Network analysis in response to calorie restriction

Émilie Montastier, Nathalie Villa-Vialaneix
et de nombreux co-auteurs !

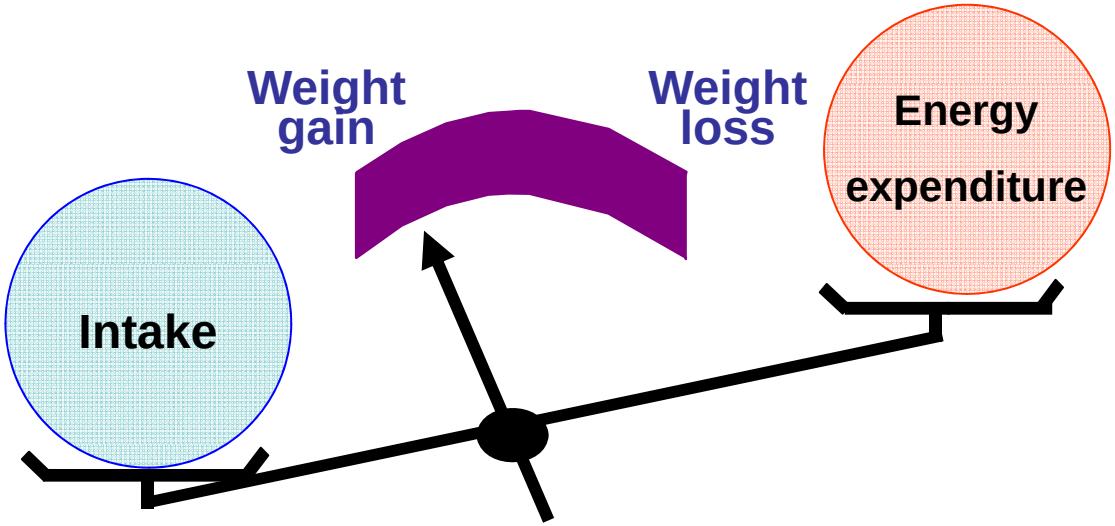

INSERM, Obesity Research Laboratory IM2C

& INRA, Unité MIA-T

Journée Régionale GenoToul BioInfo & BioStats

13 Juin 2014

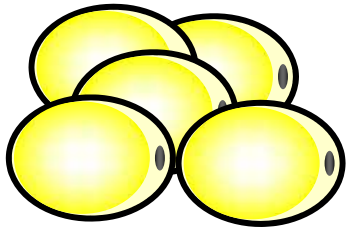
Maintenance of weight loss: an obstacle in successful treatment of obese individuals



Overweight /obesity

Medical treatment of obesity: 80% failled after one year *(Wing RR, Am J Clin Nutr, 2005)*: it doesn't work!

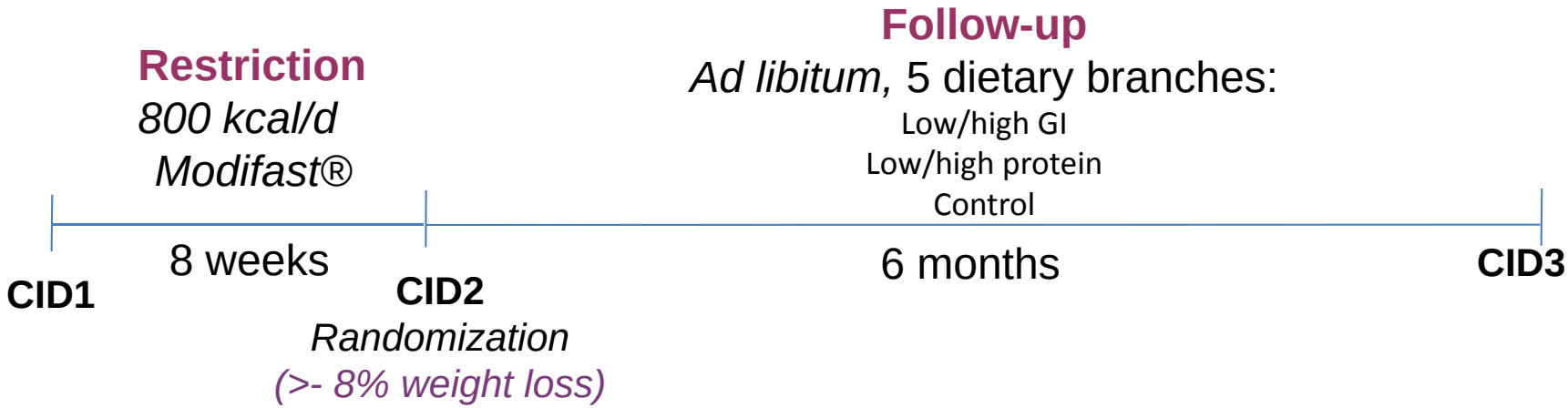
Surplus energy storage in adipose tissue



Weight follow-up after energy restriction induced weight loss

Effect of glycemic index and protein content

EU project, 8 centres, 450 families



CID: « Clinical Intervention Day »

- *Anthropometry*
- *Blood and urine sampling*
- *Adipose tissue biopsies*

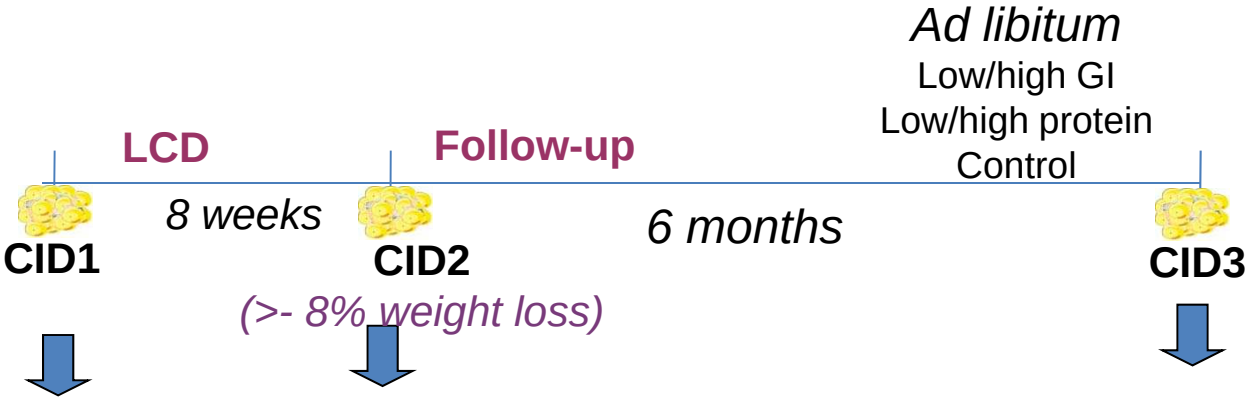
3 types of data in 135 women

Bio-clinical data:

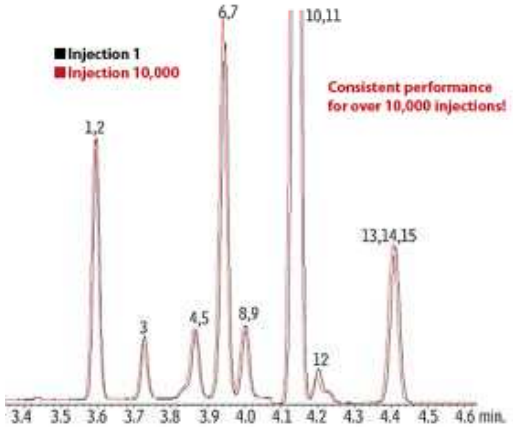
Body composition, RMR

MetS, insulin sensitivity

Plasma measurements



Adipose tissue fatty acid composition



221 genes:

Metabolism:

Glycolysis

Citrate cycle

Lipogenesis, fatty acid transport

Immunity: chemokines, receptors, complement

Fluidigm

Standard issues in network analysis

Inference

Giving expression data, how to build a graph whose edges represent the direct links between genes?

