Sex-and Age-Specific Genetic Analysis of Chronic Back Pain

Freidin, Maxim B.; Tsepilov, Yakov A.; Stanaway, Ian B.; Meng, Weihua; Hayward, Caroline; Smith, Blair H.

Published in:
Pain

DOI:
10.1097/j.pain.0000000000002100

Publication date:
2021

Sex- and age-specific genetic analysis of chronic back pain

Maxim B. Freidin[1], Yakov A. Tsepilov[2,3], Ian B. Stanaway[4], Weihua Meng[5], Caroline Hayward[6], Blair H. Smith[5], Samar Khoury[7,8,9], Marc Parisien[7,8,9], Andrey Bortsov[10], Luda Diatchenko[7,8,9], Sigrid Børte[11,12,13], Bendik S. Winsvold[12,13], Ben M. Brumpton[13], John-Anker Zwart[11,12,13], HUNT All-In Pain[14], Yurii S. Aulchenko[2,3,15], Pradeep Suri[16,17,18], Frances M.K. Williams[1]

1 – Department of Twin Research and Genetic Epidemiology, School of Life Course Sciences, King's College London, London, UK;

2 – Laboratory of Theoretical and Applied Functional Genomics, Novosibirsk State University, Novosibirsk, Russia;

3 – Laboratory of Recombination and Segregation Analysis, Institute of Cytology and Genetics, Novosibirsk, Russia;

4 – Harborview Medical Center, Kidney Research Institute, Division of Nephrology, School of Medicine, University of Washington, Seattle, USA;

5 – Division of Population Health and Genomics, Medical Research Institute, Ninewells Hospital and School of Medicine, University of Dundee, Dundee, UK;

6 – Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, UK;

7 – The Alan Edwards Centre for Research on Pain, McGill University, Montreal, Quebec, Canada;

8 – Department of Anesthesia, McGill University, Montreal, Quebec, Canada;

9 – Faculty of Dentistry, McGill University, Montreal, Quebec, Canada;

10 – Center for Translational Pain Medicine, Department of Anesthesiology, Duke University, Durham, USA;

11 – Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway;

12 – Department of Research, Innovation and Education, Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway;

13 – K. G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway;

14 – Full list of co-authors and affiliations are in Akcnowledgements section;

15 – PolyOmica, 's-Hertogenbosh, the Netherlands;

16 – VA Puget Sound Health Care System, Seattle, USA;

17 – Department of Rehabilitation Medicine, University of Washington, Seattle, USA.

18 – Clinical Learning, Evidence, and Research (CLEAR) Center, University of Washington, Seattle, USA.

Total number of pages: 30

Number of tables: 6

Number of figures: 4

Supplementary tables: 2

Supplementary figures: 1

Corresponding author: Prof Frances M.K. Williams; Department of Twin Research and Genetic Epidemiology, King's College London, St Thomas's Hospital, Westminster Bridge Road, London, SE1 7EH, UK; frances.williams@kcl.ac.uk;

Abstract

Sex differences for chronic back pain (cBP) have been reported, with females usually exhibiting greater morbidity, severity and poorer response to treatment. Genetic factors acting in an age-specific manner have been implicated but never comprehensively explored. We performed sex- and age-stratified GWAS and SNP-by-sex interaction analysis for cBP defined as "Back pain for 3+ months" in 202,077 males and 237,754 females of European ancestry from UK Biobank. Two and seven non-overlapping genome-wide significant loci were identified for males and females, respectively. A male-specific locus on chromosome 10 near *SPOCK2* gene was replicated in four independent cohorts. Four loci demonstrated SNP-by-sex interaction, although none of them were formally replicated. SNP-explained heritability was higher in females (0.079 vs 0.067, p = 0.006). There was a high, although not complete, genetic correlation between the sexes (r = 0.838±0.041, different from 1 with p = 7.8E-05). Genetic correlation between the sexes for cBP decreased with age (0.858±0.049 in younger people vs 0.544±0.157 in older people; p = 4.3E-05). There was a stronger genetic correlation of cBP with self-reported diagnosis of intervertebral disc degeneration in males than in females (0.889 vs 0.638; p = 3.7E-06). Thus, the genetic component of cBP in the UK Biobank exhibits a mild sex- and age-dependency. This provides an insight into the possible causes of sex- and age-specificity in epidemiology and pathophysiology of cBP and chronic pain at other anatomical sites.

INTRODUCTION

Epidemiological studies provide evidence for different risk of back pain (BP) between the sexes, with women usually demonstrating greater prevalence, severity and chronicity than men (2, 6, 10, 23, 36, 41, 42, 45, 50, 55). This may be explained in part by variation in socioeconomic, biological and psychological factors (gender disparities, sex hormones, BMI, depression, pain behaviours) (6, 42, 45, 55). However even after adjustment for these factors, females remain more affected by BP (42, 45). Sex differences also exist for the prevalence and severity of other chronic pain conditions (e.g. fibromyalgia, migraine, and irritable bowel syndrome), and their response to pain treatment (3) suggesting a general propensity for women to develop (or perceive and report) chronic pain rather than structural or anatomical differences.

The phenomenon of sex-specificity in complex disease is well known and various factors have been implicated, including hormone profiles and behavioural factors (40, 49, 52). Apart from these, genetic factors have also been considered as one of the possible contributors (20, 24, 35, 39, 51). While mechanisms of sex-specificity in chronic pain have been rigorously studied with respect to hormone levels, pain perception, psychosocial and behavioural factors (comprehensively reviewed in (3, 16)), few studies have explored the sex-specific impact of genetics on pain (4, 31, 33, 53). Classical twin studies provide some evidence for differential contribution of genetic factors to BP in males and females. Even though in a younger sample (16-41 years) no differences in heritability for lifetime risk of BP was observed between the sexes (22), different heritability estimates for BP have been obtained in people aged 70 years and older with modest additive genetic effects in men, but not in women (21). The same trend was observed for chronic neck pain (15). Using both a SNP-based approach and classical twin modelling, differential heritability estimates for the sexes have been demonstrated in a number of traits having high genetic or phenotypic correlation with BP such as waist circumference (19), obesity-related anthropometric traits (12, 19), subjective well-being (34), and insomnia (13). This raises the interesting possibility that sex-specific genetic risk factors for BP may explain sex-specificity in BP.

We have previously examined the UK Biobank dataset (47) to study genetic associations with BP, identifying three genome-wide significant loci (18). In the current study we set out to carry out SNP-by-sex interaction analysis for chronic BP (cBP), defined as BP lasting at least 3 months. We also explored age-specificity of genetic factors in cBP in males and females by the analysis of age-stratified groups.

METHODS

Study overview

The study pipeline is overviewed on Figure 1. First, we carried out sex- and age-specific GWASs for cBP in the discovery sample of Europeans from UK Biobank. Second, we carried out SNP-by-sex interaction analysis for sex-specific genetic markers. Third, we performed replication analysis of sex-specific loci and significant SNP-by-sex interaction loci in four independent cohorts. Then, we estimated and compared SNP-based heritability between the sex- and age-specific groups. Finally, we carried out a comparative sex- and age-specific analysis of genetic correlations for cBP and a wide range of complex traits from a publicly available database.

Sample and phenotype definition

The discovery sample for the study has been taken from UK Biobank, a resource following health and well-being of over 500,000 volunteer participants. The details of recruitment and assessment of the participants are described in full elsewhere (47). Briefly, the participants comprise people aged 40-69 year at the time of recruitment who were registered with a general practitioner in the UK. The participants were enrolled in 2006-2010 in 22 assessment centres in England, Wales, and Scotland and completed detailed touch-screen questionnaires on their demographics, lifestyle, health, and environment. Invitations to take part in UK Biobank have been issued to about 9.2 million individuals, of which about 5.5% accepted. Among other items, self-reported ethnic background has been assessed (White, Asian or Asian British, Black or Black British, Chinese, Mixed, and other). Additionally, genetic principal components have been used by UK Biobank to identify a more genetically homogeneous group of white Europeans (Caucasians) among those who self-classified as "White British".

For the purpose of the current study, we used the discovery sample comprised 451,324 European participants of UK Biobank who self-classified as "White British" ethnicity and were genetically similar by genetic principal components, as well as individuals who belonged to the same genetic PC cluster as the "White British" despite self-reporting other ethnic ancestry as described previously (18).

The phenotype of cBP was defined through a combination of the UK Biobank data-fields 6159 accumulating responses to the question: "Pain type(s) experienced in last month" and the UK Biobank data-field 3571: "Back pain for 3+ months". Those who indicated "Back pain" in response to the data-field 6159 (Pain types(s)) question and also replied "Yes" to the data-field 3571 (Back pain for 3+ months) question, were classified as cases. Those who did not indicate "Back pain" in response to the data-field 6159 or replied "No" to the data field 3571 question, were classified as

controls. Individuals who replied “Do not know” or “Prefer not to answer” to any of the questions were excluded. We also excluded those who reported the presence of “General pain for 3+ months” (data-field 2956) as this may represent chronic widespread musculoskeletal pain or fibromyalgia. The remaining sample comprised 439,831 individuals including 202,077 males (35,705 cases, 17.7%) and 237,754 females (43,230 cases, 18.2%). Mean age (±SD) in males and females was 57.5±8.1 and 57.1±7.9 years, respectively, and mean BMI was 27.9±4.2 and 27.0±5.1 kg/m^2, respectively.

Replication was carried out using European individuals from four datasets: Generation Scotland: Scottish Family Health Study (UK); the Orofacial Pain: Prospective Evaluation and Risk Assessment (OPPERA) cohort (US); the Nord-Trøndelag Health Study (HUNT) cohort (Norway); and the English Longitudinal Study of Aging (ELSA) cohort (UK) (Table 1).

Generation Scotland is a family-based genetic epidemiology study of 24,000 volunteers across Scotland aged 18-98 years (44). The replication sample for the current study comprised 19,599 individuals including 8,023 males and 11,576 females. The phenotype of chronic back pain was defined as the following. Participants first answer a question: “Are you currently troubled by pain or discomfort?”; those who reply positively are queried: “Have you had this pain or discomfort for more than 3 months?". To those who reply positively, the questionnaire gives specific sites participants can select: back pain; neck or shoulder pain; headache; facial or dental pain; stomach ache or abdominal pain; pain in your arms, hands, hips, legs or feet; chest pain; and other pain. Accordingly, the definition of chronic back pain cases in GS: those who selected Back pain option, while the controls – all other participants.

