BRIEF REPORT

Sequence variation at 8q24.21 and risk of back pain [version 1; peer review: 1 approved]

Frances M.K. Williams¹, Yakov A. Tsepilov²,³, Maxim B. Freidin¹, Tatiana I. Shashkova²,⁴,⁵, Pradeep Suri⁶,⁷, Yurii S. Aulchenko²,³,⁸

¹Department of Twin Research and Genetic Epidemiology, King's College London, London, UK
²Laboratory of Theoretical and Applied Functional Genomics, Novosibirsk State University, Novosibirsk, Russian Federation
³Laboratory of Recombination and Segregation Analysis, Institute of Cytology and Genetics SD RAS, Novosibirsk, Russian Federation
⁴Department of Molecular and Biological Physics, Moscow Institute of Physics and Technology (State University), Moscow, Russian Federation
⁵Research and Training Center on Bioinformatics, A.A. Kharkevich Institute for Information Transmission Problems RAS, Moscow, Russian Federation
⁶Medical Genetics, Department of Medicine, University of Washington, Seattle, Washington, USA
⁷Division of Rehabilitation Care Services, VA Puget Sound Health Care System, Seattle, Washington, USA
⁸Polyomica, 's-Hertogenbosch, The Netherlands

Abstract
Back pain (BP) is a common condition of major social importance and poorly understood pathogenesis. Intervertebral lumbar disc degeneration in all its guises is one of the major biological risk factors for BP. Previously, we identified the locus at 8q24.21 associated with chronic BP, which has been found elsewhere associated with sciatica after surgery for lumbar disc herniation. In the current study we used co-localisation methods to identify the gene most likely to harbor the causal variant. We show that the same functional variant at the 8q24.21 locus is responsible for both lumbar disc degeneration and BP, and we also studied the effects of this locus on related phenotypes. Our results link the locus to intervertebral disc and bone mineral density, but not to anthropometric measurements, thus corroborating the epidemiological evidence. Moreover, the same functional variant at the locus is more likely to affect the expression of the nearby FAM49B gene, rather than the GSDMC gene, which was previously proposed as a causative one for BP.

Keywords
Lumbar disk degeneration, back pain, transcriptomics, genome wide association study

Open Peer Review

Invited Reviewers

1

version 1
22 May 2020

1. Lisbet Haglund, McGill University, Montreal, Canada

Any reports and responses or comments on the article can be found at the end of the article.
Corresponding authors: Frances M.K. Williams (frances.williams@kcl.ac.uk), Yurii S. Aulchenko (y.s.aulchenko@polyomica.com)

Author roles: Williams FMK: Conceptualization, Supervision, Writing – Original Draft Preparation; Tsepilov YA: Data Curation, Formal Analysis, Investigation, Writing – Review & Editing; Freidin MB: Data Curation, Formal Analysis, Writing – Review & Editing; Shashkova TI: Formal Analysis, Writing – Review & Editing; Suri P: Data Curation, Formal Analysis, Writing – Review & Editing; Aulchenko YS: Conceptualization, Supervision, Writing – Original Draft Preparation

Competing interests: YSA is a co-owner of Maatschap PolyOmica and PolyKnomics BV, private organizations, active in re-search and development in the field of computational and statistical (gen)omics. Other authors declare no competing interest.

Grant information: This work was funded by the EU FP7 PainOMICS project [grant agreement # 602736]. TIS was sup-ported by the Russian Ministry of Science and Education under the 5-100 Excellence Programme. YSA was supported by the Federal Agency of Scientific Organizations via the Institute of Cytology and Genet-ics [project 0324-2019-0040-C-01]. YAT was supported by the Federal Agency of Scientific Organiza-tions via the Institute of Cytology and Genetics [project 0324-2019-0040-C-01] and by the Russian Foundation for Basic Research [project 19-015-00151]. The Genotype-Tissue Expression (GTEx) Pro-ject was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from the GTEx Portal on February 2017. The CEDAR was supported by grants to Michel Georges from WELBIO (CAUSIBD), BELSPO (BeMGI), and Horizon 2020 (SYS-CID).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2020 Williams FMK et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Williams FMK, Tsepilov YA, Freidin MB et al. Sequence variation at 8q24.21 and risk of back pain [version 1; peer review: 1 approved] F1000Research 2020, 9:424 https://doi.org/10.12688/f1000research.22725.1

Introduction

Back pain (BP) is a common debilitating condition with a lifetime prevalence of 40% and a major socioeconomic impact. According to the Global Burden of Disease 2016 study, it leads the list of disabling conditions in many parts of the world. The greatest biological risk for episodes of severe BP is thought to be lumbar disc degeneration (LDD). Rather than a discrete disease, LDD is now considered a continuous lifelong aging process that usually starts in the third decade of life but may also be seen in children. It is highly heritable, as well as being influenced by smoking, age and body mass index. The overall burden of lumbar disc degeneration on magnetic resonance imaging is a strong predictor of BP episodes. There is a clear genetic predisposition to both BP and LDD with estimates of heritability in the range of 30%–70%. Also, these traits have been shown to share genetic risk factors. However, little is known about the genetic architecture of BP and only three associated genetic loci have been identified.