Example: co-expression networks built from microarray/RNAseq data (nodes = genes; edges = significant “direct links” between expressions of two genes)

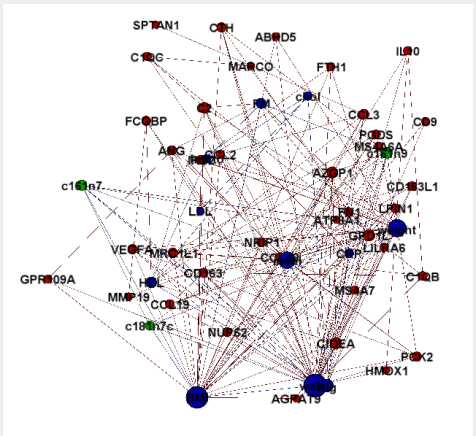
Standard issues in network analysis

Inference

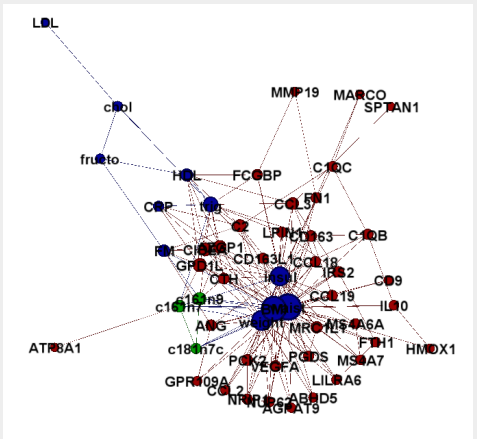
Giving expression data, how to build a graph whose edges represent the direct links between genes?

Graph mining (examples)

1. Network visualization: nodes are not a priori given a position.



Random positions



Positions aiming at representing connected nodes closer

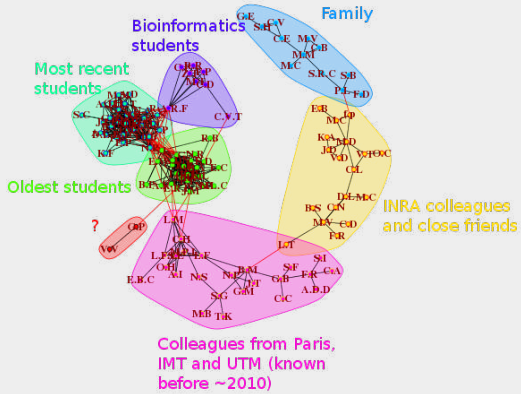
Standard issues in network analysis

Inference

Giving expression data, how to build a graph whose edges represent the direct links between genes?

Graph mining (examples)

1. Network visualization: nodes are not a priori given a position.
2. Network clustering: identify “communities”



Data

Experimental protocol

135 obese women and 3 times: before LCD, after a 2-month LCD and 6 months later (between the end of LCD and the last measurement, women are randomized into one of 5 recommended diet groups).
At every time step, 221 gene expressions, 28 fatty acids and 15 clinical variables (i.e., weight, HDL, ...)

Data

Experimental protocol

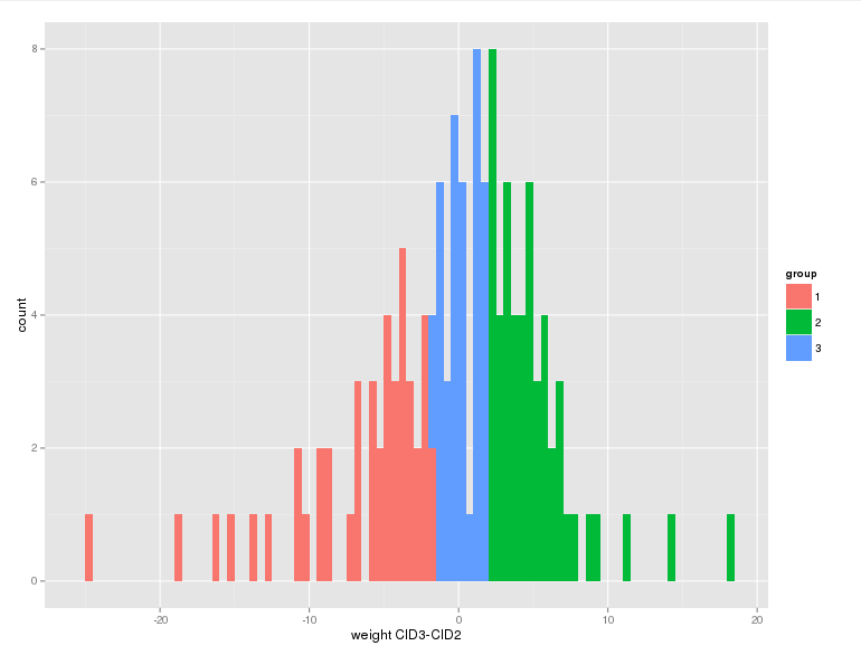
135 obese women and 3 times: before LCD, after a 2-month LCD and 6 months later (between the end of LCD and the last measurement, women are randomized into one of 5 recommended diet groups).
At every time step, 221 gene expressions, 28 fatty acids and 15 clinical variables (i.e., weight, HDL, ...)

Correlations between gene expressions and between a gene expression and a fatty acid levels are not of the same order: inference method must be different inside the groups and between two groups.

Data

Data pre-processing

At CID3, individuals are split into three groups: weight loss, weight regain and stable weight (groups are not correlated to the diet group according to χ^2-test).



Method for CID1, CID2 and 3×CID3

Network inference → Clustering → Mining

3 intra-dataset networks
sparse partial correlation

merge into one network

3 inter-dataset networks
rCCA

5 networks
CID1 **CID2**
3×CID3

Study/Compare clusters

Extract important nodes

Inference

Intra-level networks: use of partial correlations and a sparse approach (graphical Lasso as in the R package **gLasso**) to select edges [Friedman et al., 2008]

Inference

Intra-level networks: use of partial correlations and a sparse approach (graphical Lasso as in the R package **gLasso**) to select edges [Friedman et al., 2008]

Inter-levels networks: use of regularized CCA (as in the R package **mixOmics**) to evaluate strength of the correlations [Lê Cao et al., 2009]

Inference

Intra-level networks: use of partial correlations and a sparse approach (graphical Lasso as in the R package **gLasso**) to select edges [Friedman et al., 2008]

Inter-levels networks: use of regularized CCA (as in the R package **mixOmics**) to evaluate strength of the correlations [Lê Cao et al., 2009]

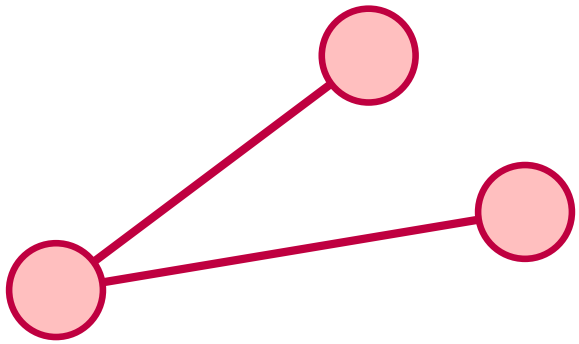
Combination of the 6 informations: tune the number of edges intra or inter-levels so that it is of the order of the number of nodes in the corresponding level(s)

Visualization

Purpose: How to display the nodes in a meaningful and aesthetic way?

Visualization

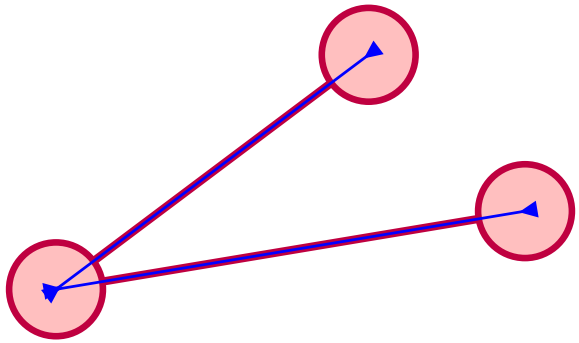
Purpose: How to display the nodes in a meaningful and aesthetic way?
Standard approach: force directed placement algorithms (FDP) (e.g., [Fruchterman and Reingold, 1991])



Visualization

Purpose: How to display the nodes in a meaningful and aesthetic way?