OPPERA is a project aiming to investigate the impact of genetic, physiological, psychological and clinical factors on the development of painful temporomandibular disorder (1). The replication sample for the current study comprised 1,584 individuals including 575 males and 1,009 females. The phenotype of chronic back pain was defined using the comprehensive pain survey questionnaire as the following: Participants that reported having more than 5 episodes of back pain in the last year and those that reported between 2-4 episodes last year and that the episode lasted more than two hours. Participants reporting chronic widespread pain and fibromyalgia were excluded.

The HUNT study is a population-based cohort of 125,000 Norwegian participants recruited during three waves between 1984-2008 (25). The replication sample for the current study comprised 66,534 individuals including 32,362 males and 34,172 females. The phenotype of cBP was defined as the following. The questionnaire data were used, with the participants who have replied “Yes” to ‘During the last year, have you suffered from pain and/or stiffness in your muscles and joints that has lasted for at least three consecutive months?’ and listed lower back or upper back pain as a relevant region

to the question 'Where did you have these complaints?', were classified as having cBP. Participants who self-reported or had hospital-diagnosed fibromyalgia were excluded both from cases and controls for greater compatibility with the UK Biobank definition of cBP used in the current study.

ELSA cohort is a longitudinal study of more than 27,000 individuals recruited during eight waves since 2002 (46). The replication sample for the current study comprised 6,115 individuals including 2,780 males and 3,335 females. The phenotype of cBP was defined as the following. The questionnaires across waves 3 to 8 were assessed and those who positively responded to the questions "Whether often troubled with pain" and "Whether feel pain in back" were considered to have BP during a particular wave, while those who replied negatively to the first and/or second question were considered not to have BP. After obtaining these data in each wave separately, those who were cases in at least two waves were defined as cBP cases, while the rest were defined as controls.

Genome-wide association study

GWAS in the discovery sample was carried out in males and females separately using BOLT-LMM v 2.3.2 (27). Linear additive genetic model was fitted adjusting for age, genotyping array type and the first 10 genetic principal components provided by UK Biobank. The following filters were applied: minor allele frequency >0.001, genotyping and individual call rates >0.98%, imputation quality score (INFO) >0.7. A total of 14,828,942 autosomal and X-chromosomal biallelic single nucleotide polymorphisms (SNPs) and short insertions/deletions remained after filters applied were analysed. The genome-wide significance threshold was taken as $p < 2.5E-08$ accounting for two GWAS studies in males and females.

Leading independent SNPs in associated loci were established by Conditional and Joint Association (COJO) analysis (57). This method tests significant SNPs in the locus of association and identifies genetic variants having the strongest effect independent of the presence of other variants in linkage disequilibrium (LD). LD score regression was applied to quantify the impact of polygenicity and unobserved confounders on the results of GWAS (9). The proportion of cBP risk variance explained by the analysed genetic factors (SNP-based heritability) was calculated using BOLT-LMM v2.3.2 and compared between males and females using z-statistic (38).

SNP-by-sex interaction analysis

SNP-by-sex interaction analysis was carried out via a comparison of SNP-effect size (regression coefficients from GWAS) in males and females using t-statistic (38). Prior to estimating the t-statistic, SNP-effects and standard errors were scaled by dividing them by the phenotype variance to account for the use of a linear regression model for a categorical phenotype having unequal prevalence in

the comparison groups. This was performed for lead SNPs from male- and female-specific GWAS only (n = 9); accordingly, the significance threshold was set as p = 0.05/9 = 0.0056.

Replication

For replication, association analyses were carried out in each cohort separately using appropriate software (Table 1). OPPERA cohort applied logistic regression adjusting for age, genetic principal components and technical covariates. Other cohorts applied linear mixed-effects models adjusting for age, cohort-specific covariates and relatedness via genetic kinship matrices. Meta-analysis of the replication cohorts was performed by Z-score approach implemented in METAL software (56) for sex-specific GWAS signals and by Fisher's combined probability test for SNP-by-sex interaction (17).

Genetic correlations

Genetic correlation is a measure of similarity between traits due to shared genetic factors. LD score regression was used to calculate the genetic correlations (8). Genetic correlations were calculated for cBP in males vs females; also they were calculated in sex-stratified groups between cBP and 832 complex traits available on LDhub (8, 59) and ten traits considered as putative risk factors for BP, which have previously been identified as having statistically significant genetic correlation with BP (18): osteoarthritis, self-reported intervertebral disc problems, scoliosis, smoking, BMI, well-being, intelligence, educational attainment, anxiety, and depression. Genetic correlations between cBP and other traits were compared between males and females using z-statistic after applying Fisher's z-transformation.

RESULTS

Age-specific prevalence of chronic back pain in UK Biobank

The prevalence of cBP in the total sample of 439,831 European individuals from UK Biobank was significantly lower in males vs females: 17.7% vs 18.2%, p = 1.4E-08. The prevalence of cBP in males remained fairly constant over age, while in females there was a gradual increase in prevalence with age, becoming significantly different between the sexes > 65 years (Figure 2). This pattern of higher cBP prevalence in older females is consistent with the results of meta-analysis of other cohorts (54). Based on this we divided the sample into two age strata "Younger than 65" (n= 158,245 males; n = 193,265 females) and "65+" (n= 43,832 males, n = 44,489 females). In the younger group, prevalence of cBP was the same in males and females (17.7%) but the older group showed statistically significantly higher prevalence among females (17.7 vs 20.2%; p = 0.002). Subsequent sex-specific genetic analysis was carried out in the total sample and within age strata.

Sex-specific genetic loci for chronic back pain

The results of GWAS for cBP in for males and females from UK Biobank are shown in Figure 3 and in Table 2. In males, 2 associated genetic loci were identified with the lead SNPs rs1678626 (10:73826335; at *SPOCK2*; β = -0.0068±0.0012; p = 2.4E-08), and rs72922230 (18:50394407; at *DCC*; β = -0.0069±0.0012; p = 2.4E-08). In females, 7 genetic loci were identified: rs367563576 (1:150495378; near *LINC00568*; β = 0.0067±0.0012; p = 7.6E-09), rs62327819 (4:147211141; at *SLC10A7*; β = -0.0070±0.0012; p = 8.1E-09), rs1039325 (5:30761421; near *RP11-136H13.2*; β = -0.0065±0.0011; p = 8.7E-09), rs116007789 (7:101223945; near *LINC01007*; β = -0.0785±0.0133; p = 3.3E-09), rs7834973 (8:69639672; at *C8orf34*; β = -0.0068±0.0012; p = 4.2E-09), rs12308843 (12:23974404; at *SOX5*; β = -0.0103±0.0013; p = 9.4E-15), and rs2391333 (13:107166694; at *EFNB2*; β = -0.0066±0.0012; p = 1.9E-08).

In both sexes, LD score regression indicated high polygenicity and no evidence for confounding in the results of GWAS (for males: λ GC = 1.1459, intercept = 1.0093±0.0074, and ratio (the impact of confounder-driven inflation) = 0.0530±0.0419; for females, λ GC = 1.2005, intercept = 1.0053±0.0068, and ratio = 0.0215±0.0273).

The genome-wide significant loci observed in males and females were mutually exclusive: i.e., genome-wide significant loci for one sex were not genome-wide significant in the other sex. Moreover, for 4 of the loci detected in females, the effect sizes were statistically significantly different from those seen in males, suggesting SNP-by-sex interaction (Table 2): rs367563576

(1:150495378; adjusted p_{int} = 0.0315); rs62327819 (4:147211141; adjusted p_{int} = 1.0E-08); rs1039325 (5:30761421; adjusted p_{int} = 0.0243); and rs2391333 (13:107166694; adjusted p_{int} = 0.0045).

Replication of sex-specific associations

Replication was attempted for the lead 9 SNPs using 4 independent cohorts (Table 3). One out of the nine loci was replicated after adjusting for multiple testing (p<0.0056=0.05/9): the locus tagged by rs1678626 on chromosome 10 near *SPOCK2* gene in males (Z = -2.992; p = 0.0028). The direction of the effect for rs1678626 was consistent across samples (rs1678626*T allele is protective). The locus tagged by rs62327819 on chromosome 4 near *SLC10A7* gene exhibited a significant p-value in females (Z = 2.818; p = 0.0048), but had opposite direction of effect (rs62327819*C allele in protective in discovery, but risk-increasing in replication). Also, in females a nominally significant replication was observed for the chromosome 13 locus near *EFNB2* gene tagged by rs2391333 (Z = 2.237; p = 0.0253). The strongest signal in the discovery in females on chromosome 12 near *SOX5* tagged by rs12308843 was not replicated (Z = -1.885; p = 0.0595) but showed the same direction of the effect (rs12308843*G allele is protective).

Replication of top SNP-by-sex interaction signals

Replication was attempted for the top SNPs in the 4 regions of the significant SNP-by-sex interactions (Table 4; Supplementary Figure 1). Either the top SNPs or proxy SNPs with LD>0.9 with the top SNP depending on the availability in replication cohorts were used. Only for the locus on chromosome 4, Fisher's combined probability test suggested a statistically significant SNP-by-sex interaction (χ^2 = 22.9, df=8, p = 0.0035; Table 4). However, despite it appearing to replicate, there is a discrepancy in the direction of effect between UK Biobank and the replication cohorts. Namely, in UK Biobank the rs7682719*T allele is positively associated with cBP in males and negatively in females, resulting in a positive sign for t-statistics for interaction (Table 2; Supplementary Figure 1), while the t-statistics in all replication studies are in the opposite direction (Table 2; Supplementary Figure 1). Also, there is a discrepancy in nominally significant associations in the replication cohorts. Namely, the SNP is nominally associated with BP in males in OPPERA, while it is nominally associated with BP in females in HUNT and ELSA (which is consistent with UK Biobank results). Thus, the results of replication do not support SNP-by-sex interaction for chromosome 4 locus established in UK Biobank.

