Genome-wide association studies (GWAS) and post GWAS analysis of PBC

Yige Sima, Rebecca Darlay, Heather Cordell

Institute of Genetic Medicine, Newcastle University

Yige Sima 170054643

y.sima@newcastle.ac.uk

INTRODUCTION

Primary biliary cirrhosis (PBC) is a rare autoimmune liver disease resulting from progressive bile duct destruction, leading to cirrhosis and eventually liver failure. PBC is more common in females with a female to male ratio of 10 to 1. There is currently no effective targeted immune therapy [Knott 2019].

GWAS is a statistical analysis method that studies single nucleotide polymorphisms (SNPs), variations of a single nucleotide, in diseased and non-diseased people, which identifies SNPs related with a particular disease. Four previous GWAS studies already identified association between the human leukocyte antigen (HLA) locus (a fixed point on chromosome), and 27 non-HLA risk loci, with PBC pathogenesis [Cordell et al. 2015].

The aim of this project is to identify new disease loci contributing to PBC with larger cohorts including Italian, UK and Canadian data using both GWAS and post-GWAS analysis.

METHODS

Study Samples

UK (1816 cases and 5155 controls), **Canadian-UK** (4615 cases and 9233 controls), **Italian** (891 cases and 621 controls).

Quality Control (QC)

QC checks ware carried out across all three data sets using the software packages *PLINK* and *R*. QC removed SNPs with high missing rate < 95%, significant deviation from Hardy Weinberg Equilibrium ($P < 10^{-5}$), minor allele frequency < 1% and large differences in case and control missing rate (> 1%). QC also removed samples having high missing data rate (> 90%) and heterozygosity rate > six s.d. from the mean, related or duplicated samples, apparent gender discrepancies and samples with divergent ancestry. SNPS and samples passing QC were used to carry out genome-wide imputation.

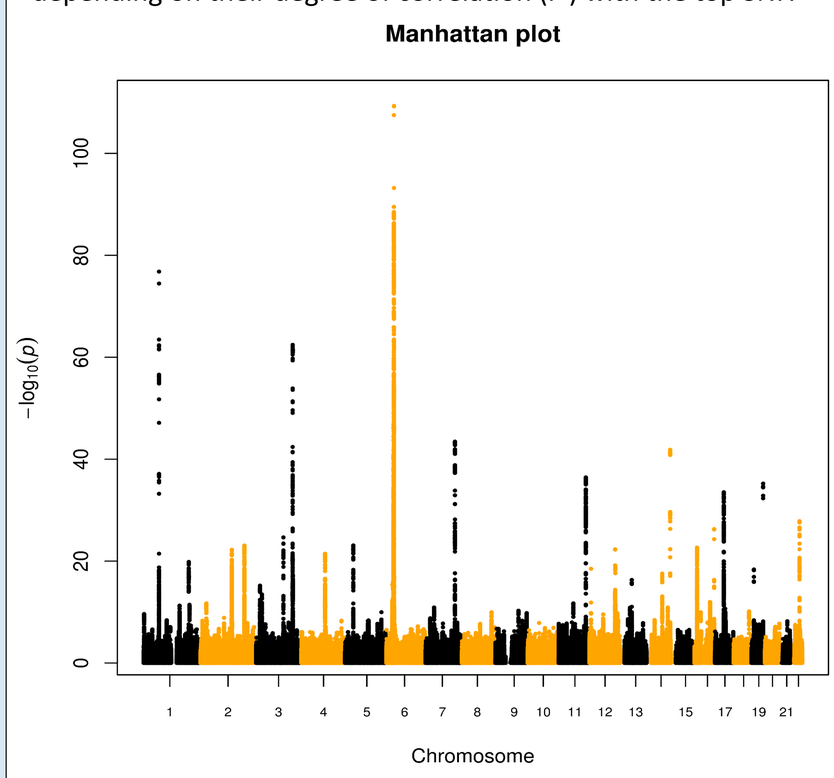
Statistical analysis

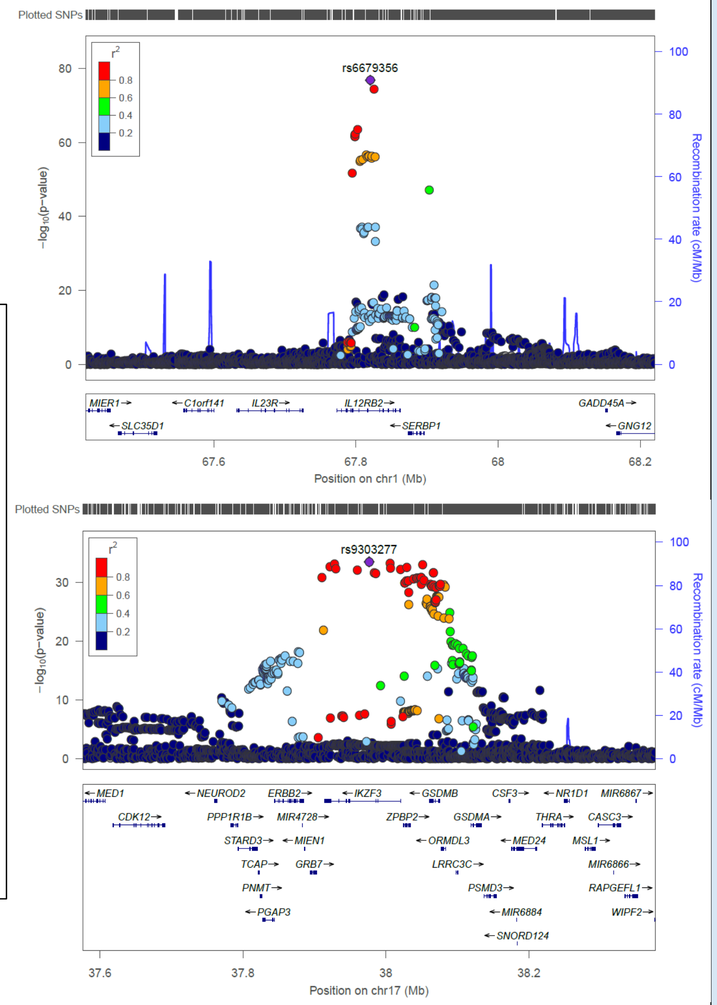
Logistic regression was performed on imputed data using *PLINK*. The results were meta-analysed using the software package *META*.

RESULTS

Figure1: The Manhattan plot shows association signals of genotyped and imputed SNPs for all three cohorts combined.

Figure 2-3: The LocusZoom plots show the top SNPs of two example PBC risk loci. Each circle represents the P value of one SNP, top SNPs are coloured in purple, other SNPs are coloured depending on their degree of correlation (r^2) with the top SNP.





DISCUSSION

The identification of multiple risk loci across the genome showed the important role of gene predisposition in the development of PBC. Similarly to other studies, the HLA locus showed significant association as it encodes molecules that play a key role in the immune system.

Further investigations are needed in order to understand the functional effects of these risk loci and to improve knowledge of PBC pathogenesis.

FUTURE WORKS

To identify specific biological pathways related to PBC development and possible drugs that target these pathways, which would help the development of new treatments that improve outcomes for PBC patients.

References:

Cordell, H. J., et al. "International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways." Nat Commun 2015;**6**:8019.

Knott, D. L. .Primary Biliary Cholangitis. 2019. [cited 2019. Available from: https://patient.info/doctor/primary-biliary-cholangitis-pro.