	0.0074
TF	Factor: ETF; motif: GVGGMGG; match class: 1	0.0193
TF	Factor: AP-2; motif: SNNNCCNAGGCN	0.0306
TF	Factor: AP-2; motif: SNNNCCNAGGCN; match class: 0	0.0306

**Table 2** Module 19 young enriched terms table. Multiple enrichment are linked to muscle development and growth

### Multiple conditions modules comparison and analysis

Differential expression analysis allowed to detect genes involved in aging in the last years (GenAge [49], Digital Ageing Atlas [50]). Such discriminant analysis is limited in helping to understand aging as this phenomenon is composed of concomitant mechanisms [51]. Understanding the interaction between the genes is therefore crucial to determine the altered functions and the changes involved. Differential GCN between conditions overcomes this problem by detecting the subtle pattern modifications. Using our previously defined young (20 to 30 years old) and old (50 to 60 years old) skeletal muscle modules, we ran GWENA's GCN differential co-expression functionality to compare the modules between these age ranges. The GCN of each module detected in the young sub-population were taken as a reference and tested against the ones detected in the old sub-population.

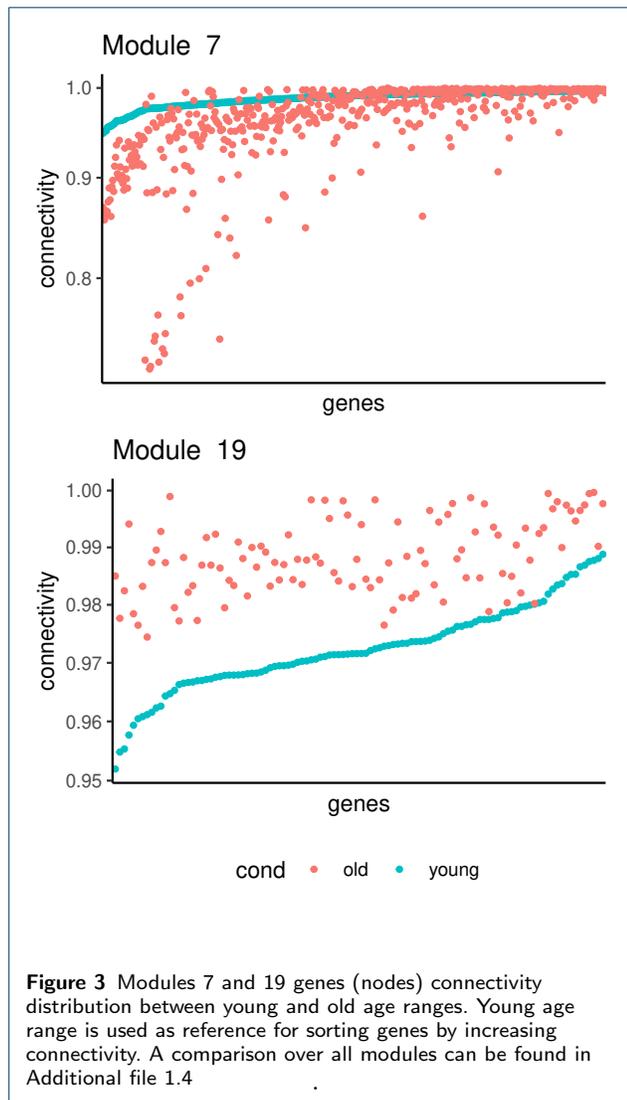
Comparison status	# modules	Modules id
preserved	11	1, 2, 3, 4, 6, 8, 9, 11, 12, 14, 19
moderately preserved	17	7, 10, 13, 17, 18, 20, 21, 22, 23, 24, 25, 27, 28, 30, 31, 33, 34
unpreserved	2	16, 32
inconclusive	4	5, 15, 26, 29

**Table 3** Modules comparison between young and old age range and their comparison status

From the 35 modules detected in the previously described single condition analysis of young muscle data, this differential GCN of young vs. old age range returned 2 modules that were unpreserved, 17 modules that were moderately preserved, 11 modules that were preserved, and 4 that were inconclusive (Table 3, Additional file 1.1). Unpreserved and moderately preserved modules are the most promising for identifying groups of genes differently expressed with age. Few and heterogeneous significant enrichment terms were associated to unpreserved modules while several moderately preserved modules had enrichment terms known to be linked to aging [52, 53, 54, 51] such as transcription regulation (module 21), cellular stress (modules 20 and 27), immune response (modules 7 and 28), cell proliferation (module 13).

In addition to this biological information, the topological comparison of these modules allows to grasp the nature of the interactions variations between the genes (nodes in the network) and their co-expression score (edge weight in the network). Connectivity, as defined by J. Dong and S. Horwath [55], is a common topological metric computed in GCN as it is representative of the network robustness and is known to be linked to network deregulation [56, 57]. Over all modules, the connectivity of the genes in module 7 was noticeably dropping between young and old age range (Figure 3, Additional file 1.4). Using a co-expression score filter of 0.95, this loss of connectivity materialized in the network through a disconnection (edge loss) of peripheral genes (genes with low degree) such as in sub-module 4 between the young and old age range (Figure 4.A and B.). Several other genes of the module 7 also showed an increased connectivity, therefore a reconnection (edge gain) to other genes. These results confirm the observations from previous studies of a connectivity loss in the network of modules linked to aging [58, 57]. Overall, they support an alteration of the transcription regulation.

The overall observation of gain and loss of connectivity with aging also suggests that there may be non significant variation in connectivity despite a reorganization of connections between genes. Since many genes are involved in different biological functions (e.g., through post-translational modification), it is



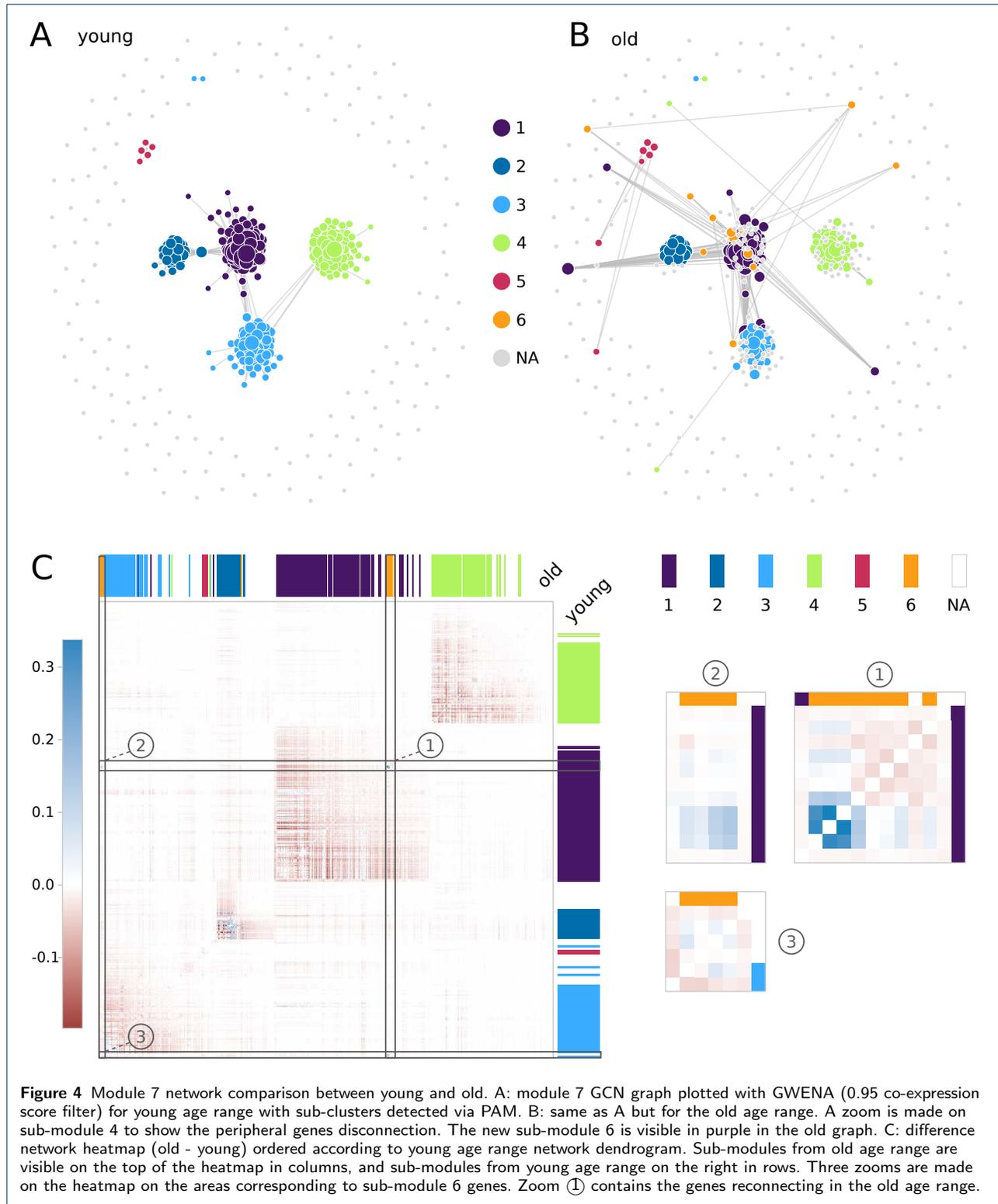
possible that with aging there may be a change in co-expression patterns between genes to activate biological functions other than the initial one without losing connectivity. Genes in sub areas of the module's network in the younger age range would then disconnect from some genes to reconnect to other genes in the old age range, forming different sub areas.

