



Article Multi-Trait Exome-Wide Association Study of Back Pain-Related Phenotypes

Irina V. Zorkoltseva¹, Elizaveta E. Elgaeva^{1,2}, Nadezhda M. Belonogova¹, Anatoliy V. Kirichenko¹, Gulnara R. Svishcheva^{1,3}, Maxim B. Freidin⁴, Frances M. K. Williams⁵, Pradeep Suri^{6,7,8,9}, Yakov A. Tsepilov¹ and Tatiana I. Axenovich^{1,*}

- ¹ Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, 630090 Novosibirsk, Russia; zor@bionet.nsc.ru (I.V.Z.); elgaeva@bionet.nsc.ru (E.E.E.); belon@bionet.nsc.ru (N.M.B.); kianvl@bionet.nsc.ru (A.V.K.); gulsvi@mail.ru (G.R.S.); tsepilov@bionet.nsc.ru (Y.A.T.)
- ² Department of Natural Sciences, Novosibirsk State University, 630090 Novosibirsk, Russia
- ³ Vavilov Institute of General Genetics, Russian Academy of Sciences, 119333 Moscow, Russia
- ⁴ Department of Biology, School of Biological and Behavioural Sciences, Queen Mary University of London, London EC1M 6BQ, UK; maxim.freydin@kcl.ac.uk
- ⁵ Department of Twin Research and Genetic Epidemiology, King's College London, London SE1 7EH, UK; frances.williams@kcl.ac.uk
- ⁶ Seattle Epidemiologic Research and Information Center, VA Puget Sound Health Care System, Seattle, WA 98108, USA
- ⁷ Division of Rehabilitation Care Services, Seattle, WA 98208, USA
- ⁸ Clinical Learning, Evidence, and Research Center, University of Washington, Seattle, WA 98195, USA
- ⁹ Department of Rehabilitation Medicine, University of Washington, Seattle, WA 98195, USA
- * Correspondence: aks@bionet.nsc.ru

Abstract: Back pain (BP) is a major contributor to disability worldwide, with heritability estimated at 40–60%. However, less than half of the heritability is explained by common genetic variants identified by genome-wide association studies. More powerful methods and rare and ultra-rare variant analysis may offer additional insight. This study utilized exome sequencing data from the UK Biobank to perform a multi-trait gene-based association analysis of three BP-related phenotypes: chronic back pain, dorsalgia, and intervertebral disc disorder. We identified the *SLC13A1* gene as a contributor to chronic back pain via loss-of-function (LoF) and missense variants. This gene has been previously detected in two studies. A multi-trait approach uncovered the novel *FSCN3* gene and its impact on back pain through LoF variants. This gene deserves attention because it is only the second gene shown to have an effect on back pain due to LoF variants and represents a promising drug target for back pain therapy.

Keywords: linear combination of traits; chronic back pain; dorsalgia; intervertebral disc disorder; *SLC13A1*; *FSCN3*; shared heritability; loss-of-function (LoF) variant; rare and ultra-rare genetic variants

1. Introduction

Back pain (BP) is the leading cause of disability worldwide [1,2]. Several BP-related phenotypes are commonly used for genetic analysis, with chronic back pain (CBP) being the most severe form, defined as self-reported BP lasting more than 3 months, and the most commonly studied to date [3–6]. Dorsalgia is an electronic health record (EHR)-based phenotype defined by clinical diagnostic codes for back pain and neck pain, but predominantly reflects the former. Intervertebral disc disorder (IDD) is another BP-related phenotype defined by EHR-based codes for intervertebral disc degeneration that are typically (but not necessarily) used when BP is present [7,8]. Several dozen loci associated with CBP, dorsalgia, and IDD have been identified by genome-wide association studies



