

# Prioritizing drug targets for non-alcoholic fatty liver disease based on comorbidity network analysis

Thanh-Phuong NGUYEN<sup>1\*</sup>, Laura CABERLOTTO<sup>2,3</sup>, Thomas SAUTER<sup>1</sup>, Frank LAMMERT<sup>4</sup> and Jochen G. SCHNEIDER<sup>4,5</sup>

<sup>1</sup>Systems Biology Group, Life Sciences Research Unit, University of Luxembourg

<sup>2</sup>Aptuit (Verona), Italy

<sup>3</sup>The Microsoft Research, University of Trento Centre for Computational Systems Biology

<sup>4</sup>Saarland University Medical Center, Dpt. of Internal Medicine II, Homburg, Germany

<sup>5</sup>Luxembourg Centre for Systems Biomedicine, University of Luxembourg

\*Corresponding author: [phuong.nguyen@uni.lu](mailto:phuong.nguyen@uni.lu)

## Introduction

Non alcoholic fatty liver disease (NAFLD) is highly associated with other components of the metabolic syndrome, in particular obesity and diabetes. More importantly, NAFLD can progress from simple steatosis to severe steatohepatitis. The exact pathophysiology is still unclear to date, but includes inflammation, immune cell infiltration, lipoapoptosis, fibrosis and cirrhosis. Cirrhosis is a risk factor for hepatocellular carcinoma. The lack of pharmacological treatment options and the mechanistic involvement of various biological processes such as oxidative stress, autophagy, unfolded protein response and inflammation suggest a systems biology approach to find novel targets for the disorder. However, systematic network medicine approach of the complex phenotypic dependencies and comorbidities has yet to be performed. Our aim was to identify novel drug targets for NAFLD treatment by a systematic *in silico* network analysis. The application of a supervised hypothesis generating methodology that considered a large datasets enabled us to detect underlying shared pathogenesis and to identify novel pathways to serve as marker or target.

## Method

There are four major steps in our proposed methods: (1) Data curation, (2) Comorbidity network construction, (3) Drug target prioritization, and (4) Clinical and biological interpretation.

1. We curated genes relevant to NAFLD and its comorbid diseases of interest (glucose metabolism disorders - GM, Lipid metabolism disorders - LM, obesity - OB, metabolic syndrome X - MS) from the Comparative Toxicogenomics Database (CTD) and the Medical Subject Headings (MESH), called "seed genes".
2. We reconstructed a protein interaction network by extracting the first neighbours of the seed genes based on the Human Protein Reference Database (HPRD).
3. We computed betweenness centrality (BC) to unravel the crucial genes in the network, which are potentially putative drug targets. A protein is a central connector mediating and maintaining the links between proteins in the network if it presents high BC. Large-scale shortest path analyses showed NAFLD connectivity to other comorbid diseases and allowed inferring the underlying molecular mechanisms. Shortest path analysis is to find a path between two proteins in a network that has the minimal number of its constituent edges. The genes in the shortest paths connecting the comorbid diseases are promisingly novel targets for drug repurposing.
4. The novel drug targets are functionally analyzed using GO-Term enrichment, pathway enrichment and literature review.

### 1. Data Curation

### 2. Comorbidity Network Construction

### 3. Drug Target Prioritization

### 4. Clinical/Biological Characterization

## Result

### 1- Central genes in the disease comorbidity network

We identified 594 disease genes, most of which (38.7%) were related to GM (n=230). LM disorders were represented by 116 genes, and OB by 114 genes; 19 genes were associated with MS, and 30 genes with NAFLD. Network construction on the 594 disease genes resulted in a network of 2,175 proteins and 4,605 interactions.

Among the most central genes in the network (Figure 1), 2 genes were directly related to NAFLD (TGFB1 and F2). 8 new candidate genes were inferred (APP, TP53, HSP90AA1, PRKACA, EP300, SRC, CSNK2A1, MAPK1).

