Analysis of genetically independent phenotypes identifies shared genetic factors associated with chronic musculoskeletal pain at different anatomic sites

Yakov A. Tsepilov[1,2*], Maxim B. Freidin[3*], Alexandra S. Shadrina[1,2], Sodbo Z. Sharapov[1,2], Elizaveta E. Elgaeva[2], Jan van Zundert[4,5], Lennart C. Karssen[6], Pradeep Suri[7,8], Frances M.K. Williams[3], Yurii S. Aulchenko[1,2,6,9^]

[1] Laboratory of Theoretical and Applied Functional Genomics, Novosibirsk State University, Novosibirsk, Russia
[2] Laboratory of Recombination and Segregation Analysis, Institute of Cytology and Genetics, Novosibirsk, Russia
[3] Department of Twin Research and Genetic Epidemiology, School of Life Course Sciences, King's College London, London, UK
[4] Department of Anesthesiology and Pain Medicine, Maastricht University Medical Centre, Maastricht, The Netherlands
[5] Department of Anesthesiology and Multidisciplinary Paincentre, ZOL Genk/Lanaken, Belgium
[6] PolyOmica, 's-Hertogenbosch, the Netherlands
[7] Division of Rehabilitation Care Services, VA Puget Sound Health Care System, Seattle, USA
[8] Seattle Epidemiologic Research and Information Center (ERIC), Department of Veterans Affairs Office of Research and Development, Seattle, USA
[9] Department of Complex Genetics, Care and Public Health Research Institute, Maastricht University, The Netherlands

Correspondence^: y.s.aulchenko@polyomica.com

Abstract

Chronic musculoskeletal pain has a negative impact on all aspects of human life. Genetic studies of pain are complicated by the high complexity and heterogeneity of pain phenotypes. In this research, we aimed to reduce phenotype heterogeneity and reveal genes and pathways shared by chronic musculoskeletal pain at four locations: back, neck/shoulder, hip, and knee. Our study was based on the results of genome-wide association studies performed using UK Biobank data with a total sample size of 456,000 individuals. We applied principal component analysis based on the matrix of genetic covariances between the studied pain traits and constructed four genetically independent phenotypes (GIPs). The leading GIP (GIP1) explains the largest proportion of the genetic variance of and covariance between the analyzed phenotypes (78.4%), and the later GIPs (GIP2-4) explain progressively less. We identified and replicated five loci associated with GIP1 and one locus associated with GIP2. The genes decisively prioritized for the GIP1-associated loci were *SLC39A8*, *ECM1*, and *FOXP2*. For the remaining two GIP1-associated loci, we proposed several candidates (*AMIGO3*, *BSN*, *RBM6*, *FAM212A*, *RNF123*, *UBA7* and *MIR7114*, *NSMF*, *NOXA1*, *GRIN1*), but were unable to prioritize any of them convincingly. The most likely causal gene in the locus associated with GIP2 was *GDF5*. For GIP1, gene set/tissue/cell type enrichment analyses identified multiple terms related to the nervous system. Genetic correlations analysis revealed a genetic overlap between GIP1 and osteoarthritis as well as a set of anthropometric (such as overweight and waist circumference), sociodemographic (such as age of first birth and college completion) and psychiatric/personality (such as depressive symptoms and neuroticism) traits. We suggest that GIP1 represents a biopsychological component of chronic musculoskeletal pain, related to physiological and psychological aspects and likely reflecting pain perception and processing.

Introduction

Chronic pain is one of the most prevalent human health problems, affecting on average 20-30% of adults [1-3], and it is one of the most challenging conditions for clinical management [4]. Often chronic pain is present without a clear pathophysiological cause such as tissue damage and cannot be attributed to a known disorder. Chronic musculoskeletal pain is the most prevalent type of chronic pain in older adults [5]. Prevalence estimates vary widely depending on the studied population and the definition used to define these conditions [6]. For instance, in a Swedish study, the prevalence was 23.9% for chronic regional musculoskeletal pain and 11.4% for chronic widespread pain [7]. In Japan, the prevalence of chronic musculoskeletal pain was found to be 15.4% in the general population and reached 18.6% among individuals aged 40-49 [8]. The most prevalent self-reported chronic musculoskeletal pain conditions are low back, neck, and shoulder pain [7, 8]. According to the Global Burden of Disease Study 2015, low back pain and neck pain were the leading causes of global years lived with disability in 1990-2015 [9].

Precise biological mechanisms underlying chronic pain are yet to be elucidated [10, 11]. There is good evidence that chronic pain disorders are complex heritable traits [12, 13]. Exploring the genetic underpinning of chronic pain phenotypes can expand basic knowledge on their etiology and biological mechanisms, improve diagnostics, and facilitate the development of effective therapies via the identification of therapeutic targets.

Genetic association studies have suggested a number of genes associated with chronic musculoskeletal pain phenotypes [14-16]. These studies were predominantly hypothesis-driven candidate-gene studies, which often had small samples sizes, and with some leading to conflicting results as has been borne out in other traits [17]. Compared to candidate-gene studies, genome-wide association studies (GWAS) offer an agnostic data-driven approach that allows identification of susceptibility genes without a prior mechanistic hypothesis. So far, only a few GWAS for forms of chronic musculoskeletal pain have been published, including chronic widespread pain [18], fibromyalgia [19], chronic back pain [20], sciatica [21], and painful temporomandibular disorder [22]. Thus, the genetic architecture of chronic musculoskeletal pain is far from being defined.

Research in chronic pain genetics faces a number of obstacles. According to the biopsychosocial model of pain, chronic pain results from a complex and dynamic interaction among biological, psychologic and social factors [23]. The extreme complexity and heterogeneity of chronic pain phenotypes complicates identification of novel loci and makes it difficult to distinguish whether identified variants affect the risk of the primary pain-causing pathology (if any) or influence the development and maintenance of the chronic pain state itself. Both the primary underlying condition and its treatment, and the treatment of chronic pain, may confound studies. A study exemplifying these challenges is our recent GWAS of chronic back pain [18]. Despite the large sample size of nearly 158,000 individuals in the discovery sample and 284,000 subjects in the replication sample, we were able to detect and replicate only one locus. Thus, new strategies are required to improve understanding of the genetic influences in chronic pain conditions.

One possible solution to the problem of clinical heterogeneity is to study endophenotypes and subgroups of patients having different characteristics [15]. A complementary approach to reducing heterogeneity is to elucidate the common pathways shared by distinct pain phenotypes. Indeed, different chronic pain conditions may have common biological pathways such as those related to pain perception and processing. Several studies have provided evidence for shared genetic factors between

conditions manifesting chronic pain [24] as well as pain at different anatomical sites [25, 26]. However, to the best of our knowledge, no study yet published has explicitly identified these genetic factors.

Here, we investigated the genetic factors underlying chronic musculoskeletal pain reported at four locations (back, neck/shoulder, hip, and knee). These anatomical sites are commonly affected by osteoarthritis (OA). Pain is the predominant symptom of OA, but its intensity may be poorly correlated with OA severity based on pathological changes revealed by radiographs. Current evidence suggests that not only structural lesions, but also neuronal pathways and alterations of pain processing contribute to maintaining pain in OA patients [27]. We assumed that studying pain at multiple sites potentially linked through OA can unravel shared musculoskeletal pathways and, more importantly, provide deeper understanding of general chronic pain mechanisms. We used a novel approach to explore the genetic background of pain traits by analyzing genetically independent phenotypes (GIPs). Using data from UK Biobank [28] we identified and replicated specific loci associated with these GIPs, followed by *in silico* functional analysis, including a search for pleiotropic effects of functional variants, prioritization of likely causal genes, analysis of gene set and tissue enrichment, and estimation of genetic correlations with other complex traits.

Methods

Study sample and phenotype definition

The study sample comprised UK Biobank participants [28]. Sociodemographic, physical, lifestyle, and health-related characteristics of this cohort have been reported elsewhere [29]. In brief, individuals enrolled in the UK Biobank study were aged 40–69 years; were less likely to be obese, to smoke, to drink alcohol; had fewer self-reported health conditions as compared to the general population. All study participants provided written informed consent, and the study was approved by the North West Multi-Centre for Research Ethics Committee (11/NW/0382).

This particular study was approved by the UK Biobank research team under project #18219. Cases and controls were defined based on questionnaire responses. First, participants responded to "Pain type(s) experienced in the last months" followed by questions inquiring if the specific pain had been present for more than 3 months. Those who reported back, neck or shoulder, hip, or knee pain lasting more than 3 months were considered chronic back, neck/shoulder, hip, and knee pain cases, respectively. Participants reporting no such pain lasting longer than 3 months were considered controls (regardless of whether they had another regional chronic pain, such as abdominal pain, or not). Individuals who preferred not to answer or reported more than 3 months of pain all over the body were excluded from the study (since these subjects met criteria for chronic widespread pain and were thought likely to have an underlying generalized propensity to pain). Further details are given in Supplementary Methods.

Overall, 456,580 individuals with imputed genotype data and phenotype data were included in the present study. Of these, 265,000 participants of European ancestry (defined by SNP-based principal component analysis) were randomly selected to provide the GWAS discovery cohort. The decision to include only Europeans was based solely on the highest representation of these individuals among the UK Biobank participants. The replication cohort (N = 191,580) comprised individuals of African (N = 7,541) and South Asian ancestry (Indian, Pakistani, and Bangladeshi; N = 9,208) as well as the remaining European ancestry participants (N = 174,831). Descriptive characteristics of the groups is provided in Table S1.

Genotyping and imputation

Genotyping and imputation data were obtained from the UK Biobank March 2018 data release. Genotyping was conducted using the Affymetrix UK BiLEVE and Affymetrix UK Biobank Axiom arrays. Imputation was performed with the IMPUTE4 program (https://jmarchini.org/impute-4/) [30] using the Haplotype Reference Consortium (HRC) [31] and merged UK10K and 1000 Genomes phase 3 reference panels. Details on DNA extraction and quantification [32] as well as on the centralized analysis of the genetic data, genotype quality, properties of population structure and relatedness of the genetic data, and efficient phasing and genotype imputation have been reported previously [33].

Genome-wide association study

GWAS were carried out using BOLT-LMM v.2.3.2 software [34]. Linear mixed-effects models were fitted to test for additive effects of the SNPs (genotype dosage) on pain phenotypes adjusting for age, sex, genotyping platform batch and the first ten genetic principal components. The following filters were applied: minor allele frequency > 0.0002 for Europeans and > 0.005 for Africans and Asians; imputation quality score > 0.7; genotyping and individual call rates > 0.98. Only biallelic autosomal SNPs and indels were analyzed. BOLT-LMM software requires LD score data for the

analysis. For Europeans, we used LD scores distributed as part of BOLT-LMM package. For Africans and South Asians, we carried out LD score estimation using LD score software [35] and data from 500 individuals randomly selected from each ethnic group. The results of GWAS were corrected for residual inflation using the LD score regression intercept [35].

Locus definition

Associated loci were defined as regions within ±250 kb around the lead SNP. Only the most significant SNP per locus was reported.

Genetically independent phenotypes

To elucidate genetic components explaining four chronic musculoskeletal pain phenotypes (chronic back, neck/shoulder, hip, and knee pain), we used a modified principal component analysis (PCA) technique that combines multiple correlated variables into a set of uncorrelated principal components (PCs). PCs are linear combinations of variables constructed such that the first PC explains the maximum proportion of the total variance of the set of traits, the second PC accounts for the largest proportion of the remaining variance, and so on. In conventional PCA of a set of traits, vectors of coefficients of orthogonal transformation are equal to the eigenvectors of the matrix of phenotypic covariance. In the present study, we used the matrix of genetic covariances between the traits of interest to decompose them into genetically independent components, that we called *genetically independent phenotypes* (GIPs). GIPs are not correlated genetically and the first GIP (GIP1) explains most of the genetic variance of -and covariance between- four musculoskeletal pain phenotypes. Technical details of our approach are described in Supplementary Methods. It should be noted that principal component analysis has already been used for studying genetic background of complex traits [36, 37], although it was applied to obtain phenotypically independent phenotypes, not GIPs. In both cases heritability of obtained principal components was not less than heritability of original traits.

The matrix of genetic covariances (estimated by LD Score regression [38]) and orthogonal transformation coefficients were obtained using the discovery cohort of European ancestry individuals. The 95% confidence intervals of these coefficients were estimated via the Monte Carlo sampling. For each resulting "discovery" GIP, GWAS results were calculated as described in Supplementary Methods.

GIPs for replication datasets were constructed using the orthogonal transformation coefficients obtained at the discovery step. GWAS results for each "replication" GIP were combined by a meta-analysis. Furthermore, GWAS for GIPs for European ancestry replication cohort (N = 439,831 in total) were meta-analyzed with GWAS for discovery GIPs, and the results were used for subsequent post-GWAS *in silico* analyses. Meta-analyses were conducted using the inverse-variance-weighted approach (fixed-effects model) with METAL software [39].