Citation: Zorkoltseva, I.V.; Elgaeva, E.E.; Belonogova, N.M.; Kirichenko, A.V.; Svishcheva, G.R.; Freidin, M.B.; Williams, F.M.K.; Suri, P.; Tsepilov, Y.A.; Axenovich, T.I. Multi-Trait Exome-Wide Association Study of Back Pain-Related Phenotypes. *Genes* 2023, 14, 1962. https://doi.org/ 10.3390/genes14101962

Academic Editor: Wenjian Bi

Received: 21 September 2023 Revised: 16 October 2023 Accepted: 18 October 2023 Published: 19 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). (GWAS) [7–10]. Genes associated with BP-related phenotypes are involved in cartilage and bone biology, neurological, and inflammatory processes, with some genes being associated with more than one phenotype [8]. In the largest genetic study of BP-related phenotypes, six loci marked by the *COL11A1*, *SPON2*, *C6orf106*, *GSDMC*, *CHST3*, and *FOXA3* genes were shown to affect both IDD and dorsalgia [8]. The *SOX5* and *PANX3* genes have been associated with CBP [9,11] and IDD [8]. In the current report, we use the term "back pain-related phenotypes" to refer collectively to these three different yet related phenotypes associated with back pain.

All BP-related traits are complex heritable traits with narrow-sense heritability estimated from twin studies in the range of 30% to 70% [12–15]. Liability-scale SNP-based heritability explained by common variants has been estimated at 12% for CBP [16], 11.4% for dorsalgia (https://nealelab.github.io/UKBB_ldsc/h2_summary_M54.html, accessed on 1 November 2022), and 8% for IDD (https://nealelab.github.io/UKBB_ldsc/h2_summary_ M51.html, accessed on 1 November 2022). These estimations are substantially lower than the heritability expected from classical twin studies. Several ways can be suggested to explain this missing heritability, such as using rare (minor allele frequency, MAF < 0.01) variants [17] and more powerful methods of association analysis.

Analysis of rare variants in order to estimate their impact on complex traits is now possible due to whole-exome sequencing technology. Exome sequencing data have already been used for the analysis of CBP, and the association with loss-of-function (LoF) variants in the SLC13A1 gene was detected [18].

In this study, we also used the exome sequencing data from the UK biobank project that has been used by Ao et al. (2023). However, our and the cited studies differ in materials and methods. Ao et al. used only one BP-related phenotype, namely, CBP, but we considered three different BP-related phenotypes: CBP, dorsalgia, and IDD. The lack of a standard definition of BP poses a significant challenge in identifying new genetic loci for BP. To address these challenges and enhance our understanding of the genetic influence on BP, we used three BP-related phenotypes. The genetic nature of all these phenotypes is common because the genetic correlations between them are about 90% [6,8].

We applied a new method of multi-trait association analysis [19]. This method combines several genetically correlated traits into a single multi-trait by maximizing its heritability. This approach allows the identification of the common genetic background for several genetically correlated traits and increases the power of association analysis due to the lower genetic heterogeneity of multi-trait in comparison with the original traits. It has been shown that such analysis can identify genes that were undetected by the separate analysis of each original trait [6,19].

In this study, our aim was to analyze the genes responsible for differences in the risk of BP-related phenotypes among individuals. We achieved this by examining rare and ultra-rare variants using exome sequencing data from the UK Biobank. To increase the power of the analysis, we employed a multi-trait approach.

2. Materials and Methods

2.1. Overview of the Study Design

This study was carried out using the UK Biobank 200k exome dataset (https://biobank. ndph.ox.ac.uk/ukb/field.cgi?id=23156, accessed on 14 November 2021) and phenotypes of three BP-related traits: CBP, dorsalgia, and IDD (https://biobank.ndph.ox.ac.uk/ukb/ field.cgi?id=41270, accessed on 3 November 2022).

The study design included three stages. During the first stage, we calculated the summary statistics (z-scores, effect sizes, *p*-values) using an exome-wide association study (EWAS) for three BP-related traits and for a new multi-trait, which condenses the genetic background common for all original traits analyzed. In the second stage, we performed gene-based association analyses of all original traits and multi-trait. In the third stage, we performed a conditional analysis of the identified genes to reduce the influence of strong association signals located outside the gene.