This work was focused on the locus at chromosome 8 (8q24.21), which was identified in our previous BP studies. The same locus has been reported associated with having lumbar disc herniation (LDH) in Icelanders undergoing surgery for sciatica. Bjornsdottir et al. additionally reported an association of this region to height, but concluded that the region does not influence LDH through height. They noted that the most highly associated SNP, rs6651255, was linked to expression of the nearby gasdermin (GSDMC) gene in GTEx data. Further, the immune-related function of gasdermins implied that the inflammatory and immunological consequences of disc herniation may be linked to sciatica symptom severity.

However, closely located SNPs associated with different traits do not always reflect pleiotropy. Furthermore, in the case of eQTL studies, causal hypotheses generated through single-variant look up are highly likely to be false positives. Consequently, a variety of co-localization methods have been introduced to distinguish the two common scenarios: pleiotropy (the same functional genetic variant is associated with both traits) and linkage (two different functional genetic variants, each associated with a different trait, that are in linkage disequilibrium).

The aim of this study was to (1) verify whether the same functional variant at 8q24.21 is responsible for BP and LDH and (2) to study the effects of this variation on related phenotypes. As a result, we offer an alternative explanation to the assertion by Bjornsdottir et al. regarding the role of immunity and inflammation in generating symptoms of sciatica and BP due to variation in expression at the locus in between genes CCDC26 and GSDMC.

Methods

GWAS data

We used the BP GWAS performed in all genetically confirmed white British UK Biobank participants (N= 453,862) as target data for future analyses. We used the summary level-data of LDH GWAS from the original paper of deCODE group. They do not provide full GWAS scan, only the statistics for the significant SNPs from the 8q24.21 region. We used information about 40 SNPs from this locus for which information about effect size, effective and reference alleles and p-value was available (Supplementary Table 1.3 of the work).

We derived summary statistic of 2,419 complex traits provided by the Neale Lab and 34 traits from the Gene ATLAS database. Summary statistics for gene expression levels were obtained from Westra Blood eQTL (peripheral blood, http://cnsgenomics.com/software/smr/#eQTLsummarydata, file: westra_eqtl_data_hg19.zip (hg19)) and the GTEx database (14 tissues, https://www.gtexportal.org/home/datasets, file: GTEx_Analysis_V6_all-snp-gene-associations.tar).

SMR-HEIDI analysis

Summary-level mendelian randomization and heterogeneity in dependent instruments (SMR-HEIDI) approach was used to test for potential pleiotropic effects of the 8q24.21 region on back pain (BP) and LDH and other complex traits including gene expression levels in certain tissues. SMR analysis provides evidence for pleiotropy, but is unable to define whether both traits are affected by the same underlying causal polymorphism. The latter is specified by a HEIDI test that distinguishes pleiotropy from linkage. Zhang F. and colleagues developed the software SMR tool implementing the SMR-HEIDI methods.

SMR tool is suitable to analyze pleiotropic association using summary-level data from GWAS and eQTL studies. We conducted two groups of analysis: target trait vs eQTL and target trait vs other complex traits. Summary-level data from BP GWAS was used as target (default option) in both cases. For eQTL analysis we used Westra Blood eQTL data (converted to suitable BSED format by developers) and GTEx v7, converted using --make-besd option (smr --eqtl-flist Gtex_tissue_1.txt --make-besd --out Gtex_tissue_1, where Gtex_tissue_1.txt file correspond to summary-level data of selected tissue). In order to apply SMR tool for complex traits we selected the region +250kb around the target SNP from each complex trait GWAS. After that, we add new columns Probe_ID with trait_name to corresponded SNP subset. Finally, we merge these SNP subsets into one file and converted it in BESD format using --make-besd option.

In both cases we set the minor allele frequency threshold as 0.03 and included in the HEIDI test SNPs inside +250kb window centered around the target SNP; other parameters used default settings. The LD matrix for the HEIDI test was estimated using PLINK 1.9 software and individual data of white UK Biobank participants, the subset of individuals used to calculate BP GWAS (N=10,000).