Additionally, we used the same methodology to obtain the first GIP for the extended set of pain traits available in the UK Biobank: chronic back, neck/shoulder, hip, knee, stomach/abdominal pain and headache. Facial pain, which is also present in the UK Biobank database, was not included in the analysis due to low prevalence (0.9% in European ancestry dataset, 4016 cases and 435815 controls) and statistically insignificant SNP-based heritability, that makes the genetic correlation analysis impossible. GIP1 for six pain phenotypes was constructed for the discovery and European ancestry replication cohort, and GWAS results for these cohorts were meta-analyzed. GIP1 for six pain phenotypes was included in the analysis of genetic correlation with GIP1 for four pain phenotypes.

Conditional analysis

Conditional and joint (COJO) analysis was carried out as previously described [40]. Calculations were performed using the GCTA software [41]. Linkage disequilibrium (LD) matrix was computed with PLINK 1.9 software (https://www.cog-genomics.org/plink2) using 100,000 individuals randomly selected from the discovery cohort. We claimed one independent signal per locus if no polymorphism other than the lead SNP passed the significance threshold of P = 5e-08. Regional association plots were generated using LocusZoom (http://locuszoom.org/) for regions within ±250 kb from the lead SNP.

Prediction of SNP effects

We analyzed the functional effects of a set of SNPs and indels in high LD ($r^2 > 0.8$) with replicated variants. LD was calculated using PLINK 1.9 [42] (*--show-tags* option) and genotype data for 503 European ancestry individuals (1000 Genomes phase 3 version 5 data). Additionally, we selected SNPs within replicated regions (± 250 kb from lead SNPs) associated with GIPs at $P \leq T$, where log10(T) = log10(P_{min}) + 1, and P_{min} is a P-value for the strongest association per locus. These SNPs were added in the analysis since genotype data for the UK Biobank samples were imputed using the Haplotype Reference Consortium (HRC) panel, and some HRC SNPs could possibly be missed in the 1000 Genomes panel. All selected variants were annotated using the Ensembl Variant Effect Predictor (VEP) [43] as well as FATHMM-XF [44] and FATHMM-INDEL [45]. In the latter two methods, predictions of variant effects were made according to scores ranging from 0 to 1, with scores above 0.5 predicted to be deleterious while those below 0.5 predicted to be neutral or benign.

DEPICT and FUMA analyses

Gene set and tissue/cell type enrichment analyses and gene prioritization were performed using the Data-driven Expression Prioritized Integration for Complex Traits (DEPICT) tool [46]. We employed the DEPICT software version 1.1, release 194 with default parameters (https://data.broadinstitute.org/mpg/depict/). Tests were conducted for both genome-wide significant SNPs (P < 5e-08) and for SNPs associated with GIPs at P < 1e-05. The MHC region was omitted. The significance threshold for DEPICT analyses was set at FDR < 0.05.

Gene set and tissue enrichment analyses were also performed using the FUMA (Functional Mapping and Annotation of Genome-Wide Association Studies) platform [47] (GENE2FUNC function, with default parameters) based on the MAGMA method [48] and the MsigDB c5 database [49]. The significance threshold for FUMA analyses was set at Bonferroni-corrected P-value < 0.05.

SMR/HEIDI analysis

Summary data-based Mendelian Randomization (SMR) analysis followed by the Heterogeneity in Dependent Instruments (HEIDI) test [50] was used to study potential pleiotropic effects of identified loci on GIPs, human complex traits, and gene expression levels in different tissues. SMR analysis provides evidence for pleiotropy (the same locus is associated with two or more traits). It cannot define whether traits in a pair are affected by the same underlying causal polymorphism, and this is specified by a HEIDI test, which distinguishes pleiotropy from linkage disequilibrium. It should be noted that SMR/HEIDI analysis does not identify which allele is causal and cannot distinguish pleiotropy from causation.

Summary statistics for gene expression levels was obtained from Westra Blood eQTL (peripheral blood, http://cnsgenomics.com/software/smr/#eQTLsummarydata) [51] and the GTEx version 7

database (48 tissues, https://gtexportal.org) [52]. Summary statistics for other complex traits were derived from the GWAS-MAP database [53] developed by our group. The GWAS-MAP platform integrates a database of summary-level GWAS results for 673 complex traits from the UK Biobank, 123 metabolomics traits, 1,206 circulating proteins, 41 cytokines and growth factors, 190 plasma protein and IgG N-glycosylation traits, inflammatory bowel disease (including Crohn's disease), and 8 traits related to coronary artery disease, myocardial infarction, and factors associated with these conditions. Summary statistics for the UK Biobank traits was provided by the Neale Lab (http://www.nealelab.is/) and the Gene ATLAS (http://geneatlas.roslin.ed.ac.uk/) [54]. In this study, we added to the GWAS-MAP database results from 18 GWAS of chronic musculoskeletal pain-related traits obtained in the present study (GWAS in the discovery dataset and the results from European ancestry meta-analysis for chronic back, neck/shoulder, knee, hip pain; GWAS in the discovery dataset and the results from European ancestry meta-analysis for GIPs constructed for these four phenotypes; European ancestry meta-analysis of GWAS for chronic stomach/abdominal pain and chronic headache). Additionally, we added the results of GWAS of osteoarthritis from the Michigan PheWeb database (http://pheweb.sph.umich.edu/SAIGE-UKB/pheno/740). This OA GWAS was performed using the UK Biobank data by the Scalable and Accurate Implementation of GEneralized mixed model (SAIGE) method [55].

Description of all 2,262 traits is provided in Table S2. The GWAS-MAP platform contains embedded software for our implementation of SMR/HEIDI analysis [50], LD Score regression [35], and 2-sample Mendelian randomization analysis (MR-Base package [56]). Further details are given in Supplementary Methods.

In gene expression analysis, the significance threshold for SMR was set at P = 3.24e-06 (0.05/15,445, where 15,445 is the total number of tests corresponding to all analyzed SNPs, expression probes, and tissues). In complex traits analysis, the significance threshold for SMR was set at P = 3.71e-06 (0.05/(6*2,244), where 6 is the number of loci, and 2,244 is the number of non-pain traits). The significance threshold for HEIDI tests in both analyses was set at P = 0.01 (P < 0.01 corresponds to the rejection of pleiotropy hypothesis). Details of data processing are given in Supplementary Methods.

Genetic correlations and heritability

SNP-captured heritability (h^2) and genetic correlations between GIPs and human complex traits were estimated using LD Score regression [38]. In total, we examined 209 non-UK Biobank traits available in the LD hub database (http://ldsc.broadinstitute.org/ldhub/). We removed duplicates and included only the most recent study for each trait (as indicated by the largest PubMed ID number). Since osteoarthritis was not present in the LD hub database, we used summary statistics for this trait obtained from the Michigan PheWeb database (http://pheweb.sph.umich.edu/SAIGE-UKB/pheno/740). The statistical significance threshold was set at 5.95e-05 (0.05/(210*4), where 210 is the number of traits and 4 is the number of GIPs).

Genetic correlations between GIPs and LD hub traits were calculated using the LD hub web interface. Genetic correlations between GIPs, osteoarthritis and chronic pain traits were calculated using the GWAS-MAP platform.

For 39 LD hub traits showing statistically significant correlations with GIP1 as well as for osteoarthritis, four chronic pain traits and four GIPs, matrices of genetic correlation were generated. Clustering and visualization were performed by the “corrplot” package for the R language (basic

"hclust" function). For clustering, we estimated squared Euclidean distances by subtracting absolute values of genetic correlation from 1 and used the Ward's clustering method.

Additionally, we estimated the genetic correlation between GIP1 for four analyzed chronic pain traits and the first GIP constructed using the same methodology for six chronic pain traits (back, neck/shoulder, knee, hip, stomach/abdominal pain, and headache) using the GWAS-MAP platform.

Results

Overview of the study design

Our study was designed to investigate the genetic components underlying chronic musculoskeletal pain at four locations: back, neck/shoulder, hip, and knee (Figure 1). Individuals who reported more than 3 months of pain all over the body were not included in the present study. All studied pain phenotypes were found to have statistically significant SNP-based heritability (2-4% on the observed scale, Table S2) and to be genetically correlated with each other (Figure 2c).

Using the matrix of genetic covariances between the studied chronic pain traits as estimated from the discovery cohort, we constructed four genetically independent pain phenotypes (GIP1 to GIP4) in the discovery and replication cohorts. GIP1, explaining most of the genetic variance and covariance between the studied pain traits, was of foremost interest in the present research. Nevertheless, we also considered the remaining GIPs, which are genetically independent contributors to chronic pain at the four studied sites.

For each GIP, GWAS results were obtained. Associations reaching the genome-wide significance threshold in the discovery cohort were considered replicated if the Bonferroni-corrected significance threshold was reached in the meta-analysis of replication cohorts. For replicated loci, gene prioritization was performed using several approaches. We conducted a functional bioinformatics analysis searching for relevant gene sets and tissues (DEPICT/FUMA analyses), analyzed pleiotropic effects (SMR/HEIDI analysis) and investigated genetic correlations with other complex traits. *In silico* functional analysis was performed using the cohort of European ancestry individuals since this subsample was the largest.

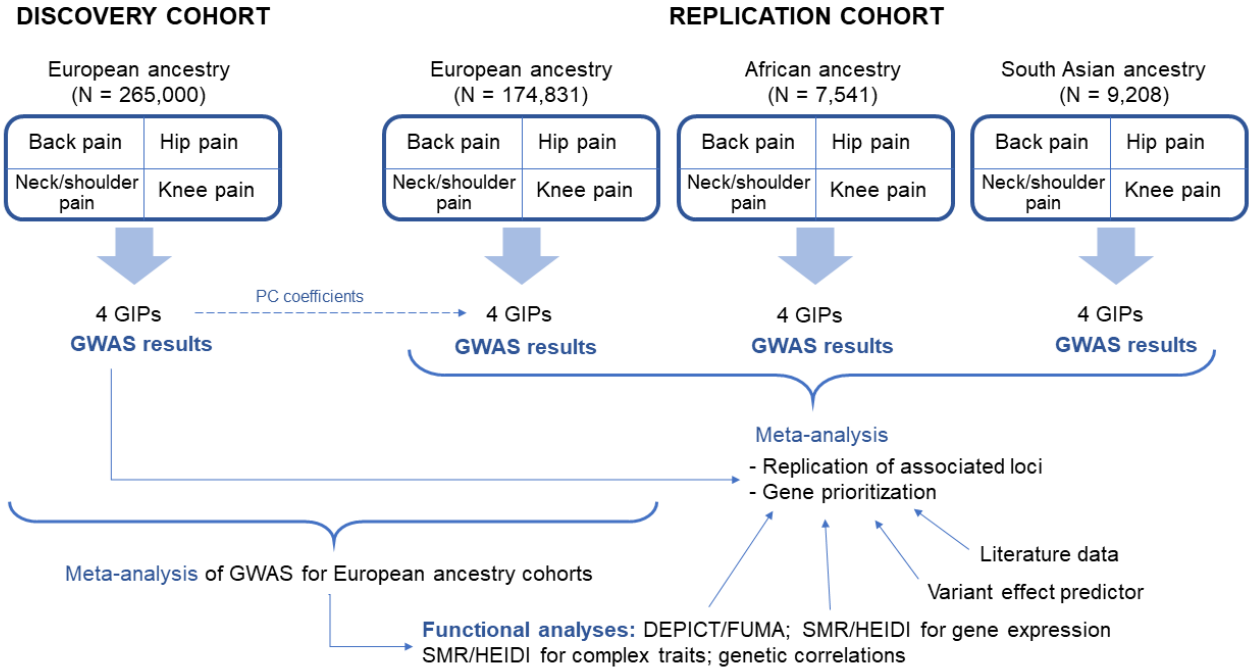

Figure 1. Overview of the study. European ancestry individuals provided the matrix of genetic covariances and orthogonal transformation coefficients. The four chronic musculoskeletal pain phenotypes were decomposed into four GIPs. Orthogonal transformation coefficients were further used to construct GIPs in the replication cohorts of European, African, and South Asian ancestry individuals. For each GIP, GWAS results were obtained. Replication of associations and *in silico* functional analyses were based on the meta-analyses of GWAS for the replication cohorts and European ancestry cohorts, respectively. For replicated loci, the most likely causal genes were prioritized.
DEPICT, Data-driven Expression Prioritized Integration for Complex Traits framework; GIP, genetically independent phenotype; PC, principal components; SMR/HEIDI, Summary data-based Mendelian Randomization analysis followed by the Heterogeneity in Dependent Instruments test; FUMA, Functional Mapping and Annotation of Genome-Wide Association Studies platform.