Going beyond GWENA's functionalities, we investigated this reorganization of the gene connections occurring between the young and old network in the module 7. The reorganizations in sub areas were investigated through a partitioning around medoids (PAM) clustering method [59, 60] on the filtered co-expression matrix (0.95 score filter). The clustering applied on both conditions returned 5 optimal sub-modules for the young, and 6 optimal sub-modules for the old. The gene composition of the sub-modules was highly similar between the condition (at least 81% common

genes). Most of the differences in the gene composition are due to the disconnection of peripheral genes as previously spotted, and a small portion of the differences are due to the reconnection of genes or their attribution to another sub-module (Figure 4.A and B).

The new sub-module (sub-module 6, 4.B) appearing in the old age range is especially interesting as its creation is at the expense of the sub-module 1 of the young age range. Of the 13 genes composing it, 8 of the genes are from the sub-module 1, 3 are reconnecting genes, and 2 are from sub-modules 2 and 3. A gene set enrichment analysis with GWENA of this sub-module 6 revealed significant enrichment in functions related to wound healing, coagulation, vessel diameter, platelet degranulation, and plasminogen activation (Additional file 1.5). This is coherent with known morphological alterations of the vascular system in the aged skeletal muscle [61, 62], and the global immune/inflammatory response increased in aging [54]. Also, 51 enrichments from young sub-module 1 were not found significant in any of the sub-modules in the old age range. These enrichments involve antibacterial humoral response and negative regulation of endopeptidase activity. These terms are known to be associated with satellite cells (muscle stem cells responsible for muscle regeneration) regulators released by the vasculature in higher quantity in young skeletal muscle [62].

To complement these analyses, we investigated the variations of co-expression scores leading to the appearance of sub-module 6 in the old age range. Using the network co-expression matrix  $\delta$  returned by GWENA for each condition, a co-expression difference matrix (Figure 4.C) was computed such as  $\delta_{old} - \delta_{young}$ . In this matrix, gene pairs with a negative score indicates a decrease in the co-expression over aging while a positive score indicates an increase in the co-expression. Among the variations, 3 genes showed a significant increase in co-expression between them but also towards other genes of sub-module 6. The pattern visible in Figure 4.C ① and ② suggest that these genes may be driving the co-expression changes occurring in this sub-module. These genes are FGG (ENSG00000171557), FGA (ENSG00000171560), and FGB (ENSG00000171564), the three fibrinogen chain coding genes involved in the polymerization of a fibrin matrix. This finding is consistent with previous studies about the increasing fibrinogen content in the elderly skeletal muscle leading to persistent fibrin deposition preventing myofiber repair [63, 64]. They also support their hypothesis of an inflammatory response triggered by a fibrin accumulation. All these results tend to support the idea of not only a global loss in connectivity in aging but also of a gene co-expression reorganization leading to the expression of other functions involved in aging.



GWENA's contribution and comparison with existing tools

Weighted GCN can be computed from existing tools such as WGCNA [3], wTO [65], CEMiTool [66]. As

both GWENA and CEMiTool use elements from WGCNA, they share notable functionalities. They use similar network construction and modules detection

Functionalities	GWENA	WGCNA	CEMiTool	wTO
Gene set enrichment				
- Gene ontology	yes	yes	no	no
- Pathways (KEGG/Reactome)	yes	no	no	no
- Regulation actors (TRANSFAC/miRTarBase)	yes	no	no	no
- Protein databases (Human Protein Atlas/CORUM)	yes	no	no	no
- Custom GMT import	yes	no	yes	no
Native network visualization	yes	no	no <sup>1</sup>	yes
Phenotype association	yes	yes	yes	no
Hub gene detection	yes	yes <sup>2</sup>	yes <sup>3</sup>	no
Igraph compatibility for extended topology metrics calculation	yes	no	no	no
Modules differential co-expression	yes	yes <sup>4</sup>	no	no <sup>5</sup>

**Table 4** Key functionalities differences for module analysis between GWENA, WGCNA, CEMiTool, and wTO. <sup>1</sup> CEMiTool allow network visualization only if a protein-protein interaction network file is provided. <sup>2</sup> Provide only a single hub gene selection by module. <sup>3</sup> Provide the top 10 hub genes independently of the module size or connectivity. <sup>4</sup> Differential co-expression without multiple test correction. <sup>5</sup> No differential co-expression method available. However provides a consensus network method.

functions from WGCNA but offer their own filter on the datasets. On its side, wTO used a different version of a topological score to construct the network as they don't perform a power law conversion on the correlation matrix and don't use the same definition of topological transformation. Therefore, the main differences between WGCNA, wTO, CEMiTool and GWENA lie in the added functionalities for module analysis.

Regarding biological integration, wTO provides neither phenotypic association nor gene set enrichment (Table 4). The other three tools allow phenotypic association but differ on gene set enrichment analysis. While CEMiTool only allows enrichment on imported GMTs, WGCNA and GWENA allow enrichment on gene ontology. GWENA also allows enrichment on other databases of pathways, regulatory agents, and proteins (in addition to imported GMTs).

Additional topological analysis functions are also available in several of these tools. The most common, hub gene detection, is present in WGCNA, CEMiTool, and GWENA in different forms. CEMiTool and WGCNA offer respectively as hub gene the top 10 most connected genes and genes with a top kME score (membership module based on eigengene). However, methods based on a fixed number of hub genes tend to bias the information since the number of hub genes can vary according to the number of genes present in the module. GWENA therefore proposes several methods (highest connectivity, superior degree, Kleinberg's score) based on a selection of genes with a hub score above a threshold. Finally, GWENA includes a differential co-expression analysis in the analysis pipeline as opposed to packages dedicated solely to it (DiffCoEx [67], CoDiNA [68], CoXpress [69]). The method in GWENA differs from the one present in WGCNA in that it includes a permutation test to prevent the problem of multi-testing. With the addition of the Z-summary score to detect unpreserved modules,

GWENA is therefore the only pipeline including a differential co-expression analysis with high confidence in modules found unpreserved.

GWENA, as other GCN analysis tool has limitations. A first one common to all GCN construction method is that the quality of input data (e.g. filtration and/or proper normalization) will inevitably bias the results, especially if it breaks the scale-free property. A second limitation is the design of the permutation test that prevents reporting a significant unpreservation. The non-rejection of the null hypothesis of unpreservation can only state a lack of evidence of preservation. Therefore, unpreserved modules are determined among these modules lacking evidence (the non-significant modules) by the calculation of Z summary which only provide a tendency in the unpreservation [40]. The present application of GWENA to skeletal muscle aging also presents its own limitation. All analyses were performed on skeletal muscle sample and results were commented regarding this context. However, to be sure of the specificity of the findings, an additional differential co-expression of the modules should be performed on samples from other tissues from subjects with similar age range. As single-cell technologies are becoming common, the differential co-expression could also be used to target the cell-to-cell specific aging variation inside a tissue. Finally, as co-expression networks were unsigned and aging is a complex phenomenon involving actors beyond gene expression, causal effect of any finding need to be experimentally verified.

