

Phenotype network exploration between SNPs, genes, and pathways

For SNPs passing previously stated p-value threshold and associated with multiple phenotypes, we explored network connections and phenotypes in the context of pathways. We created three separate tables. We started by filtering association results to include only SNPs associated with more than one phenotype. We then used Biofilter to annotate these SNPs with information regarding gene locations, maintaining only SNPs located within canonical gene boundaries defined by Entrez [47]. Finally, we annotated these genes with pathway information from KEGG, allowing inference of connections between genes sharing the same pathway.

To visualize PheWAS results linking SNPs and phenotypes in a set of networks, the SNP-phenotype table was imported into Cytoscape 3.0 [48,49]. The SNP-gene table was then imported into Cytoscape, and this network was merged with our initial network. To explore various association patterns in the context of gene and pathway interactions, gene-pathway networks were integrated. The *yFiles orthogonal layout* was used to show network patterns.

We explored potential connections between SNPs and multiple phenotypes linked via gene information to biological pathways using KEGG, and visualized these networks. Correlated phenotypes tended to cluster together (e.g. hematocrit with hemoglobin; HDL-cholesterol with LDL-cholesterol and total cholesterol) as shown in **Figure 1**. This was because we had SNPs associating with correlated phenotypes. Integration of gene and KEGG pathway information provided insights into biological interactions. **Figure 2** (which zooms in on a region of **Figure 1**) shows a cluster of lipid phenotypes in the network, with many connections to various metabolic pathways. For example, *PGLS* (a

pentose phosphate pathway gene) connected with total cholesterol and high-density lipoprotein, because *PGLS* SNPs were associated with both phenotypes. Similarly, *NT5M* (a gene in purine metabolism and nicotinate and nicotinamide metabolism pathways) connected with all three phenotypes.

Figure 1

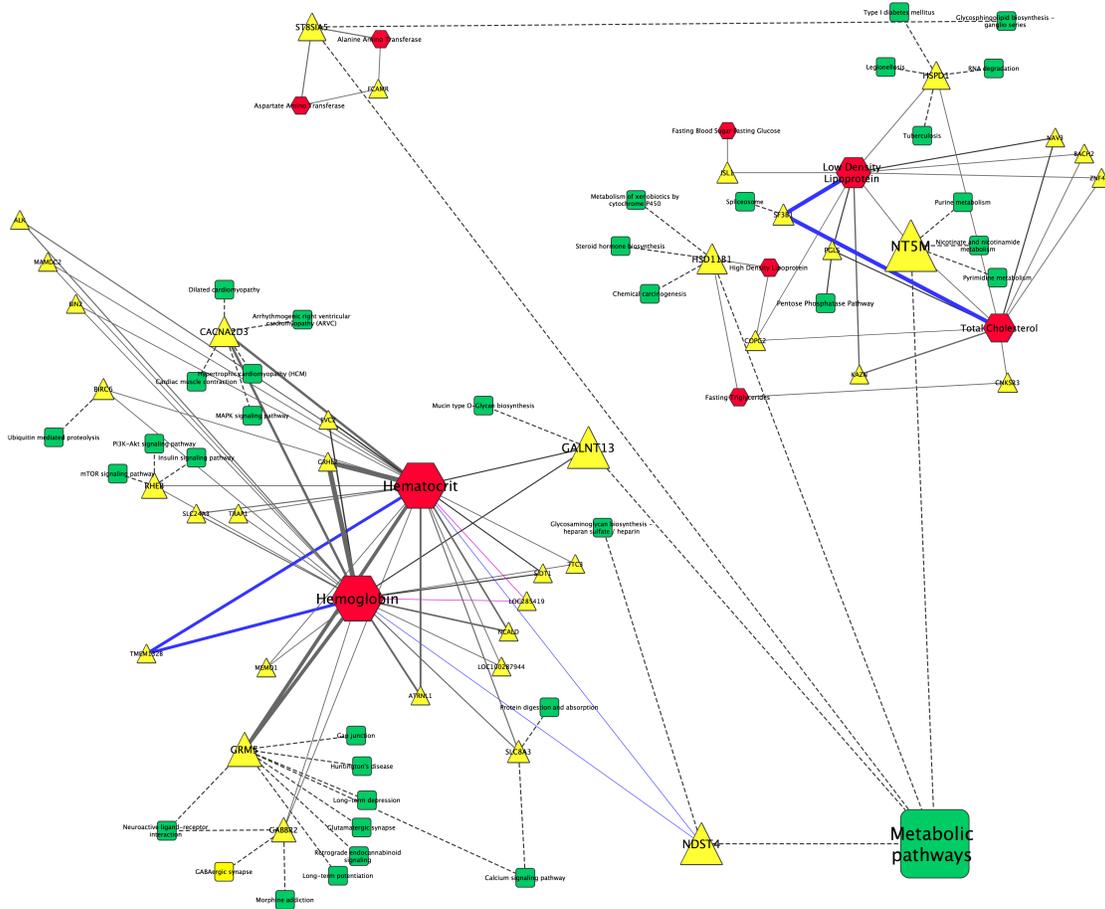


Figure 1. Pathway Connections visualized via Cytoscape. In the network red hexagons indicate associated clinical lab variables, and the yellow triangles are the genes the associated SNPs of this study are within. The thickness of the edge between gene and phenotype is scaled based on number of SNPs within an LD block. The color of edges

