

# Metabolite signatures of organ dysfunction in sickle cell disease patients

Yann Ilboudo<sup>1</sup>; Melanie Garrett<sup>2</sup>, Allison Ashley-Koch<sup>2</sup>, Marilyn Telen<sup>3</sup>; Guillaume Lettre<sup>1</sup>



Department of Biochemistry and Molecular Medicine, University of Montreal, Montreal Heart Institute, Montreal, Québec (1)

Center for Human Disease Modeling, Duke University Medical Center, Durham, North Carolina, United States of America (2)

Department of Medicine, Division of Hematology, Duke University Medical Center, Durham, North Carolina, United States of America (3)

Contact: yann.ilboudo@umontreal.ca



INSTITUT DE  
CARDIOLOGIE  
DE MONTREAL

## Introduction

Sickle cell disease (SCD) is a monogenic disease caused by mutations in the  $\beta$ -globin gene. The disease stems from a point mutation leading to a Glu  $\rightarrow$  Val substitution at the 6<sup>th</sup> codon of the protein. The complications related to the disease are systemic as they impact multiple organ systems. Our goal in this study was to identify metabolome changes contributing to SCD-related severity.

## Methods

### Study Design

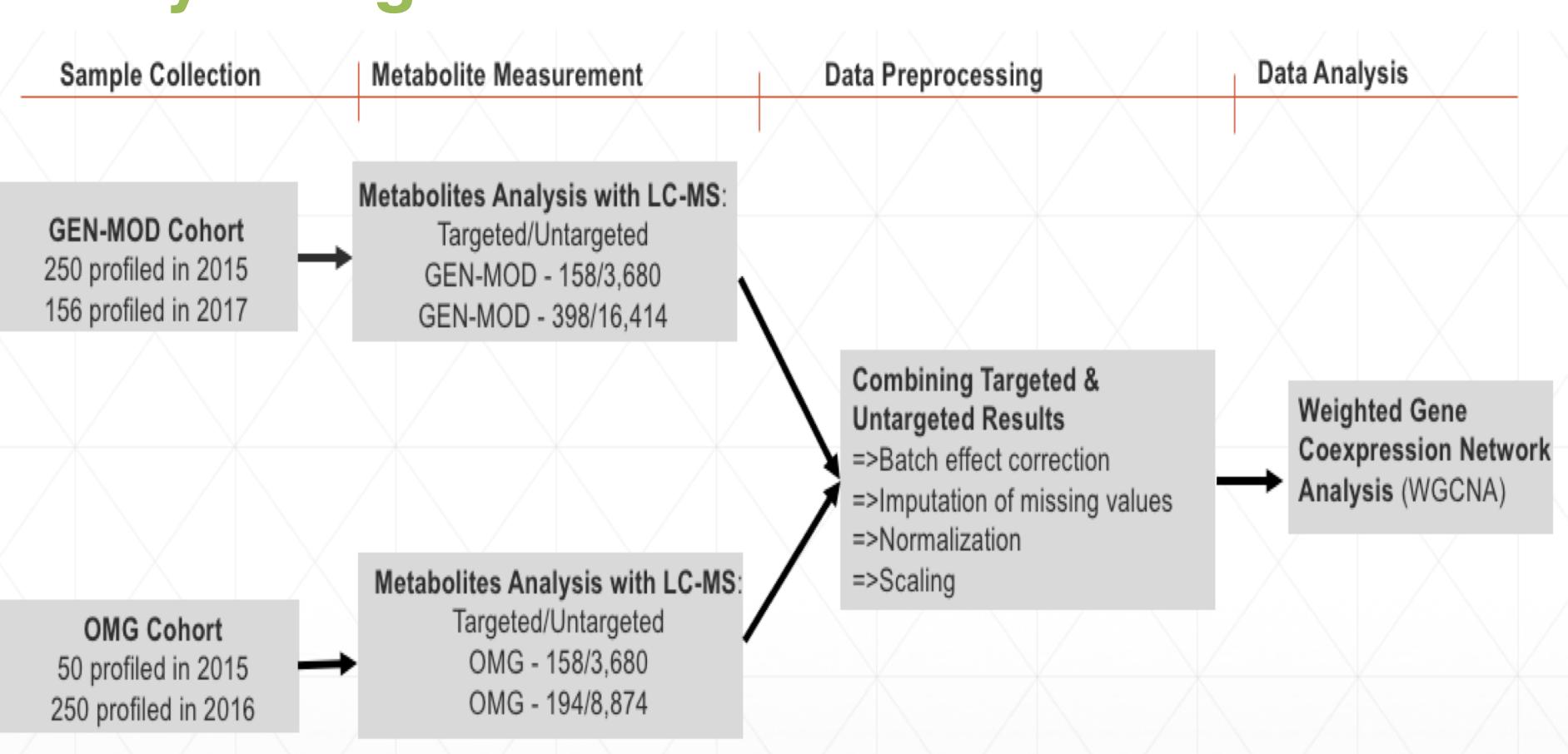


Figure 1. Study Design: Employing both targeted and untargeted approaches, we profiled the plasma of 706 SCD patients using liquid chromatography tandem mass spectrometry. The cohort included 406 French patients (GEN-MOD cohort) of recent African descent and 300 African Americans (OMG cohort) from southeastern US. In total, we measured the levels of 233 known and 1,880 unknown metabolites.

### Combining Datasets

#### • Post-QC

- 2,113 metabolites remained across cohorts
- 688 is the final sample size after removing outliers

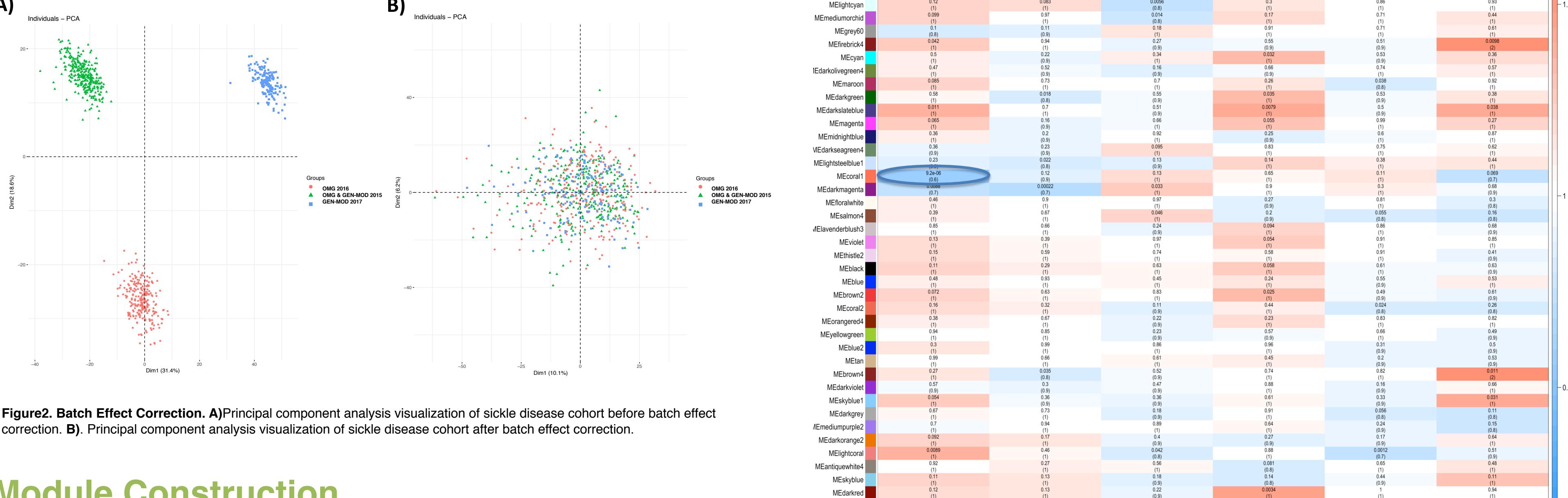


Figure 2. Batch Effect Correction. A) Principal component analysis visualization of sickle disease cohort before batch effect correction. B) Principal component analysis visualization of sickle disease cohort after batch effect correction.