## Conclusion

In this paper, we introduced GWENA, an R package on Bioconductor to construct and analyze GCN in a single pipeline through a whole range of tools from biological integration, topological analysis, and differential co-expression. The package eases the GCN analysis

through simple input and output functions combined to a set of visualizations to explore the results. The separation of each step of the analysis in one function also allows quick and easy replacement if users wish to use another method for this block.

GWENA demonstrated its performances on both single and multiple condition analysis through an exploration of variations of skeletal muscle function and processes in aging. The single condition analysis showed it is possible to find new genes potentially involved in an existing GO annotation using hub genes, network neighboring genes and enrichments. The differential co-expression analysis between young and old samples isolated modules specifically linked to aging and detected the rearrangement in connectivity related to aging. Additional analysis supported the observed genes co-expression reorganization beyond simple connectivity loss. This resulted in a reinforcement of previous supposition on inflammatory response to fibrin increases in skeletal muscle aging.

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable

#### Availability of data and materials

- Project name : Gene Whole co-Expression Network Analysis (GWENA)
- Operating system(s): Platform independent
- Programming language: R
- Package on Bioconductor : <https://bioconductor.org/packages/devel/bioc/html/GWENA.html>.
- Package GitHub repository (development, issues, and pull requests) : <https://github.com/Kumquat/GWENA>.
- Licence : GPL-3
- Source code for GTEx analysis : [https://github.com/Kumquat/GTEEx\\_v8\\_GWENA/](https://github.com/Kumquat/GTEEx_v8_GWENA/).
- GTEx v8 public data (RNA-seq and anonymized phenotype) : <https://gtexportal.org/home/datasets>.
- GTEx detailed phenotype requires a dbGaP request.

#### Competing interests

The authors declare that they have no competing interests.

#### Funding

This work was supported by the L'Oreal Research and Innovation chair in Digital Biology.

#### Author's contributions

Conceptualization: LGG. Data curation: LGG. Formal analysis: LGG. Methodology: LGG, MPSB. Software: LGG. Supervision: MPSB. Validation: LGG, MPSB, OP, AD. Funding acquisition: AD. Writing: LGG. Proofreading and correction: MPSB, AB, OP, AD. All authors read and approved the final manuscript

#### Acknowledgements

We thanks Régis Ongaro-Carcy, Julien Prunier, and Audrey Lemaçon for the enriching discussions on the conceptualization and suggestions on the package.

#### List of abbreviations

- **GWENA** : Gene Whole co-Expression Network Analysis
- **GCN** : Co-expression Network
- **GO** : Gene Ontology
- **GO:MF** : GO Molecular Function

- **GO:BP** : GO Biological Process
- **GO:CC** : GO Cellular Compartment
- **KEGG** : Kyoto Encyclopedia of Genes and Genomes
- **REAC** : Reactome
- **WP** : WikiPathwaya
- **TF** : TRANSFAC
- **MIRNA** : mirTarBase
- **HPA** : Human Protein Atlas
- **CORUM** : Comprehensive Resource of Mammalian protein complexes
- **HP** : Human Phenotype Ontology
- **GMT** : Gene Matrix Transposed file format
- **DE** : Differential Expression
- **TOM** : Topological Overlap Matrix
- **ORA** : Over Representation Analysis
- **BMI** : Body Mass Index
- **ENSG** : Ensemble Gene ID (human)

#### Competing interests

The authors declare that they have no competing interests.

#### Funding

This work was supported by L'Oreal Research and Innovation Chair in Digital Biology.

#### Author details

<sup>1</sup>Département de médecine moléculaire, Faculté de médecine, Université Laval, 2325 rue de l'Université, G1V 0A6 Québec, Canada. <sup>2</sup>Centre de recherche du Chu de Québec-Université Laval, 2705 boulevard Laurier Québec, G1V 4G2 Québec, Canada. <sup>3</sup>L'Oréal Research and Innovation, 15 rue Pierre Dreyfus, 92110 Clichy, France.

#### References

1. Barabási, A.L., Oltvai, Z.N.: Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics* **5**(2), 101–113 (2004). doi:[10.1038/nrg1272](https://doi.org/10.1038/nrg1272)
2. Hartwell, L.H., Hopfield, J.J., Leibler, S., Murray, A.W.: From molecular to modular cell biology. *Nature* **402**(6761 SUPPL. 1), 47–52 (1999). doi:[10.1038/35011540](https://doi.org/10.1038/35011540)
3. Langfelder, P., Horvath, S.: WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**(1), 559 (2008). doi:[10.1186/1471-2105-9-559](https://doi.org/10.1186/1471-2105-9-559)
4. Mao, L., Van Hemert, J.L., Dash, S., Dickerson, J.A.: Arabidopsis gene co-expression network and its functional modules. *BMC Bioinformatics* **10**, 1–24 (2009). doi:[10.1186/1471-2105-10-346](https://doi.org/10.1186/1471-2105-10-346)
5. Tang, J., Kong, D., Cui, Q., Wang, K., Zhang, D., Gong, Y., Wu, G.: Prognostic genes of breast cancer identified by gene co-expression network analysis. *Frontiers in Oncology* **8**(SEP), 1–13 (2018). doi:[10.3389/fonc.2018.00374](https://doi.org/10.3389/fonc.2018.00374)
6. van Dam, S., Vösa, U., van der Graaf, A., Franke, L., de Magalhães, J.P.: Gene co-expression analysis for functional classification and gene–disease predictions. *Briefings in Bioinformatics* (December 2016), 139 (2017). doi:[10.1093/bib/bbw139](https://doi.org/10.1093/bib/bbw139)
7. Zhang, B., Gaiteri, C., Bodea, L.G., Wang, Z., McElwee, J., Podtelezchnikov, A.A., Zhang, C., Xie, T., Tran, L., Dobrin, R., Fluder, E., Clurman, B., Melquist, S., Narayanan, M., Suver, C., Shah, H., Mahajan, M., Gillis, T., Mysore, J., MacDonald, M.E., Lamb, J.R., Bennett, D.A., Molony, C., Stone, D.J., Gudnason, V., Myers, A.J., Schadt, E.E., Neumann, H., Zhu, J., Emilsson, V.: Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* **153**(3), 707–720 (2013). doi:[10.1016/j.cell.2013.03.030](https://doi.org/10.1016/j.cell.2013.03.030)
8. Tseng, G.C., Sibille, E., Gaiteri, C., Ding, Y., French, B.: Beyond modules and hubs: the potential of gene coexpression networks for investigating molecular mechanisms of complex brain disorders. *Genes, Brain and Behavior* **13**(1), 13–24 (2013). doi:[10.1111/gbb.12106](https://doi.org/10.1111/gbb.12106)
9. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G.: Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**(1), 25–29 (2000). doi:[10.1038/75556](https://doi.org/10.1038/75556)
10. Fabregat, A., Sidiropoulos, K., Garapati, P., Gillespie, M., Hausmann, K., Haw, R., Jassal, B., Jupe, S., Korninger, F., McKay, S., Matthews,