Standard approach: force directed placement algorithms (FDP) (e.g., [Fruchterman and Reingold, 1991])

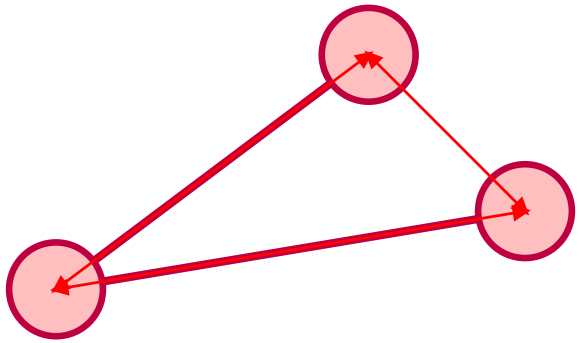


- attractive forces: similar to springs along the edges

Visualization

Purpose: How to display the nodes in a meaningful and aesthetic way?

Standard approach: force directed placement algorithms (FDP) (e.g., [Fruchterman and Reingold, 1991])

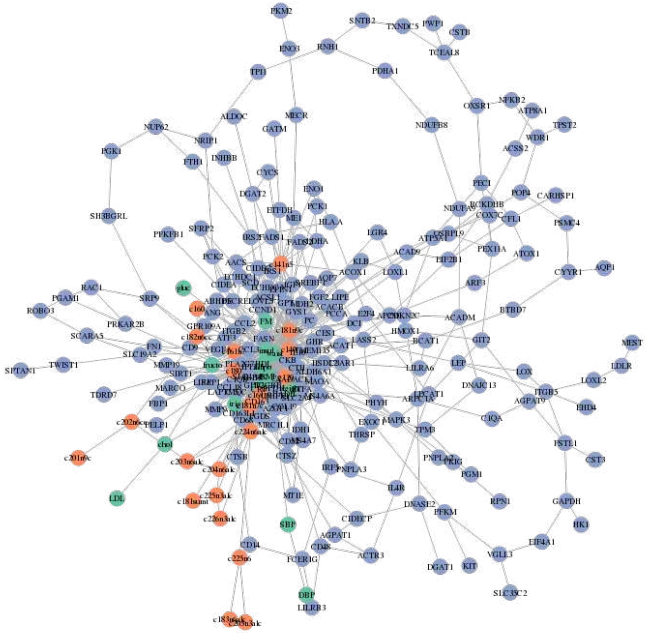


- attractive forces: similar to springs along the edges
- repulsive forces: similar to electric forces between all pairs of vertices

Visualization

Purpose: How to display the nodes in a meaningful and aesthetic way?

Standard approach: force directed placement algorithms (FDP) (e.g., [Fruchterman and Reingold, 1991])



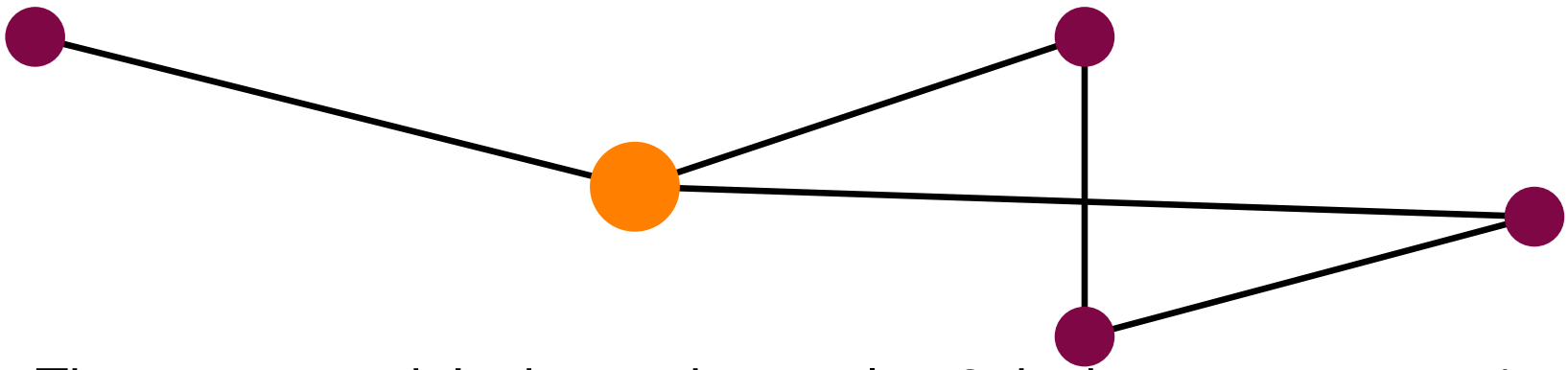
iterative algorithm until stabilization of the vertex positions.

Important node extraction

1. vertex degree: number of edges adjacent to a given vertex. Vertices with a high degree are called hubs: measure of the vertex popularity.

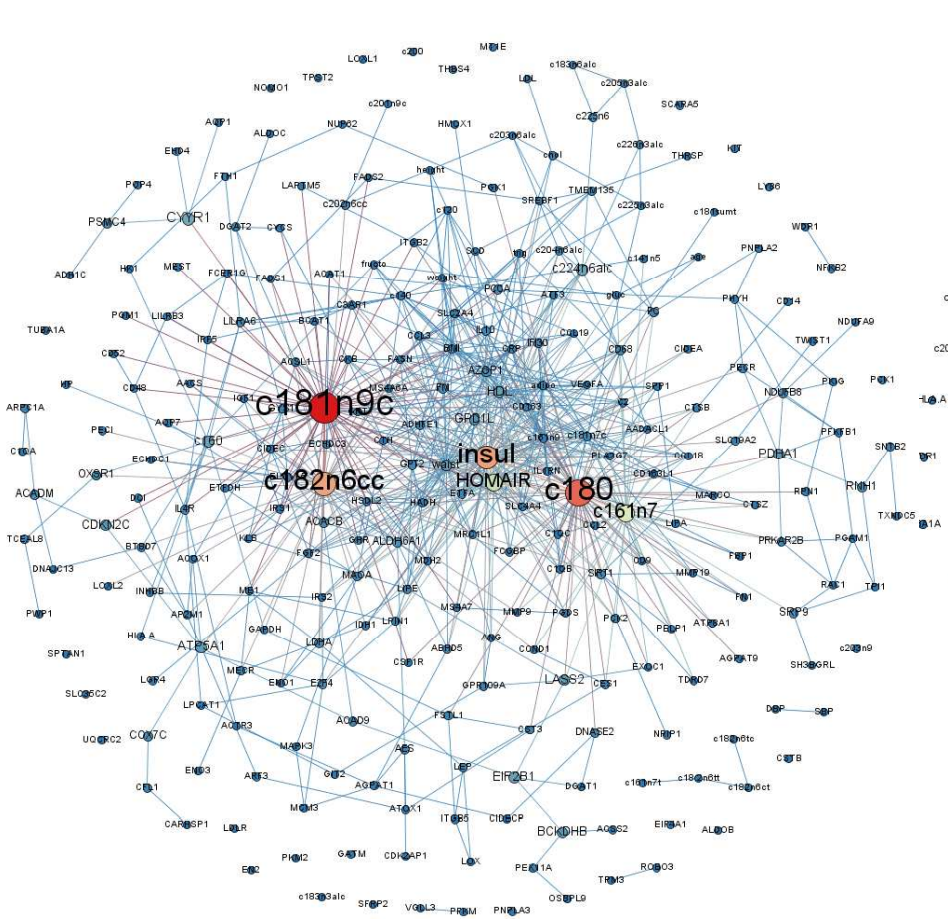
Important node extraction

1. vertex degree: number of edges adjacent to a given vertex. Vertices with a high degree are called hubs: measure of the vertex popularity.
2. vertex betweenness: number of shortest paths between all pairs of vertices that pass through the vertex. Betweenness is a centrality measure (vertices with a large betweenness that are the most likely to disconnect the network if removed).