SNP-heritability of chronic back pain by sex and age

In both age groups the SNP-based heritability was higher in females, and this effect was the most pronounced in the "65+" age group (Table 5). Trends towards lower heritability in males over age 65 and higher heritability in females over age 65 was observed, though these did not reach statistical significance. Genome-wide summary statistics were also available for Generation Scotland which allowed estimating SNP-based heritability. Consistent with the findings in UK Biobank, estimated heritability was higher in females compared to males in Generation Scotland: 0.129±0.020 vs 0.064±0.027, respectively, with a borderline statistical significance of the difference (p = 0.053).

Sex-specific genetic correlation

We estimated and compared the genetic correlation for cBP between males and females and also compared the genetic correlation for cBP and its risk factors between males and females. We applied both a hypothesis-driven approach using 10 traits considered to be risk factors for cBP and an agnostic approach using all complex traits available on LDHub.

The sexes were highly genetically correlated for risk of cBP, particularly in the young: total sample, 0.8377±0.0406 (p = 1.8E-94); "Younger than 65" age group, 0.8582±0.0494 (p = 1.3E-67); and "65+" age group, 0.5444±0.1565 (p = 5.0E-04). Between-sex genetic correlations within the age groups were statistically significantly different (p = 4.3E-05) (Figure 4, left panel). At the same time, between-age genetic correlation within sex strata did not differ significantly: 0.7656±0.1209 vs. 0.8585±0.0948 in males and females, respectively (p = 0.072) (Figure 4, right panel).

In females there was a significant genetic correlation between cBP and 9 of the 10 risk factors (except anxiety/panic attacks) (Table 6). For males genetic correlation was seen between cBP with most risk factors except anxiety/panic attacks and scoliosis (Table 6). A significant difference in genetic correlation in males and females was observed for self-reported intervertebral disc problems: r_g = 0.889 vs r_g = 0.638 in males and females, respectively (p_{adj} = 6.7E-06). Also, nominally significant differences of genetic correlation were found for cBP with peripheral joint osteoarthritis and BMI: p = 0.042 and 0.045, respectively.

Next, we analysed the genetic correlation between cBP and 10 risk factors as above by age groups followed by a comparison between them. In the "Younger than 65" group the genetic correlation in males and females were similar to that of the whole sample (Supplementary Table 1) including the difference in genetic correlation for self-reported intervertebral disc problems: 0.866 vs 0.659 in males and females, respectively (p_{adj} = 1.3E-04). However, in the "65+" group the following

differences compared to the whole sample were observed: no statistically significant genetic correlations of chronic BP with depression and smoking in males, no correlation with scoliosis in females, and no correlation with anxiety in either sex group (Supplementary Table 1). Again, in the "65+ group" there was a significant difference in genetic correlation between the sexes for self-reported intervertebral disc problems: 0.806 vs 0.480 in males and females, respectively (p_{adj} = 0.040). There was a nominally significant difference in correlation with depression: 0.137 vs 0.439 in males and females, respectively (p = 0.008).

Genetic correlation of cBP with 832 traits available on LDHub were filtered via removal of traits that did not pass LDHub internal quality criteria and those involving back pain definitions (Back pain, chronic back pain, dorsalgia, and "None of the above" response to question "Pain type(s) experienced in last months"). We adjusted the p-value for the number of the remaining traits (n = 747) and additionally removed traits that were non-significant after adjustment for multiple testing in both sexes. A total of 297 traits remained that were significantly correlated with cBP in at least one sex (Supplementary Table 2), of which 2 traits exhibited statistically significantly different genetic correlations in males vs. females after accounting for multiple comparisons: "Neck or shoulder pain experienced in last months" (0.7405 vs 0.8349 in males and females, respectively; p_{adj} = 8.1E-04); and "Serious illness/injury or assault to yourself in last 2 years" (0.3793 vs 0.6404; p_{adj} = 5.5E-03).

DISCUSSION

The current study tested the hypothesis that sex- and age-specificity exists in the genetic predisposition to cBP. The results of the study suggest small but statistically significant differences in SNP-based heritability in males and females (consistently observed in UK Biobank and Generation Scotland). We identified two and seven genome-wide significant loci in males and females, respectively. Comparing the results of the current study with our previous findings (18, 48) reveals that the loci *SPOCK2* and *DCC* found previously are driven by males, while the loci *C8orf346* and *SOX5* seen in previous studies are driven by females.

Only one locus was replicated in the four independent cohorts: *SPOCK2* in males. The lack of replication for other loci may in part be explained by the low power attributable to the rather small effect sizes that achieve significance only upon the use of such a large sample as UK Biobank. However, for the strongest locus in females on chromosome 12 near *SOX5* the power for replication was estimated at 95% considering a single locus and 81% considering 9 loci. For the *SLC10A7* locus in females we detected significant by p-value replication signal although the direction of the effect was opposite in replication compared to UK Biobank. The opposite effect of the *SLC10A7* locus between UK Biobank and replication cohorts raises a possibility of a flip-flop effect, reported seen in a number of genetic studies of complex human traits (26, 60). The phenomenon is thought to be based on the variable patterns of LD between the causal and marker SNPs and/or by the variation in the prevalence of the causal SNP (58). In particular, the effect direction of a weak marker SNP may be driven by the direction of the effect of linked strong causal variants, not explicitly analyzed in a GWAS (e.g. rare variants). At the same time, in another population these strong causal variants may be absent, too rare, or have a different LD pattern, so the direction of effect will be specific for the weak SNP and may be opposite for strong variants resulting in the observed flip-flop in the effect direction.

In total, four of nine sex-specific genomic loci exhibited significant SNP-by-sex interaction. For the locus on chromosome 4 near *SLC10A7* gene, effect direction in replication cohorts was opposite to the UK Biobank; as above. Considering a possibility of a flip-flop effect, the *SLC10A7* locus may be of interest for an in-depth analysis for SNP-by-sex interaction in cBP. None of the other loci were replicated in the independent cohorts. Thus, it is difficult to conclude if the observed SNP-by-sex interactions are specific to the UK Biobank dataset, or replication cohorts were not of sufficient size to detect an association.

Potentially interesting observations include the larger genetic correlation of cBP with self-reported intervertebral disc problems in males vs. females, and larger genetic correlations of cBP with serious

illness/injury or assault and neck or shoulder pain in females vs. males. The situation with intervertebral disc problems and neck or shoulder pain is consistent with the different prevalence of these traits in males and females: in people with cBP, the prevalence of self-reported disc problems was higher in males than in females (0.076 vs 0.063; p = 5.5E-13) consistent with higher genetic correlations in males compare to females; the prevalence of neck pain was lower in males than in females (0.420 vs 0.471; p < 2.2E-16) consistent with higher genetic correlations in females compared to males. However, the prevalence of serious illness/injury or assault to yourself was higher in males with cBP than in females (0.162 vs 0.133; p < 2.2E-16), opposite to the expectation based on the differences in genetic correlations that was higher in females. The difference in genetic correlation between males and females may in part reflect the different impact of other factors related to cBP, rather than only differential genetic background between the sexes. For example, one such factor may be doctors' diagnoses: UK General Practitioners may more readily assign a diagnosis of disc degeneration to males rather than females in the presence of cBP, due to a greater social acceptability of cBP in males attributable to an underlying structural problem. This type of social desirability bias could be driven by the attitudes or beliefs of both practitioners and patients and may reflect referral bias for imaging studies. Alternatively, pleiotropy of cBP with disc degeneration, neck pain and injury history may truly differ between the sexes.

Important observations were made examining the genetic background of cBP by age. Namely, we found variation in SNP-based heritability including its trend to decrease with age in males and increase in females. For many complex traits, heritability tends to decline with age but for some traits the opposite trend has also been reported (5, 7, 32, 37, 43). Explanations of this phenomenon include varying contribution of environmental influence with age as well as different genetic factors contributing to the risk of diseases or phenotypes manifestation in different age groups (37). Similarly, the same disease phenotype may reflect different underlying pathology at different ages (32).

Genetic correlation for cBP between males and females fell with age, too. Finally, the structure of genetic correlations of cBP and its putative risk factors changed with age. In particular, there was a decline in genetic correlation between cBP and depression in males with age, which was not seen in females, leading to a large magnitude difference in cBP vs depression genetic correlation between males and females in the "65+" age group: 0.137 vs 0.439; p = 0.008. Interestingly, the prevalence of depression among people with cBP was lower in males than females in both age groups: 0.077 vs 0.123, p < 2.2E-16, in males and females, respectively, in the "Younger than 65" group; and 0.046 vs 0.078, p < 2.2E-16, in males and females, respectively, in the "65+" group. Overall, this suggests changes in the relative contribution of genetic factors in males and females with age.

It should be noted that previously genetic correlation between the same trait estimated in family members in different environments (in our cases environment is the gender and age) have been used as an indicator of the gene-environment interaction (28). Falconer (14) suggested that the same trait measured in two different environments can be treated as two different traits. If the family genetic effects do not change across environments or if they are related such that performance of any genotype in environment 2 is proportional to that in environment 1, the genetic correlation of family members across environments is equal to one. The null hypothesis of no significant gene-environment interaction is rejected whenever the genetic correlation across environments is significantly less than one. In our case we used the ideologically similar approach for calculation of genetic correlation using GWAS results calculated in different environments. So, given the genetic correlation significantly less than one between males and females and between young and old, we suggest the existence of gene-by-sex and gene-by- age interactions. These estimates may still be biased by confounders having differential influence on the trait in different environments.

The study has several limitations. Most of the phenotypes explored were based on self-assessment and are inherently subject to recall bias. This is especially the case for a loosely defined phenotype of cBP. However, as has been seen with other complex phenotypes, self-reported measures may be fully comparable with objective measures (11). Concerning pain phenotypes specifically, UK Biobank data have been shown to be representative of general population and consistent with other studies in terms of chronic pain prevalence and its association with social, demographic, and psychological risk factors (29). Another limitation is the lack of consistency in cBP phenotype definition between the cohorts; this, further complicated by the different prevalence of the phenotype in replication cohorts, might have been one of the reasons we did not replicate the majority of sex-specific genomic loci and SNP-by-sex findings. Finally, we chose to focus only on the loci that were significant in sex-specific GWASs, while a genome-wide SNP-by-sex interaction analysis is warranted. However, the methodology of such the analysis is not yet fully developed and inherently low power for GxE analysis remains a major challenge (30), especially for such a complex and heterogeneous phenotype as BP.