- L., May, B., Milacic, M., Rothfels, K., Shamovsky, V., Webber, M., Weiser, J., Williams, M., Wu, G., Stein, L., Hermjakob, H., D'Eustachio, P.: The reactome pathway knowledgebase. *Nucleic Acids Research* **44**(D1), 481–487 (2016). doi:[10.1093/nar/gkv1351](https://doi.org/10.1093/nar/gkv1351)
11. Pierson, E., Koller, D., Battle, A., Mostafavi, S.: Sharing and Specificity of Co-expression Networks across 35 Human Tissues. *PLoS Computational Biology* **11**(5), 1–19 (2015). doi:[10.1371/journal.pcbi.1004220](https://doi.org/10.1371/journal.pcbi.1004220)
  12. Hahn, M.W., Kern, A.D.: Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Molecular Biology and Evolution* **22**(4), 803–806 (2005). doi:[10.1093/molbev/msi072](https://doi.org/10.1093/molbev/msi072)
  13. Chowdhury, H.A., Bhattacharyya, D.K., Kalita, J.K.: (Differential) Co-Expression Analysis of Gene Expression: A Survey of Best Practices. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **1**(c), 1–1 (2019). doi:[10.1109/tcbb.2019.2893170](https://doi.org/10.1109/tcbb.2019.2893170)
  14. Gov, E., Arga, K.Y.: Differential co-expression analysis reveals a novel prognostic gene module in ovarian cancer. *Scientific Reports* **7**(1), 1–10 (2017). doi:[10.1038/s41598-017-05298-w](https://doi.org/10.1038/s41598-017-05298-w)
  15. Bhuva, D.D., Cursons, J., Smyth, G.K., Davis, M.J.: Differential co-expression-based detection of conditional relationships in transcriptional data: Comparative analysis and application to breast cancer. *Genome Biology* **20**(1), 1–21 (2019). doi:[10.1186/s13059-019-1851-8](https://doi.org/10.1186/s13059-019-1851-8)
  16. Yan, Q., Wu, F., Yan, Z., Li, J., Ma, T., Zhang, Y., Zhao, Y., Wang, Y., Zhang, J.: Differential co-expression networks of long non-coding RNAs and mRNAs in *Cleistogenes songorica* under water stress and during recovery. *BMC Plant Biology* **19**(1), 1–19 (2019). doi:[10.1186/s12870-018-1626-5](https://doi.org/10.1186/s12870-018-1626-5)
  17. Bulut, E.A., Soysal, P., Aydin, A.E., Dokuzlar, O., Kocyigit, S.E., Isik, A.T.: Vitamin B12 deficiency might be related to sarcopenia in older adults. *Experimental Gerontology* **95**, 136–140 (2017). doi:[10.1016/j.exger.2017.05.017](https://doi.org/10.1016/j.exger.2017.05.017)
  18. Santilli, V., Bernetti, A., Mangone, M., Paoloni, M.: Clinical definition of sarcopenia. *Clinical Cases in Mineral and Bone Metabolism* **11**(3), 177–180 (2014). doi:[10.11138/ccmbm/2014.11.3.177](https://doi.org/10.11138/ccmbm/2014.11.3.177)
  19. Janssen, I., Heymsfield, S.B., Ross, R.: Low relative skeletal muscle mass (sarcopenia) in older persons is associated with functional impairment and physical disability. *Journal of the American Geriatrics Society* **50**(5), 889–896 (2002). doi:[10.1046/j.1532-5415.2002.50216.x](https://doi.org/10.1046/j.1532-5415.2002.50216.x)
  20. Sakuma, K., Aoi, W., Yamaguchi, A.: Molecular mechanism of sarcopenia and cachexia: recent research advances. *Pflügers Archiv European Journal of Physiology* **469**(5-6), 573–591 (2017). doi:[10.1007/s00424-016-1933-3](https://doi.org/10.1007/s00424-016-1933-3)
  21. Jiao, J., Demontis, F.: Skeletal muscle autophagy and its role in sarcopenia and organismal aging. *Current Opinion in Pharmacology* **34**, 1–6 (2017). doi:[10.1016/j.coph.2017.03.009](https://doi.org/10.1016/j.coph.2017.03.009)
  22. Morgan, M., Obenchain, V., Hester, J., Pagès, H.: SummarizedExperiment: Summarized-Experiment container, 2018 (2018)
  23. Langfelder P, H.S.: Frequently Asked Questions. Accessed: 2020-08-26 (2014). <https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/faq.html>
  24. Miller, J.A., Cai, C., Langfelder, P., Geschwind, D.H., Kurian, S.M., Salomon, D.R., Horvath, S.: Strategies for aggregating gene expression data: the collapseRows function. *BMC bioinformatics* **12**(1), 1–13 (2011)
  25. Parsana, P., Ruberman, C., Jaffe, A.E., Schatz, M.C., Battle, A., Leek, J.T.: Addressing confounding artifacts in reconstruction of gene co-expression networks. *Genome Biology* **20**(1), 94 (2019). doi:[10.1186/s13059-019-1700-9](https://doi.org/10.1186/s13059-019-1700-9)
  26. Song, L., Langfelder, P., Horvath, S.: Comparison of co-expression measures: Mutual information, correlation, and model based indices. *BMC Bioinformatics* **13**(1) (2012). doi:[10.1186/1471-2105-13-328](https://doi.org/10.1186/1471-2105-13-328)
  27. Yip, A.M., Horvath, S.: Gene network interconnectedness and the generalized topological overlap measure. *BMC Bioinformatics* **8**, 1–14 (2007). doi:[10.1186/1471-2105-8-22](https://doi.org/10.1186/1471-2105-8-22)
  28. Ravasz, E., Barabási, A.L.: Hierarchical organization in complex networks. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* **67**(2), 7 (2003). doi:[10.1103/PhysRevE.67.026112](https://doi.org/10.1103/PhysRevE.67.026112). [0206130](https://doi.org/10.206112)
  29. Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., Vilo, J.: g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research* **47**(W1), 191–198 (2019). doi:[10.1093/nar/gkz369](https://doi.org/10.1093/nar/gkz369)
  30. New approach for understanding genome variations in KEGG. *Nucleic Acids Research* **47**(D1), 590–595 (2019). doi:[10.1093/nar/gky962](https://doi.org/10.1093/nar/gky962)
  31. Matys, V., Kel-Margoulis, O.V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K., Voss, N., Stegmaier, P., Lewicki-Potapov, B., Saxel, H., Kel, A.E., Wingender, E.: TRANSFAC and its module TRANSCmpel: transcriptional gene regulation in eukaryotes. *Nucleic acids research* **34**(Database issue), 108–110 (2006). doi:[10.1093/nar/gkj143](https://doi.org/10.1093/nar/gkj143)
  32. Chou, C.H., Shrestha, S., Yang, C.D., Chang, N.W., Lin, Y.L., Liao, K.W., Huang, W.C., Sun, T.H., Tu, S.J., Lee, W.H., Chiew, M.Y., Tai, C.S., Wei, T.Y., Tsai, T.R., Huang, H.T., Wang, C.Y., Wu, H.Y., Ho, S.Y., Chen, P.R., Chuang, C.H., Hsieh, P.J., Wu, Y.S., Chen, W.L., Li, M.J., Wu, Y.C., Huang, X.Y., Ng, F.L., Buddhakosai, W., Huang, P.C., Lan, K.C., Huang, C.Y., Weng, S.L., Cheng, Y.N., Liang, C., Hsu, W.L., Huang, H.D.: MiRTarBase update 2018: A resource for experimentally validated microRNA-target interactions. *Nucleic Acids Research* **46**(D1), 296–302 (2018). doi:[10.1093/nar/gkx1067](https://doi.org/10.1093/nar/gkx1067)
  33. Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigartyo, C.A.-K., Odeberg, J., Djureinovic, D., Takanen, J., Hober, S., Alm, T., Edqvist, P.-H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J.M., Hamsten, M., von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., von Heijne, G., Nielsen, J., Ponten, F.: Tissue-based map of the human proteome. *Science* **347**(6220), 1260419–1260419 (2015). doi:[10.1126/science.1260419](https://doi.org/10.1126/science.1260419). [0208024](https://doi.org/10.208024)
  34. Ruepp, A., Brauner, B., Dunger-Kaltenbach, I., Frishman, G., Montrone, C., Strassky, M., Waagele, B., Schmidt, T., Doudieu, O.N., Stumpflen, V., Mewes, H.W.: CORUM: The comprehensive resource of mammalian protein complexes. *Nucleic Acids Research* **36**(SUPPL. 1), 646–650 (2008). doi:[10.1093/nar/gkm936](https://doi.org/10.1093/nar/gkm936)
  35. Köhler, S., Carmody, L., Vasilevsky, N., Jacobsen, J.O.B., Danis, D., Gouridine, J.P., Gargano, M., Harris, N.L., Matentzoglou, N., McMurry, J.A., Osumi-Sutherland, D., Cipriani, V., Balhoff, J.P., Conlin, T., Blau, H., Baynam, G., Palmer, R., Gratian, D., Dawkins, H., Segal, M., Jansen, A.C., Muaz, A., Chang, W.H., Bergerson, J., Laulederkind, S.J.F., Yüksel, Z., Beltran, S., Freeman, A.F., Sergouniotis, P.I., Durkin, D., Storm, A.L., Hanauer, M., Brudno, M., Bello, S.M., Sincan, M., Rageth, K., Wheeler, M.T., Oegema, R., Lourghi, H., Della Rocca, M.G., Thompson, R., Castellanos, F., Priest, J., Cunningham-Rundles, C., Hegde, A., Lovering, R.C., Hajek, C., Oly, A., Notarangelo, L., Similuk, M., Zhang, X.A., Gómez-Andrés, D., Lochmüller, H., Dollfus, H., Rosenzweig, S., Marwaha, S., Rath, A., Sullivan, K., Smith, C., Milner, J.D., Leroux, D., Boerkoel, C.F., Klion, A., Carter, M.C., Groza, T., Smedley, D., Haendel, M.A., Mungall, C., Robinson, P.N.: Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Research* **47**(D1), 1018–1027 (2019). doi:[10.1093/nar/gky1105](https://doi.org/10.1093/nar/gky1105)
  36. Slenter, D.N., Kutmon, M., Hanspers, K., Riutta, A., Windsor, J., Nunes, N., Mélius, J., Cirillo, E., Coort, S.L., Dlgles, D., Ehrhart, F., Giesbertz, P., Kalafati, M., Martens, M., Miller, R., Nishida, K., Rieswijk, L., Waagmeester, A., Eijssen, L.M.T., Evelo, C.T., Pico, A.R., Willighagen, E.L.: WikiPathways: A multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Research* **46**(D1), 661–667 (2018). doi:[10.1093/nar/gkx1064](https://doi.org/10.1093/nar/gkx1064)
  37. Gabor, C., Tamas, N.: The igrph software package for complex network research. *InterJournal Complex Systems* **1695** (2006)
  38. Ritchie, S.C., Watts, S., Fearnley, L.G., Holt, K.E., Abraham, G., Inouye, M.: A Scalable Permutation Approach Reveals Replication and Preservation Patterns of Network Modules in Large Datasets. *Cell Systems* **3**(1), 71–82 (2016). doi:[10.1016/j.cels.2016.06.012](https://doi.org/10.1016/j.cels.2016.06.012)
  39. Phipson, B., Smyth, G.K.: Permutation p-values should never be zero: calculating exact p-values when permutations are randomly drawn. *Statistical applications in genetics and molecular biology* **9**(1) (2010)
  40. Langfelder, P., Luo, R., Oldham, M.C., Horvath, S.: Is my network module preserved and reproducible? *PLoS Computational Biology* **7**(1)

