**Abstract**

**Background:** Cystic fibrosis (CF) is the most common fatal genetic disease affecting Canadian children and young adults. At present, there is no cure. Further, non-CF genes have been identified that affect the severity of the symptoms of CF. These genes are called modifier genes. Our goal is to study the effects of three such modifier genes in the general public.

**Methods:** A PheWAS study was performed for the following modifier genes and their respective SNP’s: SCL26A9 and rs4077468 on chromosome 1, SLC6A14 and rs3788766 on Chromosome X and lastly, SCL9A3 and rs57221529 on Chromosome 5. We used the UK biobank data registry, which has over 500,000 participants, to perform the PheWAS study using the ICD10 codes hat were converted into phenotypes. A logistic regression model was used for finding associations between phenotypes and the modifier genes with phenotypes modeled as binary outcome variables and the minor allele count as the predictor (e.g. minor allele T count for rs4077469 C/T, being either 0, 1 or 2). The allele count was modeled as an additive model. Adjusted and unadjusted analysis were both performed. The model was adjusted for covariates age, age-square and sex. Since, there were 1511 phenotypes, 1511 logistic regression were performed for each gene and a Bonferroni correction was applied to determine statistically significant associations giving a corrected p-value of 3.309x. Further, separate analysis was performed for males and females for the gene SLC6A14 as it is found on X chromosome.

**Findings:** For the adjusted analysis for SCL9A3 with SNP rs57221529, we was found it to be statistically associated with having esophagitis, GERD and related diseases (OR = 1.064, S.E. = 0.013, p-value = 1.79x). Further, for each allele count (0, 1, 2) the number of cases were 7% to 8% of the controls with the minimum number of cases being 784 for allele count 2. With a total of 19,687 cases and 243,236 controls. For males the gene SLC6A14 at SNP rs3788766 was statistically associated with having Urinary obstruction (OR = 1.68, S.E. = 0.127, p-value = 4.24x). Further, for each allele count (0 or 2 as males cannot have one allele as this gene is on the X chromosome) there were only a total of 64 cases and 117,334 contols, with 37 cases for allele count 2 vs. 27 for allele count 0. For females no statistically significant association was found. Lastly, for the modifier gene SLC26A9 at SNP rs4077469 no association with any of the phenotypes was found. Very similar results were attained for the unadjusted analysis.

**Interpretation:** For the gene SCL9A3 and SNP rs57221529, C is the risk allele with OR = 1.064 which means that with every additional C allele the odds of having esophagitis, GERD and related diseases increases by 6.4%. Further, since there are a significant amount of cases and controls for each allele count this suggests that this is a actual association and not a sporadic one. For males even though the gene SLC6A14 was found to be statistically significantly associated with urinary obstruction, there were only 64 cases in total versus 117334 control suggesting this is possibly a sporadic relationship and not a real association. Hence, it is possible that there is an association between the gene SCL9A3 near the location SNP rs57221529 with having esophagitis, GERD and related diseases.

**key words**: Cystic fibrosis, UK biobank, modifier genes, SCL26A9, SLC6A14, SCL9A3