2.2. Data

We used information about 188,237 white participants from the UK Biobank 200k exome dataset. Information on self-reported ethnic background was captured in data field 21000. Three BP-related phenotypes were analyzed: CBP (N = 178,963), dorsalgia (N = 183,652), and IDD (N = 183,652). The characteristics of the patients with the different BP-related phenotypes are presented in Supplementary Table S1. The prevalence of three BP-related phenotypes was 17.97% for CBP, 3.48% for dorsalgia, and 1.76% for IDD in a large sample of white people from the UK Biobank (N = 459,161). Many individuals have two or three of these BP-related phenotypes simultaneously (Supplementary Figure S1). In particular, 68% of individuals with IDD also have CBP and/or dorsalgia.

For CBP, cases and controls were defined based on questionnaire responses (data field 3571 "Back pain for 3+ months"). 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 pain lasting more than 3 months were considered chronic back pain cases, while participants reporting no such pain were considered controls. Individuals who preferred not to answer or reported more than 3 months of pain all over the body were excluded from the study.

For dorsalgia and IDD, EHR-based phenotypes were defined by the level 2 ICD-10 codes M54 and M51, respectively. All participants with either of these level 2 medical codes (M54 and M51 and/or codes of a lower level (level 3 and/or 4, which are subsumed under the level 2 code)) were considered as cases. For an overview of ICD-10 codes in the UK Biobank, see https://biobank.ndph.ox.ac.uk/ukb/field.cgi?id=41202 (accessed on 3 November 2022).

We used exome sequencing data from the UK Biobank 200k release. We carried out quality control as advised in https://biobank.ndph.ox.ac.uk/showcase/refer.cgi?id=914 (accessed on 3 November 2022). We used BCFtools 1.16.1 (https://samtools.github.io/bcftools/, accessed on 5 November 2022) to filter single nucleotide variants (SNV) directly on the multi-sample VCF (data field 23156). After quality control, there were 11,375,237 autosomal SNVs with MAF < 0.01.

2.3. EWAS of Original Traits

A single-variant association analysis for each BP-related trait was performed using the fastGWA-GLMM tool, version 1.94.0 beta [20]. We conducted the analysis using a sparse genomic relationship matrix reflecting genomic relatedness between individuals to account for random effects mediated by relationships between individuals. This matrix was calculated using fastGWA-GLMM (-make-bK-sparse option with default parameters) for the whole UK Biobank sample (N = 487,000). EWAS was performed using the –fastGWAmlm-binary option and filters, keeping SNVs with MAF 5 × 10⁻⁶ and an SNV messiness rate of 0.02. We included sex, age, batch, and the first ten genetic principal components provided by UK Biobank as covariates in the association analysis to correct for fixed effects. The summary statistics for exome data were calculated on the GRCh38/hg38 genomic build and included 7,890,268 SNVs for CBP and 7,998,110 SNVs for dorsalgia and IDD.

2.4. EWAS of Multi-Trait

We used the SHAHER framework, which aims to construct a multi-trait from a group of genetically correlated traits and identify genetic variants controlling this trait [19]. The method suggests that the genetic background of each of the genetically correlated traits can be decomposed into two components: one common to all traits (the shared genetic impact, or SGI) and one specific for each given trait. To find shared genetic variants that explain SGI, a new trait called the shared genetic impact trait (SGIT) is created as a linear combination of the original traits. The coefficients in this combination are defined by maximizing the proportion of SGI in the genetic background of SGIT. To estimate the SGIT coefficients, we utilized the heritability estimates, phenotypic correlation matrices, and genetic correlation matrices assessed beforehand for the original BP-related traits. Then, EWAS summary statistics for SGIT were calculated using the summary statistics calculated for the original traits and the SGIT coefficients. In particular, the z-score for SGIT was calculated as a weighted sum of z-scores for individual traits. The scheme of the SHAHER analysis of the BP-related traits is shown in Scheme 1. The SHAHER method has been described by Svishcheva et al. [19]. The current SHAHER analysis is described in detail in the Supplementary Methods.