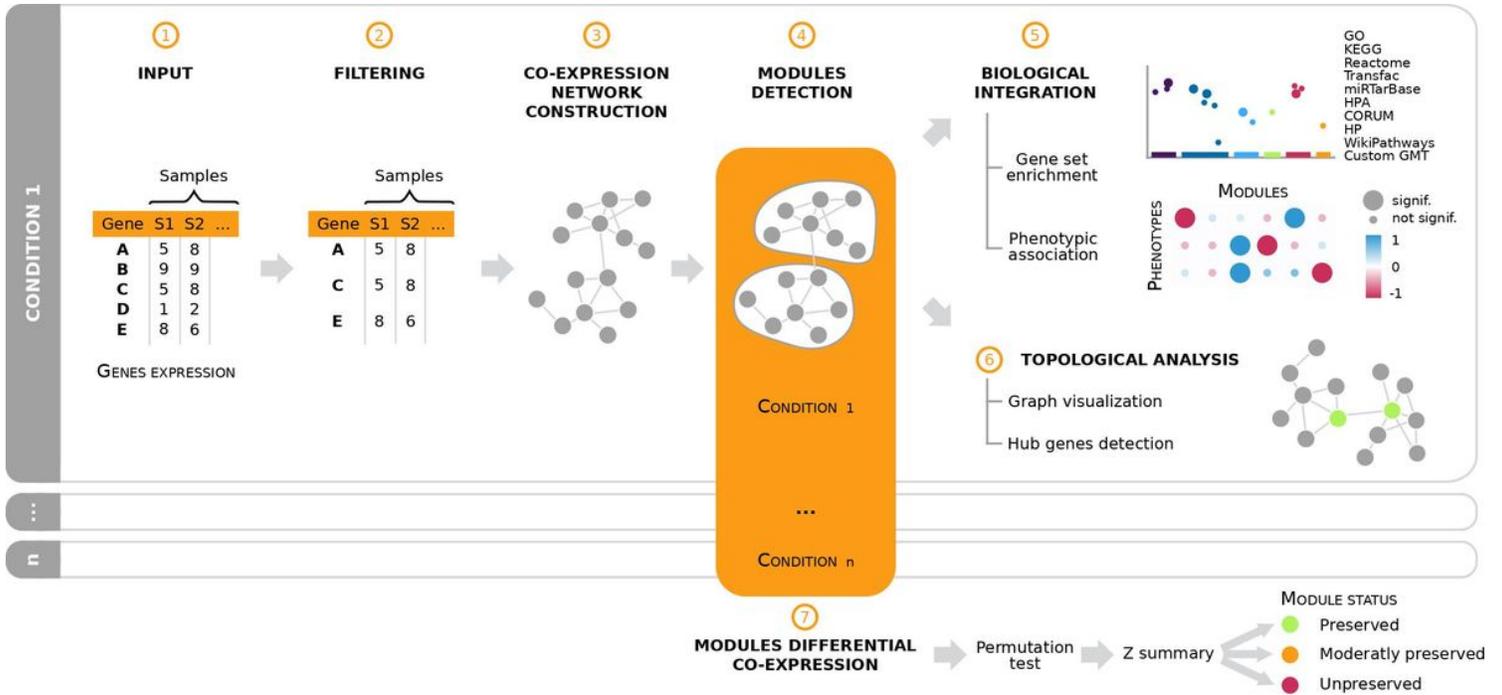
- (2011). doi:[10.1371/journal.pcbi.1001057](https://doi.org/10.1371/journal.pcbi.1001057)
41. Li, B., Zhang, Y., Yu, Y., Wang, P., Wang, Y., Wang, Z., Wang, Y.: Quantitative assessment of gene expression network module-validation methods. *Scientific Reports* **5**(1), 15258 (2015). doi:[10.1038/srep15258](https://doi.org/10.1038/srep15258)
  42. Ardlie, K.G., DeLuca, D.S., Segrè, A.V., Sullivan, T.J., Young, T.R., Gelfand, E.T., Trowbridge, C.A., Maller, J.B., Tukiainen, T., Lek, M., Ward, L.D., Kheradpour, P., Iriarte, B., Meng, Y., Palmer, C.D., Esko, T., Winckler, W., Hirschhorn, J.N., Kellis, M., MacArthur, D.G., Getz, G., Shabalin, A.A., Li, G., Zhou, Y.H., Nobel, A.B., Rusyn, I., Wright, F.A., Lappalainen, T., Ferreira, P.G., Ongen, H., Rivas, M.A., Battle, A., Mostafavi, S., Monlong, J., Sammeth, M., Melé, M., Reverter, F., Goldmann, J.M., Koller, D., Guigó, R., McCarthy, M.I., Dermitzakis, E.T., Gamazon, E.R., Im, H.K., Konkashbaev, A., Nicolae, D.L., Cox, N.J., Flutre, T., Wen, X., Stephens, M., Pritchard, J.K., Tu, Z., Zhang, B., Huang, T., Long, Q., Lin, L., Yang, J., Zhu, J., Liu, J., Brown, A., Mesticelli, B., Tidwell, D., Lo, E., Salvatore, M., Shad, S., Thomas, J.A., Lonsdale, J.T., Moser, M.T., Gillard, B.M., Karasik, E., Ramsey, K., Choi, C., Foster, B.A., Syron, J., Fleming, J., Magazine, H., Hasz, R., Walters, G.D., Bridge, J.P., Miklos, M., Sullivan, S., Barker, L.K., Traino, H.M., Mosavel, M., Siminoff, L.A., Valley, D.R., Rohrer, D.C., Jewell, S.D., Branton, P.A., Sobin, L.H., Barcus, M., Qi, L., McLean, J., Hariharan, P., Um, K.S., Wu, S., Tabor, D., Shive, C., Smith, A.M., Buia, S.A., Undale, A.H., Robinson, K.L., Roche, N., Valentino, K.M., Britton, A., Burges, R., Bradbury, D., Hambright, K.W., Seleski, J., Korzeniewski, G.E., Erickson, K., Marcus, Y., Tejada, J., Taherian, M., Lu, C., Basile, M., Mash, D.C., Volpi, S., Struewing, J.P., Temple, G.F., Boyer, J., Colantuoni, D., Little, R., Koester, S., Carithers, L.J., Moore, H.M., Guan, P., Compton, C., Sawyer, S.J., Demchok, J.P., Vaught, J.B., Rabiner, C.A., Lockhart: The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**(6235), 648–660 (2015). doi:[10.1126/science.1262110](https://doi.org/10.1126/science.1262110)
  43. Nieuwenhuis, T.O., Yang, S.Y., Verma, R.X., Pillalamarri, V., Arking, D.E., Rosenberg, A.Z., McCall, M.N., Halushka, M.K.: Consistent RNA sequencing contamination in GTEx and other data sets. *Nature Communications* **11**(1) (2020). doi:[10.1038/s41467-020-15821-9](https://doi.org/10.1038/s41467-020-15821-9)
  44. Somekh, J., Shen-Orr, S.S., Kohane, I.S.: Batch correction evaluation framework using a-priori gene-gene associations: Applied to the GTEx dataset. *BMC Bioinformatics* **20**(1), 1–10 (2019). doi:[10.1186/s12859-019-2855-9](https://doi.org/10.1186/s12859-019-2855-9)
  45. Murray, J., Whitson, R.H., Itakura, K.: Reduced prostaglandin I2 signaling in Arid5b2/2 primary skeletal muscle cells attenuates myogenesis. *FASEB Journal* **32**(4), 1868–1879 (2018). doi:[10.1096/fj.201700453RR](https://doi.org/10.1096/fj.201700453RR)
  46. Okazaki, Y., Murray, J., Ehsani, A., Clark, J., Whitson, R.H., Hirose, L., Yanaka, N., Itakura, K.: Increased glucose metabolism in Arid5b -/- skeletal muscle is associated with the down-regulation of TBC1 domain family member 1 (TBC1D1). *Biological Research* **53**(1), 1–14 (2020). doi:[10.1186/s40659-020-00313-3](https://doi.org/10.1186/s40659-020-00313-3)
  47. Gray, S., Feinberg, M.W., Hull, S., Kuo, C.T., Watanabe, M., Sen, S., Depina, A., Haspel, R., Jain, M.K.: The Krüppel-like factor KLF15 regulates the insulin-sensitive glucose transporter GLUT4. *Journal of Biological Chemistry* **277**(37), 34322–34328 (2002). doi:[10.1074/jbc.M201304200](https://doi.org/10.1074/jbc.M201304200)
  48. Fan, L., Hsieh, P.N., Sweet, D.R., Jain, M.K.: Krüppel-like factor 15: Regulator of BCAA metabolism and circadian protein rhythmicity. *Pharmacological Research* **130**, 123–126 (2018). doi:[10.1016/j.phrs.2017.12.018](https://doi.org/10.1016/j.phrs.2017.12.018)
  49. Tacutu, R., Thornton, D., Johnson, E., Budovsky, A., Barardo, D., Craig, T., Dlana, E., Lehmann, G., Toren, D., Wang, J., Fraifeld, V.E., De Magalhães, J.P.: Human Ageing Genomic Resources: New and updated databases. *Nucleic Acids Research* **46**(D1), 1083–1090 (2018). doi:[10.1093/nar/gkx1042](https://doi.org/10.1093/nar/gkx1042)
  50. Craig, T., Smelick, C., Tacutu, R., Wuttke, D., Wood, S.H., Stanley, H., Janssens, G., Savitskaya, E., Moskalev, A., Arking, R., De Magalhães, J.P.: The Digital Ageing Atlas: Integrating the diversity of age-related changes into a unified resource. *Nucleic Acids Research* **43**(D1), 873–878 (2015). doi:[10.1093/nar/gku843](https://doi.org/10.1093/nar/gku843)
  51. Zierer, J., Menni, C., Kastenmüller, G., Spector, T.D.: Integration of 'omics' data in aging research: From biomarkers to systems biology. *Aging Cell* **14**(6), 933–944 (2015). doi:[10.1111/acer.12386](https://doi.org/10.1111/acer.12386)