Overall, our study suggests that the genetic component of cBP in the UK Biobank exhibits a mild sex- and age-dependency raising implications for age- and sex-stratified analysis of cBP. Such analyses may be fruitful for other types of chronic pain, musculoskeletal and somatic, given prior suggestions of sex- and age-specificity in other pain types and locations.

ACKNOWLEDGEMENTS

The study was carried out under UK Biobank approved project #18219. We are grateful to participants from all cohorts as well as the studies' clerical, clinical and support staff. All cohorts received ethical approvals from respective organisations

YSA is supported by the Russian Ministry of Education and Science under the 5-100 Excellence Programme and by the Federal Agency of Scientific Organizations via the Institute of Cytology and Genetics (project 0324-2019-0040-C-01 / AAAA-A17-117092070032-4). YAT is supported by the Russian Foundation for Basic Research (project 19-015-00151) and by the Russian Ministry of Education and Science under the 5-100 Excellence Programme. LD is supported by the Canadian Excellence Research Chairs Program (CERC9). CH is supported by an MRC University Unit Programme Grant MC_UU_00007/10 (QTL in Health and Disease).

Generation Scotland: Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006]. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Clinical Research Facility, University of Edinburgh. Edinburgh, Scotland and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award "Stratifying Resilience and Depression Longitudinally" (STRADL) Reference 104036/Z/14/Z).

OPPERA: OPPERA was supported by the National Institute of Dental and Craniofacial Research (NIDCR) grant number U01DE017018. The authors thank the OPPERA program staff and participants at the participating universities, and for resources specifically provided for OPPERA by these institutions: University at Buffalo, University of Florida, University of Maryland–Baltimore, and University of North Carolina–Chapel Hill. The OPPERA program also acknowledges resources specifically provided for this project by the participating universities: University at Buffalo, University of Florida, University of Maryland–Baltimore, and University of North Carolina–Chapel Hill. Funding for genotyping was provided by NIDCR through a contract to the Center for Inherited Disease Research at Johns Hopkins University (HHSN268201200008I). Data from the OPPERA study are available through the NIH dbGaP: phs000796.v1.p1 and phs000761.v1.p1

ELSA: The English Longitudinal Study of Ageing is jointly run by University College London, Institute for Fiscal Studies, University of Manchester and National Centre for Social Research. Genetic analyses have been carried out by UCL Genomics and funded by the Economic and Social Research Council and the National Institute on Ageing. All ELSA GWAS data have been deposited in the European Genome-phenome Archive (Database: EGAS00001001036). Data governance was provided

by the METADAC data access committee, funded by ESRC, Wellcome, and MRC (2015-2018: Grant Number MR/N01104X/1 2018-2020: Grant Number ES/S008349/1).

HUNT: The Nord-Trøndelag Health Study (The HUNT Study) is a collaboration between HUNT Research Centre, (Faculty of Medicine and Health Sciences, NTNU, Norwegian University of Science and Technology), Trøndelag County Council, Central Norway Regional Health Authority, and the Norwegian Institute of Public Health. The genotyping was financed by the National Institute of health (NIH), University of Michigan, The Norwegian Research council, and Central Norway Regional Health Authority and the Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU). The genotype quality control and imputation has been conducted by the K.G. Jebsen center for genetic epidemiology, Department of public health and nursing, Faculty of medicine and health sciences, Norwegian University of Science and Technology (NTNU).

HUNT All-In Pain authors:

Amy E Martinsen[1], Anne Heidi Skogholt[2,3], Cristen Willer[4], Egil Andreas Fors[5], Ingrid Heuch[6], Ingunn Mundal[7], Jonas Bille Nielsen[2,4,8], Knut Hagen[9], Kristian Bernhard Nilsen[10], Kristian Hveem[2,11], Lars Fritsche[12], Laurent F. Thomas[2,3], Linda M Pedersen[6], Maiken E Gabrielsen[2], Marianne Bakke Johnsen[1,2,13], Marie Udnesseter Lie[1,13], Oddgeir Holmen[11], Synne Øien Stensland[1,14] , Wei Zhou[15,16]

1 Research and Communication Unit for Musculoskeletal Health (FORMI), Department of Research, Innovation and Education, Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway.

2 K. G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway.

3 Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway.

4 Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan, Ann Arbor, 48109, MI, USA.

5 Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway.

6 Department of Research, Innovation and Education, Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway.

7 Department of Health Science, Molde University College, Molde, Norway.

8 Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark.

9 Department of Neuromedicine and Movement Science, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU), Trondheim, Norway.

10 Department of Neurology, Oslo University Hospital, Oslo, Norway.

11 HUNT Research Center, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway.

12 Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, 48109, MI, USA.

13 Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway.

14 NKVTS, Norwegian Centre for Violence and Traumatic Stress Studies, Oslo, Norway.

15 Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA.

16 Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA.

DATA AVAILABILITY

Summary statistics for GWAS and SNP-by-sex interaction analysis for UK biobank have been deposited on zenodo.com under doi:10.5281/zenodo.3560890.

COI

YSA is a founder and co-owner of PolyOmica and PolyKnomics, private organisations providing services, research, and development in the field of quantitative and statistical genetics and computational genomics. Other authors declare no conflicts of interest.

Figures legends

Figure 1 – The study pipeline overview. Sex- and age-specific GWASs for cBP in the discovery sample of Europeans from UK Biobank was carried out. Replication analysis for sex-specific genome-wide significant loci was performed in four independent European cohorts: HUNT (Norway), Generation Scotland (GS, UK), OPPERA (USA), and ELSA (UK). Next, we carried out SNP-by-sex interaction analysis for sex-specific genetic markers followed by replication analysis in the same cohorts. Then, we estimated and compared SNP-based heritability between the sex- and age-specific groups. Finally, we carried out a comparative sex- and age-specific analysis of genetic correlations for cBP and a wide range of complex traits. Details of the samples and methods used are provided in the Table 1 and the Methods section.

Figure 2 – Prevalence of chronic BP by age in males and females in UK Biobank. The dataset has been split into equally sized bins based on quantiles of age distribution. Whiskers indicate 95% CI. * $p < 0.05$. The plot was produced using epiDisplay package for R (https://cran.r-project.org/web/packages/epiDisplay/epiDisplay.pdf).

Figure 3 – Miami plot for sex-stratified GWAS for chronic back pain in Northern European sample from UK Biobank. Red line depicts the genome-wide significance threshold ($p < 2.5E-08$). Top panel shows the results for women, bottom panel shows the results for men. The plot was produced using EasyStrata package for R (www.genepi-regensburg.de/easystrata).

Figure 4 – Genetic correlation between males and females by age group (left panel) and between the age groups by sex (right panel). Whiskers represent 95% CI. Between-sex genetic correlations were significantly different in the age groups ($p = 4.3E-05$), while between-age genetic correlations within same sex group were not ($p = 0.072$).

Supplementary Figure 1 – Forests plots for sex-stratified effect sizes in UK Biobank and replication cohorts. Effect sizes in all cohorts except for OPPERA are transformed to log OR using the transformation $\beta/(\mu*(1-\mu))$, where β is linear regression effect size and μ is proportion of cases in the sample. This was done for compatibility with OPPERA cohort that was analysed using logistic regression. Details of cohorts are provided in Table 1. The plots were produced using foresplot package for R (https://cran.r-project.org/web/packages/forestplot/vignettes/forestplot.html)

REFERENCES

1. Bair E, Brownstein NC, Ohrbach R, Greenspan JD, Dubner R, Fillingim RB, Maixner W, Smith SB, Diatchenko L, Gonzalez Y, Gordon SM, Lim PF, Ribeiro-Dasilva M, Dampier D, Knott C, Slade GD. Study protocol, sample characteristics, and loss to follow-up: the OPPERA prospective cohort study. J Pain. 2013;14(12 Suppl):T2-19.
2. Barrero LH, Hsu YH, Terwedow H, Perry MJ, Dennerlein JT, Brain JD, Xu X. Prevalence and physical determinants of low back pain in a rural Chinese population. Spine (Phila Pa 1976). 2006;31(23):2728-34.
3. Bartley EJ, Fillingim RB. Sex differences in pain: a brief review of clinical and experimental findings. Br J Anaesth. 2013;111(1):52-8.
4. Belfer I, Segall SK, Lariviere WR, Smith SB, Dai F, Slade GD, Rashid NU, Mogil JS, Campbell CM, Edwards RR, Liu Q, Bair E, Maixner W, Diatchenko L. Pain modality- and sex-specific effects of COMT genetic functional variants. Pain. 2013;154(8):1368-76.
5. Bergen SE, Gardner CO, Kendler KS. Age-related changes in heritability of behavioral phenotypes over adolescence and young adulthood: a meta-analysis. Twin Res Hum Genet. 2007;10(3):423-33.
6. Bingefors K, Isacson D. Epidemiology, co-morbidity, and impact on health-related quality of life of self-reported headache and musculoskeletal pain--a gender perspective. Eur J Pain. 2004;8(5):435-50.
7. Brown WM, Beck SR, Lange EM, Davis CC, Kay CM, Langefeld CD, Rich SS, Framingham Heart S. Age-stratified heritability estimation in the Framingham Heart Study families. BMC Genet. 2003;4 Suppl 1:S32.
8. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, ReproGen C, Psychiatric Genomics C, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control C, Duncan L, Perry JR, Patterson N, Robinson EB, Daly MJ, Price AL, Neale BM. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015;47(11):1236-41.
9. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics C, Patterson N, Daly MJ, Price AL, Neale BM. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015;47(3):291-5.
10. Carmona L, Ballina J, Gabriel R, Laffon A, Group ES. The burden of musculoskeletal diseases in the general population of Spain: results from a national survey. Ann Rheum Dis. 2001;60(11):1040-5.
11. Cherny SS, Livshits G, Wells HRR, Freidin MB, Malkin I, Dawson SJ, Williams FMK. Self-reported hearing loss questions provide a good measure for genetic studies: a polygenic risk score analysis from UK Biobank. Eur J Hum Genet. 2020.
12. Chiu YF, Chuang LM, Kao HY, Shih KC, Lin MW, Lee WJ, Quertermous T, Curb JD, Chen I, Rodriguez BL, Hsiung CA. Sex-specific genetic architecture of human fatness in Chinese: the SAPPHIRe Study. Hum Genet. 2010;128(5):501-13.
13. Drake CL, Friedman NP, Wright KP, Jr., Roth T. Sleep reactivity and insomnia: genetic and environmental influences. Sleep. 2011;34(9):1179-88.
14. Falconer DS. The problem of environment and selection. American Naturalist. 1952;830:293-8.
15. Fejer R, Hartvigsen J, Kyvik KO. Heritability of neck pain: a population-based study of 33,794 Danish twins. Rheumatology (Oxford). 2006;45(5):589-94.
16. Fillingim RB, King CD, Ribeiro-Dasilva MC, Rahim-Williams B, Riley JL, 3rd. Sex, gender, and pain: a review of recent clinical and experimental findings. J Pain. 2009;10(5):447-85.
17. Fisher RA. Statistical Methods for Research Workers (4th ed.) Edinburgh: Oliver & Boyd; 1934.
18. Freidin MB, Tsepilov YA, Palmer M, Karssen LC, Suri P, Aulchenko YS, Williams FMK, Group CMW. Insight into the genetic architecture of back pain and its risk factors from a study of 509,000 individuals. Pain. 2019;160(6):1361-73.