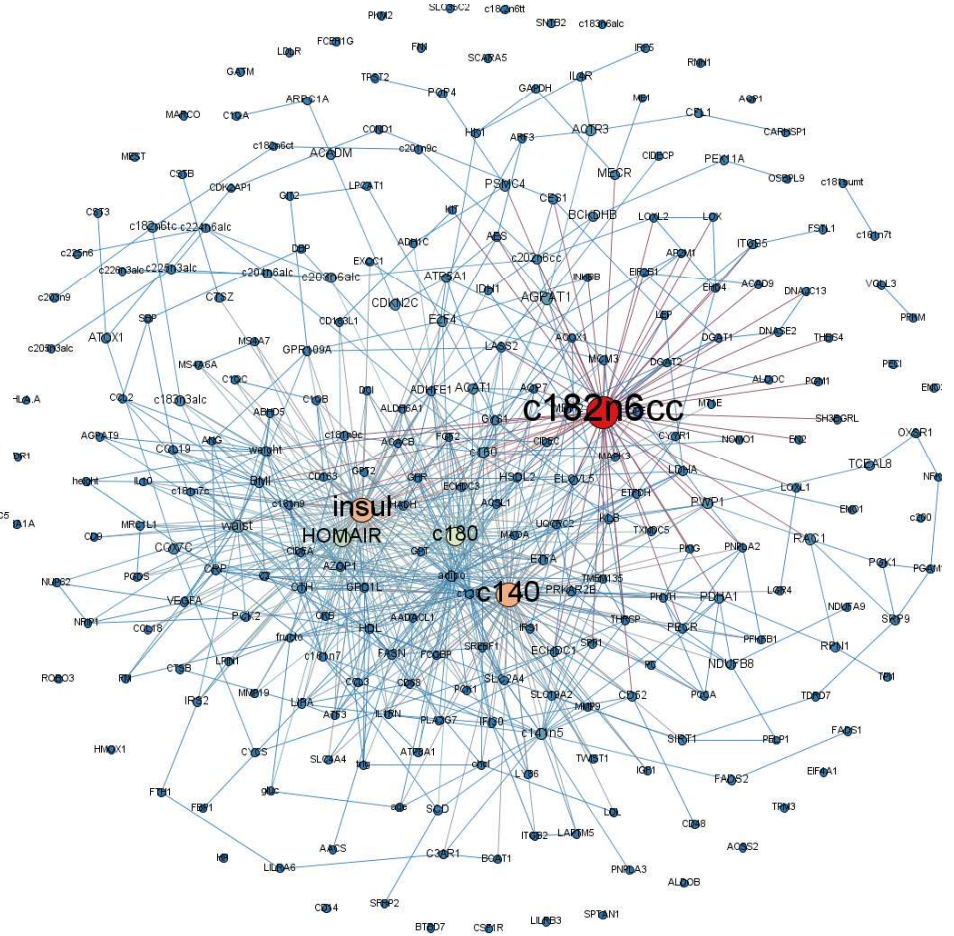


The orange node's degree is equal to 2, its betweenness to 4.

Fatty acids are highest centrality hubs

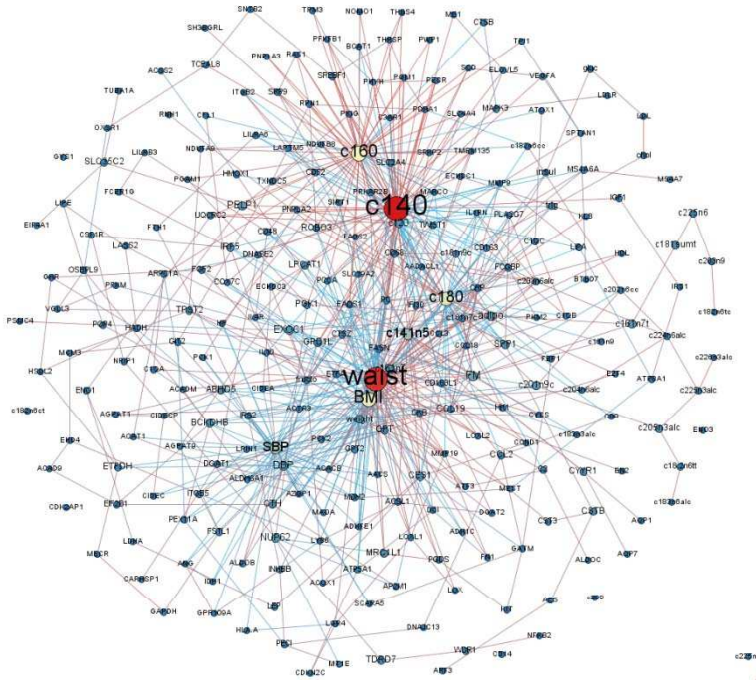


Baseline (CID1)

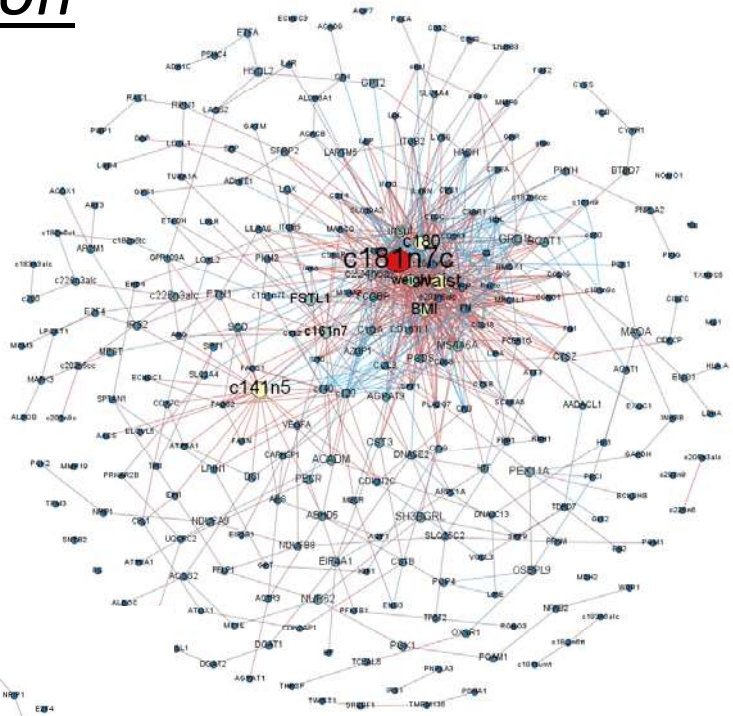


After restriction (CID2)

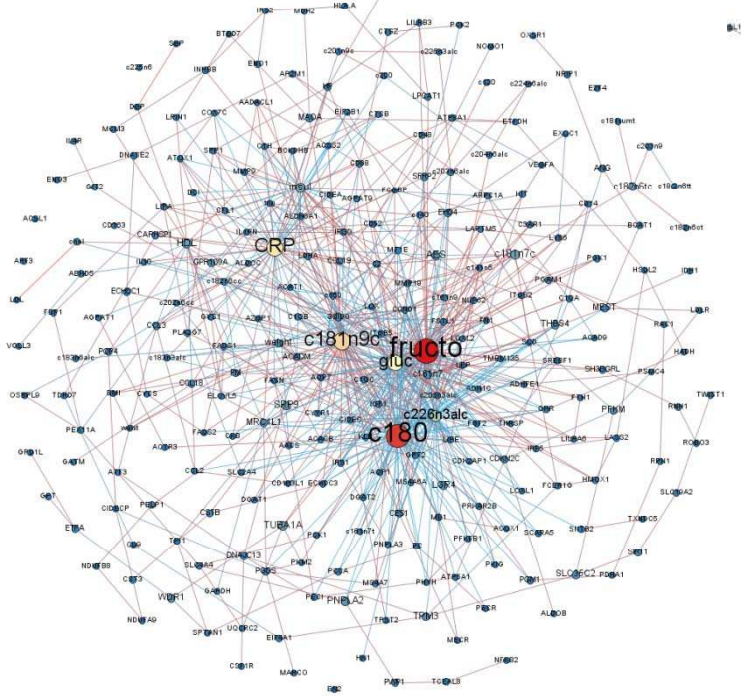
End of intervention (CID 3)



"Weight regain" group



"Weight loss" group



"Weight maintain" group

Vertex clustering

Cluster vertexes into groups that are densely connected and share a few links (comparatively) with the other groups. Clusters are often called communities (social sciences) or modules (biology). Node modules are known to be more robust and meaningful than individual relationships between pairs of nodes.