Genetically independent phenotypes

The four original chronic musculoskeletal pain phenotypes were converted into GIPs using the coefficients of orthogonal transformation generated in the principal component analysis based on the matrix of genetic covariances. Coefficients of orthogonal transformation represent contribution of each pain phenotype on each GIP, while genetic variance explained by GIPs approximates contribution of each GIP to each pain phenotype. A graphical representation of orthogonal transformation coefficients, as well as the genetic variance of chronic musculoskeletal pain phenotypes explained by each GIP, is shown in Figures 2a and 2b, respectively.

The contributions of all pain phenotypes to GIP1 had the same direction and approximately the same magnitude. GIP1 showed the best stability based on the narrow 95% confidence intervals of orthogonal transformation coefficients. As expected, GIP1 explained the largest proportion of genetic variance (78.4%) of the four investigated musculoskeletal pain traits (the formula for calculating this value is provided in Supplementary Methods, page 9). SNP-based heritability of GIP1 was 7% and was found to be substantially larger than the heritability of the four individual pain phenotypes (2-4%, Figure 2c).

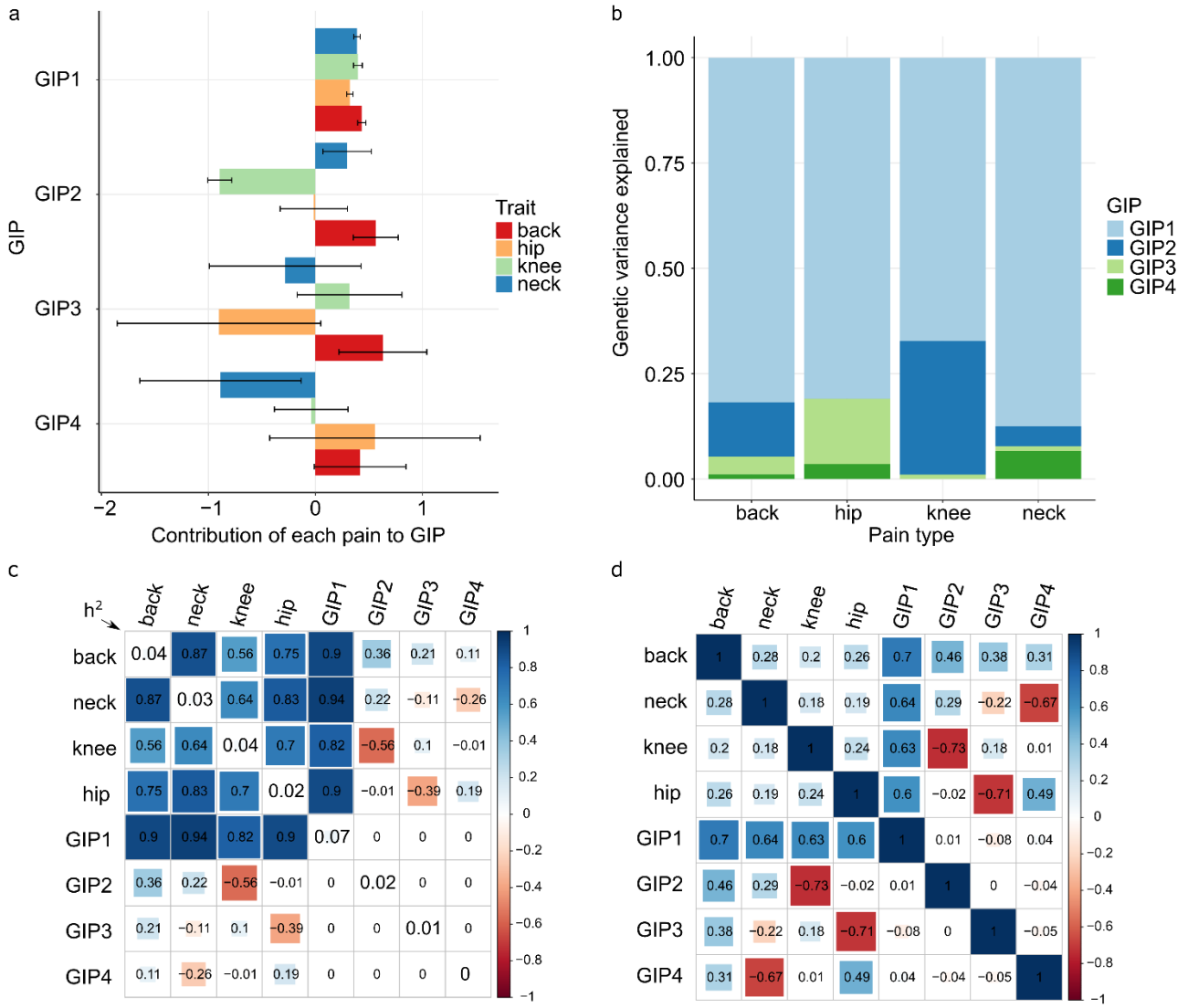

Figure 2. Genetically independent phenotypes (GIP) for chronic musculoskeletal pain. **a.** Barplots depicting the contribution of the four chronic musculoskeletal pain traits to each GIP. The bars represent orthogonal transformation coefficients, and the whiskers indicate their 95% confidence intervals. **b.** Genetic variance of the studied chronic musculoskeletal pain explained by four GIPs. **c.** Estimated matrix of genetic correlations between the four chronic musculoskeletal pain phenotypes and GIPs. The diagonal elements represent estimates of SNP-based heritability (h^2) on the observed scale for each trait. **d**. Matrix of phenotypic correlations between the four chronic musculoskeletal pain phenotypes and GIPs (estimated for pain phenotypes and predicted for GIPs). Estimates for **c, d** were obtained using the discovery cohort of European ancestry individuals (N = 265,000).

GWAS for genetically independent phenotypes

At the discovery stage, 9 loci passed the study-level threshold of statistical significance set at $P < 1.3e-08$ (5e-08/4, where 4 is the number of GIPs) after correction for the LD Score regression intercept (1.016 for GIP1, 1.001 for GIP2, 1.013 for GIP3, and 1.021 for GIP4). Six of the loci were associated with GIP1, and three with GIP2 (Table 1). Conditional and joint analysis showed single association signals per locus (Table S3). Manhattan plots of $-\log_{10}(P)$ are given in Figure S1, quantile-quantile plots in Figure S2, and regional association plots in Figure S3.

Associations of six loci (five associated with GIP1 and one with GIP2) were replicated at $P < 5.6e-03$ (0.05/9, where 9 is the number of loci identified in the discovery stage). Full results of associations with each GIP and studied chronic musculoskeletal pain phenotype are provided in Table S4.

Two of the six replicated loci showed genome-wide significant associations with chronic pain at specific location in the discovery cohort (P < 5e-08, Table S4). These included the GIP1-associated locus near the *EXD3* gene (tagged by rs73581580 and associated with chronic back pain) and the GIP2-associated locus near the *GDF5* gene (tagged by rs143384 and associated with chronic knee pain). In the meta-analysis of European ancestry discovery and replication cohorts, two additional loci reached a genome-wide significance for association with pain at specific location: the GIP1-associated locus near the *SLC39A8* gene (tagged by rs13107325 and associated with chronic neck/shoulder pain) and the GIP1-associated locus near the *ECM1* gene (tagged by rs3737240 and associated with chronic hip pain) (Table S4).

Functional effects of SNPs rs13107325, rs3737240, and rs143384 and/or their associations with complex traits and diseases have been described previously (Table S5). In brief, a missense polymorphism rs13107325 in the divalent cation transporter gene *SLC39A8* is one of the most pleiotropic variants in the human genome, associated with multiple traits including spine conditions (Table S5). The allele T, associated with GIP1 in our study, was associated in prior studies with decreased height, greater spinal curvature, increased risk of severe adolescent idiopathic scoliosis [57], osteoarthritis [58], Crohn's disease [59], and schizophrenia [60].

A missense SNP rs3737240 is located in the *ECM1* (extracellular matrix protein 1) gene encoding a protein involved in negative regulation of endochondral bone formation and chondrogenesis [61-63]. Previous studies reported association of the variant T allele (or the tightly linked rs13294 A allele (r^2 = 0.97 in European ancestry populations), which is also a missense *ECM1* gene variant), inversely associated with GIP1 in our study, with the increased risk of ulcerative colitis [64-66]. GIP1-associated allele rs3737240 C is in high LD with the allele rs12040949 C (r^2 = 0.94 in European ancestry populations), which showed an association with the increased risk of hip osteoarthritis in a recent GWAS [58].

Polymorphism rs143384 in the 5'-untranslated region of the growth differentiation factor 5 gene (*GDF5)* is in high LD with rs143383 (rs143384 T allele is positively correlated with rs143383 T allele). In multiple previous studies, rs143383 T allele was associated with decreased expression of *GDF5* in the joint [67], increased risk of osteoarthritis [58, 67-71], lumbar disc degeneration [72], and congenital dislocation of the hip [73, 74]. In our study, association of rs143383 with GIP2 had the same magnitude of effect as that of rs143384, and also passed the study-level statistical significance threshold (discovery cohort: P = 8.53e-12 after correction for residual inflation). It should be noted that both variant alleles (rs143384 T and rs143383 T) were inversely associated with GIP2, consistent with the negative coefficient of knee pain phenotype observed for GIP2 (Figure 2a).

Table 1. Top SNPs associated with GIPs.

GIP*	Lead SNP	Chr:position**	RefA/Eff A†	Nearest gene§	Discovery cohort (N = 265,000)					Meta-analysis of 3 replication cohorts¥				
					β	SE	*P*	$P^{‡}_{(GC)}$	EAF	*β*	SE	*P*	EAF	N
GIP2	**rs143384**	**20:34025756**	**C/T**	*GDF5*	**-0.020**	**0.003**	**4.87e-13**	**7.40e-13**	**59.8%**	**-0.022**	**0.003**	**1.65e-10**	**58.5%**	**191,580**
GIP1	**rs7628207**	**3:49754970**	**T/C**	*AMIGO3*	**-0.023**	**0.004**	**1.71e-10**	**2.37e-10**	**82.3%**	**-0.012**	**0.004**	**4.92e-03**	**81.8%**	**191,580**
GIP1	**rs13107325**	**4:103188709**	**T/C**	*SLC39A8*	**-0.032**	**0.005**	**8.78e-10**	**1.19e-09**	**92.6%**	**-0.035**	**0.007**	**4.21e-08**	**92.6%**	**191,580**
GIP1	**rs3737240**	**1:150483355**	**T/C**	*ECM1*	**0.017**	**0.003**	**2.01e-09**	**2.69e-09**	**60.4%**	**0.010**	**0.003**	**3.17e-03**	**61.1%**	**191,580**
GIP1	**rs73581580**	**9:140251458**	**G/A**	*EXD3*	**0.025**	**0.004**	**3.89e-09**	**5.15e-09**	**12.4%**	**0.030**	**0.005**	**9.54e-09**	**12.3%**	**174,831**
GIP1	**rs12705966**	**7:114248851**	**G/A**	*FOXP2*	**0.018**	**0.003**	**5.71e-09**	**7.52e-09**	**66.7%**	**0.012**	**0.004**	**1.70e-03**	**67.2%**	**191,580**
GIP2	rs4985445	16:69867835	G/A	*WWP2*	0.017	0.003	1.56e-09	2.09e-09	54.3%	0.007	0.003	0.0371	53.2%	191,580
GIP2	rs548227718	5:175902724	G/A	*FAF2*	-0.283	0.048	3.02e-09	4.01e-09	0.1%	0.096	0.060	0.1056	0.1%	174,831
GIP1	rs111368900	1:53084695	G/A	*GPX7*	0.242	0.041	5.01e-09	6.60e-09	0.2%	0.089	0.048	6.55e-02	0.2%	174,831

Replicated associations are shown in bold. EAF, effect allele frequency; SE, standard error; SNP, single nucleotide polymorphism

*Genetically independent phenotype with which the locus is associated

**Chromosome: position on chromosome according to GRCh37.p13 assembly

†Reference allele/effective allele

§Nearest gene according to the NCBI dbSNP database (https://www.ncbi.nlm.nih.gov/snp/)

‡*P*-value corrected for residual inflation using the LD Score regression intercept

¥Cohorts of individuals of African, South Asian and European ancestry from the UK Biobank (3.9%, 4.8%, and 91.3% in the total replication cohort, N = 191,580)

Functional annotation of the revealed signals

Literature-based gene prioritization

For genes located near the lead SNPs (± 250 kb) associated with GIPs, we performed a search in the Online Mendelian Inheritance in Man database (OMIM, https://www.omim.org/), Google Scholar, the NCBI Gene (https://www.ncbi.nlm.nih.gov/gene), and the Pubmed database (https://www.ncbi.nlm.nih.gov/pubmed) to infer whether the biological functions of these genes may better explain their involvement in chronic musculoskeletal pain. The list of genes in the studied regions was based on regional association plots (Figure S3) and is given in Table S4. Summary information on the genes that we considered most likely to be causal (literature data with references to corresponding sources) is provided in Table S6. In brief, we found 13 genes with plausible roles in pain phenotypes or related conditions: *GDF5* and *MMP24* (near rs143384), *AMIGO3* and *BSN* (rs7628207), *SLC39A8* (rs13107325), *MIR6878*, *ECM1* and *CTSS* (rs3737240), *MIR7114*, *NSMF*, *NOXA1*, and *GRIN1* (rs73581580), and *FOXP2* (rs12705966). Some of these genes have been linked to neuropathic pain (*MMP24* [75, 76], *CTSS* [77-81], *NOXA1* [82]) and recovery after central nervous system (CNS) injury (*AMIGO3* [83, 84]). Genes *GDF5* [85, 86], *SLC39A8* [57, 87, 88], *ECM1* [61-63], *MIR6878* [89], and *MIR7114* [90] were shown to be related to musculoskeletal disorders (osteoarthritis, ankylosing spondylitis) and/or skeletal development. Other genes (*BSN* [91], *NSMF* [92, 93], *GRIN1* [94], and *FOXP2* [95, 96]) are involved in nervous system development or synaptic transmission. In particular, the *FOXP2* gene product is required for proper development of speech and language regions of the brain during embryogenesis [95, 96].