### Module Construction

- Calculate Pearson correlation for metabolites across samples
- Transform correlations with “Power Adjacency Function”
  - Amplifies strong connections and dampens weak connections
  - This results in a more “hub and spoke” type of network
- Measure “Topological Overlap” dissimilarity (TOM) using Adjacency
  - Who is connected to who and by how much
- Perform hierarchical clustering based on TOM
  - Dendrogram
  - Define modules by cutoff height in the dendrogram
  - Modules are identified by color
- The R package “WGCNA” and custom functions available at [https://github.com/yilboudo/SCD\\_WGCNA](https://github.com/yilboudo/SCD_WGCNA) were used for all analyses
- We allowed at least 7 metabolites per module

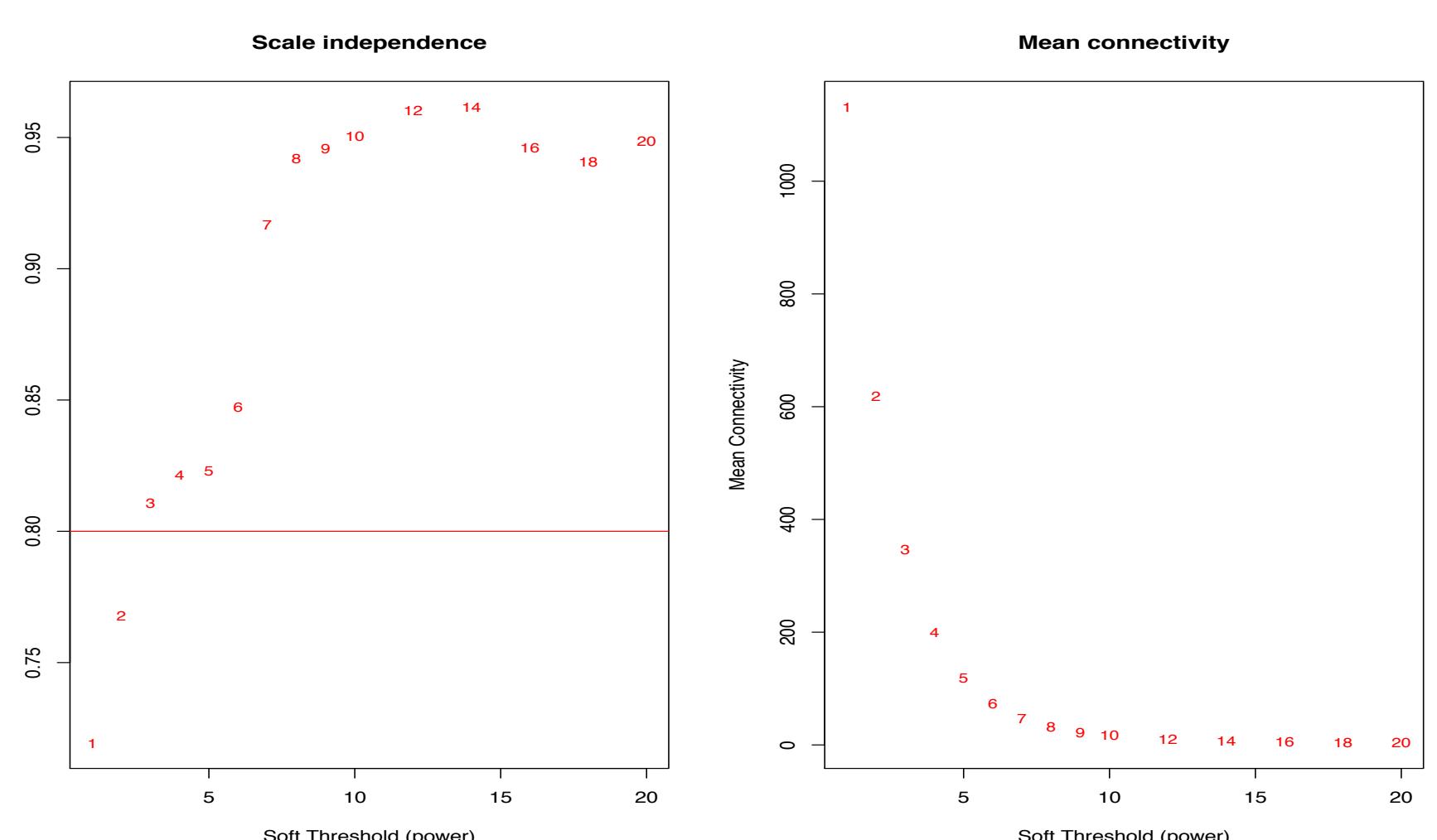


Figure 3. Analysis of Network Topology. Analysis of network topology for various soft-thresholding powers. The left panel shows the scale-free fit index (y-axis) as a function of the soft-thresholding power (x-axis). The right panel displays the mean connectivity (degree, y-axis) as a function of the soft-thresholding power (x-axis).