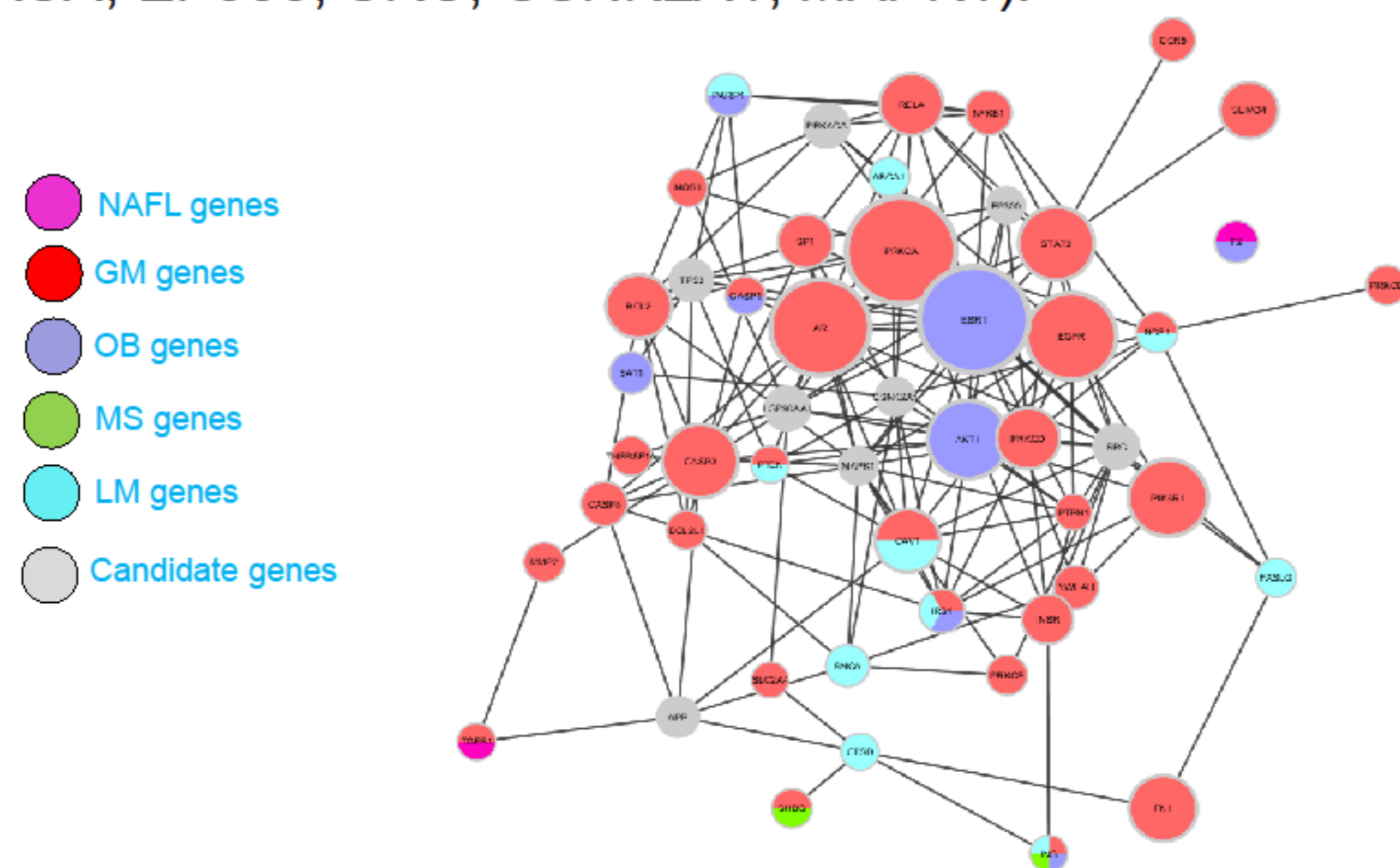


Figure 1: Network of top 50 central genes ranked by Betweenness centrality. Genes related to different diseases are presented in different colours. The multi-colour genes are genes shared by two or three or diseases. The bigger nodes obtain bigger BC.

### 2- Inferred genes underlying the comorbidities

We identified 10 key genes (Table 1) that were related to at least two diseases and then consecutively regulated the comorbidities between the diseases. Of note, most of the 10 genes were related to lipid and glucose metabolism, and two (INS, IRS1) were involved in aldosterone-regulated sodium absorption and diabetes.

Table 1: Ten key genes shared in the five diseases, NAFLD, LM, MS, GM, OB, LM.

	Lipid Metabolism	Metabolic Syndrome	Glucose Metabolism	Obesity
TGFB1	✓			✓
F2	✓			✓
INS		✓	✓	✓
IRS1		✓	✓	✓
CAV1	✓	✓		✓
NCF1	✓	✓	✓	✓
PTEN	✓	✓	✓	✓
SHBG		✓	✓	✓
PARP1	✓			✓
CASP1			✓	✓

### 3- Comorbidity between NAFLD and other diseases

We found a large number of shortest paths linking the NAFLD-related genes to the genes related to glucose and lipid metabolism disorders and obesity (46,143, 22,000 and 21,516 paths, respectively). Figure 2 shows that the five diseases are well-connected through sharing and mediating genes.

Investigating the sharing genes, we found that NAFLD shares 9 genes (SREBF1, TGFB1, PPARA, LEP, GSTM1, NFE2L2, TNFRSF1B, ACE, FAS) with GM, 4 genes (LDLR, SCARB1, PPARA, ACE) with LM, 2 genes (LEP, PPARA) with MS, and 2 genes (SREBF1, PPARA, PPARA, LEP, F2, TNFRSF1B) with OB.