Prediction of SNP effects

Variant Effect Predictor (VEP) identified four missense variants: rs13107325 in the *SLC39A8* gene, rs3737240 and rs13294 in the *ECM1* gene, and rs79140116 in the *EXD3* gene. SIFT and PolyPhen tools predicted possibly damaging/deleterious effects only for rs13107325 and rs13294, while the remaining SNPs were designated as benign/tolerated (Table S7a). Polymorphism rs13107325 is a triallelic SNP (C>T, A), and possibly damaging effects were predicted for both minor alleles T and A. Allele A is extremely rare and was not analyzed in the present study. Allele T was pain-predisposing (and positively associated with GIP1). Polymorphism rs13294 is also a triallelic SNP (G>A, T) and the extremely rare allele T was not covered by our GWAS. SIFT and PolyPhen tools predicted possibly damaging/deleterious effects only for the rare T variant, while allele A (inversely associated with GIP1 in our study) was attributed as benign/tolerated. However, it is still possible that in the case of a large effect of the rare allele rs13294T on GIP1, lead SNP rs3737240 only tags this rare variant (rs3737240 and rs13294 are located 1.6 kb from each other and are in high LD, $r^2 = 0.97$ in European ancestry populations). FATHMM-XF and FATHMM-INDEL identified a potentially pathogenic intronic SNP rs28535523 in the *UBA7* gene and an intronic indel rs34291892 in the *FOXP2* gene (Tables S7b, c). Potentially pathogenic variants rs28535523 T and rs34291892 insertion A were positively associated with GIP1. Data on matching the possibly damaging/deleterious/pathogenic alleles with the effects on GIPs, amino acid changes (where appropriate), and lead SNP alleles are presented in Table S7d.

Pleiotropic effects on gene expression

Summary data-based Mendelian Randomization (SMR) analysis followed by the Heterogeneity in Dependent Instruments (HEIDI) test provided evidence that the same causal SNP in the locus tagged by rs143384 is associated with GIP2 and the expression of *GDF5*, *UQCC1* and *RP3-477O4.16*

(the gene encoding long intergenic non-coding RNA) in different tissues including brain caudate basal ganglia (Table S8). Pleiotropic effects were also found for the locus tagged by rs3737240 associated with GIP1 and *MRPS21* gene expression in blood, and for the locus tagged by rs7628207 associated with GIP1 and expression levels of the genes *RBM6*, *FAM212A*, *RNF123* and pseudogene *ACTBP13* (mainly in nervous tissues). It is likely that the locus tagged by rs7628207 contains regulatory elements that influence transcription of adjacent genes. Interestingly, *RNF123* gene expression has been linked to the risk of major depression [97], and major depressive disorders are genetically correlated with pain [25]. As the *AMIGO3* gene transcript (the CNS-related gene bearing the GIP1-associated SNP rs7628207 in its intron) was not present among the list of probes analyzed in the GTEx [52] and Westra projects [51], we could not infer pleiotropy. Other genes found in the literature-based and SNP effect analyses did not passed thresholds in SMR and HEIDI tests, signifying that we have no support to claim that their expression is influenced by causal variants associated with GIPs.

DEPICT gene prioritization

Statistically significant results of DEPICT gene prioritization (FDR < 0.05) were observed only for GIP1 and only when the *P*-value threshold for input SNPs was set at 1e-05. The list of prioritized genes is provided in Table S9a. Of the genes identified in previous analyses only *BSN* and *FOXP2* were found to be prioritized by DEPICT.

Summary of gene prioritization

A summary list of prioritized genes is presented in Table 2. For each locus tagged by rs143384, rs13107325, rs3737240, and rs12705966, two or more lines of evidence support a role for *GDF5*, *SLC39A8*, *ECM1*, and *FOXP2* genes, respectively, providing solid ground for their prioritization. Single candidate genes could not be suggested for loci tagged by rs7628207 and rs73581580 since different approaches yielded different results. The nearest gene to rs7628207 is *AMIGO3*, which has been shown to participate in inhibition of axon regeneration in the damaged CNS [83, 84]. Five more genes are present in this region that were prioritized by *in silico* methods and/or based on prior literature data (in particular, the *BSN* gene encoding Bassoon presynaptic cytomatrix protein). Lead SNP rs73581580 is located in the intron of the *EXD3* gene, an ortholog of *C. elegans* mut-7 gene required for transposon silencing and RNA interference in that organism. Nevertheless, results from other studies suggest four genes with more plausible effects on chronic musculoskeletal pain (*MIR7114* [90], *NOXA1* [82], *NSMF* [92, 93], and *GRIN1* [94], Table S6).

Table 2. Summary of gene prioritization.

Lead SNP	Locus*	GIP**	Number of genes in the locus†	Prioritized gene	Nearest gene, *yes/no* (lead SNP location)	Evidence for prioritization
rs143384	20:34025756	GIP2	15	***GDF5***	*yes* (5' UTR)	L, S
rs7628207	3:49754970	GIP1	18	*AMIGO3*	*yes* (intronic)	L
				BSN	*no*	L, D
				RBM6	*no*	S
				FAM212A	*no*	S
				RNF123	*no*	S
				UBA7	*no*	V
rs13107325	4:103188709	GIP1	3	***SLC39A8***	*yes* (missense)	L, V
rs3737240	1:150483355	GIP1	19	***ECM1***	*yes* (missense)	L, V
rs73581580	9:140251458	GIP1	32	*MIR7114*	*no*	L
				NSMF	*no*	L
				NOXA1	*no*	L
				GRIN1	*no*	L
rs12705966	7:114248851	GIP1	2	***FOXP2***	*yes* (intronic)	L, V, D

Genes with strong evidence for prioritization are indicated in bold.
D, DEPICT analysis; L, literature-based prioritization (see Table S6); S, SMR/HEIDI analysis; V, Variant Effect Predictor/FATHMM analysis; UTR, untranslated region
*Chromosome: position on chromosome according to GRCh37.p13 assembly
**Genetically independent phenotype with which the locus is associated
†Calculated based on regional association plots generated with LocusZoom tool (http://locuszoom.org/) in a 500 kb window (±250 kb around the lead SNP, Figure S3)

Gene set and tissue/cell type enrichment

DEPICT gene set and tissue/cell type enrichment analyses provided statistically significant results only for GIP1 (Table S9c-f). For SNP sets associated with GIP1 with $P < 5e\text{-}08$, tissue/cell type enrichment with FDR < 0.05 was found for two terms: the "Neural Stem Cells" cell type and "Retina" tissue. However, relaxing the significance threshold of input SNPs to $P < 1e\text{-}05$ led to identification of 24 additional tissues, all of which were related to CNS. The same pattern was observed for gene set enrichment (for SNPs with $P < 1e\text{-}05$), revealing 462 terms mainly involved in nervous system function, development and morphology (e.g. "regulation of nervous system development", "axonogenesis", "synapse", "regulation of transmission of nerve impulse").

FUMA gene set and tissue enrichment analyses for GIP1 detected 9 gene categories (6 of them were nervous system-related) and 12 brain tissues, respectively (Table S10, Figure S4). For GIP2 and GIP3, a total of three gene sets were found by FUMA analysis, although we considered them as non-specific (e.g. "nikolsky_breast_cancer_20q11_amplicon"; Table S10). No statistically significant gene sets were revealed for GIP4, and no statistically significant tissue types were identified for GIP2, GIP3, and GIP4.

Pleiotropic effects on complex traits

Five out of six replicated loci demonstrated pleiotropic effects on human complex traits in the SMR/HEIDI analysis (Table S11, Figure 3). As expected, the GIP1-associated locus rs13107325 (known as one of the most pleiotropic variants of the genome) was associated with the greatest number of diverse phenotypes, which included anthropometric traits (weight, height, and BMI), fluid intelligence score, prospective memory and education, sleep duration, Crohn's disease, self-reported osteoarthritis, diastolic blood pressure, blood cell traits, and alcohol intake frequency. Traits linked

with the GIP2-associated locus rs143384 were mainly related to anthropometry and knee-related conditions (gonarthrosis and internal derangement of knee). The locus tagged by the missense SNP rs3737240 (*ECM1* gene) showed pleiotropic effects on platelet count and plasma level of extracellular matrix protein 1 (ECM1) measured with the SOMAscan platform [98]. The same pain-promoting allele in this locus that was positively associated with GIP1 was linked to an increase in ECM1 level, reinforcing the role of *ECM1* as the candidate in this region. In the locus tagged by rs73581580, GIP1-associated alleles were linked to higher frequency of tiredness and difficulty of getting up in the morning. In the locus tagged by rs7628207, GIP1-associated variants were related to decreased plasma level of thioredoxin domain-containing protein 12 (TXNDC12), decreased overall health rating, decreased age at first live birth, decreased educational attainment, increased basal metabolic rate, and increased hip circumference. Interestingly, rs7628207 is adjacent to the *AMIGO3* gene prioritized by us based on the literature data (Table 2, Table S6) which is linked to the gene encoding *TXNDC12* via a trans-protein QTL rs4688759 [99].

Hospital-diagnosed osteoarthritis (the UK Biobank trait for which GWAS summary statistics was downloaded from the Michigan PheWeb database, see Methods section) was not revealed in the SMR/HEIDI analysis for any of the analyzed loci. However, for rs13107325, rs3737240, and rs143384, we can speculate that this could be due to the limited statistical power of the analysis. The SMR test *P*-values for these loci were quite low, although did not reach the Bonferroni-corrected significance threshold of $P = 3.71e\text{-}06$ (rs13107325: $P_{SMR} = 1.14e\text{-}05$, $beta_{SMR} = 0.63$; rs3737240: $P_{SMR} = 1.68e\text{-}05$, $beta_{SMR} = 0.89$; rs143384: $P_{SMR} = 6.13e\text{-}04$, $beta_{SMR} = -0.40$; $P_{HEIDI} \geq 0.01$ for all these loci). Thus, we cannot rule out a hypothesis that the same causal SNPs within the loci tagged by rs13107325 and rs3737240 may be associated with GIP1 and the increased risk of osteoarthritis, and the same causal SNPs within the locus tagged by rs143384 can be associated with GIP2 and the decreased risk of osteoarthritis.