### Correlation with clinical phenotypes

- Correlated modules' first principal component with 15 clinical phenotypes:
  - Logistic regression for 6 complications (e.g. stroke, acute chest syndrome).
  - Linear regression for 11 hematological traits (hemoglobin concentration, mean cell volume).

## Results

### Network Dendogram

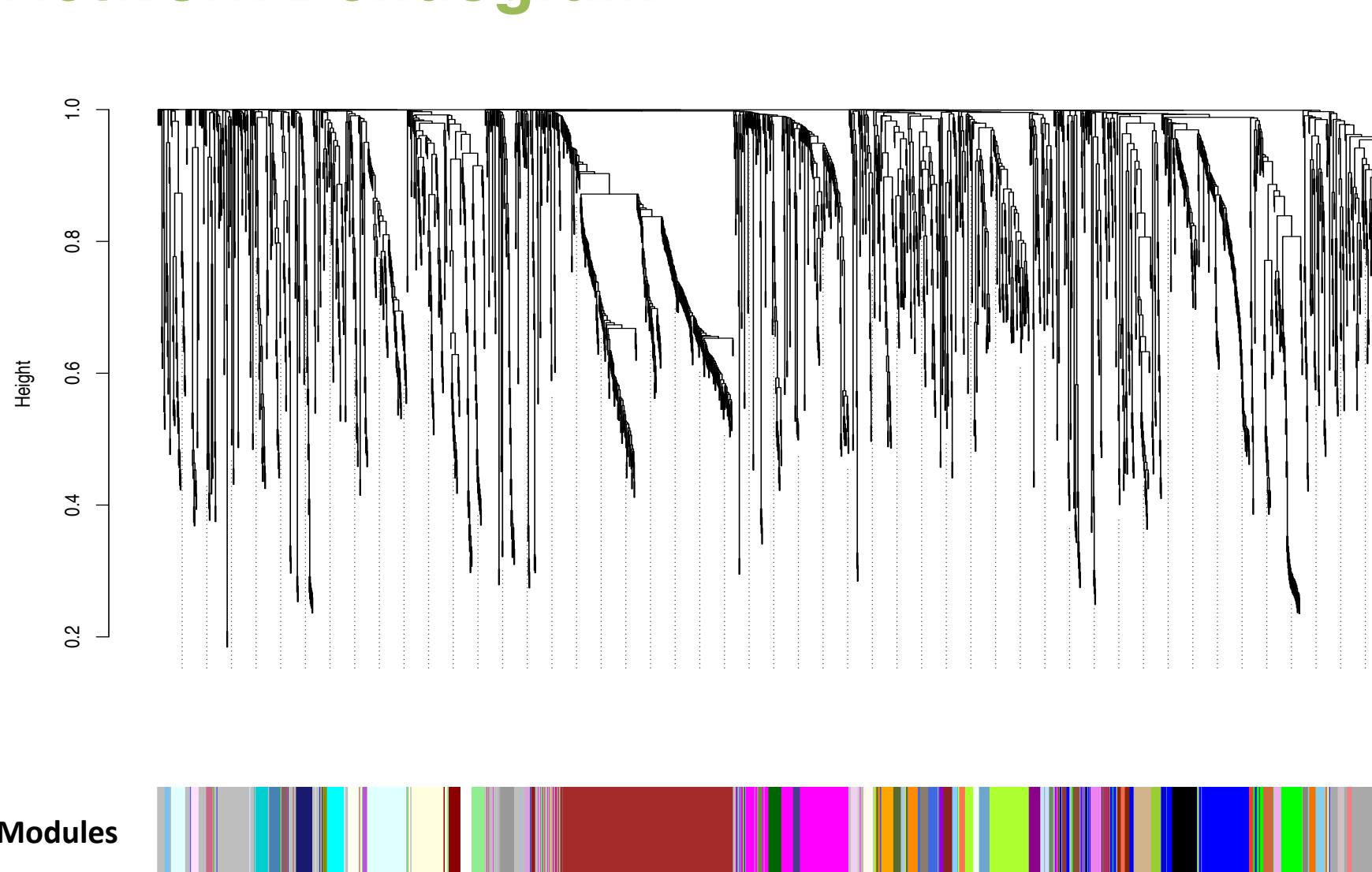


Figure 4. Network dendrogram. Modules are represented by their different colors. A total of 66 modules were constructed allowing 7 metabolites per module. They may represent highly interconnected metabolites; belonging to the same family and acting together in the same pathway.

