EBioMedicine 28 (2018) 180–186

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

Research Paper

A Genome-Wide Association Study Finds Genetic Associations with Broadly-Defined Headache in UK Biobank (N = 223,773)

Weihua Meng [a,*,1], Mark J. Adams [b,1], Harry L. Hebert [a], Ian J. Deary [c], Andrew M. McIntosh [b,c], Blair H. Smith [a]

[a] Division of Population Health Sciences, School of Medicine, University of Dundee, Dundee DD2 4BF, UK
[b] Division of Psychiatry, Edinburgh Medical School, University of Edinburgh, Edinburgh EH10 5HF, UK
[c] Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology, University of Edinburgh, Edinburgh EH8 9JZ, UK

ARTICLE INFO

Article history:
Received 5 December 2017
Received in revised form 20 January 2018
Accepted 20 January 2018
Available online 31 January 2018

Keywords:
Headache
Genome-wide association study
LRP1
UK biobank
Tissue expression

ABSTRACT

Background: Headache is the most common neurological symptom and a leading cause of years lived with disability. We sought to identify the genetic variants associated with a broadly-defined headache phenotype in 223,773 subjects from the UK Biobank cohort.
Methods: We defined headache based on a specific question answered by the UK Biobank participants. We performed a genome-wide association study of headache as a single entity, using 74,461 cases and 149,312 controls.
Results: We identified 3343 SNPs which reached the genome-wide significance level of $P < 5 \times 10^{-8}$. The SNPs were located in 28 loci, with the top SNP of rs11172113 in the *LRP1* gene having a *P* value of 4.92×10^{-47}. Of the 28 loci, 14 have previously been associated with migraine. Among 14 new loci, rs77804065 with a *P* value of 5.87×10^{-15} in the *LINC02210-CRHR1* gene was the top SNP. Significant relationships between multiple brain tissues and genetic associations were identified through tissue expression analysis. We also identified significant positive genetic correlations between headache and many psychological traits.
Conclusions: Our results suggest that brain function is closely related to broadly-defined headache. In addition, we found that many psychological traits have genetic correlations with headache.

1. Introduction

Headache is the most common neurological symptom, with a life time prevalence of over 90% in the general population in the UK (Boardman et al., 2003). It represents 4.4% of consultations in primary care and 30% of outpatient consultations in neurology (Larner, 2006; Latinovic et al., 2006).

According to the International Headache Society, headache can be generally divided into two categories: primary headache, if not associated with another disorder; and secondary headache if associated with an underlying medical illness (Headache Classification Committee of the International Headache Society, 2013). Primary headaches mainly include migraine, tension-type, and cluster headaches. Secondary headaches include any head pain caused by infection, neoplasm, head injury, some metabolic disorders, or drugs (Headache Classification Committee of the International Headache Society, 2013).

In a comprehensive review of population-based epidemiological studies of headache, the global prevalence of recurrent headache in all ages was found to be 46% for all headaches, including 11% for migraine and 42% for tension-type headache (Stovner et al., 2007). Tension-type headache is the most prevalent type of headache, whereas migraine is the most disabling (Riesco et al., 2007).

Migraine affects around 6 million people in England in the age range 16–65 and it is the sixth cause in terms of years of life lost to disability according to the Global Burden of Diseases 2013 (Global Burden of Disease Study 2013 Collaborators, 2015). Migraine costs the National Health Service almost £2 billion per year (Steiner et al., 2003). It presents with recurrent headache attacks and/or hypersensitivity to light and sound. Around one third of migraineurs experience an aura, which are transient neurological symptoms mostly involving the visual system (Silberstein, 2004).

Family studies and twin studies have suggested that both migraine and tension-type headache are heritable traits with a heritability over 40% (Russell et al., 2007; Schürks, 2012). Recently, genome-wide association studies (GWAS) have identified many genetic loci associated with migraine (Anttila et al., 2010; Anttila et al., 2013; Chasman et al., 2011; Freilinger et al., 2012; Ligthart et al., 2011). A GWAS meta-analysis of 375,000 patients involving 22 centers has identified 38 genetic susceptibility loci for migraine with the *LRP1* region in chromosome 12 being the most strongly associated (Gormley et al., 2016a). Along with other GWAS on migraine, the total number of loci identified to be associated with migraine is currently 47 (Gormley et al., 2016b). No GWAS have been performed for tension-type headache so far.

* Corresponding author.
E-mail address: w.meng@dundee.ac.uk (W. Meng).
[1] Weihua Meng and Mark J Adams contributed equally to this paper.

https://doi.org/10.1016/j.ebiom.2018.01.023

There are several phenotypic associations between headache and metabolic, psychological, and other factors such as obesity (Scher et al., 2003; Waldie and Poulton, 2002). Genome-wide association studies provide a potential route to discover genetic correlations with other complex traits and diseases that in turn may provide clues to shared genetic architectures and etiologies (Bulik-Sullivan et al., 2015a).

To identify the genetic variants associated with headache, we conducted this GWAS using the UK Biobank cohort which has never been contributed to genetic studies of headache including migraine. We used a broadly-defined headache phenotype, the one available in the UK Biobank dataset. Secondly, we sought to test for shared genetic associations with other complex traits and diseases using linkage-disequilibrium score regression (Bulik-Sullivan et al., 2015b).

