**NAPS2 Cohort**

The NAPS2 cohort has been described previously (Whitcomb et al. 2008). Briefly, the NAPS2 cohort was designed to be a well powered cohort of RAP and CP individuals for the purpose of investigating the genetic, environmental, and metabolic components of RAP and CP. 1492 individuals with pancreatitis (3 AP, 510 RAP, and 979 CP) and 869 controls.

**Inclusion criteria**

AP was defined as at least two of the following three criteria: abdominal pain, elevated amylase or lipase > 3 times the upper limit of the normal range, or abdominal imaging findings consistent with AP. RAP was defined as two or more documented episodes of AP without evidence of CP on imaging. CP was defined as definitive evidence of CP by imaging using endoscopic retrograde cholangiopancreatography according to the Cambridge classification or by CT scan (Sarner and Cotton 1984). Subjects matching the criteria for AP, RAP, or CP were designated as cases, subjects that did not match the criteria for AP, RAP, or CP were designated as controls.

**Exclusion criteria**

Thirty-six Individuals were duplicates and were excluded from the analysis. Five individuals were excluded from the analysis after demonstrating cryptic relatedness with >50 individuals with a relatedness score > 0.2.

**Genotyping**

Individuals’ single nucleotide polymorphisms (SNPs) were genotyped using the Illumina HumanOmniExpress BeadChip and Illumina HumanCoreExome BeadChip, Illumina, San Diego, CA.

**Imputation**

Genotyped SNPs from the Illumina HumanOmniExpress BeadChip and Illumina HumanCoreExome BeadChips were used with reference genome GRCh37 in imputation separately using the Sanger Imputation Service (McCarthy et al. 2016) with EAGLE2 (Loh et al. 2016) used for phasing and PBWT (Durbin 2014) used for imputation. Imputed SNPs with an imputation quality score > 0.5 were included in further analyses.

**Principal Components of Ancestry and Genetic Relatedness Matrix**

Principal components of ancestry (PCA) and the genetic relatedness matrix (GRM) were generated from the autosomes using GCTA (Yang et al. 2011).

**Stepwise regression**

Stepwise regression was performed on the full model of age, age2, sex, BMI, alcohol use, and PCAs 1-10 to determine the appropriate model using the MASS package (Venables et al. 2003) in R v3.5.2 (R Core Team 2018). BMI was calculated using the subject’s height and weight. Alcohol use and smoking were self-reported by study participants. Alcohol use was defined as “yes” if the study participant reported current alcohol consumption, otherwise alcohol use was designated as “no”. Study participants were classified as a smoker if the study participant reported current smoking, otherwise the study participant was classified as a nonsmoker. Stepwise regression identified sex, BMI, smoking, PCA 1, and PCA 8 as contributing to the model of pancreatitis.

**Genome-wide association study**

GWAS was performed on 9,838,266 imputed SNPs with genotyping on the autosomes with a minor allele frequency ≥ 0.01 using GCTA (Yang et al. 2011). A multiple linear model controlled for age, sex, BMI, smoking, and PCAs 1-10 as covariates was used for the analysis along with a genetic relatedness matrix. Genome-wide significant (p ≤ 5 x 10-8) loci were defined using PLINK v1.90 (Purcell et al. 2007).

Bibliography

Durbin, R. 2014. Efficient haplotype matching and storage using the positional Burrows-Wheeler transform (PBWT). *Bioinformatics* 30(9), pp. 1266–1272.

Loh, P.-R., Danecek, P., Palamara, P.F., et al. 2016. Reference-based phasing using the Haplotype Reference Consortium panel. *Nature Genetics* 48(11), pp. 1443–1448.

McCarthy, S., Das, S., Kretzschmar, W., et al. 2016. A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics* 48(10), pp. 1279–1283.

Purcell, S., Neale, B., Todd-Brown, K., et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* 81(3), pp. 559–575.

R Core Team 2018. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Sarner, M. and Cotton, P.B. 1984. Classification of pancreatitis. *Gut* 25(7), pp. 756–759.

Venables, William N., Venables, W N and Ripley, B.D. 2003. *Modern Applied Statistics with S*. Text is Free of Markings. New York: Springer-Verlag New York, LLC.

Whitcomb, D.C., Yadav, D., Adam, S., et al. 2008. Multicenter approach to recurrent acute and chronic pancreatitis in the United States: the North American Pancreatitis Study 2 (NAPS2). *Pancreatology* 8(4–5), pp. 520–531.

Yang, J., Lee, S.H., Goddard, M.E. and Visscher, P.M. 2011. GCTA: a tool for genome-wide complex trait analysis. *American Journal of Human Genetics* 88(1), pp. 76–82.