### Results (continued)

### Modules Correlation with Blood Parameters

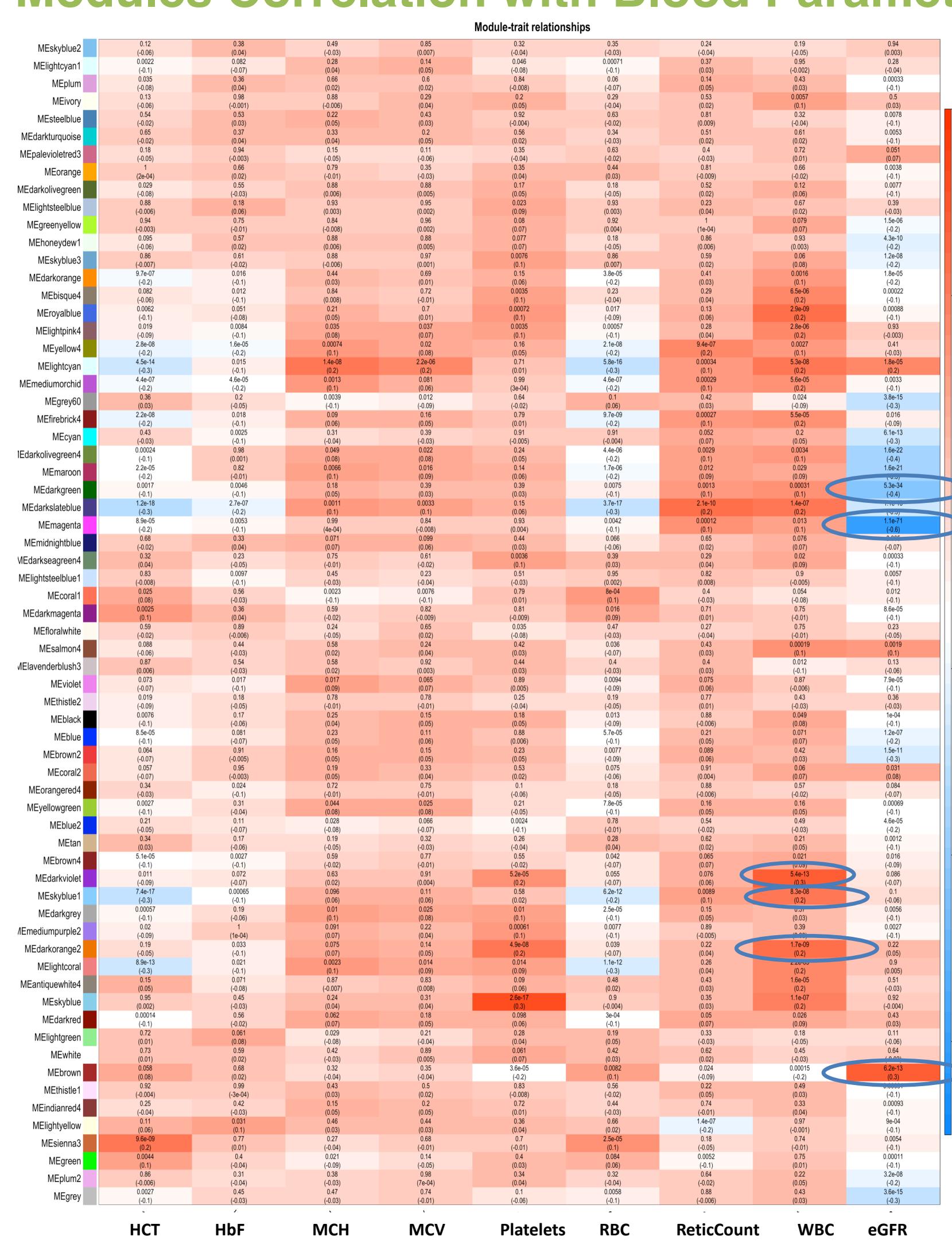


Figure 5. Module-blood traits correlation. Each row corresponds to a module eigenvalue, column to a trait. Each cell contains the corresponding regression coefficient (beta) and p-value. The table is color-coded by correlation according to the color legend.