  52. Kuehne, A., Hildebrand, J., Soehle, J., Wenck, H., Terstegen, L., Gallinat, S., Knott, A., Winnefeld, M., Zamboni, N.: An integrative metabolomics and transcriptomics study to identify metabolic alterations in aged skin of humans in vivo. *BMC Genomics* **18**(1), 169 (2017). doi:[10.1186/s12864-017-3547-3](https://doi.org/10.1186/s12864-017-3547-3)
  53. López-Otín, C., Blasco, M.A., Partridge, L., Serrano, M., Kroemer, G.: The hallmarks of aging. *Cell* **153**(6) (2013). doi:[10.1016/j.cell.2013.05.039](https://doi.org/10.1016/j.cell.2013.05.039). NIHMS150003
  54. de Magalhães, J.P., Curado, J., Church, G.M.: Meta-analysis of age-related gene expression profiles identifies common signatures of aging. *Bioinformatics* **25**(7), 875–881 (2009). doi:[10.1093/bioinformatics/btp073](https://doi.org/10.1093/bioinformatics/btp073)
  55. Dong, J., Horvath, S.: Understanding network concepts in modules. *BMC Systems Biology* **1**, 1–20 (2007). doi:[10.1186/1752-0509-1-24](https://doi.org/10.1186/1752-0509-1-24)
  56. Anglani, R., Creanza, T.M., Liuzzi, V.C., Piepoli, A., Panza, A., Andriulli, A., Ancona, N.: Loss of connectivity in cancer co-expression networks. *PLoS ONE* **9**(1) (2014). doi:[10.1371/journal.pone.0087075](https://doi.org/10.1371/journal.pone.0087075)
  57. Bormann, F., Rodríguez-Paredes, M., Hagemann, S., Manchanda, H., Kristof, B., Gutekunst, J., Raddatz, G., Haas, R., Terstegen, L., Wenck, H., Kaderali, L., Winnefeld, M., Lyko, F.: Reduced DNA methylation patterning and transcriptional connectivity define human skin aging. *Aging Cell* **15**(3), 563–571 (2016). doi:[10.1111/acer.12470](https://doi.org/10.1111/acer.12470)
  58. Southworth, L.K., Owen, A.B., Kim, S.K.: Aging Mice Show a Decreasing Correlation of Gene Expression within Genetic Modules. *PLoS Genetics* **5**(12), 1000776 (2009). doi:[10.1371/journal.pgen.1000776](https://doi.org/10.1371/journal.pgen.1000776)
  59. Kaufmann, L., Rousseeuw, P.: Clustering by Means of Medoids. *Data Analysis based on the L1-Norm and Related Methods*, 405–416 (1987)
  60. Schubert, E., Rousseeuw, P.J.: Faster k-Medoids Clustering: Improving the PAM, CLARA, and CLARANS Algorithms. In: Amato, G., Gennaro, C., Oria, V., Miloš, R. (eds.) *Similarity Search and Applications*, pp. 171–187. Springer, Cham (2019). doi:[10.1007/978-3-030-32047-8\\_16](https://doi.org/10.1007/978-3-030-32047-8_16). [http://link.springer.com/10.1007/978-3-030-32047-8\\_16](http://link.springer.com/10.1007/978-3-030-32047-8_16)
  61. El Assar, M., Angulo, J., Rodríguez-Mañas, L.: Oxidative stress and vascular inflammation in aging. *Free Radical Biology and Medicine* **65**, 380–401 (2013). doi:[10.1016/j.freeradbiomed.2013.07.003](https://doi.org/10.1016/j.freeradbiomed.2013.07.003)
  62. Gopinath, S.D., Rando, T.A.: Stem Cell Review Series: Aging of the skeletal muscle stem cell niche. *Aging Cell* **7**(4), 590–598 (2008). doi:[10.1111/j.1474-9726.2008.00399.x](https://doi.org/10.1111/j.1474-9726.2008.00399.x)
  63. Mann, C.J., Perdiguero, E., Kharraz, Y., Aguilar, S., Pessina, P., Serrano, A.L., Muñoz-Cánoves, P.: Aberrant repair and fibrosis development in skeletal muscle. *Skeletal Muscle* **1**(1), 1–20 (2011). doi:[10.1186/2044-5040-1-21](https://doi.org/10.1186/2044-5040-1-21)
  64. Gligoričević, N., Zámorová Križáková, M., Penezić, A., Katrlík, J., Nedić, O.: Structural and functional changes of fibrinogen due to aging. *International Journal of Biological Macromolecules* **108**, 1028–1034 (2018). doi:[10.1016/j.ijbiomac.2017.11.016](https://doi.org/10.1016/j.ijbiomac.2017.11.016)
  65. Gysi, D.M., Voigt, A., Fragoso, T.D.M., Almaas, E., Nowick, K.: wTO: An R package for computing weighted topological overlap and a consensus network with integrated visualization tool. *BMC Bioinformatics* **19**(1), 1–16 (2018). doi:[10.1186/s12859-018-2351-7](https://doi.org/10.1186/s12859-018-2351-7)
  66. Russo, P.S.T., Ferreira, G.R., Cardozo, L.E., Bürger, M.C., Arias-Carrasco, R., Maruyama, S.R., Hirata, T.D.C., Lima, D.S., Passos, F.M., Fukutani, K.F., Lever, M., Silva, J.S., Maracaja-Coutinho, V., Nakaya, H.I.: CEMiTool: A Bioconductor package for performing comprehensive modular co-expression analyses. *BMC Bioinformatics* **19**(1), 1–13 (2018). doi:[10.1186/s12859-018-2053-1](https://doi.org/10.1186/s12859-018-2053-1)
  67. Tesson, B.M., Breitling, R., Jansen, R.C.: DiffCoEx: A simple and sensitive method to find differentially coexpressed gene modules. *BMC Bioinformatics* **11** (2010). doi:[10.1186/1471-2105-11-497](https://doi.org/10.1186/1471-2105-11-497)
  68. Gysi, D.M., de Miranda Fragoso, T., Zebardast, F., Bertoli, W., Busskamp, V., Almaas, E., Nowick, K.: Whole transcriptomic network analysis using Co-expression Differential Network Analysis (CoDiNA). *PLoS ONE* **15**(10 October), 1–28 (2020). doi:[10.1371/journal.pone.0240523](https://doi.org/10.1371/journal.pone.0240523)
  69. Watson, M.: CoXpress: Differential co-expression in gene expression data. *BMC Bioinformatics* **7**, 1–12 (2006). doi:[10.1186/1471-2105-7-509](https://doi.org/10.1186/1471-2105-7-509)