19. Ge T, Chen CY, Neale BM, Sabuncu MR, Smoller JW. Phenome-wide heritability analysis of the UK Biobank. PLoS Genet. 2017;13(4):e1006711.
20. Gilks WP, Abbott JK, Morrow EH. Sex differences in disease genetics: evidence, evolution, and detection. Trends Genet. 2014;30(10):453-63.
21. Hartvigsen J, Christensen K, Frederiksen H, Petersen HC. Genetic and environmental contributions to back pain in old age: a study of 2,108 danish twins aged 70 and older. Spine (Phila Pa 1976). 2004;29(8):897-901; discussion 2.
22. Hestbaek L, Iachine IA, Leboeuf-Yde C, Kyvik KO, Manniche C. Heredity of low back pain in a young population: a classical twin study. Twin Res. 2004;7(1):16-26.
23. Ihlebaek C, Hansson TH, Laerum E, Brage S, Eriksen HR, Holm SH, Svendsrod R, Indahl A. Prevalence of low back pain and sickness absence: a "borderline" study in Norway and Sweden. Scand J Public Health. 2006;34(5):555-8.
24. Khramtsova EA, Davis LK, Stranger BE. The role of sex in the genomics of human complex traits. Nat Rev Genet. 2019;20(3):173-90.
25. Krokstad S, Langhammer A, Hveem K, Holmen TL, Midthjell K, Stene TR, Bratberg G, Heggland J, Holmen J. Cohort Profile: the HUNT Study, Norway. Int J Epidemiol. 2013;42(4):968-77.
26. Lin PI, Vance JM, Pericak-Vance MA, Martin ER. No gene is an island: the flip-flop phenomenon. Am J Hum Genet. 2007;80(3):531-8.
27. Loh PR, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed-model association for biobank-scale datasets. Nat Genet. 2018;50(7):906-8.
28. Lynch M, Walsh B. Genetics and analysis of quantitative traits. Sunderland, Mass.: Sinauer; 1998. xvi, 980 p. p.
29. Macfarlane GJ, Beasley M, Smith BH, Jones GT, Macfarlane TV. Can large surveys conducted on highly selected populations provide valid information on the epidemiology of common health conditions? An analysis of UK Biobank data on musculoskeletal pain. Br J Pain. 2015;9(4):203-12.
30. McAllister K, Mechanic LE, Amos C, Aschard H, Blair IA, Chatterjee N, Conti D, Gauderman WJ, Hsu L, Hutter CM, Jankowska MM, Kerr J, Kraft P, Montgomery SB, Mukherjee B, Papanicolaou GJ, Patel CJ, Ritchie MD, Ritz BR, Thomas DC, Wei P, Witte JS. Current Challenges and New Opportunities for Gene-Environment Interaction Studies of Complex Diseases. Am J Epidemiol. 2017;186(7):753-61.
31. Meng W, Deshmukh HA, Donnelly LA, Wellcome Trust Case Control C, Surrogate markers for M, Macro-vascular hard endpoints for Innovative diabetes Tools study g, Torrance N, Colhoun HM, Palmer CN, Smith BH. A Genome-wide Association Study Provides Evidence of Sex-specific Involvement of Chr1p35.1 (ZSCAN20-TLR12P) and Chr8p23.1 (HMGB1P46) With Diabetic Neuropathic Pain. EBioMedicine. 2015;2(10):1386-93.
32. Menni C, Mangino M, Zhang F, Clement G, Snieder H, Padmanabhan S, Spector TD. Heritability analyses show visit-to-visit blood pressure variability reflects different pathological phenotypes in younger and older adults: evidence from UK twins. J Hypertens. 2013;31(12):2356-61.
33. Mogil JS, Wilson SG, Chesler EJ, Rankin AL, Nemmani KV, Lariviere WR, Groce MK, Wallace MR, Kaplan L, Staud R, Ness TJ, Glover TL, Stankova M, Mayorov A, Hruby VJ, Grisel JE, Fillingim RB. The melanocortin-1 receptor gene mediates female-specific mechanisms of analgesia in mice and humans. Proc Natl Acad Sci U S A. 2003;100(8):4867-72.
34. Nes RB, Czajkowski N, Tambs K. Family matters: happiness in nuclear families and twins. Behav Genet. 2010;40(5):577-90.
35. Ober C, Loisel DA, Gilad Y. Sex-specific genetic architecture of human disease. Nat Rev Genet. 2008;9(12):911-22.
36. Oksuz E. Prevalence, risk factors, and preference-based health states of low back pain in a Turkish population. Spine (Phila Pa 1976). 2006;31(25):E968-72.
37. Ortega-Alonso A, Sipila S, Kujala UM, Kaprio J, Rantanen T. Genetic influences on change in BMI from middle to old age: a 29-year follow-up study of twin sisters. Behav Genet. 2009;39(2):154-64.

38. Pulit SL, Stoneman C, Morris AP, Wood AR, Glastonbury CA, Tyrrell J, Yengo L, Ferreira T, Marouli E, Ji Y, Yang J, Jones S, Beaumont R, Croteau-Chonka DC, Winkler TW, Consortium G, Hattersley AT, Loos RJF, Hirschhorn JN, Visscher PM, Frayling TM, Yaghootkar H, Lindgren CM. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. Hum Mol Genet. 2019;28(1):166-74.
39. Rawlik K, Canela-Xandri O, Tenesa A. Evidence for sex-specific genetic architectures across a spectrum of human complex traits. Genome Biol. 2016;17(1):166.
40. Regitz-Zagrosek V. Sex and gender differences in health. Science & Society Series on Sex and Science. EMBO Rep. 2012;13(7):596-603.
41. Schmidt CO, Raspe H, Pfingsten M, Hasenbring M, Basler HD, Eich W, Kohlmann T. Back pain in the German adult population: prevalence, severity, and sociodemographic correlates in a multiregional survey. Spine (Phila Pa 1976). 2007;32(18):2005-11.
42. Schneider S, Randoll D, Buchner M. Why do women have back pain more than men? A representative prevalence study in the federal republic of Germany. Clin J Pain. 2006;22(8):738-47.
43. Schousboe K, Visscher PM, Erbas B, Kyvik KO, Hopper JL, Henriksen JE, Heitmann BL, Sorensen TI. Twin study of genetic and environmental influences on adult body size, shape, and composition. Int J Obes Relat Metab Disord. 2004;28(1):39-48.
44. Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM, Deary IJ, Macintyre DJ, Campbell H, McGilchrist M, Hocking LJ, Wisely L, Ford I, Lindsay RS, Morton R, Palmer CN, Dominiczak AF, Porteous DJ, Morris AD. Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. Int J Epidemiol. 2013;42(3):689-700.
45. Smith BH, Elliott AM, Chambers WA, Smith WC, Hannaford PC, Penny K. The impact of chronic pain in the community. Fam Pract. 2001;18(3):292-9.
46. Steptoe A, Breeze E, Banks J, Nazroo J. Cohort profile: the English longitudinal study of ageing. Int J Epidemiol. 2013;42(6):1640-8.
47. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, Liu B, Matthews P, Ong G, Pell J, Silman A, Young A, Sprosen T, Peakman T, Collins R. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015;12(3):e1001779.
48. Suri P, Palmer MR, Tsepilov YA, Freidin MB, Boer CG, Yau MS, Evans DS, Gelemanovic A, Bartz TM, Nethander M, Arbeeva L, Karssen L, Neogi T, Campbell A, Mellstrom D, Ohlsson C, Marshall LM, Orwoll E, Uitterlinden A, Rotter JI, Lauc G, Psaty BM, Karlsson MK, Lane NE, Jarvik GP, Polasek O, Hochberg M, Jordan JM, Van Meurs JBJ, Jackson R, Nielson CM, Mitchell BD, Smith BH, Hayward C, Smith NL, Aulchenko YS, Williams FMK. Genome-wide meta-analysis of 158,000 individuals of European ancestry identifies three loci associated with chronic back pain. PLoS Genet. 2018;14(9):e1007601.
49. Taneja V. Sex Hormones Determine Immune Response. Front Immunol. 2018;9:1931.
50. Thomas E, Silman AJ, Croft PR, Papageorgiou AC, Jayson MI, Macfarlane GJ. Predicting who develops chronic low back pain in primary care: a prospective study. BMJ. 1999;318(7199):1662-7.
51. Traglia M, Bseiso D, Gusev A, Adviento B, Park DS, Mefford JA, Zaitlen N, Weiss LA. Genetic Mechanisms Leading to Sex Differences Across Common Diseases and Anthropometric Traits. Genetics. 2017;205(2):979-92.
52. Ullah MF, Ahmad A, Bhat SH, Abu-Duhier FM, Barreto GE, Ashraf GM. Impact of sex differences and gender specificity on behavioral characteristics and pathophysiology of neurodegenerative disorders. Neurosci Biobehav Rev. 2019;102:95-105.
53. Wang C, Cheng Y, Liu T, Li Q, Fillingim RB, Wallace MR, Staud R, Kaplan L, Wu R. A computational model for sex-specific genetic architecture of complex traits in humans: implications for mapping pain sensitivity. Mol Pain. 2008;4:13.