In particular, those genes related to adipocytokine and PPAR signalling and cytokine-cytokine receptor interaction.

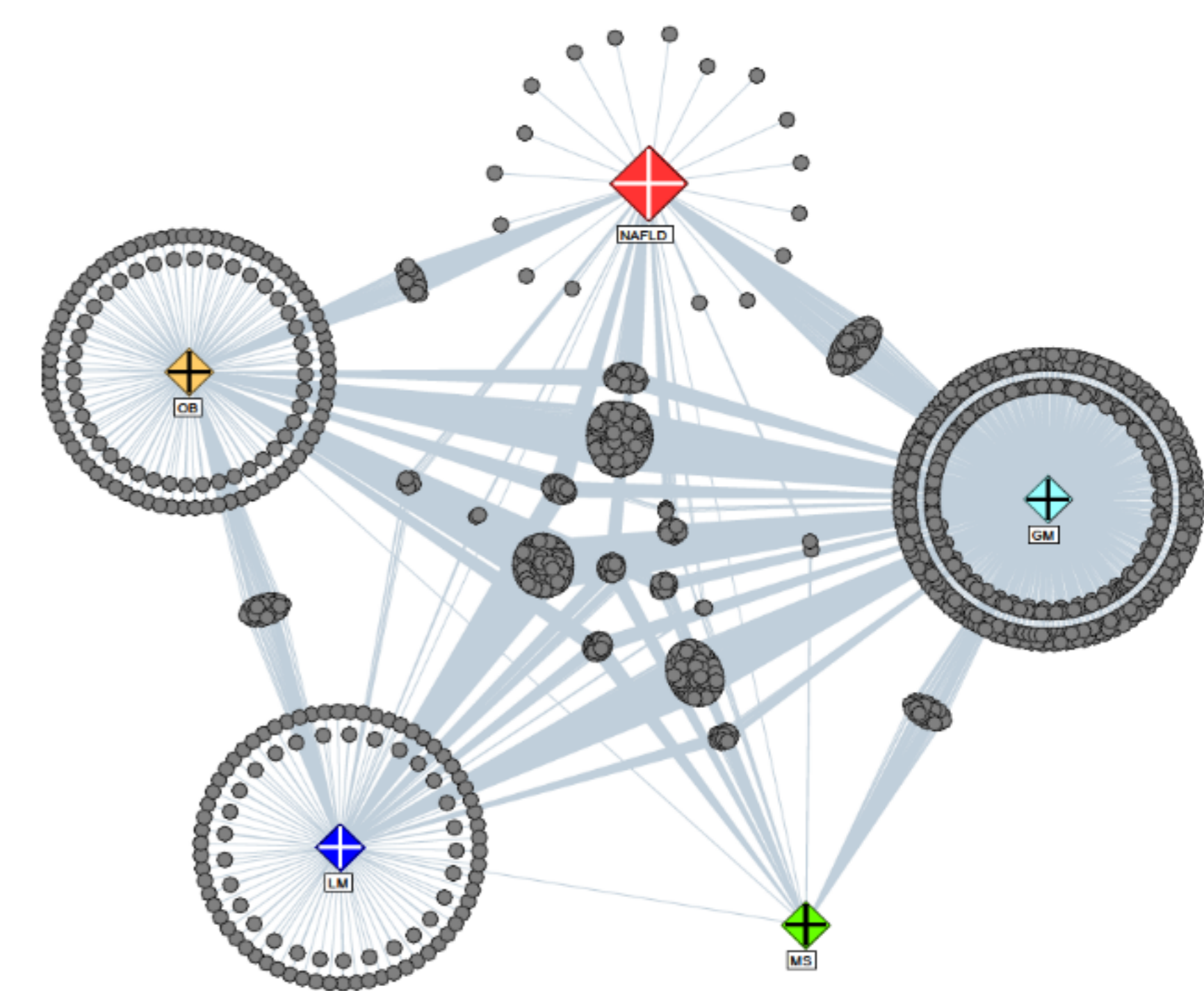


Figure 2: Comorbidity network collapsed 5 meta-nodes of the five diseases (NAFLD, LM, GM, OB, MS). Disease nodes are the diamond nodes and proteins interacting with disease proteins are the circular nodes.

## Conclusion

The comorbidity network analysis of NAFLD allowed the identification of genes that could represent the most promising molecular targets for prioritization of drug therapy in NAFLD. Our results confirm a potential use of PPAR $\alpha$  (PPARA) agonists in the treatment of NAFLD because of its effects on mitochondrial beta oxidation and reduced lipogenesis in the liver and for adipocytokines such as leptin (LEP) as a marker for the disease. In addition, we were able to demonstrate that thrombin (F2) and transforming growth factor 1 (TGFB1) related directly to NAFLD, and therapeutic administration of direct thrombin inhibitors may be used to reduce hepatic inflammation in humans as already demonstrated in mice.