2. Materials and Methods

2.1. Participants and Genetic Information of Participants

The UK Biobank is a health research resource that aims to improve the prevention, diagnosis and treatment of a wide range of illnesses. The UK Biobank cohort recruited over 500,000 people aged between 40 and 69 years in 2006–2010 across the UK. Participants completed a detailed clinical, demographic, and lifestyle questionnaire, underwent clinical measures, provided biological samples (blood, urine and saliva) for future analysis, and agreed to have their health records accessed. The informed consent of all participants has been obtained. Details of the UK Biobank resource can be found at www.ukbiobank.ac.uk. UK Biobank received ethical approval from the National Health Service National Research Ethics Service (reference 11/NW/0382). The current analyses were conducted under approved UK Biobank data application number 4844.

The detailed methods of DNA extraction and quality control can be found at http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/DNA-Extraction-at-UK-Biobank-October-2014.pdf. Participants' DNA was genotyped by bespoke Affymetrix UK Biobank chips. Standard QC steps were performed by the Wellcome Trust Centre for Human Genetics at Oxford University. The detailed QC steps can be found at http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=155580.

In July 2017, the genetic information (including directly genotyped genotypes and imputed genotypes) from 501,708 samples was released to UK Biobank project research collaborators. The detailed QC steps of imputation are described by Bycroft et al. (Bycroft et al., 2017).

2.2. Phenotypic Information on Pain

The UK Biobank participants were offered a pain-related questionnaire, which included the question: 'in the last month have you experienced any of the following that interfered with your usual activities?'. The options were: 1. Headache; 2. Facial pain; 3. Neck or shoulder pain; 4. Back pain; 5. Stomach or abdominal pain; 6. Hip pain; 7. Knee pain; 8. Pain all over the body; 9. None of the above; 10. Prefer not to say. Participants could select more than one option. (UK Biobank Questionnaire field ID: 6159).

The headache cases in this study were those who selected the 'Headache' option for the above question, regardless of whether they had selected other options.

The controls in this study were those who selected the 'None of the above' option.

2.3. Statistical Analysis

In this study, we used BGENIE (https://jmarchini.org/bgenie/) to be the main GWAS software and removed single nucleotide polymorphism (SNPs) with INFO scores < 0.1, with minor allele frequency < 0.5%, or those that failed Hardy-Weinberg tests $P < 10^{-6}$. SNPs on the X and Y chromosomes and mitochondrial SNPs as well as imputed SNPs that were not in the Haplotype Reference Consortium panel were excluded from analyses. Standard Frequentist association tests using BGENIE was used to perform association studies adjusting for age, sex, body mass index (BMI), 9 population principal components, genotyping arrays, and assessment centers. Gender difference between cases and controls was compared using chi-square testing. Age and BMI were compared using independent t testing in IBM SPSS 22 (IBM Corporation, New York). SNP associations were considered significant if they had a P value $< 5 \times 10^{-8}$. GCTA was used to calculate SNP-based or narrow-sense heritability (Yang et al., 2011) using a genomic relationship matrix calculated from genotyped autosomal SNPs.

SNP functional annotations were applied by the FUMA web application and a Manhattan plot was generated by R (Watanabe et al., 2017). R was also used to generate the corresponding Q-Q plot, a tool to evaluate differences between cases and controls caused by potential confounders.

The gene analysis and gene-set analysis were performed with MAGMA v1.6, which was integrated in FUMA (de Leeuw et al., 2015). Both analyses were based on GWAS summary statistics. In gene analysis, summary statistics of SNPs are aggregated to the level of whole genes, testing the joint association of all SNPs in the gene with the phenotype. In gene-set analysis, individual genes are aggregated to groups of genes sharing certain biological, functional or other characteristics. This will provide insight into the involvement of specific biological pathways or cellular functions in the genetic etiology of a phenotype. Tissue expression analysis was obtained from GTEx (https://www.gtexportal.org/home/) which was also integrated in FUMA. The purpose of using FUMA web application was to provide extra information to visualize and interpret GWAS results.

In order to identify genetic correlations between headache and other complex traits, we used linkage disequilibrium score regression through LD Hub v1.4.1 (available at http://ldsc.broadinstitute.org/ldhub/) (Zheng et al., 2017). This web-tool uses individual SNP allele effect sizes and the average linkage disequilibrium in a region to estimate the bivariate genetic correlations of headache with 234 traits. Those with P values of 2.1×10^{-4} (0.05/234) or less should be regarded as surviving Bonferroni adjustment for multiple testing.

3. Results

3.1. GWAS Results

During the initial assessment visit (2006–2010), at which participants were recruited and consent was given by 501,708 UK Biobank participants, the specific pain question (see the Materials and Methods section for details) received 775,252 responses to all options. Among these responses, the number of participants who selected the 'Headache' option was 102,994 (cases), and the number of participants who selected the 'None of the above' option was 197,149 (controls). We further removed those whose ancestry was not white British ($n = 22{,}694$) based on principal component analysis, those who were related to one or more others in the cohort (a cut-off value of 0.025 in the generation of the genetic relationship matrix) (n = 52,166), those who were also participants in a Psychiatric Genomics Consortium Major Depressive Disorder cohort (n = 597), and those who failed quality-control (QC) (n = 913). Thus we finally identified 74,461 cases (27,350 males and 47,111 females) and 149,312 controls (71,480 males and 77,832 females) for the GWAS association analysis. After quality control, there were 9,304,965 SNPs for the GWAS analysis.

The clinical characteristics of these cases and controls are summarized in Table 1. There were statistical differences ($P < 0.001$) in age, sex and BMI between cases and controls.

We identified 3343 SNPs which reached GWAS significance of $P < 5 \times 10^{-8}$ (Fig. 1, Supplementary Table 1). These SNPs represented 28 independent loci including 14 newly-identified loci (Table 2).