**Identifying and confirming the functional validity of genetic loci in alcohol consumption using UK Biobank**

Andrew Thompson1,2, James Cook3, Tarja Kinnunen4, Jeff Barclay5, Andrew P Morris3, Munir Pirmohamed1,2

1 Wolfson Centre for Personalised Medicine, Institute of Translational Medicine, University of Liverpool, United Kingdom

2Liverpool Centre for Alcohol Research University of Liverpool, United Kingdom

3Biostatistics, Institute of Translational Medicine, University of Liverpool, United Kingdom

4Institute of Integrative Biology, Institute of Translational Medicine, University of Liverpool, United Kingdom

5Cellular and Molecular Physiology, Institute of Translational Medicine, University of Liverpool, United Kingdom

**Purpose:** Determine genetic factors associated with extreme population-level alcohol consumption in white British individuals from UK Biobank. Examine the functional validity of genome-wide significant variants using model organisms and data mining techniques.

**Methods:** UK Biobank is a large population cohort of ~502,000 individuals from the UK aged 40-69 years at recruitment. Each participant completed a comprehensive demographic, lifestyle and, health questionnaire, underwent clinical measures, and provided biological samples. Questions from the UK Biobank baseline assessment were used to develop two study groups: heavy drinkers (cases: males consuming >50 UK units/week; females >35 units/week) and drinkers not reaching criteria for cases (controls). Genetic association analysis was conducted using a linear mixed model. The threshold for GWAS significance was set at *P* < 5 × 10–8, and distance based clustering was used for defining loci. Phenotypic and RNA interference experiments were performed at 20°C in a temperature-controlled room on young adult *C. elegans* selected from sparsely populated NGM plates. The external ethanol was 400 mM and location rate was the outcome of interest and was quantified by thrashing in Dent’s solution.

**Results:** Application of the phenotype resulted in the identification of 21,967 cases and 103,282 controls. The top SNP in the GWAS was the well-documented missense variant rs1229984 in *ADH1B*. Genome-wide significant signals were also identified at loci; *KLB*, (index SNP rs13130794), *FTO* (rs7206790), *KANSL1* (rs7225002), *CRHR1* (rs1635291), *ANKK1* (rs10891547), and LINC01833 (rs1004787). PheWAS outcomes provided evidence of these SNPs contributing to alcohol dependence, hypertension, skeletal disorders, gout, diseases/disorders of the peritoneum, haemorrhoids, obesity and diabetes mellitus. Through genetic correlation analysis of the entire genome, 26 significant correlations that survived multiple testing correction. The traits with the strongest correlations included several smoking variables (Ever vs never smoked [rg=-0.45, *PFDR* =1.06x10-13] and age of smoking initiation [rg=-0.36, *PFDR*=0.01]), several lung cancer outcomes (Squamous cell lung cancer [rg=0.37, *PFDR*=7.12x10-3] and lung cancer [rg =0.35, *PFDR*=1.96x10-5]) and Mothers age at death (rg=-0.43, *PFD* =3.89x10−5). Investigations on the acute effects of ethanol using *C.elegans* suggested a conserved role phenotypic responses to alcohol of the genetic targets identified from the GWAS.

**Conclusion:** These results offer insight into novel genes, pathways, and causal relationships for disease risk associated with heavy alcohol consumption.