54. Wang YX, Wang JQ, Kaplar Z. Increased low back pain prevalence in females than in males after menopause age: evidences based on synthetic literature review. Quant Imaging Med Surg. 2016;6(2):199-206.
55. Webb R, Brammah T, Lunt M, Urwin M, Allison T, Symmons D. Prevalence and predictors of intense, chronic, and disabling neck and back pain in the UK general population. Spine (Phila Pa 1976). 2003;28(11):1195-202.
56. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26(17):2190-1.
57. Yang J, Ferreira T, Morris AP, Medland SE, Genetic Investigation of ATC, Replication DIG, Meta-analysis C, Madden PA, Heath AC, Martin NG, Montgomery GW, Weedon MN, Loos RJ, Frayling TM, McCarthy MI, Hirschhorn JN, Goddard ME, Visscher PM. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet. 2012;44(4):369-75, S1-3.
58. Zaykin DV, Shibata K. Genetic flip-flop without an accompanying change in linkage disequilibrium. Am J Hum Genet. 2008;82(3):794-6; author reply 6-7.
59. Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, Hemani G, Tansey K, Laurin C, Early G, Lifecourse Epidemiology Eczema C, Pourcain BS, Warrington NM, Finucane HK, Price AL, Bulik-Sullivan BK, Anttila V, Paternoster L, Gaunt TR, Evans DM, Neale BM. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics. 2017;33(2):272-9.
60. Zorina-Lichtenwalter K, Lichtenwalter RN, Zaykin DV, Parisien M, Gravel S, Bortsov A, Diatchenko L. A study in scarlet: MC1R as the main predictor of red hair and exemplar of the flip-flop effect. Hum Mol Genet. 2019;28(12):2093-106.

SUMMARY

Genetic factors of chronic back pain exhibit mild sex- and age-specificity. This raises the need for sex- and age-stratified analyses of chronic pain in future studies.

Figure 1

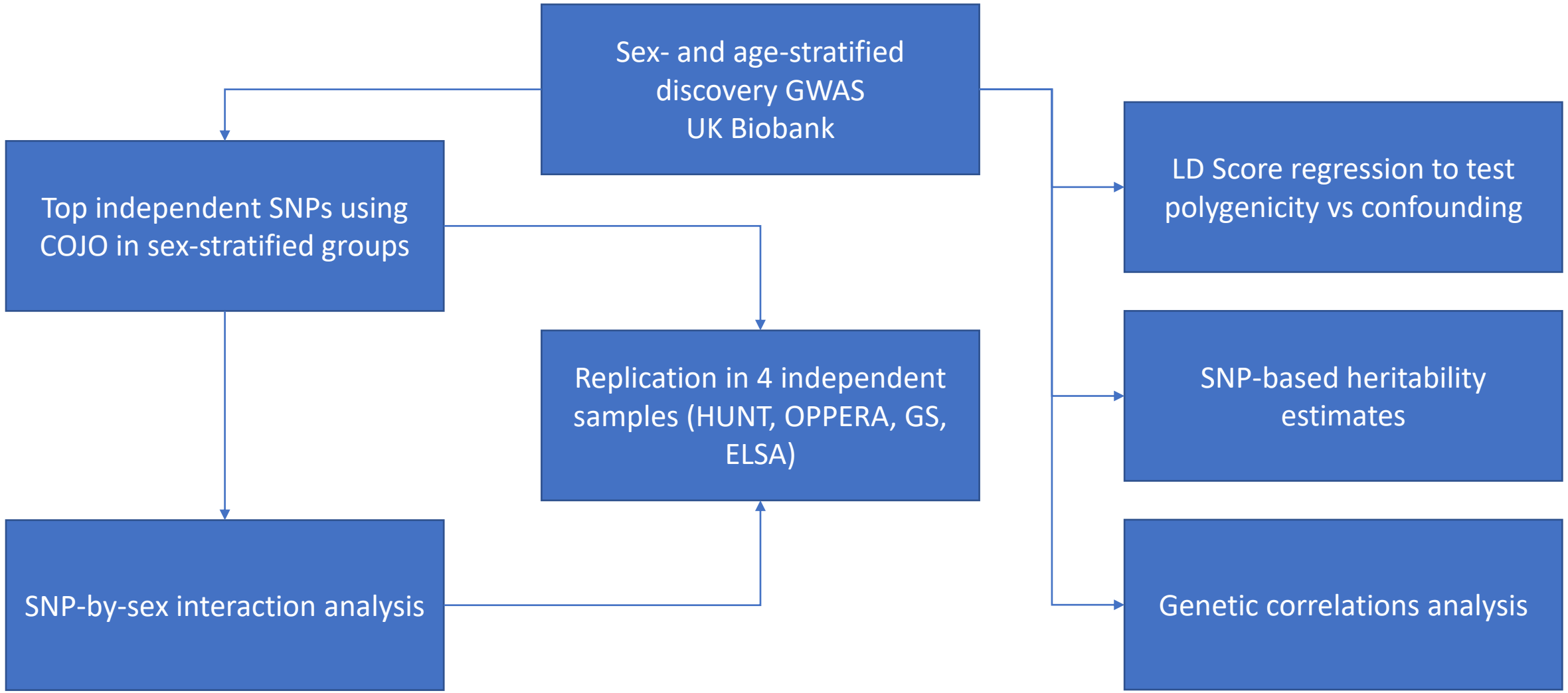
Sex- and age-stratified discovery GWAS UK Biobank
Top independent SNPs using COJO in sex-stratified groups
Replication in 4 independent samples (HUNT, OPPERA, GS, ELSA)
SNP-by-sex interaction analysis
LD Score regression to test polygenicity vs confounding
SNP-based heritability estimates
Genetic correlations analysis

Figure 2

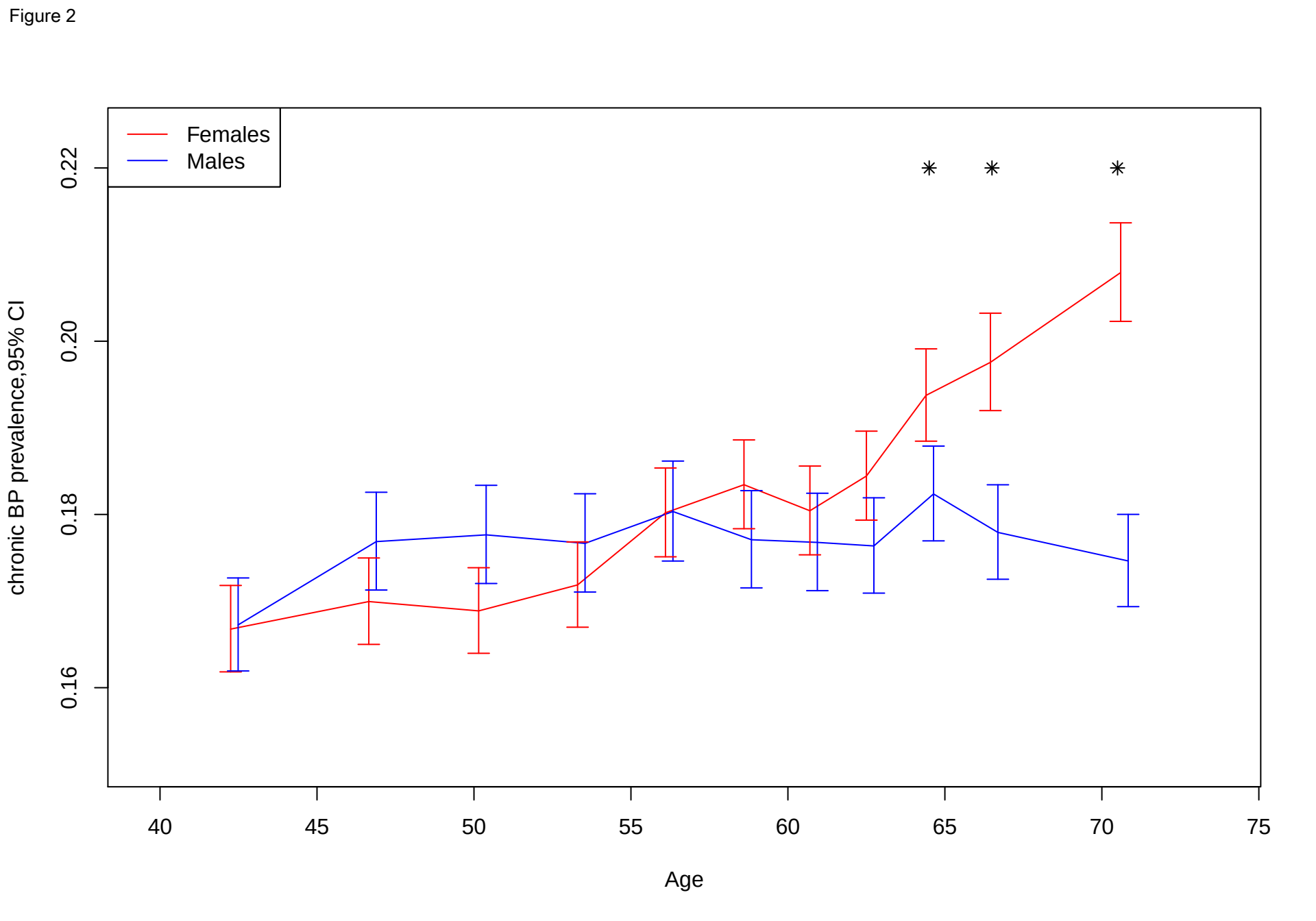
Females
Males
chronic BP prevalence,95% CI
0.16
0.18
0.20
0.22
40
45
50
55
60
65
70
75
Age
*
*
*