Vertex clustering

Cluster vertexes into groups that are densely connected and share a few links (comparatively) with the other groups. Clusters are often called communities (social sciences) or modules (biology). Node modules are known to be more robust and meaningful than individual relationships between pairs of nodes.

Nodes were clustered using modularity maximization [Newman and Girvan, 2004] performed with a deterministic annealing algorithm as described in [Reichardt and Bornholdt, 2006] (after comparison of several approaches) and implemented in the function `spinglass.community` of the R package **igraph**.

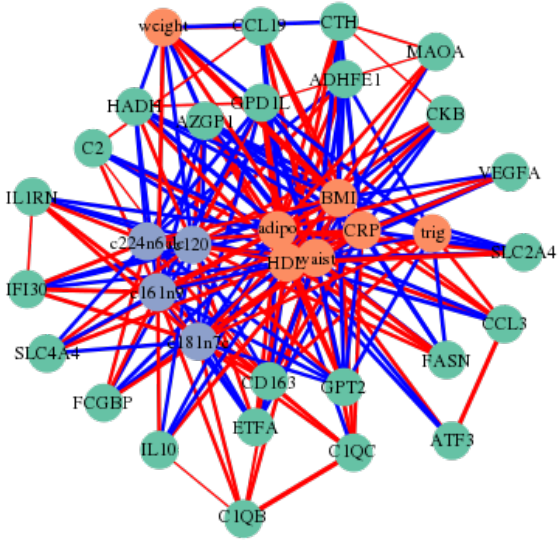
Brief overview on results

5 networks inferred with 264 nodes each:

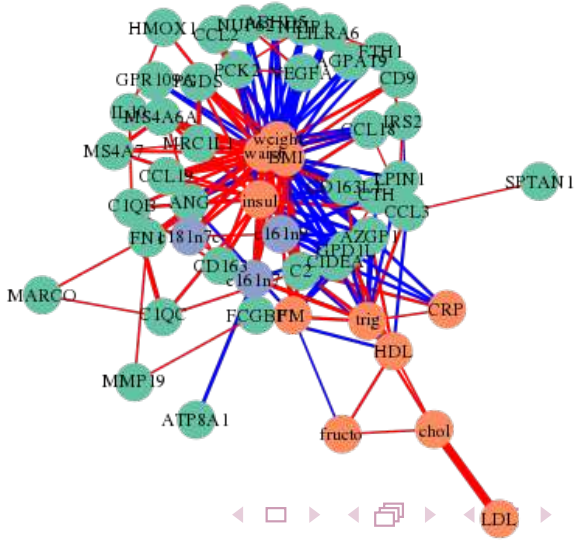
	CID1	CID2	CID3g1	CID3g2	CID3g3
size LCC	244	251	240	259	258
density	2.3%	2.3%	2.3%	2.3%	2.3%
transitivity	17.2%	11.9%	21.6%	10.6%	10.4%
nb clusters	14 (2-52)	10 (4-52)	11 (2-46)	12 (2-51)	12 (3-54)

clusters were visualized and analyzed for important node extraction

CID 1 - Cluster 4

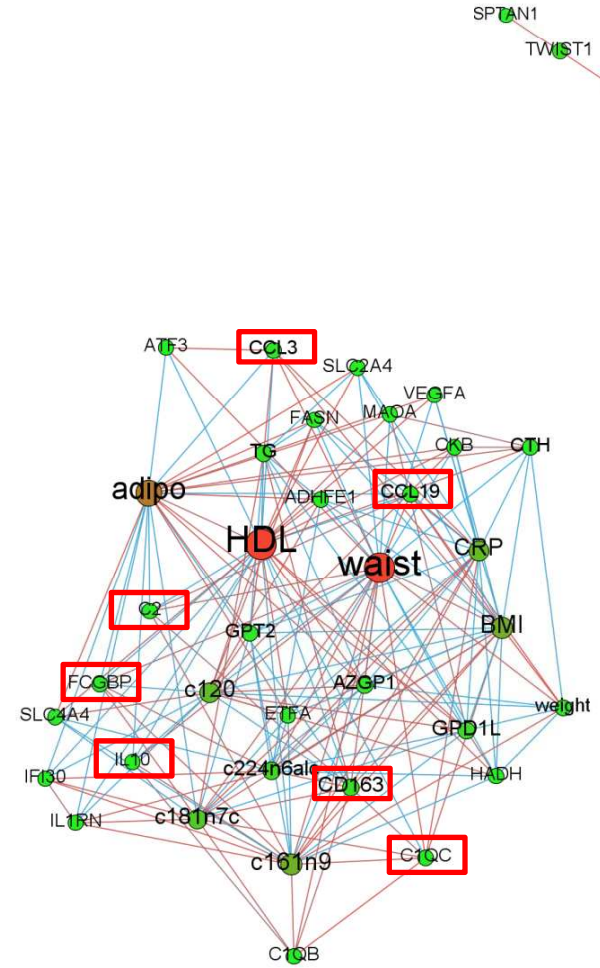


CID 2 - Cluster 5

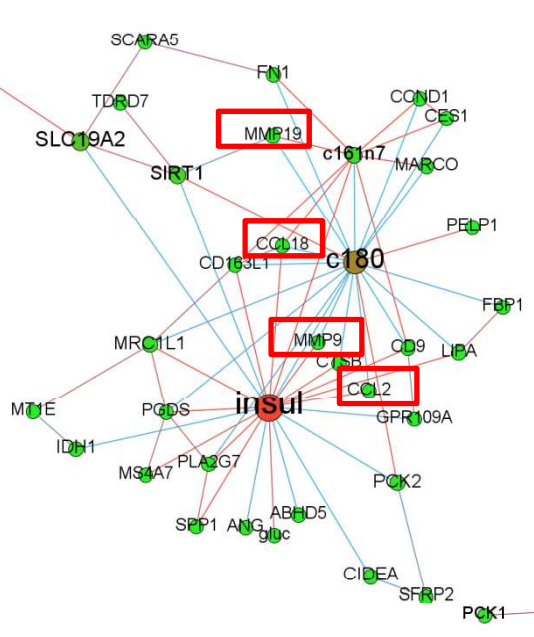


Spin glass vertexes classification

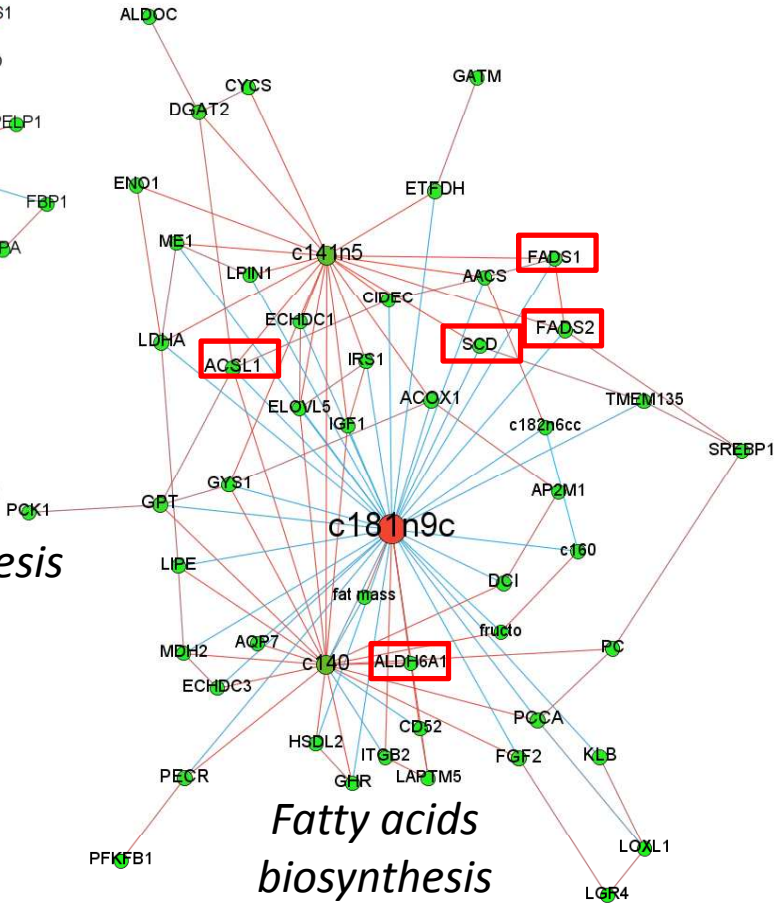
At Baseline: 14 clusters, 3 of them with at least 2 types of variables