- The first eigenvalue, PC1, from several modules were strongly correlated to blood parameters, to name a few:

- Brown module with eGFR
- Magenta module with eGFR
- Darkgreen with eGFR
- Darkviolet with Whiteblood cell
- Darkorange2 with Whiteblood cell
- Skyblue1 with Whiteblood cell

- Further inspection of the magenta module, associated with eGFR revealed carboxylic acids metabolites (citrulline, creatinine, symmetric dimethylarginine) and multiple unknown compounds

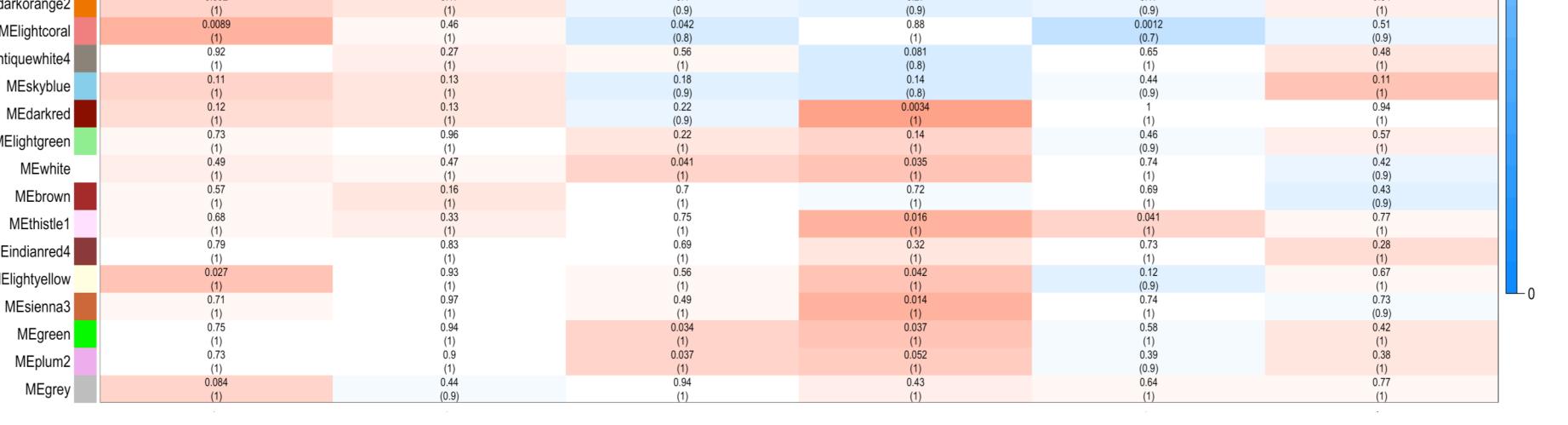


Figure 5. Module-Complications Correlation. Each row corresponds to a module eigenvector, column to a trait. Each cell contains the corresponding regression coefficient (beta) and p-value. The table is color-coded by correlation according to the color legend.

- The first eigenvalue, PC1, from two modules were strongly correlated to SCD-related complications:

- Lightcyan1 module with Cholecystectomy
- Coral1 module with Leg ulcer

• Upon further inspection, the Lightcyan1 module is composed of 4 known metabolites (glycocholate, glycocodeoxycholate, taurocholate, and taurodeoxycholate) all involved in bile acid metabolism, and 20 unknown metabolites. We also found the module to be well-defined and reproducible after robustness analysis.

• Coral1 is a module composed solely of unknown metabolites, and although its correlation to leg ulcer complications was appealing, the robustness analysis indicated that the module is spurious.