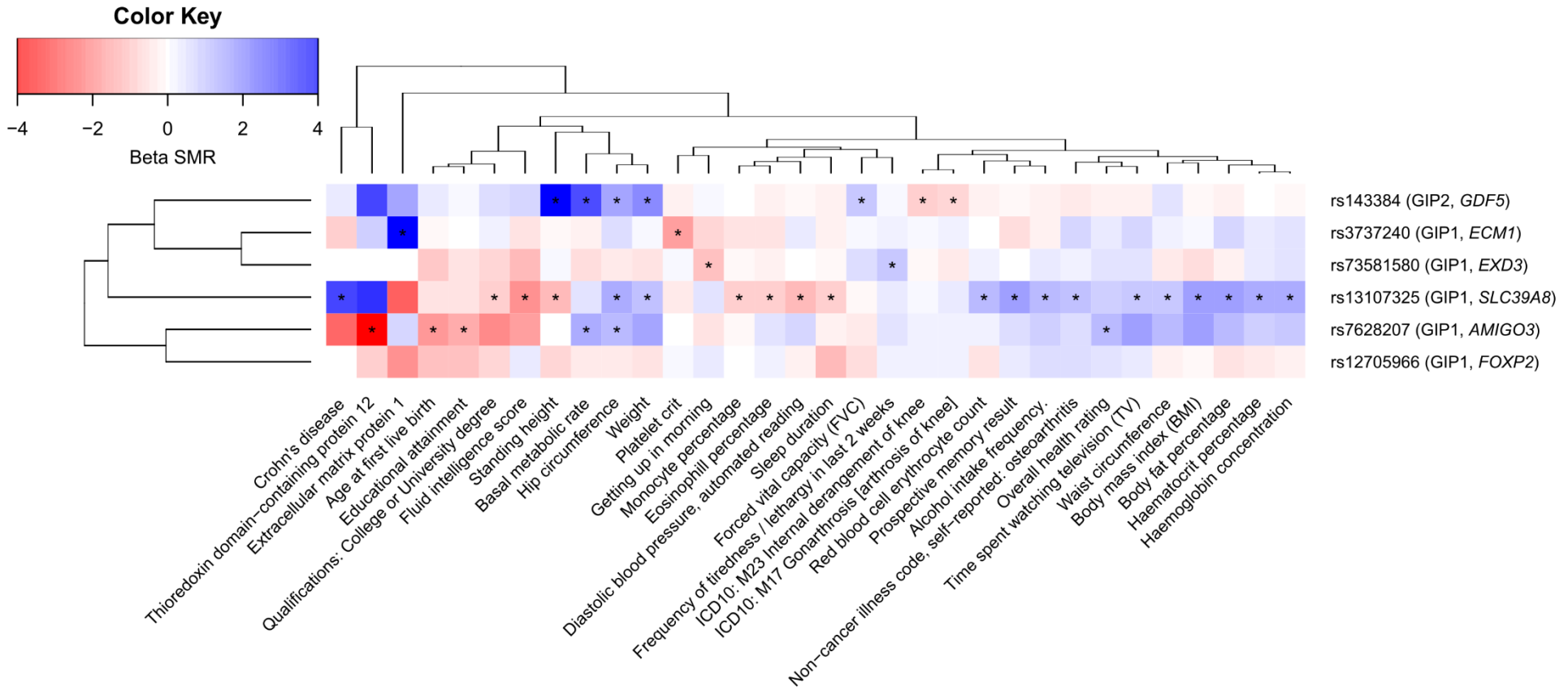

Figure 3. Pleiotropic effects of identified loci on human complex traits.
Color depicts the sign and the magnitude of SMR beta coefficient. Negative sign (red) means opposed effects on the corresponding GIP and the trait, and positive sign (blue) means the same direction of effect. |beta SMR| > 4 are depicted as |beta SMR| = 4. For "Prospective memory result" and "Overall health rating" trait, high scores correspond to poor performance. For "Getting up in morning" trait, high score corresponds to easy getting up. Traits that passed both SMR and HEIDI tests (P_{SMR} < 3.71e-06 and $P_{HEIDI} \geq 0.01$) are marked with an asterisk. Data on 45 out of 78 revealed traits are not shown. Full results are given in Table S11. GIPs associated with the loci and genes nearest to lead SNPs are indicated in parentheses. Dendrograms represent clustering based on complete linkage hierarchical clustering method.

Genetic correlations between GIPs and complex traits

GIP1 showed statistically significant genetic correlations with 40 complex traits (Table S12a, Figure 4). Among them, 11 traits were directly linked to excess weight (BMI, overweight, obesity, waist circumference), that is in line with known epidemiological associations between chronic pain and obesity-related traits [100]. Five more traits fell in the same cluster: HDL cholesterol (negative correlation with GIP1), triglycerides, HOMA-IR, leptin, and fasting insulin. Strong genetic correlations ($|r_g|$ ranging between 0.31 and 0.54) were also revealed between GIP1 and the cluster of psychiatric/personality traits (major depressive disorder, depressive symptoms, subjective well-being, and neuroticism). This finding is in accord with previous twin and family studies demonstrating a common genetic background for pain and depression [101-103]. Other traits included sociodemographic, reproductive, education-related and smoking-related traits, osteoarthritis, rheumatoid arthritis, coronary artery disease, and sleep duration.

Traits that displayed the strongest genetic correlations with GIP1 were osteoarthritis (rg = 0.65), age of first birth (rg = -0.56), depressive symptoms (rg = 0.54), and college completion (rg = 0.54). Overall, the pattern of genetic correlations with GIP1 was very similar to that observed for back pain in our previous study [104]. GIP2 was genetically correlated only with osteoarthritis (inverse genetic correlation, rg = -0.30) and obesity-related traits, and GIP4 only with hip circumference. No statistically significant genetic correlations with complex traits were found for GIP3 (Table S12 b-d).

Furthermore, we analyzed the genetic correlation between GIP1 and the first GIP constructed using the same methodology for a broader range of chronic pain traits (back, neck/shoulder, knee, hip, stomach/abdominal pain, and headache). We found out that these GIPs were almost genetically equivalent (rg = 0.99).

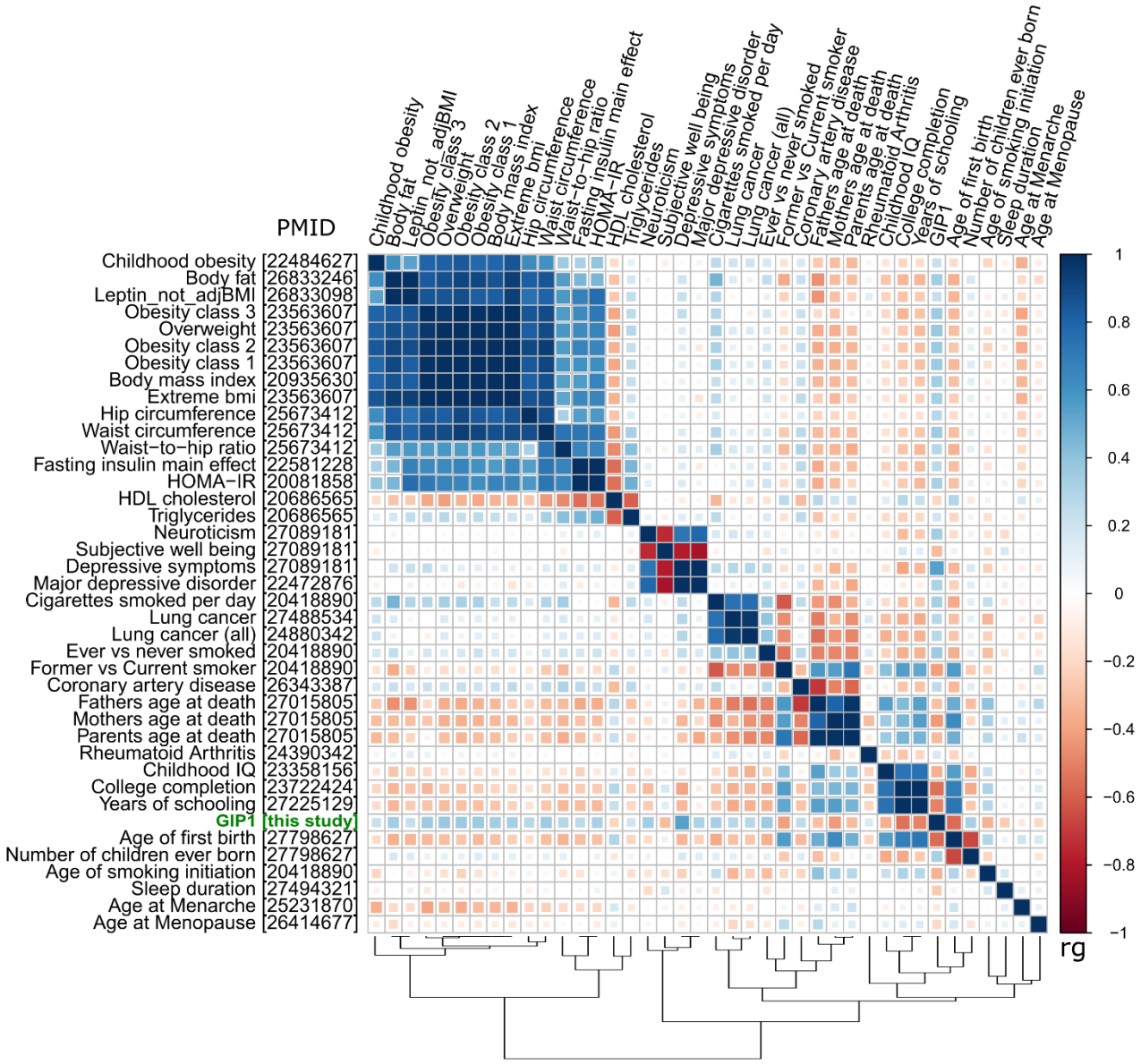

Figure 4. Matrix of genetic correlations between GIP1 and human complex traits.
Color depicts the sign and absolute value of the genetic correlation coefficients (rg). Genetic correlations between GIP1 and all presented traits were statistically significant (P < 5.98e-05). Osteoarthritis is not shown on this plot since genetic correlations analysis for this trait was performed using the GWAS-MAP platform, whereas for other traits, LD hub web interface was used.
HDL, high density lipoprotein; HOMA-IR, Homeostatic Model Assessment for Insulin Resistance; PMID, PubMed ID number of the literature source providing GWAS summary statistics.

Discussion

The genetic control of chronic musculoskeletal pain is complex, with each of very many genetic variants contributing a small effect. As a result, even very large genome-wide association studies provide only a limited number of replicated loci and rather low SNP-based heritability. Evidence from recent studies indicates that pain at different anatomical sites shares a common genetic component [24-26]. This suggests that combining several pain phenotypes in a single analytical framework may facilitate the discovery of common genetic factors – chronic musculoskeletal pain genes and pathways.

In the present study, we applied an approach that allowed us to detect genes shared between four common chronic musculoskeletal pains: back, neck/shoulder, knee and hip. Our approach relies on capturing heredity of a set of genetically correlated traits via constructing genetically independent phenotypes (GIPs) (Figure 2a). The GIPs are defined as a weighted sum of the original phenotypes, with weights selected in such a way that the first GIP (GIP1) explains most genetic variance of and covariance between the studied traits, with the later GIPs (GIP2-4) explaining progressively less. The four weights defining GIP1 based on the four chronic pain traits (back, neck/shoulder, hip, and knee pain) turned out to be approximately the same (Figure 1a). This means that GIP1, the genetic component explaining most of the cases of chronic musculoskeletal pain at the studied sites, affects the risk of chronic musculoskeletal pain to approximately the same degree, irrespective of pain's location. Unlike the first GIP, the second GIP is site-specific and reflects a genetic propensity for knee pain, but not the back or neck/shoulder pain.

We mapped and replicated six genomic loci (five associated with GIP1 and one with GIP2). Importantly, in the discovery sample, only two out of six replicated loci were genome-wide significantly associated with the individual pain phenotypes: rs73581580 with chronic back pain and rs143384 with chronic knee pain. Also, as expected, the SNP-based heritability of GIP1 was substantially higher than for any of separate pain traits (7% vs 2-4%). These results highlight the improved power of the GIP approach for identifying genetic predictors of chronic pain predisposition. It should be noted that phenotypic correlations between the traits were much lower than the genetic correlations (Figure 2c, d; pairwise phenotypic correlations ranged from 0.18 to 0.28, while pairwise genetic correlations ranged from 0.56 to 0.87). In this scenario, we can speculate that conventional multivariate approaches based on phenotypic correlations like MANOVA [105] or MultiPhen [106] would have been less powerful than our method based on genetic correlations. Moreover, while estimation of phenotypic correlation is impossible for non-overlapping samples, genetic correlations can be calculated for both overlapping samples and independent cohorts [38]. This makes our approach applicable to the traits measured within the frameworks of different genomics consortia.

Among the six replicated loci, three were well-studied polymorphisms associated with different traits and conditions in previous works (rs13107325, rs3737240 and rs143384, Table S5). In the present study, we performed a hypothesis-free analysis of pleiotropic effects of six GIP-associated loci on 2,243 complex human traits. Our analysis revealed 78 phenotypes influenced by the same causal polymorphisms that are associated with GIPs (Table S11, Figure 3). These phenotypes included a broad variety of anthropometric, sociodemographic, behavior and personality traits, diseases (such as Crohn's disease, gonarthrosis and osteoarthritis), and laboratory parameters. Interestingly, GIP1-associated alleles in the locus tagged by rs73581580 were also associated with higher frequency of tiredness and difficulty of getting up in the morning. Our results demonstrate

diversity of effects of the GIP-associated loci and suggest the presence of common pathways underlying chronic musculoskeletal pain and multiple other human traits.