Nominal P for SMR test was set at 6e-4 (0.05/89, where 89 is the total number of probes for the gene expression analysis) and 2e-5 (0.05/2,453, where 2,453 is the total number of complex traits used in analysis). For HEIDI analysis we used...
Results and discussion

We had previously performed the largest-to-date GWAS for chronic BP using the UK Biobank and CHARGE consortium cohorts, with total of 119384 cases and 334478 controls\(^1\). Cases were defined as those having BP of duration >3-6 months; controls included those without back pain or with back pain of <3-6 months. We found variants at 8q41.21 locus that showed a genome-wide significant association (OR=1.05, \(p = 4.4 \times 10^{-13}\), after genomic control correction)\(^2\). The location chr8:130,717,716 is tagged by intergenic variant rs10956487 which is in perfect LD with rs6651255 (\(r^2=1\) according to 1000 Genomes EUR samples), the lead SNP in the Icelandic study of LDH by Bjornsdottir et al.\(^3\). The strongest association signal for BP is situated near GSDMC gene; yet another protein-coding gene, FAM49B, is located within a 500 kbp window. The summary of the previously reported results and our findings for locus 8q24.21 are shown in Figure 1.

We used the SMR-HEIDI approach\(^4\) to test for potential pleiotropic effects of the 8q24.21 region (rs6651255) on BP and other complex traits including gene expression levels in relevant tissues. SMR analysis provides evidence for pleiotropy, but is unable to determine whether both traits are influenced by the same underlying causal polymorphism. The latter question was addressed by the HEIDI test which distinguishes pleiotropy from linkage. For SMR-HEIDI testing we used as the target trait the BP GWAS previously performed in all genetically confirmed white British UK Biobank participants\(^5\).

First of all, we checked relationships between BP and LDH. We found strong evidence that co-association of these traits to 8q24.21 is due to the pleiotropic action of the same functional variant (\(P_{\text{SMR}}=2.8 \times 10^{-7}\) and \(P_{\text{HEIDI}} = 0.87\); extended data\(^6\)). We further applied SMR-HEIDI to test for pleiotropic effects of the region in the collection of 2,453 traits available from the UK biobank study and GeneAtlas. We found that the BP risk allele of rs6651255 was negatively associated with heel bone mineral density t-score and positively associated with self-reported prolapsed or slipped disc (both FDR\(_{\text{SMR}}<0.01\), underlying data\(^7\) and extended data\(^8\)) and also strongly associated with standing height (underlying data\(^9\) and extended data\(^10\)) and several measures of fat-free mass (all FDR\(_{\text{SMR}}<0.001\)). However, HEIDI tests showed strong evidence against pleiotropic effects of the same functional variant in this region for BP and these anthropometric traits (all \(P_{\text{HEIDI}}<0.001\)), while associations with heel bone mineral density t-score and self-reported prolapsed or slipped disc were likely the result of pleiotropy, controlled by the same functional variant (all \(P_{\text{HEIDI}}>0.05\)). This suggests that genetic variants influencing structural spine features – the vertebral body and intervertebral disc – provide the link to BP symptoms, supporting previous work in twins highlighting shared genetic influence between spine degeneration and BP\(^11\).

The unanticipated results were obtained in the study of expression data. We have previously\(^12\) demonstrated that variants in LD with the lead SNP at 8q24.1 were located in potential regulatory regions in chondrocytes and osteoblasts (Roadmap Epigenomics Consortium\(^13\)). In the current study, we tested expression of genes in the 8q24.21 region in all available tissues considered potentially relevant to development of pain (underlying data\(^14\)), including musculoskeletal, brain and nerve tissue (from GTEx,\(^15\)). Given the suggestion by Bjornsottir et al. that the locus may act through regulation of immune function\(^16\), we also investigated expression in whole blood using the blood eQTL database\(^17\) and five immune cell subsets from CEDAR\(^18\). Similar to the Bjornsottir et al. study\(^19\), we observed that rs6651255 is co-associated with expression