Scheme 1. Calculation of EWAS summary statistics for the original traits and multi-trait.

Pairwise phenotypic correlations between the three original BP traits were assessed in a subsample of non-relatives (307,876 white individuals with CBP, dorsalgia, and IDD status information). Phenotypic correlations of SGIT with CBP, dorsalgia, and IDD were evaluated using summary-level data from imputed genotypes according to the approach proposed by Stephens et al. [21]. We estimated both genetic correlations between the traits and the trait's SNP-based heritabilities utilizing the LD Score regression tool [22] and summary statistics estimated on imputed variant data.

2.5. Gene-Based Association Analyses

For gene-based association analysis, EWAS summary statistics (z-scores and effect sizes) and the matrices of correlations between genotypes of all SNVs within a gene were used.

2.5.1. Matrices of Genotype Correlations

The correlation between every pair of variants within a gene was estimated using the sequencing genotypes of the 129,807 unrelated white UK Biobank participants (data field 22021) by plink v2.00a3.7LM (https://www.cog-genomics.org/plink/2.0/, accessed on 30 October 2022) with options –mac 3 –geno 0.02.

2.5.2. Variant Annotations

We performed a gene-based association analysis for each of four different variant annotations: LoF, LoF + missense, LoF + protein coding, and all intragenic. Variants were annotated using the Ensembl Variant Effect Predictor (VEP) [23].

The LoF variants include those annotated as frameshift variant, splice acceptor variant, splice donor variant, start loss, stop gain, stop loss, and transcript ablation. The LoF + missense variants additionally include those annotated as transcript amplification, inframe insertion, inframe deletion, missense variant, and protein-altering variant. The LoF + protein coding variants additionally include those annotated as synonymous variant, start retained variant, stop retained variant, coding sequence variant, and incomplete terminal codon variant. The all intragenic variants annotation includes all variants within a gene from 5'UTR to 3'UTR.

For gene-based analysis, we used SNVs matching the following criteria: $MAF \le 0.01$ and minor allele count (MAC) > 10. The genotypes of ultra-rare variants with $MAC \le 10$ were collapsed into a single variant (see below). The numbers of collapsed and other non-ultra-rare (MAC > 10) variants are shown in Supplementary Table S2.

Additionally, we used LoF burden tests of association for rare predicted LoF variants [24]. We collapsed all LoF variants within a gene into a single variant using the same technique as for collapsing ultra-rare variants (see below).

2.5.3. Collapsing

For the analysis of ultra-rare variants, we used a technique proposed by Zhou et al. [25]. Before testing each variant annotation, we collapsed the genotypes of variants with MAC \leq 10 (MAC cutoff details are in Supplementary Methods). In short, the procedure is as follows: individuals with a homozygous rare allele in at least one of the collapsing variants were given a genotype coding of 2, while those with a heterozygous rare allele in at least one of the collapsing variants were given a genotype coding variants were given a genotype coding of 1. Individuals with no rare allele in any of the collapsing variants were given a genotype coding of 0. We calculated the summary statistics for collapsed genotypes using the same method as for EWAS and then used the collapsed variants together with all other variants with MAC > 10 for gene-based analysis.

2.5.4. Methods of Gene-Based Analysis

The sumSTAAR framework was used for the gene-based association analysis [26]. Two regression-based methods based on the summary statistics were applied: SKAT-O [27] and PCA [28]. These methods were implemented in the sumFREGAT R-package [29]. The SKAT-O and PCA results were combined using the aggregated Cauchy omnibus test, ACAT-O [30]. The analysis was limited to protein-coding genes with at least two variations that had the EWAS summary statistics.