GIP1-associated pathways and tissues were mostly related to CNS development and functioning, suggesting that GIP1 depicts neurological and psychological components of chronic pain. Consistent with this, one of the genes prioritized for GIP1-associated loci based on multiple lines of evidence was *FOXP2*, whose product is a transcription factor expressed in fetal and adult brain and required for the development of speech and language regions [95, 96]. Involvement of psychological component in chronic pain was additionally supported by the finding of a very strong positive genetic correlation between GIP1 and depressive symptoms. Having said that, it is equally important that GIP1 was associated also with traits reflecting general health and risk factors for musculoskeletal pain: sociodemographic, reproductive, education- and smoking-related traits, and sleep duration. Importance of morphological factors for chronic musculoskeletal pain was also demonstrated by revealing of GIP1-associated genes *SLC39A8* and *ECM1*, which are known to be implicated in the development and functioning of the musculoskeletal system. *ECM1* gene encodes a negative regulator of bone mineralization and chondrogenesis [61-63]. GIP1-associated ("pain-promoting") variant in this gene showed an association with the increased level of ECM1 protein in our SMR/HEIDI analysis. GIP1-associated *ECM1* allele rs3737240 C is in a high LD ($r^2 = 0.94$ in European ancestry populations) with the allele rs12040949 C, which was associated with the increased risk of hip osteoarthritis in a recent study [58]. The product of the *SLC39A8* gene was shown to participate in osteoarthritis cartilage destruction [87, 88]. *Slc39a8* mutant zebrafish exhibit vertebral abnormalities, impaired growth, and decreased motor activity, and a missense GIP1-associated polymorphism rs13107325 in the *SLC39A8* gene has previously been associated with the increased risk of osteoarthritis [58] and severe adolescent idiopathic scoliosis [57]. Thus, similar to findings from our recent study of back pain [104], genetic factors underlying chronic musculoskeletal pain comprise biological, social and psychological components.

Since our study was aimed at investigating chronic musculoskeletal pains at anatomical sites commonly affected by osteoarthritis, it was not surprising that we found loci and genes associated with this condition and found high genetic correlation between osteoarthritis and GIP1 (rg = 0.65). Note, that this genetic correlation is similar in magnitude to correlation between GIP1 and age of first birth (-0.56), indicating that although similarities are high, there exist substantial differences between OA and GIP1. Furthermore, for GIP1, gene/tissue enrichment analysis revealed a plethora of CNS-related terms. In a recent large-scale genetic study for OA, enriched terms were not directly linked to the nervous system ("anatomical structure morphogenesis", "ion channel transport", "histidine metabolism", etc.) [58]. Finally, we performed genetically independent phenotype analysis for the extended set of chronic traits, which include not only musculoskeletal pain (six traits: back, neck/shoulder, knee, hip pain as well as stomach/abdominal pain and headache). Genetic correlation between GIP1 for four pain traits and GIP1 for six pain traits was extremely high (rg = 0.99) providing strong evidence that, despite high genetic overlap with OA, GIP1 for musculoskeletal pain likely reflects chronic pain *per se*.

It is noteworthy that pain is the main symptom and clinical outcome of osteoarthritis. In the UK Biobank study which provided GWAS summary statistics for OA [55], phenotypes were defined according to ICD-9/ICD-10 codes (electronic medical record data), so whether the study participants were examined radiographically or not is unknown. Thus, genetic overlap between GIP1 and OA can

be actually biased by a genetic correlation between GIP1 and not the OA, but pain in OA. Besides this, a study by Valdes et al. [107] obtained interesting results on the inverse relationship between preoperative radiographic severity and postoperative pain in OA patients who have undergone total joint replacement (TJR). The authors hypothesized that in OA patients with low preoperative radiographic damage, pain leading to TJR can be caused not entirely by a joint damage, but also by other factors such as central sensitization. It is possible that these factors have common genetic background with GIP1 constructed in our study.

Given that GIP1 essentially contrasts chronic musculoskeletal pain (in general) with an unpainful state, the other GIPs might be expected to account for musculoskeletal pain at specific anatomical locations. This was indeed the case with GIP2, which had the greatest impact on knee pain (Figure 2b). The only gene found to be associated with GIP2 at the genome-wide significance level was *GDF5*, a gene with well-established associations with peripheral osteoarthritis and intervertebral disc degeneration [58, 67-72]. These results are consistent with the fact that the knee is one of the most common sites of osteoarthritis. For GIP3 and GIP4, no firm conclusions can be drawn regarding what component of pain they might represent, but, as can be seen from Figure 2b, GIP3 makes a substantial contribution to hip pain and GIP4 to neck/shoulder pain.

Another approach recently applied in GWAS of chronic pain is based on obtaining a phenotype of multisite chronic pain (MCP) as a sum of the number of anatomical sites affected by pain (a study by Johnston et al. [108]). The MCP phenotype may seem similar to our GIP1 trait at first glance; however, an important potential drawback of the MCP approach is that it mixes up phenotypes that do not necessarily have the same genetic background. The summing of different pain sites into a quantitative MCP phenotype assumes equivalence between the genetic predictors of musculoskeletal pain conditions (such as back and knee pain) and the genetic predictors of non-musculoskeletal pain conditions that may include substantial components of pain due to other causes, such as migraine (in the case of headache), dental or neuropathic pain (in the case of facial pain), or visceral pain (in the case of stomach/abdominal pain). Such equivalence may be too strong an assumption to make without empirical justification. Our approach is empirical, with definition of GIPs driven by the data; another strength of our approach is its ability to reveal pain type specific genetic loci as exemplified by *GDP5* associated with GIP2 representing knee pain. Comparing with a direct knee pain GWAS, GIP2 may provide a more knee-specific phenotype from which general propensity to pain is subtracted. This claim requires experimental validation, though.

Nevertheless, our study, together with that by Johnston et al. [108], is among the first to use a GWAS framework to address the genetics of chronic pain at multiple sites. Despite the difference in methodology and the phenotypes involved, our study identified five loci also reported by Johnston et al.: *AMIGO3* (tagged by rs7628207 in [108]), *SLC39A8* (tagged by rs13135092 in [108]), *ECM1* (tagged by rs59898460 in [108]), *EXD3* (tagged by rs73581580 in both [108] and our study), and *FOXP2* (tagged by rs12537376 in [108]). It should be noted that in our study, in contrast to the study by Johnston et al., these loci have been replicated. However, both discovery and replication stages in our study as well as analyses conducted by Johnston et al. were based on the UK Biobank data only, highlighting the need to replicate these findings in independent cohorts.

Our study has limitations. The first general limitation is related to a questionnaire-based approach to phenotyping, which may lead to heterogeneous pain phenotypes. Our methods attempted to overcome this by constructing genetically independent phenotypes whose genetic basis

approximates the genetic background of distinct phenotypes and likely represents the "general pain" component of analyzed musculoskeletal pain traits. Second, in our study, we focused only on chronic musculoskeletal pain at anatomical sites potentially linked through osteoarthritis, so one must be cautious generalizing our results to other chronic pain conditions. Third, even though we carried out replication analysis, the replication cohorts were drawn from the same source dataset (UK Biobank), so sampling bias cannot be excluded. Finally, for two out of six identified loci (tagged by rs7628207 and rs73581580), we were not able to prioritize a single causal gene, and candidate genes suggested for the locus tagged by rs73581580 were selected based only on data from available literature sources.

In summary, our study of genetically independent components of chronic musculoskeletal pain phenotypes revealed hereditary factors shared by chronic back, neck/shoulder, hip, and knee pain and identified loci and genes relevant for these conditions. Our results provided further support that neurological and psychological components are important contributors to chronic pain. Using this approach may facilitate discovery of chronic pain mechanisms.

Acknowledgements

Funding

The work of YSA and SZS was supported by by the Russian Ministry of Education and Science under the 5-100 and by the Federal Agency of Scientific Organizations via the Institute of Cytology and Genetics (project 0324-2019-0040). The work of YAT, ASSh, and EEE was supported by the Russian Foundation for Basic Research (project 19-015-00151). The contribution of LCK was funded by PolyOmica. Dr. Suri was supported by VA Career Development Award # 1IK2RX001515 from the United States (U.S.) Department of Veterans Affairs Rehabilitation Research and Development (RR&D) Service. Dr. Suri is a Staff Physician at the VA Puget Sound Health Care System. The contents of this work do not represent the views of the U.S. Department of Veterans Affairs or the United States Government.

Author contributions

YSA, FMKW, and PS and conceived and oversaw the study. YAT, MBF, ASSh, YSA, FMKW, and PS contributed to the design of the study and interpretation of the results. YAT, MBF, SZS, and EEE carried out statistical analysis. ASSh performed literature analysis and literature-based gene prioritization and produced the first draft of the manuscript. LCK provided statistical and computational support. All co-authors discussed the results and contributed to preparing the final version of the manuscript.

Conflict of interest

YSA is a founder and co-owner of Maatschap PolyOmica and PolyKnomics BV, private organizations, providing services, research and development in the field of computational and statistical, quantitative and computational (gen)omics.

Supplementary Table Legends

Table S1. Descriptive characteristics of the study cohorts.
Table S2. Description of traits included in the GWAS-MAP database.
Table S3. Results of conditional and joint analysis.
Table S4. Top loci associated with GIPs at a study-level threshold of statistical significance (P < 1.25e-08).
Table S5. Literature data on well-studied SNPs associated with GIP1 and GIP2.
Table S6. Gene prioritization based on a literature review.

Table S7a. Results of the VEP analysis.
Table S7b. Results of the FATHMM-XF analysis for SNPs.
Table S7c. Results of the FATHMM-INDEL analysis for indels.
Table S7d. Matching alleles with predicted detrimental effects with their effects on GIPs and with lead SNP alleles.
Table S7e. SNP set for the VEP and FATHMM analyses.

Table S8a. Results of SMR/HEIDI analysis. Searching for pleiotropic effects on GIPs and gene expression. Associations that passed both SMR and HEIDI analyses (P_{SMR} < 3.24e-06 and $P_{HEIDI} \geq 0.01$).
TableS8b. Results of SMR/HEIDI analysis. Searching for pleiotropic effects on GIPs and gene expression.

Table S9a. GIP1. Results of DEPICT analysis for SNPs with P < 1e-05. Gene prioritization.
Table S9b. GIP1. Results of DEPICT analysis for SNPs with P < 5e-08. Gene prioritization.
Table S9c. GIP1. Results of DEPICT analysis for SNPs with P < 1e-05. Tissue enrichment analysis.
Table S9d. GIP1. Results of DEPICT analysis for SNPs with P < 5e-08. Tissue enrichment analysis.
Table S9e. GIP1. Results of DEPICT analysis for SNPs with P < 1e-05. Gene set enrichment analysis.
Table S9f. GIP1. Results of DEPICT analysis for SNPs with P < 5e-08. Gene set enrichment analysis.

Table S10. Statistically significant results of gene set enrichment analysis conducted using the FUMA platform.
Table S11. Results of SMR/HEIDI analysis. Searching for pleiotropic effects on GIPs and other complex traits. Associations that passed both SMR and HEIDI analyses (P_{SMR} < 3.71e-06 and $P_{HEIDI} \geq 0.01$).

Table S12a. Genetic correlations between GIP1 and human complex traits.
Table S12b. Genetic correlations between GIP2 and human complex traits.
Table S12c. Genetic correlations between GIP3 and human complex traits.
Table S12d. Genetic correlations between GIP4 and human complex traits.

Supplementary Figure Legends

Figure S1. Graphical summary of the discovery GWAS stage after the genomic control correction using LD Score regression intercept.
Figure S2. Quantile-quantile plots for observed vs. expected distribution of P-values for χ^2 statistics.
Figure S3. Regional association plots of $-\log_{10}(P)$ for SNPs located at the distance of ≤ 250 kb from lead SNPs.
Figure S4. Results of tissue enrichment analysis for GIP1 performed using the FUMA platform
Figure S5. Matrix of genetic correlations between GIPs, chronic musculoskeletal pain traits and hospital-diagnosed osteoarthritis (the UK Biobank trait for which GWAS summary statistics was downloaded from the Michigan PheWeb database, http://pheweb.sph.umich.edu/SAIGE-UKB/pheno/740).