**Additional Files**

Additional file 1 — Text

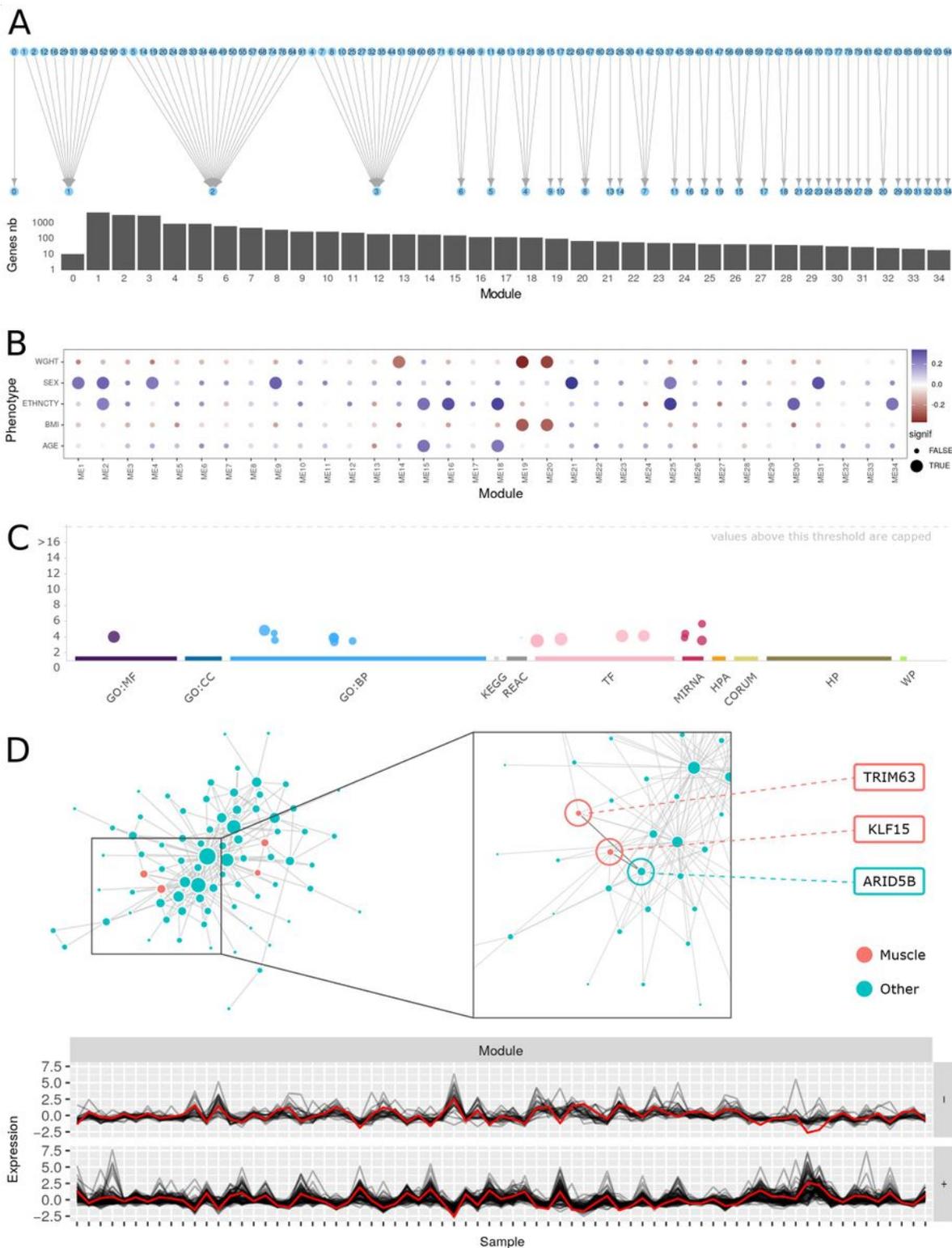
Details on data used, description of methods, and supplementary results  
(pdf)