The Bonferroni adjusted significance and suggestive levels for the total number of genes (20,000) were defined as 2.5×10^{-6} and 2.5×10^{-5} , respectively.

2.6. Conditional Analysis

Conditional analysis was performed using the GCTA-COJO tool. [31,32] The COJO selection procedure was applied to SNVs located within 1 Mb of the gene borders. The threshold *p*-value for index variants (–cojo-p) was set as the minimum *p*-value within the gene region being tested. Independent variants selected at this stage were considered conditional if they were located outside of the analyzed gene region. Conditional summary statistics (*p*-values, betas) calculated for the variants within the analyzed gene region were then used as input for gene-based analysis. To take intergenic variants into account, we merged the genotypes and summary statistics of exome data with those of imputed data. CBP summary statistics were calculated for imputed SNVs using the same sample. For SGIT, GWAS summary statistics were calculated on imputed genotypes using the whole sample of white British individuals (N = 449,136).

3. Results

3.1. Single Variants Association Analysis

We performed EWAS for the three original traits using all genetic variants that passed quality control. The EWAS summary statistics are available at https://mga.icgbio.ru/BP_phens_ewas/ (accessed on 13 October 2023). Manhattan and QQ plots are presented in Supplementary Figure S2. No significant association signal was detected. Only one variant showed an association at the suggestive level of significance: an ultra-rare variant, 12:14,507,063 (*p*-value = 2.74×10^{-7} ; MAF = 9×10^{-5}), in the PLBD1 gene for CBP.

To identify genetic variants shared by BP-related traits, we used the SHAHER approach. Using the phenotypic and genetic correlations between the original traits and their SNP-based heritability (Figure 1, first three columns and rows), we estimated the coefficients of an optimal linear combination of the original traits to build SGIT as 0.69, 0.44, and 0.31 for CBP, dorsalgia, and IDD, respectively. Using these coefficients and EWAS summary statistics for the original BP-related traits, we calculated EWAS summary statistics for SGIT. Manhattan and QQ plots are presented in Supplementary Figure S2. There were three

variants significantly associated with SGIT and ten variants suggestively associated with SGIT. However, none of these variants was included in the gene-based analysis because some of them were located outside the genes, while others had MAC \leq 10 and collapsed with other ultra-rare variants.



Figure 1. SNP-based heritability (diagonal elements), phenotypic correlations (upper triangular elements), and genetic correlations (lower triangular elements).

Then, we calculated the phenotypic and genetic correlations between the original traits and SGIT and estimated the SNP-based heritability of SGIT (Figure 1, fourth column and row). As can be seen, the phenotypic correlations between SGIT and the original traits are stronger than those between the original traits. All genetic correlations between SGIT and the original traits are no less than 0.93. The SNP-based heritability of SGIT is higher than that of original traits.

3.2. Gene-Based Association Analysis

We formed the four different variant annotations (LoF, LoF + missense, LoF + protein coding, and all intragenic) within every gene. Then, we collapsed variants with MAC \leq 10 and calculated their summary statistics. Association results for collapsed variants are freely available in the ZENODO database (https://doi.org/10.5281/zenodo.8118630, accessed on 31 March 2023). No significant or suggestive association signals were detected for the collapsed variants.

The results of gene-based association analysis with *p*-value $\leq 2.5 \times 10^{-5}$ are shown in Table 1. Full results are available in the ZENODO database (https://doi.org/10.5281/zenodo.8118630, accessed on 31 March 2023). QQ plots for gene-based analyses are shown in Supplementary Figure S3.