Figure 3

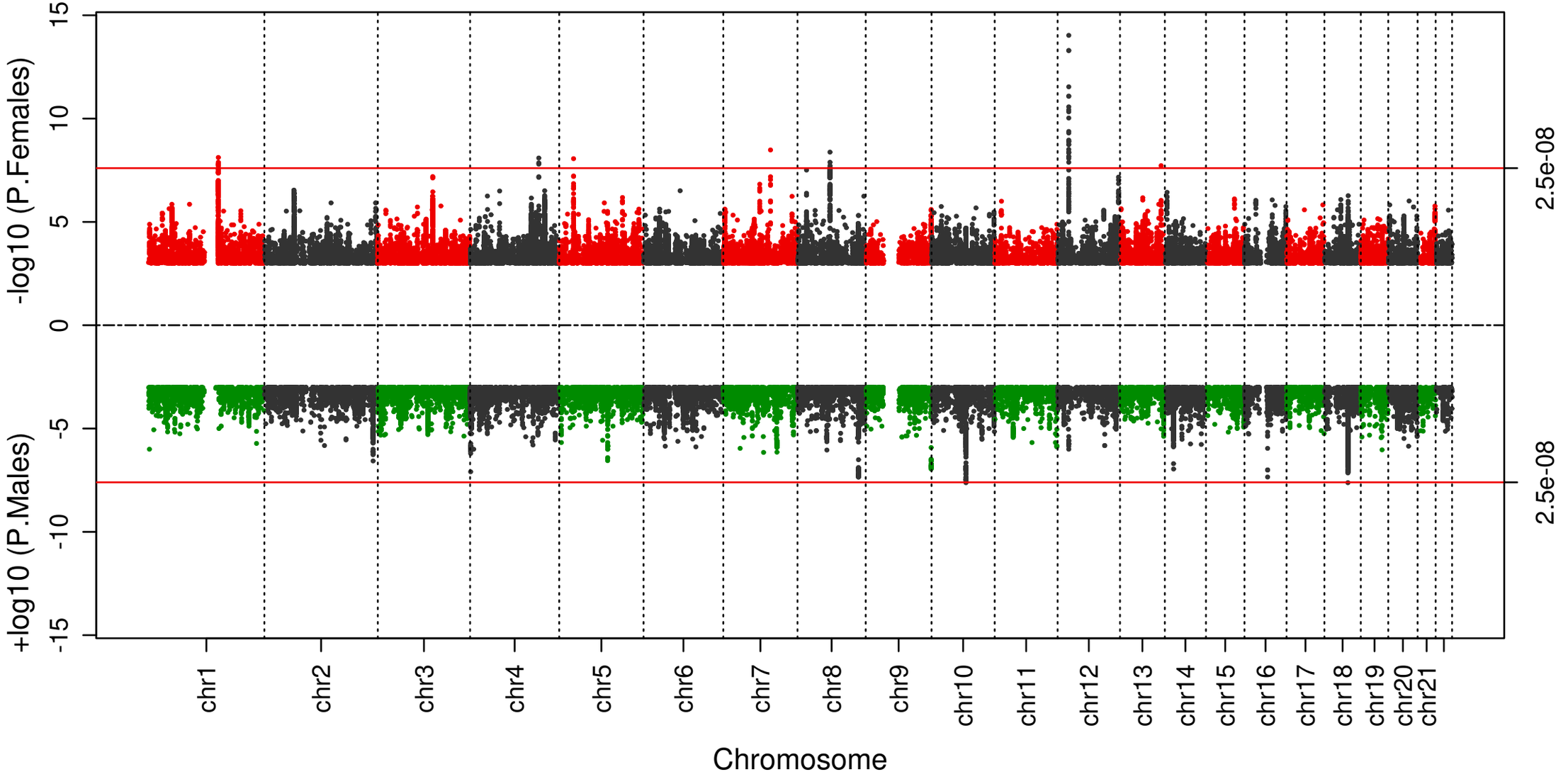
-log10 (P.Females)
+log10 (P.Males)
15
10
5
0
-5
-10
-15
2.5e-08
2.5e-08
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
Chromosome

Figure 4

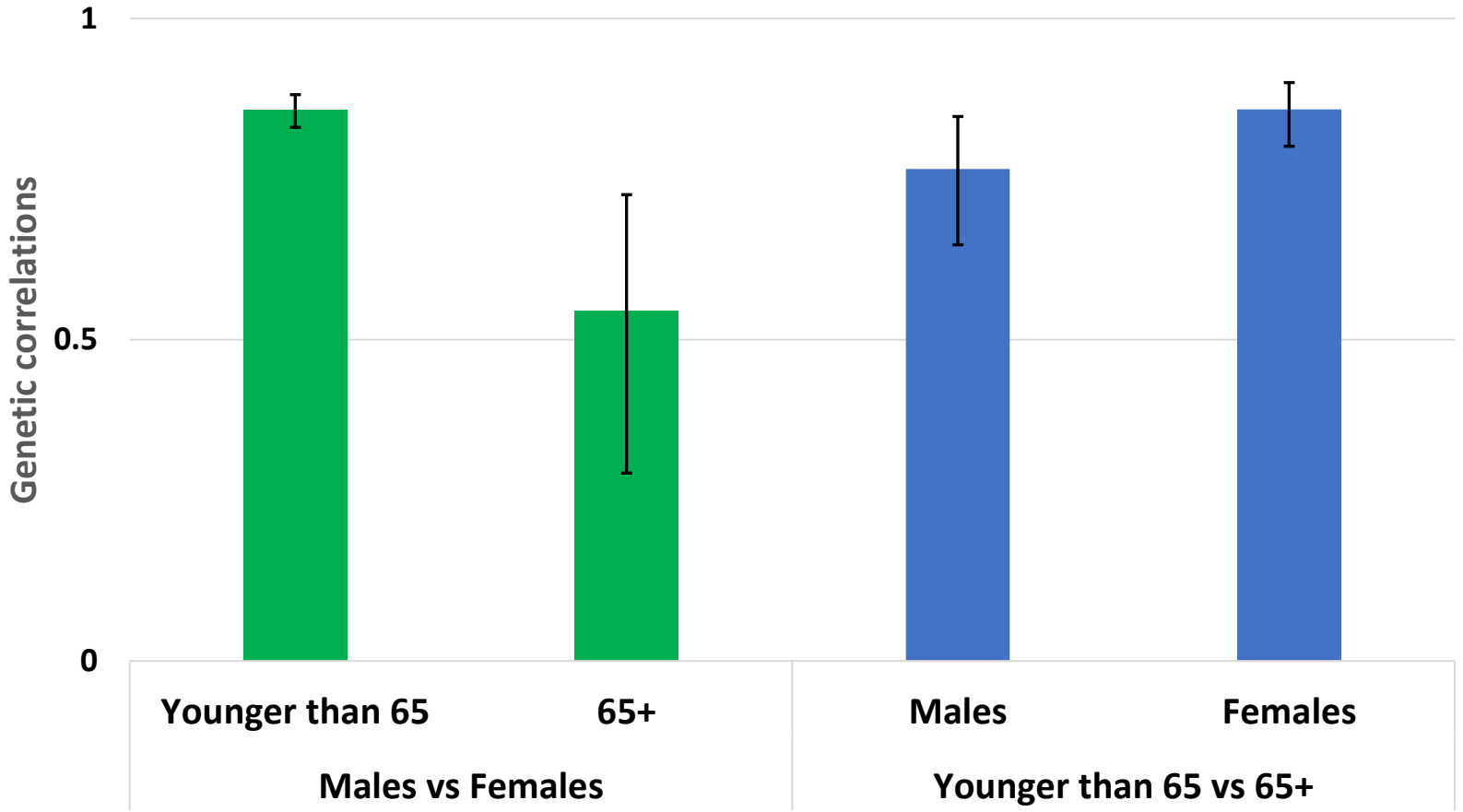
1
0.5
0
Genetic correlations
Younger than 65
65+
Males
Females
Males vs Females
Younger than 65 vs 65+

Table 1 – Discovery and replication cohorts

Trait	Group	UK Biobank (Discovery cohort)	Replication Cohorts			
			Generation Scotland (Scottish Family Health Study)	**OPPERA (Orofacial Pain: Prospective Evaluation and Risk Assessment)**	**HUNT (The Nord-Trøndelag Health Study)**	**ELSA (English Longitudinal Study of Aging)**
Sample size	Males	202,077	8,023	575	32,362	2,780
	Females	237,754	11,576	1,009	34,172	3,335
Cases / controls	Males	35,705 / 166,372	1,413 / 6,610	207 / 368	8,569 / 23,793	534 / 2,246
	Females	43,230 / 194,524	2,454 / 9,122	478 / 531	11,247 / 22,898	960 / 2,375
Age ± SD, years	Males	57.5±8.1	47.32±15.19	27.44±7.26	48.1±16.50	64.85±21.29
	Females	57.1±7.9	47.60±14.83	28.02±7.54	47.7±17.50	64.94±20.61
BMI± SD, kg/m^2	Males	27.9±4.2	26.93±4.52	25.68±4.72	26.6±3.60	25.85±7.79
	Females	27.0±5.1	26.55±5.68	24.93±5.98	26.1±4.60	25.37±8.81
Genotyping array		UK Biobank Affymetrix Axiom and UK BiLEVE Affymetrix Axiom array	Illumina HumanOmniExpressExome8v1-2_A or OmniExpressExome-8v1_A	Illumina HumanOmni 2.5M	Illumina HumanCoreExome	Illumina HumanOmni 2.5M
Imputation panel		UK10K, 1000 Genomes phase 3, HRC.r1-1	HRC.r1-1	1000 Genomes phase I	Customized reference panel*	1000 Genomes phase I
Association analysis software		BOLT-LMM 2.3.2	GCTA64	PLINK1.9	SAIGE 0.35.8.3	GCTA64

Legend to table 1: *The customized reference panel represented the merged panel of two reciprocally imputed reference panels: (1) 2,201 low-coverage whole-genome sequences samples from the HUNT study and (2) HRC v1.1 with 1,023 HUNT WGS samples removed before merging.