# Figures



**Figure 1**

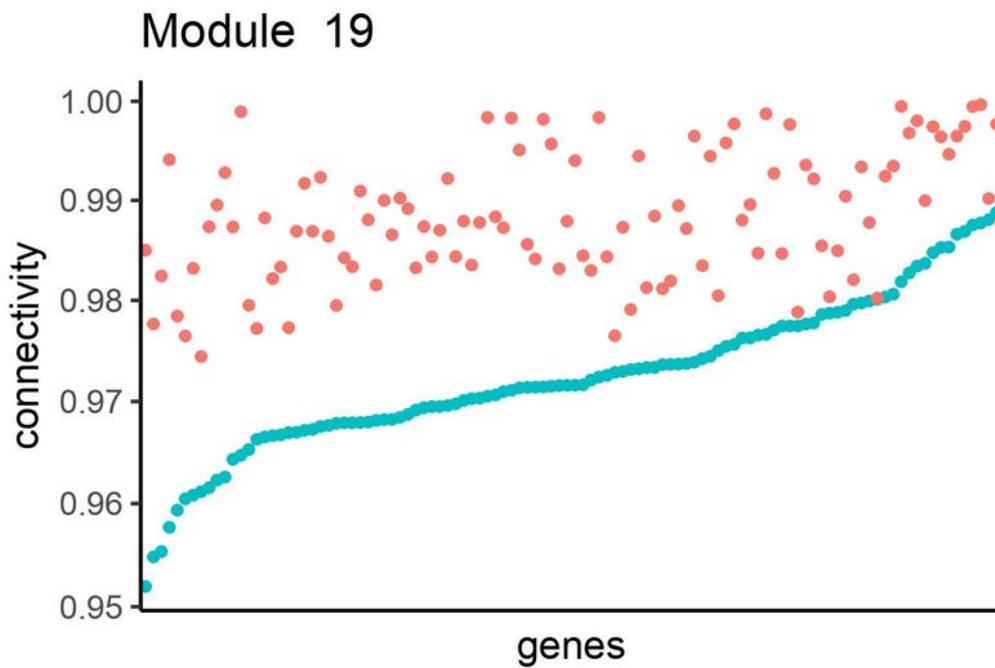
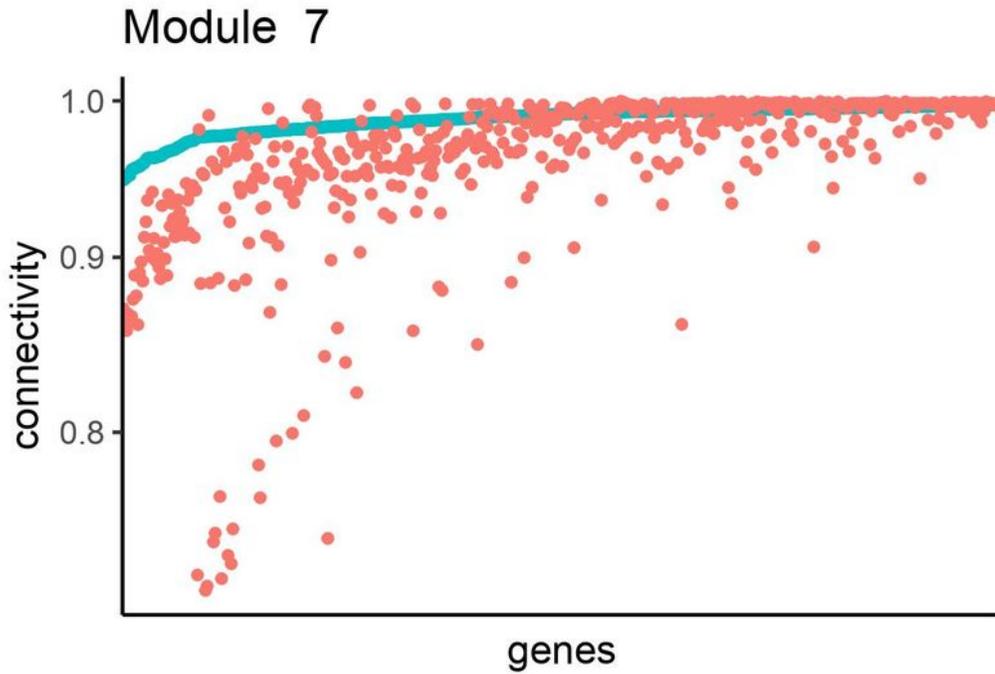
Detailed steps of analysis performed in GWENA's pipeline, from expression data to characterization of the modules and comparison of conditions. 1  $\boxtimes$  Input : expression matrix pre-normalized and aggregated to gene level if it is a transcript matrix. 2  $\boxtimes$  Filtering : optional genes filtration according to transcriptomic input technology. 3  $\boxtimes$  Co-expression network construction : computation through modified WGCNA function of a correlation matrix on the gene expression matrix, then transformation into an adjacency matrix, and finally into a topological overlap matrix (TOM). 4  $\boxtimes$  Modules detection : genes clusterization over the TOM with another modified WGCNA function. 5  $\boxtimes$  Biological integration : gene set enrichment of each module using g:Profiler services, and phenotypic association if a phenotype matrix is provided to describe the samples. 6  $\boxtimes$  Graph analysis : transformation of the TOM in a graph to compute different topological metrics and detect the hub genes. 7  $\boxtimes$  Modules differential co-expression over N conditions : permutation test using NetRep combined with a Z summary to detect preserved or unpreserved modules.



**Figure 2**

Available visualizations in GWENA along the pipeline applied to the aging study on the whole age range. A : modules merge as a bipartite graph from plot modules merge function and the genes distribution inside each of them (log scale). B : phenotypic association between the 35 modules and age, sex, BMI, ethnicity, weight. C : Manhattan-like enrichment plot (interactive in GWENA) of module 10 on GO, KEGG, Reactome, Transfac, miRTarBase, Human Protein Atlas, CORUM, Human Phenotype ontology,

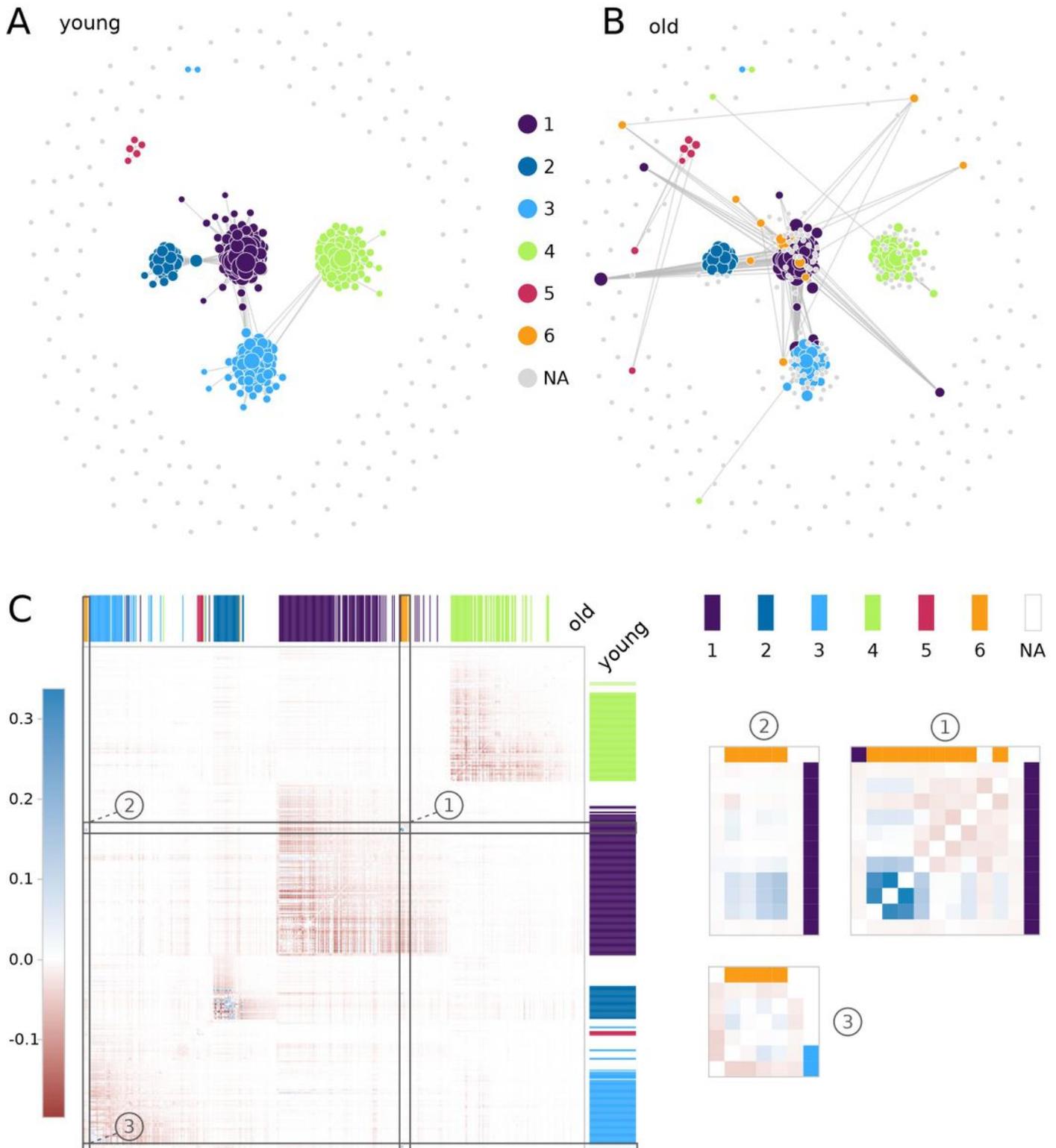
WikiPathways. D: module 19's network visualization as a graph with muscle enrichment genes colored in red and others in blue. The zoom focus on ENSG00000158022 / ENSG00000107372 / ENSG00000265972 and related hub genes. E : expression profile of module 19 split depending on the correlation sign to the eigengene.



cond    ● old    ● young

Figure 3

Modules 7 and 19 genes (nodes) connectivity distribution between young and old age ranges. Young age range is used as reference for sorting genes by increasing connectivity. A comparison over all modules can be found in Additional file 1.4



**Figure 4**

Module 7 network comparison between young and old. A: module 7 GCN graph plotted with GWENA (0.95 co-expression score filter) for young age range with sub-clusters detected via PAM. B: same as A but for

the old age range. A zoom is made on sub-module 4 to show the peripheral genes disconnection. The new sub-module 6 is visible in purple in the old graph. C: difference network heatmap (old - young) ordered according to young age range network dendrogram. Sub-modules from old age range are visible on the top of the heatmap in columns, and sub-modules from young age range on the right in rows. Three zooms are made on the heatmap on the areas corresponding to sub-module 6 genes. Zoom 1  contains the genes reconnecting in the old age range.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [GWENAwithBMCTemplateadditionalfile1.pdf](#)