References

1. Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: Prevalence, impact on daily life, and treatment. Eur J Pain. 2006;10: 287–287. doi:10.1016/j.ejpain.2005.06.009
2. Saxena AK, Jain PN, Bhatnagar S. The prevalence of chronic pain among adults in India. Indian J Palliat Care. 2018;24: 472–477. doi:10.4103/IJPC.IJPC_141_18
3. Jackson T, Thomas S, Stabile V, Han X, Shotwell M, McQueen K. Prevalence of chronic pain in low-income and middle-income countries: a systematic review and meta-analysis. Lancet. 2015;385: S10. doi:10.1016/S0140-6736(15)60805-4
4. Crofford LJ. Chronic pain: Where the body meets the brain. Trans Am Clin Climatol Assoc. 2015;126: 167–183.
5. Dieppe P. Chronic musculoskeletal pain. BMJ. 2013;346: f3146. doi:10.1136/bmj.f3146
6. Cimmino MA, Ferrone C, Cutolo M. Epidemiology of chronic musculoskeletal pain. Best Pract Res Clin Rheumatol. 2011;25: 173–183. doi:10.1016/j.berh.2010.01.012
7. Bergman S, Herrström P, Högström K, Petersson IF, Svensson B, Jacobsson LT. Chronic musculoskeletal pain, prevalence rates, and sociodemographic associations in a Swedish population study. J Rheumatol. 2001;28: 1369–1377.
8. Nakamura M, Toyama Y, Nishiwaki Y, Ushida T. Prevalence and characteristics of chronic musculoskeletal pain in Japan. J Orthop Sci. 2011;16: 424–432. doi:10.1007/s00776-011-0102-y
9. Vos T, Allen C, Arora M, Barber RM, Bhutta ZA, Brown A, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016;388: 1545–1602. doi:10.1016/S0140-6736(16)31678-6
10. Freidin MB, Lauc G, Allegri M, Primorac D, Williams FM. Using omics in chronic pain conditions to delineate mechanisms and provide new therapeutic strategies. Pain Manag. 2016;6: 211–215. doi:10.2217/pmt.16.2
11. Scholz J. Mechanisms of chronic pain. Mol Pain. 2014;10: O15. doi:10.1186/1744-8069-10-S1-O15
12. Holliday KL, McBeth J. Recent advances in the understanding of genetic susceptibility to chronic pain and somatic symptoms. Curr Rheumatol Rep. 2011;13: 521–527. doi:10.1007/s11926-011-0208-4
13. Zorina-Lichtenwalter K, Meloto CB, Khoury S, Diatchenko L. Genetic predictors of human chronic pain conditions. Neuroscience. 2016;338: 36–62. doi:10.1016/j.neuroscience.2016.04.041
14. Diatchenko L, Fillingim RB, Smith SB, Maixner W. The phenotypic and genetic signatures of common musculoskeletal pain conditions. Nat Rev Rheumatol. 2013;9: 340–350. doi:10.1038/nrrheum.2013.43
15. Holliday KL, Thomson W, McBeth J, Nair N. Genetics of chronic musculoskeletal pain. In: Watts RA, Conaghan PG, Denton C, Foster H, Isaacs J, Müller-Ladner U, editors. Oxford Textbook of Rheumatology. Oxford University Press; 2015. doi:10.1093/med/9780199642489.001.0001
16. Smith SB, Maixner DW, Fillingim RB, Slade G, Gracely RH, Ambrose K, et al. Large candidate gene association study reveals genetic risk factors and therapeutic targets for fibromyalgia. Arthritis Rheum. 2012;64: 584–593. doi:10.1002/art.33338
17. Richards JB, Kavvoura FK, Rivadeneira F, Styrkársdóttir U, Estrada K, Halldórsson B V, et al. Collaborative meta-analysis: associations of 150 candidate genes with osteoporosis and osteoporotic fracture. Ann Intern Med. 2009;151: 528–537. doi:10.7326/0003-4819-151-8-200910200-00006
18. Peters MJ, Broer L, Willemen HLDM, Eiriksdottir G, Hocking LJ, Holliday KL, et al. Genome-wide association study meta-analysis of chronic widespread pain: evidence for

involvement of the 5p15.2 region. Ann Rheum Dis. 2013;72: 427–436. doi:10.1136/annrheumdis-2012-201742

19. Docampo E, Escaramís G, Gratacòs M, Villatoro S, Puig A, Kogevinas M, et al. Genome-wide analysis of single nucleotide polymorphisms and copy number variants in fibromyalgia suggest a role for the central nervous system. Pain. 2014;155: 1102–1109. doi:10.1016/j.pain.2014.02.016
20. Suri P, Palmer MR, Tsepilov YA, Freidin MB, Boer CG, Yau MS, et al. Genome-wide meta-analysis of 158,000 individuals of European ancestry identifies three loci associated with chronic back pain. PLOS Genet. 2018;14: e1007601. doi:10.1371/journal.pgen.1007601
21. Lemmelä S, Solovieva S, Shiri R, Benner C, Heliövaara M, Kettunen J, et al. Genome-wide meta-analysis of sciatica in Finnish population. PLoS One. 2016;11: e0163877. doi:10.1371/journal.pone.0163877
22. Sanders AE, Jain D, Sofer T, Kerr KF, Laurie CC, Shaffer JR, et al. GWAS identifies new loci for painful temporomandibular disorder: Hispanic community health study/study of Latinos. J Dent Res. 2017;96: 277–284. doi:10.1177/0022034516686562
23. Gatchel RJ. The biopsychosocial model of chronic pain. Chronic Pain. Unitec House, 2 Albert Place, London N3 1QB, UK: Future Medicine Ltd; 2013. pp. 5–17. doi:10.2217/ebo.13.469
24. Vehof J, Zavos HMS, Lachance G, Hammond CJ, Williams FMK. Shared genetic factors underlie chronic pain syndromes. Pain. 2014;155: 1562–1568. doi:10.1016/j.pain.2014.05.002
25. Meng W, Adams M, Deary I, Palmer C, McIntosh AM, Smith B. Genetic correlations between pain phenotypes and depression and neuroticism. bioRxiv. 2018; doi:10.1101/362574 [PREPRINT]
26. Williams FMK, Spector TD, MacGregor AJ. Pain reporting at different body sites is explained by a single underlying genetic factor. Rheumatology (Oxford). 2010;49: 1753–5. doi:10.1093/rheumatology/keq170
27. O'Neill TW, Felson DT. Mechanisms of Osteoarthritis (OA) Pain. Current Osteoporosis Reports. 2018;16: 611-616. doi:10.1007/s11914-018-0477-1
28. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. 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: e1001779. doi:10.1371/journal.pmed.1001779
29. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. Am J Epidemiol. 2017;186: 1026–1034. doi:10.1093/aje/kwx246
30. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. Genome-wide genetic data on ~500,000 UK Biobank participants. bioRxiv. 2017; 166298. doi:10.1101/166298 [PREPRINT]
31. Consortium the HR, McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. 2016;48: 1279–1283. doi:10.1038/ng.3643
32. Welsh S, Peakman T, Sheard S, Almond R. Comparison of DNA quantification methodology used in the DNA extraction protocol for the UK Biobank cohort. BMC Genomics. 2017;18: 26. doi:10.1186/s12864-016-3391-x
33. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562: 203–209. doi:10.1038/s41586-018-0579-z
34. Loh P-R, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed-model association for biobank-scale datasets. Nat Genet. 2018;50: 906–908. doi:10.1038/s41588-018-0144-6
35. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium SWG of the PG, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet.

2015;47: 291–295. doi:10.1038/ng.3211

36. Axenovich T, Zorkoltseva I, Belonogova N, van Koolwijk L, Borodin P, Kirichenko A, et al. Linkage and association analyses of glaucoma related traits in a large pedigree from a Dutch genetically isolated population. J Med Genet. 2011;48: 802–809. doi:10.1136/jmedgenet-2011-100436
37. Ried JS, Jeff JM, Chu AY, Bragg-Gresham JL, Van Dongen J, Huffman JE, et al. A principal component meta-analysis on multiple anthropometric traits identifies novel loci for body shape. Nat Commun. 2016;7. doi:10.1038/ncomms13357
38. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015;47: 1236–1241. doi:10.1038/ng.3406
39. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26: 2190–2191. doi:10.1093/bioinformatics/btq340
40. Yang J, Ferreira T, Morris AP, Medland SE, Madden PAF, Heath AC, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet. 2012;44: 369–375. doi:10.1038/ng.2213
41. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88: 76–82. doi:10.1016/j.ajhg.2010.11.011
42. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4: 7. doi:10.1186/s13742-015-0047-8
43. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016;17: 122. doi:10.1186/s13059-016-0974-4
44. Rogers MF, Shihab HA, Mort M, Cooper DN, Gaunt TR, Campbell C. FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. Bioinformatics. 2018;34: 511–513. doi:10.1093/bioinformatics/btx536
45. Ferlaino M, Rogers MF, Shihab HA, Mort M, Cooper DN, Gaunt TR, et al. An integrative approach to predicting the functional effects of small indels in non-coding regions of the human genome. BMC Bioinformatics. 2017;18: 442. doi:10.1186/s12859-017-1862-y
46. Pers TH, Karjalainen JM, Chan Y, Westra H-J, Wood AR, Yang J, et al. Biological interpretation of genome-wide association studies using predicted gene functions. Nat Commun. 2015;6: 5890. doi:10.1038/ncomms6890
47. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017;8: 1826. doi:10.1038/s41467-017-01261-5
48. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized gene-set analysis of GWAS data. Tang H, editor. PLOS Comput Biol. 2015;11: e1004219. doi:10.1371/journal.pcbi.1004219
49. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011;27: 1739–1740. doi:10.1093/bioinformatics/btr260
50. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet. 2016;48: 481–487. doi:10.1038/ng.3538
51. Westra H-J, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet. 2013;45: 1238–1243. doi:10.1038/ng.2756
52. GTEx Consortium TGte. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45: 580–5. doi:10.1038/ng.2653
53. Gorev DD, Shashkova TI, Pakhomov E, Torgasheva A, Klaric L, Severinov A, et al. GWAS-MAP: a platform for storage and analysis of the results of thousands of genome-wide association scans. Bioinformatics of Genome Regulation and Structure/Systems Biology

(BGRS/SB-2018) The Eleventh International Conference. Novosibirsk: ICG SB RAS; 2018. p. 43. doi:10.18699/BGRSSB-2018-020

54. Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. Nat Genet. 2018;50: 1593–1599. doi:10.1038/s41588-018-0248-z

55. Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat Genet. 2018;50: 1335–1341. doi:10.1038/s41588-018-0184-y

56. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018;7: e34408. doi:10.7554/eLife.34408

57. Haller G, McCall K, Jenkitkasemwong S, Sadler B, Antunes L, Nikolov M, et al. A missense variant in SLC39A8 is associated with severe idiopathic scoliosis. Nat Commun. 2018;9: 4171. doi:10.1038/s41467-018-06705-0

58. Tachmazidou I, Hatzikotoulas K, Southam L, Esparza-Gordillo J, Haberland V, Zheng J, et al. Identification of new therapeutic targets for osteoarthritis through genome-wide analyses of UK Biobank data. Nat Genet. 2019;51: 230–236. doi:10.1038/s41588-018-0327-1

59. Li D, Achkar J-P, Haritunians T, Jacobs JP, Hui KY, D'Amato M, et al. A pleiotropic missense variant in SLC39A8 is associated with Crohn's disease and human gut microbiome composition. Gastroenterology. 2016;151: 724–732. doi:10.1053/j.gastro.2016.06.051

60. Carrera N, Arrojo M, Sanjuán J, Ramos-Ríos R, Paz E, Suárez-Rama JJ, et al. Association study of nonsynonymous single nucleotide polymorphisms in schizophrenia. Biol Psychiatry. 2012;71: 169–177. doi:10.1016/j.biopsych.2011.09.032

61. Deckers MM, Smits P, Karperien M, Ni J, Tylzanowski P, Feng P, et al. Recombinant human extracellular matrix protein 1 inhibits alkaline phosphatase activity and mineralization of mouse embryonic metatarsals in vitro. Bone. 2001;28: 14–20. doi:10.1016/S8756-3282(00)00428-2

62. Kong L, Zhao Y-P, Tian Q-Y, Feng J-Q, Kobayashi T, Merregaert J, et al. Extracellular matrix protein 1, a direct targeting molecule of parathyroid hormone-related peptide, negatively regulates chondrogenesis and endochondral ossification via associating with progranulin growth factor. FASEB J. 2016;30: 2741–2754. doi:10.1096/fj.201600261R

63. Kong L, Tian Q, Guo F, Mucignat MT, Perris R, Sercu S, et al. Interaction between cartilage oligomeric matrix protein and extracellular matrix protein 1 mediates endochondral bone growth. Matrix Biol. 2010;29: 276–286. doi:10.1016/j.matbio.2010.01.007

64. Fisher SA, Tremelling M, Anderson CA, Gwilliam R, Bumpstead S, Prescott NJ, et al. Genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease. Nat Genet. 2008;40: 710–712. doi:10.1038/ng.145

65. Festen EA, Stokkers PC, van Diemen CC, van Bodegraven AA, Boezen HM, Crusius BJ, et al. Genetic analysis in a Dutch study sample identifies more ulcerative colitis susceptibility loci and shows their additive role in disease risk. Am J Gastroenterol. 2010;105: 395–402. doi:10.1038/ajg.2009.576