Table 2 – Genome-wide significant (p < 2.5E-08) loci associated with chronic BP in males and females in the total UK Biobank sample

Group	SNP	CHR:BP	Nearest / Overlapping gene*	Effect allele	Other allele	Effect allele frequency	Univariate analysis			Conditional and joint analysis			P_{int}**	
							Effect size	SE	P-value	Effect size	SE	P-value	Raw	Adjusted
Males	rs1678626	10:73826335	*SPOCK2*	T	C	0.445	-0.0068	0.0012	2.40E-08	-0.0068	0.0012	2.37E-08	0.1857	1.000
	rs72922230	18:50394407	*DCC*	A	G	0.598	-0.0069	0.0012	2.40E-08	-0.0069	0.0012	2.41E-08	0.1143	1.000
Females	**rs367563576**	**1:150495378**	***LINC00568***	**T**	**TAC**	**0.609**	**0.0067**	**0.0012**	**7.60E-09**	**0.0067**	**0.0012**	**7.66E-09**	**0.0035**	**0.0315**
	rs62327819	**4:147211141**	***SLC10A7***	**C**	**T**	**0.322**	**-0.0070**	**0.0012**	**8.10E-09**	**-0.0070**	**0.0012**	**8.10E-09**	**1.15E-09**	**1.04E-08**
	rs1039325	**5:30761421**	***RP11-136H13.2***	**T**	**G**	**0.417**	**-0.0065**	**0.0011**	**8.70E-09**	**-0.0065**	**0.0011**	**8.77E-09**	**0.0027**	**0.0243**
	rs116007789	7:101223945	*LINC01007*	C	T	0.998	-0.0785	0.0133	3.30E-09	-0.0785	0.0133	3.32E-09	0.0057	0.0513
	rs7834973	8:69639672	*C8orf34*	T	G	0.609	-0.0068	0.0012	4.20E-09	-0.0068	0.0012	4.26E-09	0.0142	0.1278
	rs12308843	12:23974404	*SOX5*	G	C	0.764	-0.0103	0.0013	9.40E-15	-0.0103	0.0013	9.44E-15	0.0091	0.0819
	rs2391333	**13:107166694**	***EFNB2***	**C**	**T**	**0.615**	**-0.0066**	**0.0012**	**1.90E-08**	**-0.0066**	**0.0012**	**1.87E-08**	**0.0005**	**0.0045**

Legend to table 2: Genome-wide association study for chronic BP in UK Biobank by sex. Conditional and joint analysis (COJO) was applied to identify conditionally independent SNPs. Bold results denote significant SNP-by-sex interactions.

* According to SNP-nexus software (https://www.snp-nexus.org/)

** P-values for SNP-by-sex interaction analysis (raw and adjusted for 9 tests)

Table 3 – Results of replication of genome-wide significant loci in males and females

Top SNP	Proxy SNP	Effect allele*	Other allele*	EAF (SE)**	Weight	Z-score	P-value	Direction***	I^2, %	χ2 (df)	P-value for χ^2
Males											
rs1678626	**rs1049269 (OPPERA, ELSA)**	**T (A)**	**C (G)**	**0.453 (0.010)**	**43,740**	**-2.992**	**0.0028**	**---+**	**58.2**	**7.2 (3)**	**0.066**
rs72922230	rs11665656 (GS)	A (G)	G (A)	0.585 (0.013)	43,740	-1.431	0.1523	-+++	26.4	4.0 (3)	0.254
Females											
rs367563576	rs7513205 (HUNT, GS)	T (A)	TAC (G)	0.631 (0.022)	50,092	1.693	0.0904	+--+	0.0	2.7 (3)	0.444
rs62327819 (LD = 1 with rs7682719 in UK Biobank)	**rs7682719 (all cohorts)**	**C (T)**	**T (C)**	**0.316 (0.005)**	**50,092**	**2.818**	**0.0048**	**++-+**	**47.8**	**5.7 (3)**	**0.125**
rs1039325		T	G	0.406 (0.007)	50,092	-1.594	0.1109	--+-	34.1	4.5 (3)	0.208
rs116007789		C	T	0.999 (0.001)	49,083	0.615	0.5385	-+?+	66.6	6.0 (2)	0.050
rs7834973		T	G	0.608 (0.003)	50,092	-0.105	0.9161	-+-+	0.0	1.3 (3)	0.736
rs12308843	rs2955526 (ELSA)	G (C)	C (A)	0.769 (0.006)	50,092	-1.885	0.0595	----	0.0	1.6 (3)	0.667
rs2391333		C	T	0.588 (0.022)	50,092	2.237	0.0253	++-+	25.9	4.0 (3)	0.256

Legend to table 3: * Effect and other alleles for replication SNP; corresponding alleles for proxy SNPs are given in brackets; ** Mean effect allele frequency (standard error); *** In the order of HUNT, GS, OPPERA, ELSA. Bolded are statistically significant results after correction for 9 tests ($p<0.0056$).

Table 4 – Replication of SNP-by-sex interaction signals

SNP	Cohort	Males				Females				t-statistic	df	P_{int}	P_{meta}
		EAF	Effect size	SE	P-value	EAF	Effect size	SE	P-value				
rs367563576	UK Biobank	0.609	0.0018	0.0012	0.1600	0.609	**0.0067**	**0.0012**	**7.6E-09**	-2.923	392750	**0.0035**	
	Replication cohorts												0.6477
	Generation Scotland (rs7513205)	0.588	-0.0066	0.0061	0.2768	0.5923	-0.0028	0.0055	0.6083	-0.502	17374	0.6157	
	OPPERA	0.605	0.0277	0.1362	0.8385	0.604	-0.0229	0.0951	0.8098	0.304	1126	0.7612	
	HUNT (rs7513205)	0.642	0.0012	0.0034	0.7325	0.644	0.0076	0.0036	**0.0327**	-1.277	59599	0.2017	
	ELSA	0.630	-0.0011	0.0110	0.9207	0.639	0.0096	0.0118	0.4143	-0.631	5912	0.5279	
rs7682719 (proxy for rs62327819)	UK Biobank	0.316	**0.0037**	**0.0013**	**0.0044**	0.314	**-0.0068**	**0.0012**	**1.40E-08**	6.056	392624	**1.40E-09**	
	Replication cohorts												**0.0035**
	Generation Scotland	0.314	-0.0083	0.0065	0.2044	0.315	0.0030	0.0059	0.6093	-1.299	17451	0.1940	
	OPPERA	0.318	**-0.4069**	**0.1393**	**0.0035**	0.292	-0.0357	0.1000	0.7210	-2.161	1153	**0.0309**	
	HUNT	0.321	-0.0043	0.0036	0.2327	0.319	**0.0084**	**0.0037**	**0.0221**	-2.515	59556	**0.0119**	
	ELSA	0.310	0.0072	0.0109	0.5077	0.305	**0.0327**	**0.0115**	**0.0045**	-1.441	5889	0.1495	
rs1039325	UK Biobank	0.415	-0.0016	0.0012	0.2000	0.417	**-0.0065**	**0.0011**	**8.7E-09**	2.996	392332	**0.0027**	
	Replication cohorts												0.6567
	Generation Scotland	0.418	-0.0080	0.0061	0.1909	0.416	-0.0112	0.0090	0.2151	0.235	19277	0.8142	
	OPPERA	0.417	**0.2957**	**0.1279**	**0.0208**	0.417	0.1211	0.0914	0.1852	1.109	1148	0.2678	
	HUNT	0.402	-0.0034	0.0034	0.3204	0.401	-0.0031	0.0035	0.3706	-0.094	59570	0.9254	
	ELSA	0.413	-0.0004	0.0104	0.9718	0.418	-0.0185	0.0107	0.0835	1.133	5832	0.2574	
rs2391333	UK Biobank	0.616	-0.0007	0.0013	0.5700	0.615	**-0.0066**	**0.0012**	**1.9E-08**	3.458	392630	**0.0005**	

Replication cohorts											0.1397
Generation Scotland	0.613	0.0062	0.0062	0.3140	0.618	-0.0054	0.0092	0.5550	1.062	19250	0.2884
OPPERA	0.625	-0.0215	0.1254	0.8635	0.615	-0.1108	0.0926	0.2314	0.5720	1180	0.5675
HUNT	0.571	-0.0032	0.0034	0.3513	0.573	**0.0071**	**0.0035**	**0.0447**	-2.124	59606	**0.0336**
ELSA	0.615	0.0052	0.0103	0.6131	0.629	0.0196	0.0109	0.0719	-0.853	5893	0.3939

Legend to table 4: SNP-by-sex interaction analysis was carried out by comparing effect sizes in males and females as detailed in Methods. P_{int}, p-value for SNP-by-sex interaction; P_{meta}, p-value for Fisher's combined probability test. Highlighted are nominally significant associations or interactions. Details of replication cohorts are provided in Supplementary Table 1.

Table 5 – SNP-based heritability of chronic BP by age and sex in the UK Biobank

Age group	All	Males	Females	*P-value for males vs females*
Total sample	0.068±0.002	0.067±0.003	0.079±0.003	0.005
"Younger than 65"	0.069±0.002	0.068±0.004	0.080±0.003	0.033
"65+"	0.066±0.007	0.046±0.013	0.098±0.013	0.005
P-value for "Younger than 65" vs "65+"	0.699	0.092	0.198	

Legend to table 5: SNP-based heritability was calculated using REML algorithm implemented in BOLT-LMM software; p-values of male vs female and between-age group differences are given.

Table 6 – Genetic correlation between cBP and 10 risk factors in males and females

Trait	Males		Females		P_{diff}	
	r_g	p-value	r_g	p-value	Raw	Adjusted
Intervertebral disc problems (self-reported)	0.889	5.6E-19	0.638	1.6E-14	**3.7E-07**	**3.7E-06**
Osteoarthritis	0.494	6.1E-19	0.599	3.4E-36	**0.042**	0.420
BMI	0.291	6.7E-28	0.357	9.9E-46	**0.045**	0.450
Scoliosis	0.220	0.089	0.433	5.9E-05	0.156	1.000
Smoking	0.343	2.2E-26	0.325	5.8E-27	0.650	1.000
Depression	0.395	7.3E-12	0.408	5.9E-16	0.835	1.000
Fluid intelligence score	-0.309	2.2E-17	-0.289	3.5E-19	0.643	1.000
Happiness/wellbeing	0.172	0.002	0.160	0.004	0.873	1.000
Anxiety/panic attacks	0.208	0.017	0.192	0.010	0.883	1.000
Educational attainment	-0.408	3.2E-43	-0.438	1.3E-70	0.319	1.000

Legend to Table 6: Genetic correlation between chronic BP and putative risk factors for BP, by sex and comparison thereof before (P_{diff} Raw) and after adjustment for 10 tests (P_{diff} Adjusted).