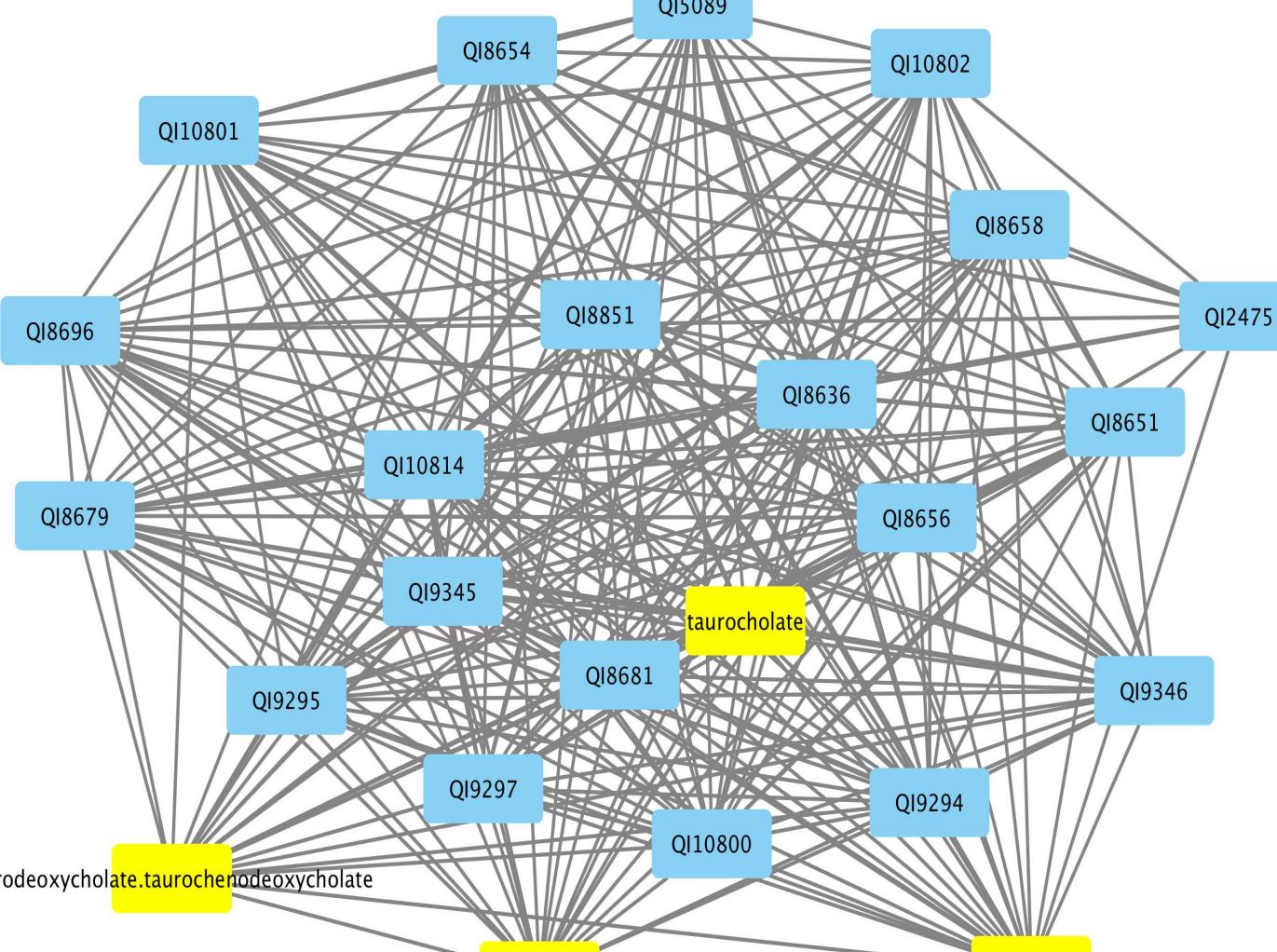
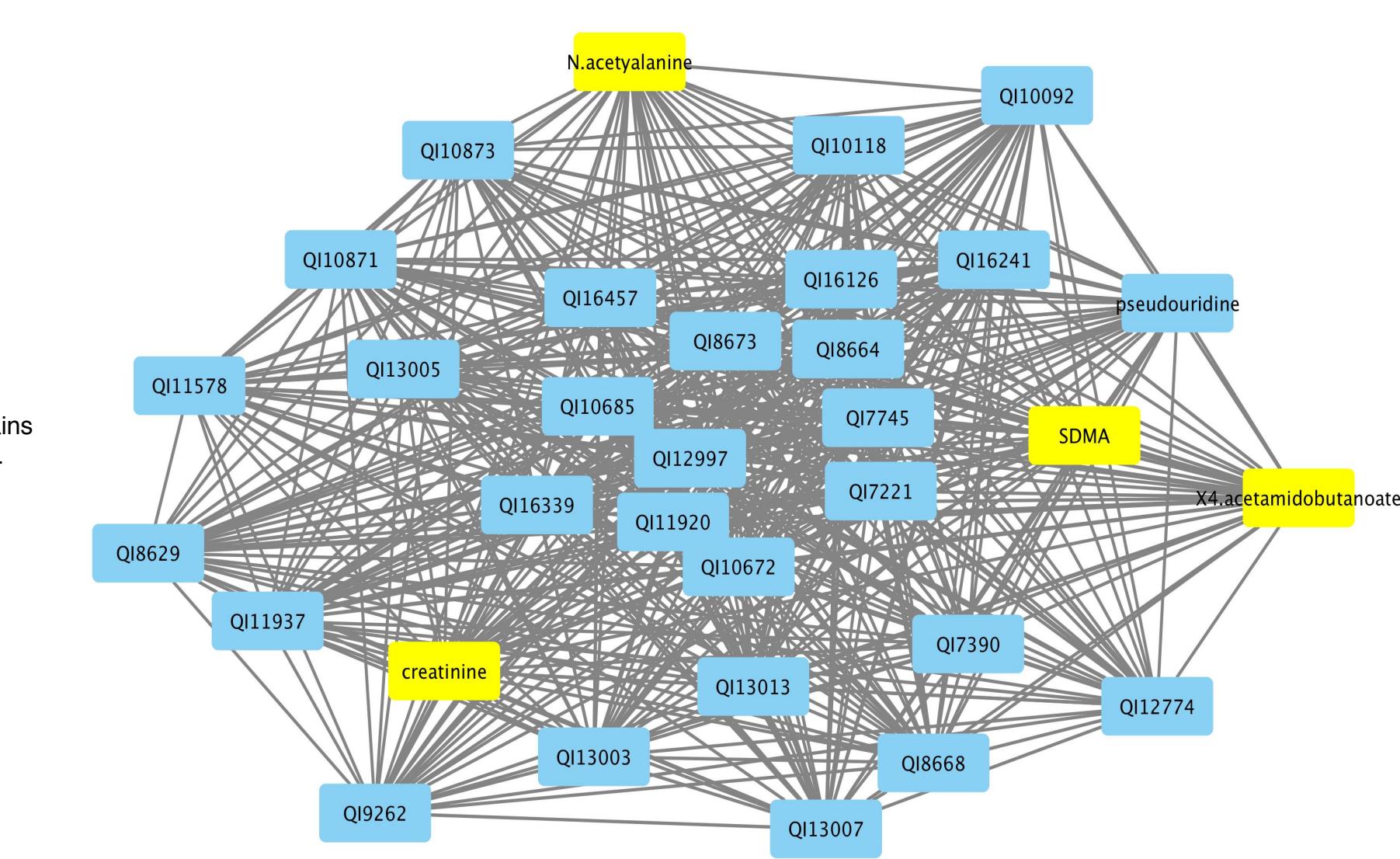