are: *blue* SNPs in LD block based on Dataset I, *purple* SNPs in LD block based on Dataset II, *grey* SNPs in LD block in both dataset. The green squares are the KEGG pathways the genes fall within. These results have been filtered only for SNPs associated with more than one phenotype.

Figure 2

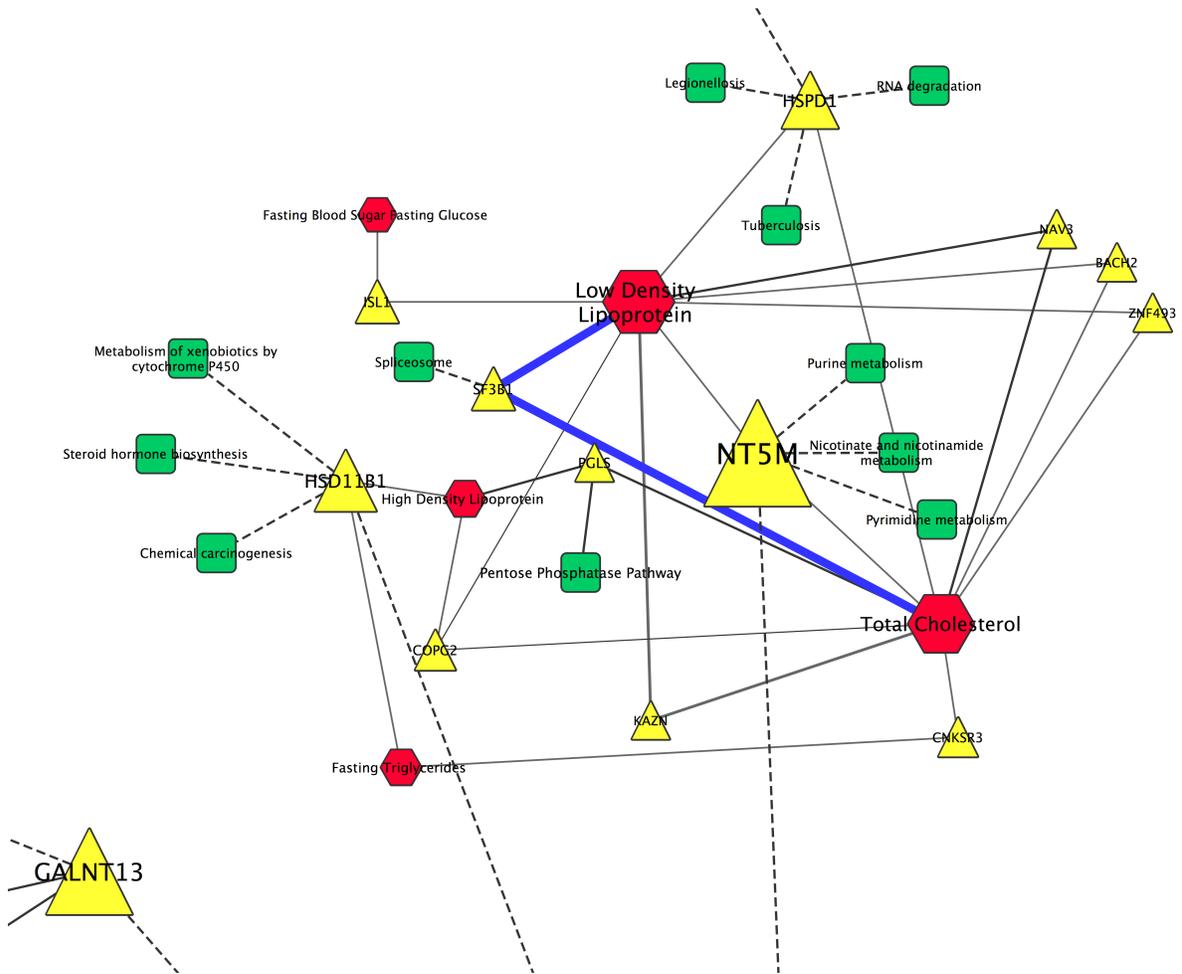


Figure 2. Cholesterol measurement links to metabolic pathways. The zoomed in section of the networks shows a cluster of all the cholesterol measurements. There are various connections between cholesterol measurements and genes that are present in metabolic pathways like *Purine metabolism*, *Pentose phosphatase pathway*, *Metabolic Pathway*, *Pyrimidine metabolism*.