---

**Figure 1. Pleiotropic effects of locus 8q24.21 on BP and related traits.** The BP-related traits and genes analyzed in our work are blue. The locus 8q24.21 is connected to associated traits (\(p\)-value < \(5 \times 10^{-8}\)) by green arrows. Possible relations between traits, as inferred from previous studies, are marked by grey arrows. Pairwise pleiotropic effects tested by SMR-HEIDI are denoted by a red cross if the test failed, and by a blue check if the test passed. A new detected pleiotropic effect on expression of FAM49B and BP based on SMR-HEIDI results is denoted by the blue arrow. LDH, lumbar disk herniation; SRDP, self-reported disk problems; HBMD, heel bone mineral density t-score.
of GSDMC in skeletal muscle (FDR_{SMR}=0.01, P_{SMR}=0.0001, underlying data)\(^9\), with the BP-associated C allele of rs6651255 being associated with increased expression of the gene. However, HEIDI results show this overlap to be unlikely due to pleiotropic action of the same functional variant (P_{HEIDI} = 3\times10^{-7}, underlying data\(^9\) and extended data\(^9\)). Our findings suggest that GSDMC is not the gene through which the functional variant acts to influence BP. The next strongest co-association was with increased expression of FAM49B in brain anterior cingulate cortex (P_{SMR}=0.004, underlying data\(^9\) and extended data\(^9\)) and decreased expression in cytotoxic T (CD8+) cells (P_{SMR}=0.02, underlying data\(^9\) and extended data\(^9\)), although these associations were only nominally significant (p<0.05), yet not statistically significant after multiple testing correction (FDR_{SMR}=0.18 and 0.28, respectively). All corresponding P_{HEIDI} were > 0.05, indicating these co-associations are likely due to pleiotropy. The product of FAM49B was found to be expressed in multiple tissues including bone, lymph node and spleen; according to Proteomics DB\(^13\), it has highest expression in lymphocytes, and was implicated as a key negative regulator of actin dynamics and T-cell activation\(^\text{19}\). The precise mechanism of which gene is involved in the development of BP is difficult to infer due to lack of eQTL data available for relevant tissues, such as intervertebral disc.

**Conclusion**

We hypothesize that the locus 8q24.21 exhibits its influence on BP through gene expression via disk and bone remodeling, and not through influence on anthropometrics or pain processing. Our findings suggest that FAM49B and not GSDMC is the most likely mechanism through which the functional variant leads to BP.

**Data availability**

**Underlying data**

Zenodo: Table S1. List of relevant studies of gene expression with available eQTL data. http://doi.org/10.5281/zenodo.3715005\(^33\)

This project contains the following underlying data:
- List of relevant studies of expression data.csv (List of relevant studies of gene expression with available eQTL data)

Zenodo: Tables S2-S3. SMR-HEIDI results for 8q24.21 locus between back pain and other phenotypes. [Data set]. http://doi.org/10.5281/zenodo.3715027\(^30\)

This project contains the following underlying data:
- SMR-HEIDI results ST2.csv (SMR-HEIDI results for 8q24.21 locus between back pain and other complex traits)
- SMR-HEIDI results ST3.csv (SMR-HEIDI results for 8q24.21 locus between back pain and expression of genes)

**Extended data**


This project contains the following extended data:
- Figure S1.png (SMR-HEIDI analysis results for rs6651255 between BP and LDH)
- Figure S2.png (SMR-HEIDI analysis results for rs6651255 between BP and GSDMC expression in skeletal muscle (GTEx v6))
- Figure S3.png (SMR-HEIDI analysis results for rs6651255 between BP and FAM49B expression in Brain anterior cingulate cortex BA24 (GTEx v6))
- Figure S4.png (SMR-HEIDI analysis results for rs6651255 between BP and FAM49B expression in CD8 cell line (CEDAR))
- Figure S5.png (SMR-HEIDI analysis results for rs6651255 between BP and heel bone mineral) density (UKBB).
- Figure S6.png (SMR-HEIDI analysis results for rs6651255 between BP and disc problem phenotype (UKBB))
- Figure S7.png (SMR-HEIDI analysis results for rs6651255 between BP and height (UKBB))

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

**Acknowledgements**

We thank Denis Gorev, Anna Torgasheva and Eugeny Pakhomov for help with summary level data collection, processing, and analyses. We thank Michel Georges for permission to use the CEDAR results for this work, and for discussion related to CEDAR. We also thank Natalia Aulchenko for help with the manuscript preparation.

**References**

5. Sambrook PN, MacGregor AJ, Spector TD: Genetic influences on cervical and lumbar disc degeneration: a magnetic resonance imaging study in twins.
Open Peer Review

Current Peer Review Status: ✔

Version 1

Reviewer Report 23 September 2020

https://doi.org/10.5256/f1000research.25092.r70799

© 2020 Haglund L. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Lisbet Haglund
Orthopaedic Research Laboratory, McGill University, Montreal, QC, Canada

The study "Sequence variation at 8q24.21 and risk of back pain" is well written and is addressing an important clinical problem. The design and data analysis are well performed.

It would be interesting to know more about the function of FAM49B and GDSMC and how a dysregulation could potentially lead to back pain. The fact that it is involved in mitochondrial dysregulation and reactive oxygen species may have implications for IVD degeneration and low back pain.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.
Reviewer Expertise: Molecular mechanisms of back pain

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

• Your article is published within days, with no editorial bias
• You can publish traditional articles, null/negative results, case reports, data notes and more
• The peer review process is transparent and collaborative
• Your article is indexed in PubMed after passing peer review
• Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com