Immune response



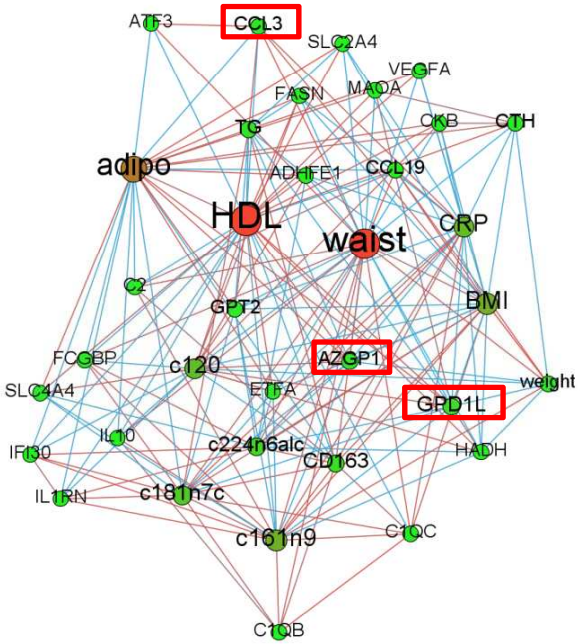
Adhesion and diapedesis



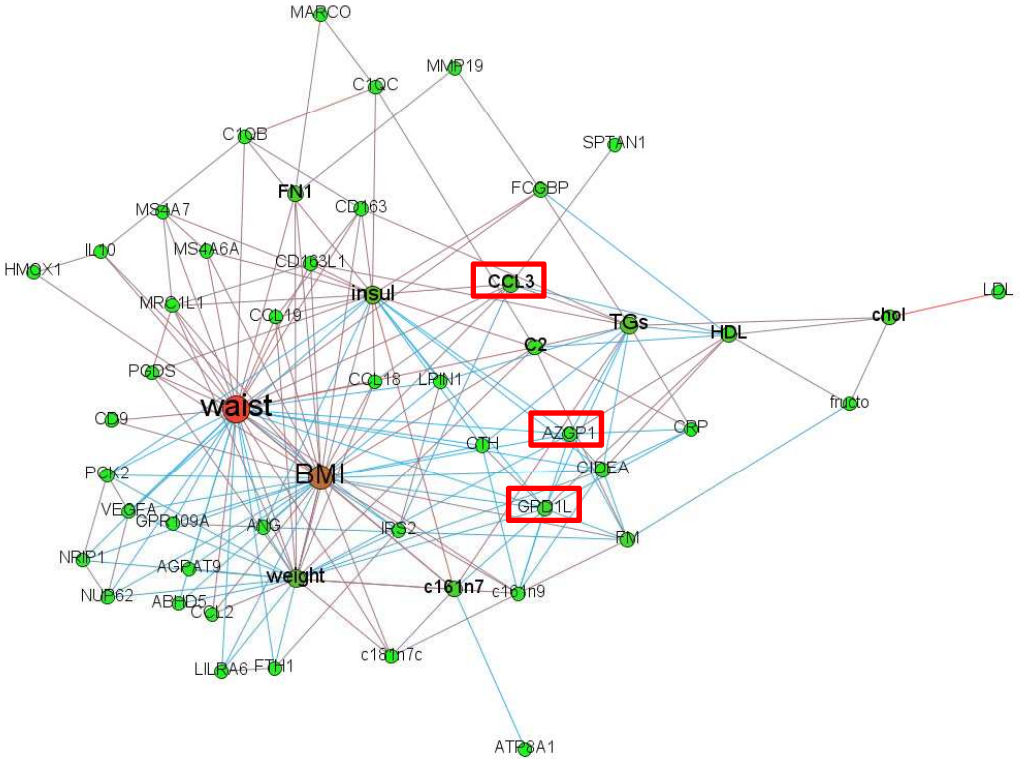
Fatty acids biosynthesis

Waist circumference is correlated with metabolic syndrome transcripts independently of weight change

At baseline (CID1)

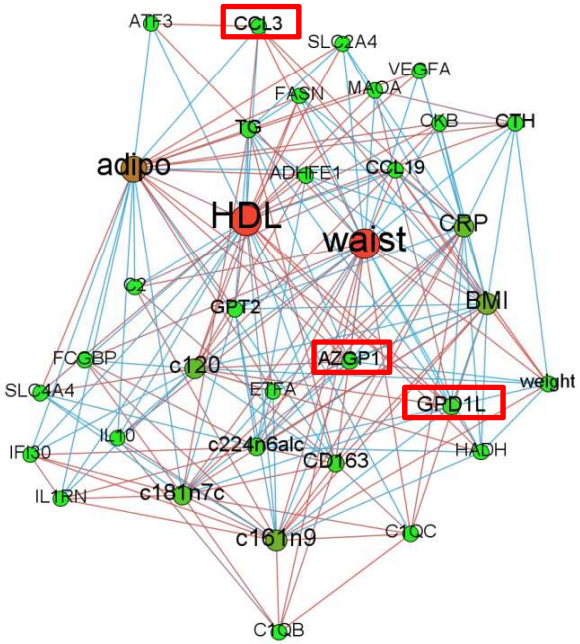


After restriction (CID 2)

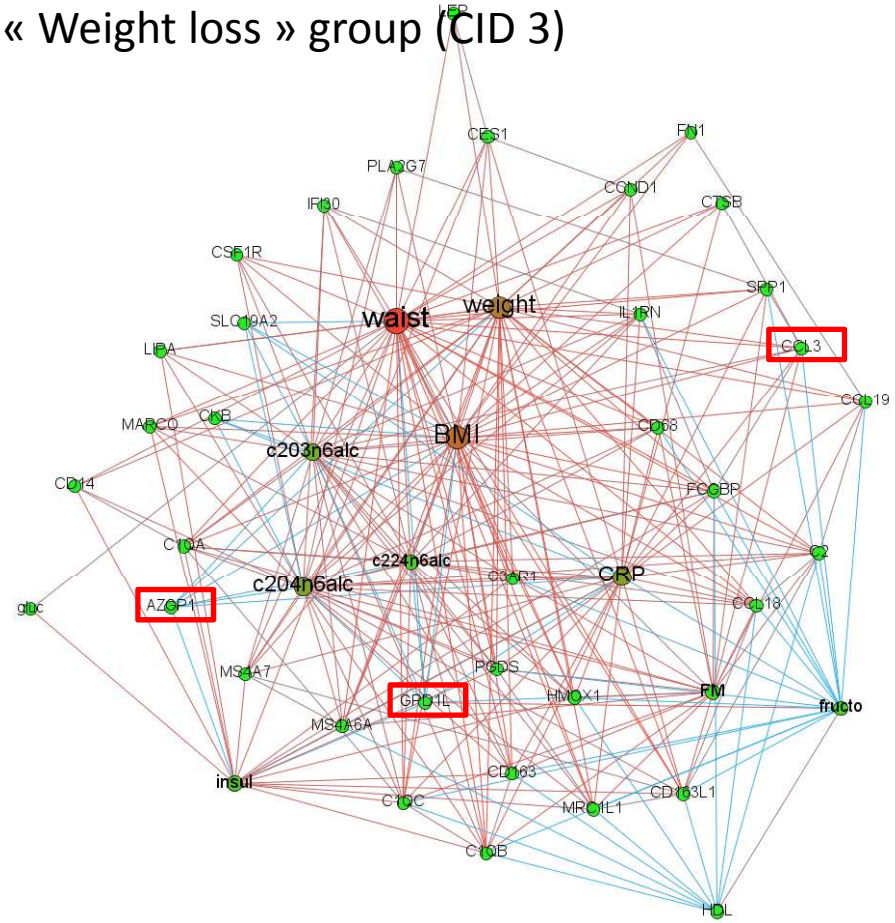


Waist circumference is correlated with metabolic syndrome transcripts independently of weight change