Trait	Variant Annotation	Gene	#MAC ^a >10	#MAC ^a \leq 10	SKATO ^b	PCA ^b	ACAT-O ^b
CBP ^c	LoF + missense	P4HTM	29	149	$6.82 imes 10^{-6}$	$6.35 imes 10^{-3}$	$1.36 imes 10^{-5}$
	LoF + protein coding	P4HTM	45	203	$1.45 imes10^{-7}$	$9.95 imes10^{-3}$	$2.89 imes10^{-7}$
	LoF + missense	SLC13A1	34	227	$1.56 imes10^{-3}$	$1.25 imes 10^{-6}$	$2.50 imes 10^{-6}$
	LoF + protein coding	KIF5B	33	302	$3.91 imes 10^{-3}$	$9.74 imes10^{-6}$	$1.94 imes10^{-5}$
	LoF + missense	TIMM44	19	134	$1.82 imes 10^{-1}$	$5.90 imes 10^{-6}$	$1.18 imes 10^{-5}$
Dorsalgia	All intragenic	RPL37	27	166	$1.07 imes 10^{-5}$	$4.28 imes10^{-2}$	$2.14 imes10^{-5}$
	LoF + missense	AVPR1A	12	174	$8.95 imes 10^{-6}$	$2.24 imes10^{-4}$	$1.72 imes 10^{-5}$
IDD ^d	LoF + missense	RNF115	12	105	$1.04 imes 10^{-5}$	$1.85 imes 10^{-3}$	$2.07 imes 10^{-5}$
	LoF + protein coding	KIAA2012	85	518	$3.99 imes10^{-6}$	$5.75 imes 10^{-2}$	$7.99 imes10^{-6}$
	All intragenic	LETMD1	37	334	$5.15 imes10^{-6}$	$5.78 imes10^{-4}$	$1.02 imes 10^{-5}$
SGIT ^e	LoF	FSCN3	5	10	$2.94 imes10^{-7}$	$6.44 imes 10^{-6}$	$5.62 imes 10^{-7}$
	LoF burden ^f	IL1R2		29 ^g			$1.14 imes10^{-5\mathrm{h}}$

Table 1. Gene-based association signals (*p*-value $\leq 2.5 \times 10^{-5}$) obtained for different traits and variant annotations.

^a Minor allele count; ^b *p*-values obtained by SKAT-O and PCA methods and combined by ACAT-O; ^c Chronic back pain; ^d Intervertebral disc disorder; ^e Linear combinations of CBP, IDD, and dorsalgia; ^f All LoF variants within gene were collapsed and analyzed as a single variant; ^g Number of LoF variants in gene; ^h *p*-value obtained for collapsed variant.

Only one of the three original traits, namely CBP, showed a significant gene-based association with two genes: *P4HTM* (LoF + protein coding) and *SLC13A1* (LoF + missense). *P4HTM* also showed a suggestive association signal for CBP on another variant annotation (LoF + missense). A few genes provided associations at the suggestive significance level: *KIF5B* and *TIMM44* for CBP; *RPL37* and *AVPR1A* for dorsalgia; and *RNF15*, *KIAA2012*, and *LETMD1* for IDD. For the LoF burdens, there was no association with a *p*-value $\leq 2.5 \times 10^{-5}$.

The gene-based association analysis of SGIT detected a significant association for the *FSCN3* gene (LoF annotation) (Table 1). Table 2 shows the *p*-values estimated for this gene on SGIT and the original traits. As can be seen, these values are low for all traits, especially for the LoF annotation.

Annotation	СВР	Dorsalgia	IDD	SGIT
LoF	$6.4 imes10^{-5}$	$7.1 imes10^{-4}$	$5.1 imes 10^{-2}$	$5.6 imes10^{-7}$
LoF + missense	$1.3 imes10^{-3}$	$1.6 imes10^{-2}$	$6.9 imes10^{-2}$	$9.2 imes10^{-5}$
LoF + protein coding	$7.2 imes10^{-3}$	$5.7 imes10^{-2}$	$1.7 imes10^{-2}$	$3.3 imes10^{-4}$
All intragenic	$3.8 imes10^{-3}$	$1.5 imes10^{-1}$	$7.4 imes10^{-1}$	$1.0 imes 10^{-3}$

Table 2. FSCN3 p-values for different trait-annotation combinations.