Figure 7. Module-blood traits correlation. Each node corresponds to a metabolite, and the edges represent link between metabolites in the module. Nodes colored in yellow are metabolites belonging to the same family. In this case, yellow nodes are bile acids. Node name starting with QI represent unknown metabolites.



## Conclusions

• Our metabolomics study in sickle cell disease patient, through the WGCNA framework, provided clusters of metabolites defined as modules. The modules can be associated with sickle related complications, and blood parameters.

• Our results identified metabolomic signatures of liver, gall bladder, and kidney complications in SCD. We shed light on key metabolites with a strong relationship with each of these organs damage.

• Our study suggest that although hemolysis is the key determinant of organ damage in SCD, understanding which specific metabolite or metabolic pathway plays a role in organ dysfunction can be exploited to predict (e.g. biomarker) or ameliorate SCD severity.

## References

- Langfelder P and Horvath S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* 2008; 9:559 doi:10.1186/1471-2105-9-559, Peter Langfelder, Steve Horvath
- Fast R Functions for Robust Correlations and Hierarchical Clustering. *Journal of Statistical Software*, 46(11), 1-17. <http://www.jstatsoft.org/v46/i11/>
- Analysis workflows were based on those contained in the excellent tutorials that can be found here: <http://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/Tutorials/index.html>