At baseline (CID1)

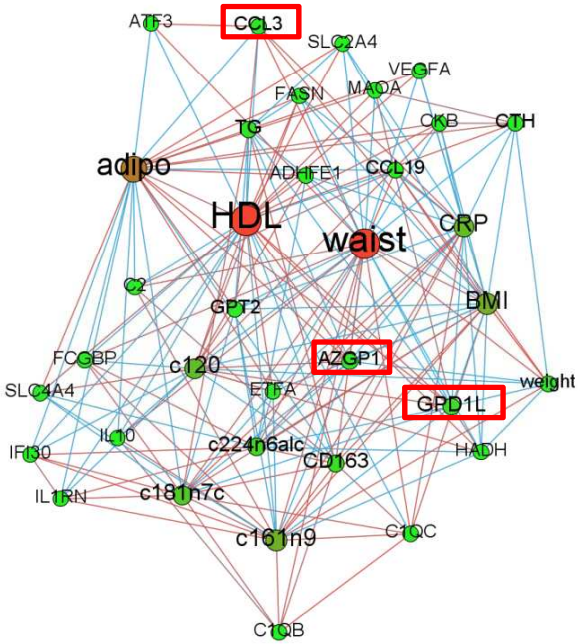


« Weight loss » group (CID 3)

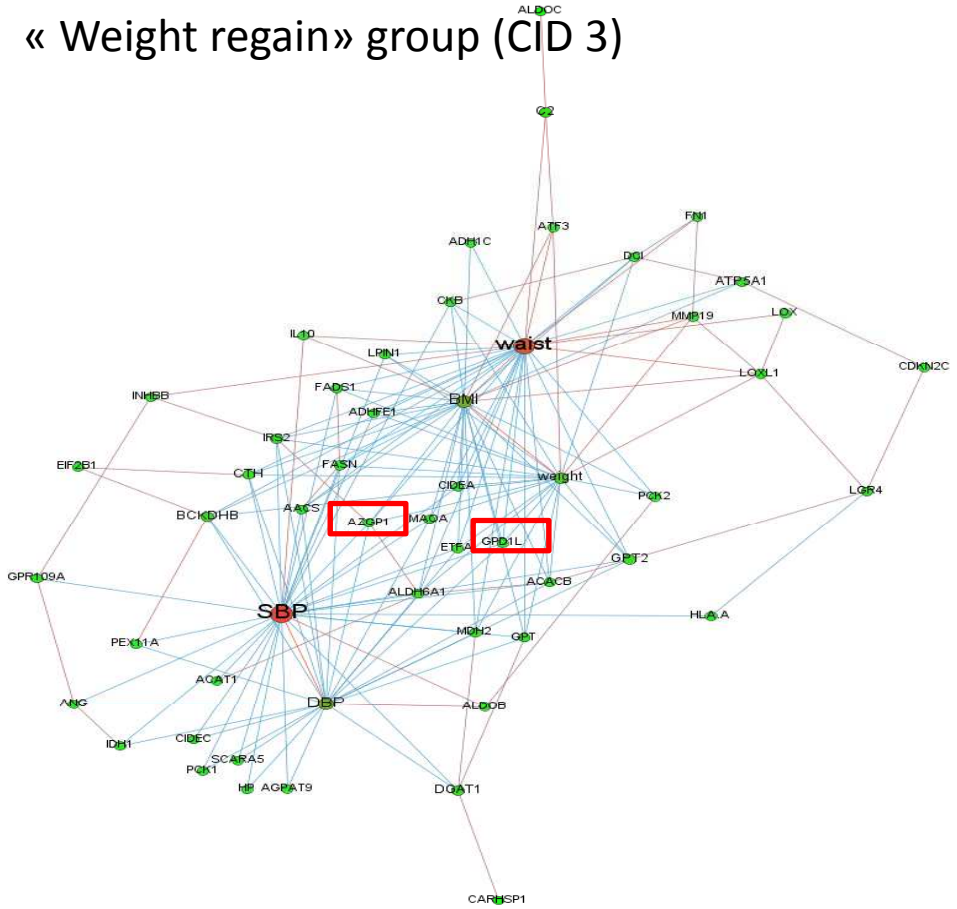


Waist circumference is correlated with metabolic syndrome transcripts independently of weight change

At baseline (CID1)



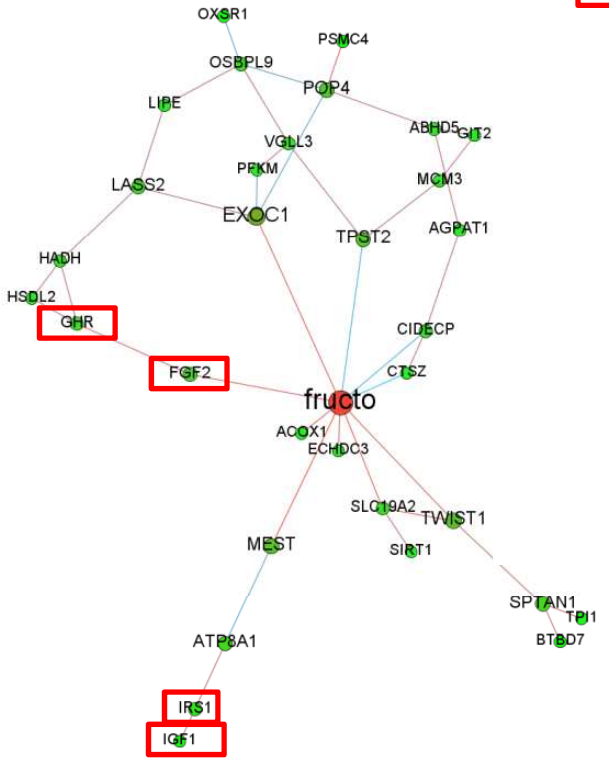
« Weight regain» group (CID 3)



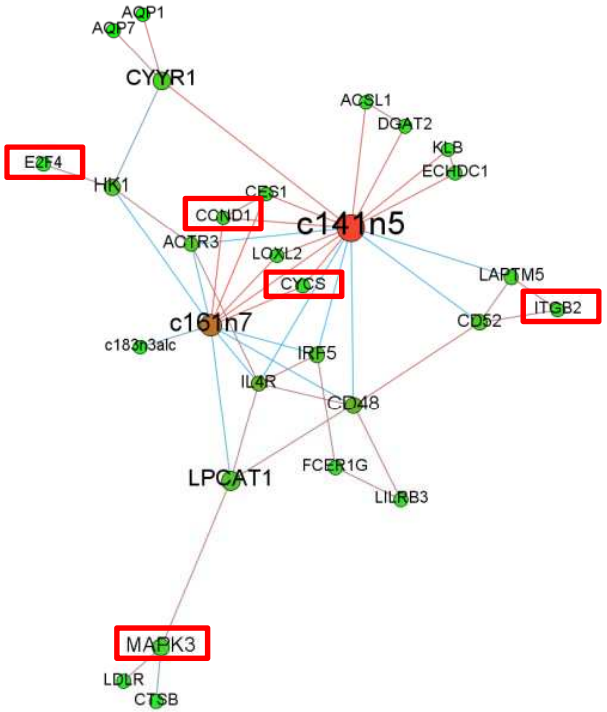
Increase in growth factors, angiogenesis and proliferation signaling in women regaining weight