66. Adali G, Ersoy Tunali N, Yorulmaz E, Tiryakioglu NO, Guray Mungan S, Ulasoglu C, et al. Extracellular matrix protein 1 gene rs3737240 single nucleotide polymorphism is associated with ulcerative colitis in Turkish patients. Turkish J Gastroenterol. 2017;28: 254–259. doi:10.5152/tjg.2017.17043

67. Egli RJ, Southam L, Wilkins JM, Lorenzen I, Pombo-Suarez M, Gonzalez A, et al. Functional analysis of the osteoarthritis susceptibility-associated *GDF5* regulatory polymorphism. Arthritis Rheum. 2009;60: 2055–2064. doi:10.1002/art.24616

68. Miyamoto Y, Mabuchi A, Shi D, Kubo T, Takatori Y, Saito S, et al. A functional polymorphism in the 5′ UTR of GDF5 is associated with susceptibility to osteoarthritis. Nat Genet. 2007;39: 529–533. doi:10.1038/2005

69. Southam L, Rodriguez-Lopez J, Wilkins JM, Pombo-Suarez M, Snelling S, Gomez-Reino JJ, et al. An SNP in the 5′-UTR of GDF5 is associated with osteoarthritis susceptibility in

Europeans and with in vivo differences in allelic expression in articular cartilage. Hum Mol Genet. 2007;16: 2226–2232. doi:10.1093/hmg/ddm174

70. Chapman K, Takahashi A, Meulenbelt I, Watson C, Rodriguez-Lopez J, Egli R, et al. A meta-analysis of European and Asian cohorts reveals a global role of a functional SNP in the 5' UTR of GDF5 with osteoarthritis susceptibility. Hum Mol Genet. 2008;17: 1497–1504. doi:10.1093/hmg/ddn038

71. Evangelou E, Chapman K, Meulenbelt I, Karassa FB, Loughlin J, Carr A, et al. Large-scale analysis of association between *GDF5* and *FRZB* variants and osteoarthritis of the hip, knee, and hand. Arthritis Rheum. 2009;60: 1710–1721. doi:10.1002/art.24524

72. Williams FMK, Popham M, Hart DJ, de Schepper E, Bierma-Zeinstra S, Hofman A, et al. GDF5 single-nucleotide polymorphism rs143383 is associated with lumbar disc degeneration in Northern European women. Arthritis Rheum. 2011;63: 708–712. doi:10.1002/art.30169

73. Dai J, Shi D, Zhu P, Qin J, Ni H, Xu Y, et al. Association of a single nucleotide polymorphism in growth differentiate factor 5 with congenital dysplasia of the hip: a case-control study. Arthritis Res Ther. 2008;10: R126. doi:10.1186/ar2540

74. Rouault K, Scotet V, Autret S, Gaucher F, Dubrana F, Tanguy D, et al. Evidence of association between GDF5 polymorphisms and congenital dislocation of the hip in a Caucasian population. Osteoarthr Cartil. 2010;18: 1144–1149. doi:10.1016/j.joca.2010.05.018

75. Komori K, Nonaka T, Okada A, Kinoh H, Hayashita-Kinoh H, Yoshida N, et al. Absence of mechanical allodynia and Abeta-fiber sprouting after sciatic nerve injury in mice lacking membrane-type 5 matrix metalloproteinase. FEBS Lett. 2004;557: 125–128. doi:10.1016/s0014-5793(03)01458-3

76. Liou J-T, Sum DC-W, Liu F-C, Mao C-C, Lai Y-S, Day Y-J. Spatial and temporal analysis of nociception-related spinal cord matrix metalloproteinase expression in a murine neuropathic pain model. J Chinese Med Assoc. 2013;76: 201–210. doi:10.1016/J.JCMA.2012.12.011

77. Clark AK, Yip PK, Grist J, Gentry C, Staniland AA, Marchand F, et al. Inhibition of spinal microglial cathepsin S for the reversal of neuropathic pain. Proc Natl Acad Sci. 2007;104: 10655–10660. doi:10.1073/pnas.0610811104

78. Clark AK, Malcangio M. Fractalkine/CX3CR1 signaling during neuropathic pain. Front Cell Neurosci. 2014;8: 121. doi:10.3389/fncel.2014.00121

79. Irie O, Kosaka T, Ehara T, Yokokawa F, Kanazawa T, Hirao H, et al. Discovery of orally bioavailable cathepsin S inhibitors for the reversal of neuropathic pain. J Med Chem. 2008;51: 5502–5505. doi:10.1021/jm800839j

80. Barclay J, Clark AK, Ganju P, Gentry C, Patel S, Wotherspoon G, et al. Role of the cysteine protease cathepsin S in neuropathic hyperalgesia. Pain. 2007;130: 225–234. doi:10.1016/j.pain.2006.11.017

81. Zhang X, Wu Z, Hayashi Y, Okada R, Nakanishi H. Peripheral role of cathepsin S in Th1 cell-dependent transition of nerve injury-induced acute pain to a chronic pain state. J Neurosci. 2014;34: 3013–3022. doi:10.1523/JNEUROSCI.3681-13.2014

82. Kallenborn-Gerhardt W, Schröder K, Geisslinger G, Schmidtko A. NOXious signaling in pain processing. Pharmacol Ther. 2013;137: 309–317. doi:10.1016/j.pharmthera.2012.11.001

83. Ahmed Z, Douglas MR, John G, Berry M, Logan A. AMIGO3 is an NgR1/p75 co-receptor signalling axon growth inhibition in the acute phase of adult central nervous system injury. PLoS One. 2013;8: e61878. doi:10.1371/journal.pone.0061878

84. Almutiri S, Berry M, Logan A, Ahmed Z. Non-viral-mediated suppression of AMIGO3 promotes disinhibited NT3-mediated regeneration of spinal cord dorsal column axons. Sci Rep. 2018;8: 10707. doi:10.1038/s41598-018-29124-z

85. Pregizer SK, Kiapour AM, Young M, Chen H, Schoor M, Liu Z, et al. Impact of broad regulatory regions on Gdf5 expression and function in knee development and susceptibility to osteoarthritis. Ann Rheum Dis. 2018;77: 450–450. doi:10.1136/annrheumdis-2017-212475

86. Chen H, Capellini TD, Schoor M, Mortlock DP, Reddi AH, Kingsley DM. Heads, shoulders, elbows, knees, and toes: Modular Gdf5 enhancers control different joints in the vertebrate skeleton. PLOS Genet. 2016;12: e1006454. doi:10.1371/journal.pgen.1006454
87. Song J, Kim D, Lee CH, Lee MS, Chun C-H, Jin E-J. MicroRNA-488 regulates zinc transporter SLC39A8/ZIP8 during pathogenesis of osteoarthritis. J Biomed Sci. 2013;20: 31. doi:10.1186/1423-0127-20-31
88. Kim J-H, Jeon J, Shin M, Won Y, Lee M, Kwak J-S, et al. Regulation of the catabolic cascade in osteoarthritis by the zinc-ZIP8-MTF1 axis. Cell. 2014;156: 730–743. doi:10.1016/j.cell.2014.01.007
89. Kolhe R, Hunter M, Liu S, Jadeja RN, Pundkar C, Mondal AK, et al. Gender-specific differential expression of exosomal miRNA in synovial fluid of patients with osteoarthritis. Sci Rep. 2017;7: 2029. doi:10.1038/s41598-017-01905-y
90. Lai N-S, Yu H-C, Tung C-H, Huang K-Y, Huang H-B, Lu M-C. Aberrant expression of interleukin-23-regulated miRNAs in T cells from patients with ankylosing spondylitis. Arthritis Res Ther. 2018;20: 259. doi:10.1186/s13075-018-1754-1
91. Altrock WD, tom Dieck S, Sokolov M, Meyer AC, Sigler A, Brakebusch C, et al. Functional inactivation of a fraction of excitatory synapses in mice deficient for the active zone protein bassoon. Neuron. 2003;37: 787–800. doi:10.1016/s0896-6273(03)00088-6
92. Spilker C, Grochowska KM, Kreutz MR. What do we learn from the murine *Jacob/Nsmf* gene knockout for human disease? Rare Dis. 2016;4: e1241361. doi:10.1080/21675511.2016.1241361
93. Kramer PR, Wray S. Novel gene expressed in nasal region influences outgrowth of olfactory axons and migration of luteinizing hormone-releasing hormone (LHRH) neurons. Genes Dev. 2000;14: 1824–1834. doi:10.1101/gad.14.14.1824
94. Hasan MT, Hernández-González S, Dogbevia G, Treviño M, Bertocchi I, Gruart A, et al. Role of motor cortex NMDA receptors in learning-dependent synaptic plasticity of behaving mice. Nat Commun. 2013;4: 2258. doi:10.1038/ncomms3258
95. Vargha-Khadem F, Gadian DG, Copp A, Mishkin M. FOXP2 and the neuroanatomy of speech and language. Nat Rev Neurosci. 2005;6: 131–138. doi:10.1038/nrn1605
96. Konopka G, Bomar JM, Winden K, Coppola G, Jonsson ZO, Gao F, et al. Human-specific transcriptional regulation of CNS development genes by FOXP2. Nature. 2009;462: 213–217. doi:10.1038/nature08549
97. Glahn DC, Curran JE, Winkler AM, Carless MA, Kent JW, Charlesworth JC, et al. High dimensional endophenotype ranking in the search for major depression risk genes. Biol Psychiatry. 2012;71: 6–14. doi:10.1016/j.biopsych.2011.08.022
98. Suhre K, Arnold M, Bhagwat AM, Cotton RJ, Engelke R, Raffler J, et al. Connecting genetic risk to disease end points through the human blood plasma proteome. Nat Commun. 2017;8: 14357. doi:10.1038/ncomms14357
99. Suhre K, Arnold M, Bhagwat AM, Cotton RJ, Engelke R, Raffler J, et al. Connecting genetic risk to disease end points through the human blood plasma proteome. Nat Commun. 2017;8: 14357. doi:10.1038/ncomms14357
100. Okifuji A, Hare BD. The association between chronic pain and obesity. J Pain Res. 2015;8: 399–408. doi:10.2147/JPR.S55598
101. Gasperi M, Herbert M, Schur E, Buchwald D, Afari N. Genetic and environmental influences on sleep, pain, and depression symptoms in a community sample of twins. Psychosom Med. 2017;79: 646–654. doi:10.1097/PSY.0000000000000456
102. Reichborn-Kjennerud T, Stoltenberg C, Tambs K, Roysamb E, Kringlen E, Torgersen S, et al. Back-neck pain and symptoms of anxiety and depression: a population-based twin study. Psychol Med. 2002;32: 1009–1020. doi:10.1017/s0033291702005950
103. McIntosh AM, Hall LS, Zeng Y, Adams MJ, Gibson J, Wigmore E, et al. Genetic and environmental risk for chronic pain and the contribution of risk variants for major depressive disorder: A Family-Based Mixed-Model Analysis. PLoS Med. 2016;13: e1002090.

1101 doi:10.1371/journal.pmed.1002090
1102 104. Freidin MB, Tsepilov YA, Palmer M, Karssen LC, Suri P, Aulchenko YS, et al. Insight into
1103 the genetic architecture of back pain and its risk factors from a study of 509,000 individuals.
1104 Pain. 2019;160: 1361-1373. doi:10.1097/j.pain.0000000000001514
1105 105. Ning Z, Tsepilov YA, Sharapov SZ, Grishenko AK, Feng X, Shirali M, et al. Beyond power:
1106 Multivariate discovery, replication, and interpretation of pleiotropic loci using summary
1107 association statistics. bioRxiv. 2019; 022269. doi:10.1101/022269 [PREPRINT]
1108 106. O'Reilly PF, Hoggart CJ, Pomyen Y, Calboli FCF, Elliott P, Jarvelin M-R, et al. MultiPhen:
1109 Joint model of multiple phenotypes can increase discovery in GWAS. PLoS One. 2012;7:
1110 e34861. doi:10.1371/journal.pone.0034861
1111 107. Valdes AM, Doherty SA, Zhang W, Muir KR, Maciewicz RA, Doherty M. Inverse
1112 relationship between preoperative radiographic severity and postoperative pain in patients
1113 with osteoarthritis who have undergone total joint arthroplasty. Semin Arthritis Rheum.
1114 2012;41: 568–575. doi:10.1016/j.semarthrit.2011.07.002
1115 108. Johnston KJA, Adams MJ, Nicholl BI, Ward J, Strawbridge RJ, Ferguson A, et al. Genome-
1116 wide association study of multisite chronic pain in UK Biobank. PLoS Genet. 2019;15:
1117 e1008164. doi:10.1371/journal.pgen.1008164
1118