Identification of key genes involved in non-alcoholic fatty liver disease progression

Aaron Welson, 210434989, Biomedical Sciences
A.Welson2@newcastle.ac.uk
Project Supervisor : **Heather Cordell**

Introduction

Non-alcoholic fatty liver disease (NAFLD) is a disease caused by excessive build-up of fat in the liver with non-alcoholic cause. NAFLD is a progressive disease and its stages are marked by the gradual development of abnormal liver phenotypes `(Figure 1)`, with later stages leaving patients no other choice than to perform a liver transplant.

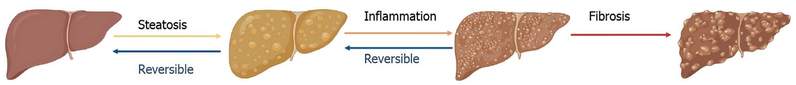

Figure 1: Stages of NAFLD development. Approximately 30% of the world population has simple steatosis, which carries a relatively benign prognosis. In 10-30% of cases however, NAFLD progresses further. This is marked by inflammation, then fibrosis. The resulting liver is susceptible to cirrhosis and hepatocellular carcinoma. (Figure adapted from [1])

Objectives

- Identification of genetic regions associated with quantitative traits in NAFLD. In the future this can help:
  - Develop genetic screens for NAFLD. This is useful since NAFLD is hard to detect due to only having little symptoms (until too late).
  - Help elucidate the mechanism of NAFLD disease progression.

Methods

Genome-wide association study (GWAS) involves using statistical techniques to determine associations between genome-wide single nucleotide polymorphisms (SNPs) and phenotypic traits. This study uses a quantitative phenotype approach as opposed to a case-control GWAS.

1. Cohort data (European) is the same one used in [2].
2. Quality control procedures and GWAS were performed with the software PLINK and GCTA respectively [3,4]
3. Four traits (Steatosis, DA, Fibrosis, NAS) of interest were studied.

   Steatosis refers to the degree of fat accumulation in the liver. DA stands for disease activity, its scoring is based upon hepatocyte ballooning and lobular inflammation. Fibrosis refers to the build up of scar tissue in the liver (usually due to inflammation). NAS stands for NAFLD activity score and involves a combination of the other three traits. See [2] for more information.

GWAS identifies **PNPLA3**, **SAMM50**, and **PARVB** as associated with quantitative traits in **NAFLD**

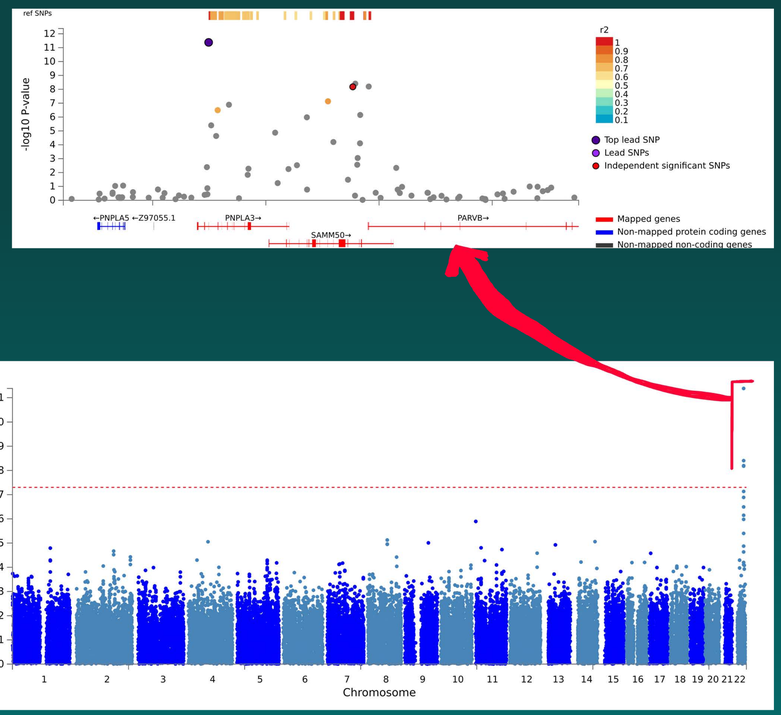

Figure 1. Manhattan plot (below) and LocusZoom plot (above) for the GWA analysis of the steatosis trait. SNPs with high association to the trait are identified with a threshold of $p<5\times10^{-8}$, which correspond to the genes: PNPLA3, SAMM50, and PARVB, on chromosome 22.

Plots made with FUMA GWAS

References

Newcastle University

Results

GWAS of steatosis consisted 2225 individuals. Significant SNPs were detected on chromosome 22 (Figure 1).

GWAS of fibrosis (2276 individuals) and NAS (2210 individuals) are similar to that of steatosis (Figures 2a, 2b). Notably, the same causative genes are involved (with exception to PARVB in fibrosis).

The lead SNP for these three traits is rs738409, with steatosis having the least p value ($p = 4.16680 \times 10^{-12}$).

GWAS of DA (2210 individuals) identified no significant SNPs.

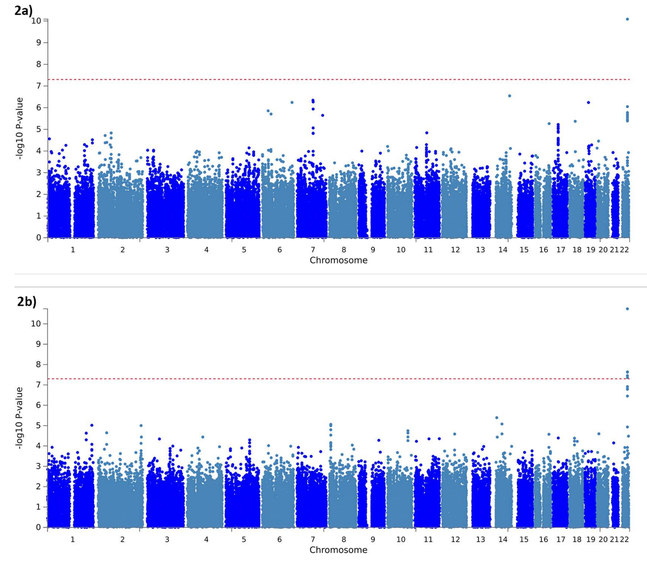

Figure 2: Manhattan plots for fibrosis (2a) and NAS (2b).

Conclusion

The results here suggests the involvement of the genes PNPLA3, SAMM50, and PARVB in NAFLD disease progression. Specifically in steatosis and fibrosis, but not inflammation (which the lack of significant SNPs for the DA trait suggests).

The same gene trio was identified in a similar study done on a Japanese population [5], corroborating the result across different ethnic groups. rs738409 has a well established association to NAFLD, it affects the PNPLA3 gene which encodes an enzyme (adiponutrin-3) involved in lipoprotein metabolism [6]. The other 2 genes do not have a clear liver function, so it is likely that the SNPs associated do not have any functional consequence and are simply in linkage with rs738409.

However, rs3788604 of the SAMM50 gene (red dot in figure 1) is independent to rs738409 and is worth looking at for future investigation.