End of intervention:

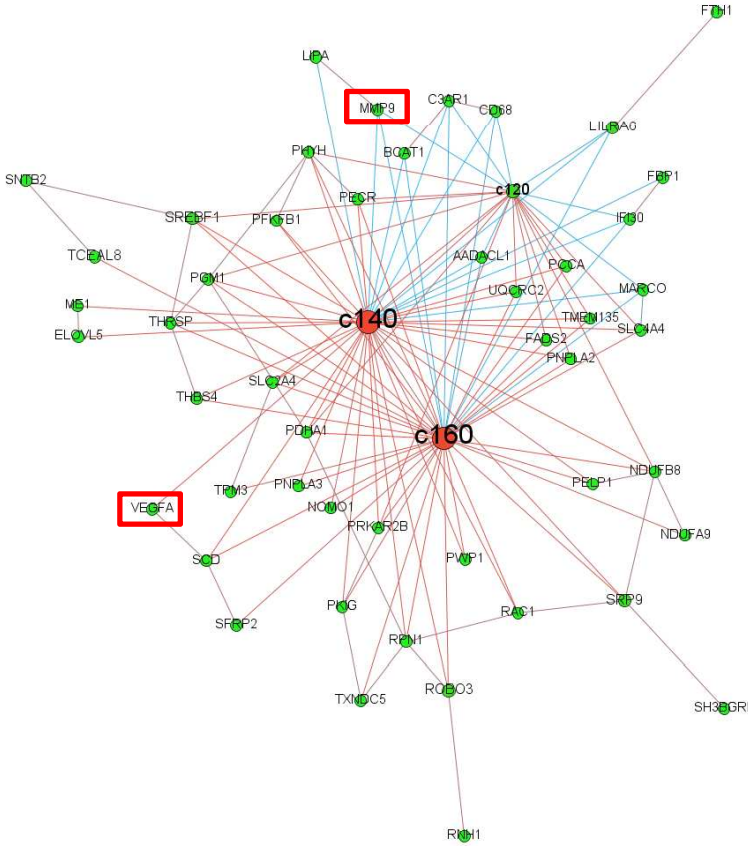
Growth Hormone signaling



Cancer signal

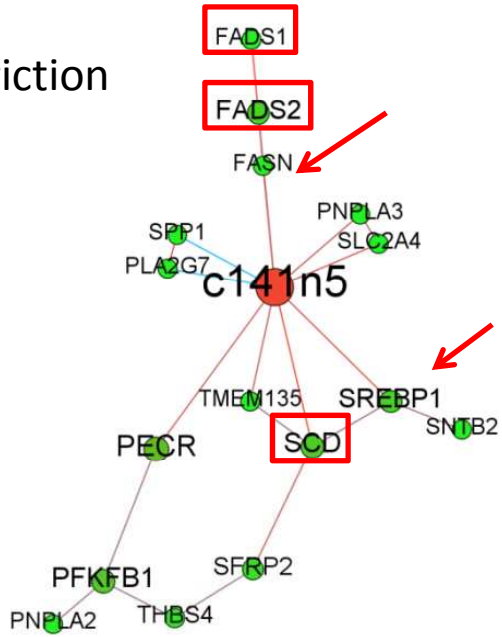


Angiogenesis inhibition by TSP1



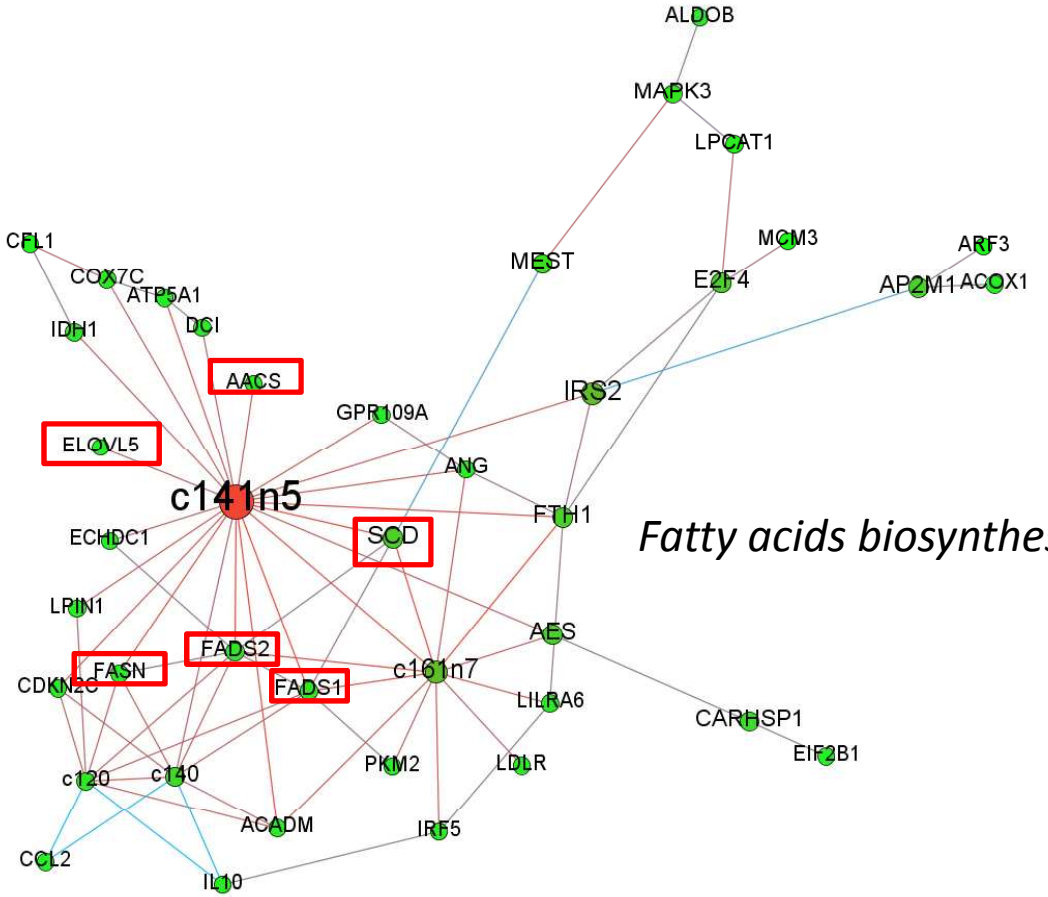
Positive relationship between AT myristoleic acid content and de novo lipogenesis mRNAs in women losing weight

After restriction (CID 2)



Fatty acids biosynthesis

End of intervention, “weight loss” group



Fatty acids biosynthesis

In conclusion:

- For the first time:
  - Integrated approach of 2 omics (from the same biopsy)
  - In adipose tissue
  - Of a large number of patients
  - In a longitudinal dietary intervention
  - Well characterized individuals

- Myristoleic acid as a main lipidic biomarkers for de novo lipogenesis: **unexpected, and quantitatively minor fatty acid in adipose tissue and plasma**

- This original approach authorizes new advances in obesity and insulin sensitivity patho-physiology understanding

- **Biostatistics post-doctoral position open!**

Thank you for your attention...

... questions?

Friedman, J., Hastie, T., and Tibshirani, R. (2008).

Sparse inverse covariance estimation with the graphical lasso.

Biostatistics, 9(3):432–441.

Fruchterman, T. and Reingold, B. (1991).

Graph drawing by force-directed placement.

Software, Practice and Experience, 21:1129–1164.

Lê Cao, K., González, I., and Déjean, S. (2009).

*****Omics: an R package to unravel relationships between two omics data sets.

Bioinformatics, 25(21):2855–2856.

Newman, M. and Girvan, M. (2004).

Finding and evaluating community structure in networks.

Physical Review, E, 69:026113.

Reichardt, J. and Bornholdt, S. (2006).

Statistical mechanics of community detection.

Physical Review, E, 74(016110).