We also performed the gene-based association analysis of SGIT using the LoF burden test. No significant association was found. The *IL1R2* gene showed an association at the suggestive level (*p*-value = 1.14×10^{-5} , Table 1). The *p*-values estimated for this gene and the original traits were 6.4×10^{-5} , 0.09, and 0.004 for CBP, dorsalgia, and IDD, respectively.

3.3. Conditional Analysis

We performed conditional gene-based association analysis for *SLC13A1* (CBP, LoF + missense), *P4HTM* (CBP, LoF + protein coding), and *FSCN3* (SGIT, LoF). *SLC13A1* and *FSCN3* harbored SNVs with the lowest *p*-values within 1 Mb from their borders; thus, no conditional SNVs were selected, and the gene-based *p*-values remained unchanged. For *P4HTM*, the strongest signal within the gene (*p*-value = 0.0037) was in high LD with an imputed SNV with the *p*-value = 0.0009. As a result, the conditional gene-based *p*-value for *P4HTM* was 0.001.

Next, we checked the influence of the genome-wide significant SNVs already known from prior studies for CBP and related traits within 1 Mb from gene borders on the association of *SLC13A1*, *P4HTM*, and *FSCN3*. We did not find such genome-wide significant

SNVs for *FSCN3*. The *SLC13A1* gene contained only one genome-wide significant SNV, rs28364172 [8]; no such variants were outside the gene. It has been shown that multiple significant GWAS signals are located around the *P4HTM* gene: rs34762726 [8], rs1491985 [33], and 14 SNVs found in the Finngen study for pain, lower back pain, dorsalgia, and other dorsopathies (https://r8.finngen.fi/, accessed on 10 December 2022). Of these 16 SNVs, 10 were present in our sample. However, their r² with the top two SNVs within the *P4HTM* gene were rather small and did not exceed 0.02. This was not surprising, as most of the GWAS signals were obtained on common variants. Being used as conditional SNVs, they did not affect the significance of *P4HTM* gene-based statistics.

4. Discussion

We performed exome-wide gene-based analysis using the UK biobank data and detected two genes, *SLC13A1* and *FSCN3*, that were significantly associated with BP-related traits. The first gene was associated with CBP, and the second gene was associated with SGIT.

The *SLC13A1* gene has been previously detected as being associated with BP-related traits due to LoF variants [8,18]. This gene encodes a protein that functions as a high-affinity sodium-dependent sulfate transmembrane transporter [34]. A deficiency of the SLC13A1 protein is associated with a reduced blood sulfate level, which plays a key role in the underlying processes leading to painful IDD [8].

To increase the power of the association analysis, we applied the new approach based on the following. A modern model of back pain suggests that this condition is influenced by a complex interplay of biological, psychological, and social factors [11,16,35,36]. The lack of a standard definition of BP poses a significant challenge in identifying new genetic loci and determining whether the identified variants contribute to the risk of underlying spine pathology or the development of persistent pain (chronicity). To address these challenges and enhance our understanding of the genetic influence on BP, we used an innovative strategy. We considered three different back pain-related phenotypes: one self-reported and two defined by EHR-based codes. EHR-based back pain-related phenotypes such as dorsalgia and IDD differ from self-reported CBP in that they reflect back pain of a severity sufficient to warrant health care seeking and are derived from longitudinal health care records, often over the span of many years, rather than point prevalence from a single survey. Nevertheless, the genetic nature of all these phenotypes is common because the genetic correlations between them are about 90%. We explored shared pathways among different pain phenotypes using multi-trait methodologies, which can help reduce heterogeneity [6,19].

Due to this approach, we identified the second gene, *FSCN3*, which is a new gene associated with BP-related traits. This gene was identified using LoF variants, making *FSCN3* a highly intriguing potential drug target. The association of a gene with disease through LoF variants presents several advantages as a drug target. Firstly, LoF variants often lead to a complete or partial loss of gene function, which can directly and measurably impact disease. This clear functional connection allows for the development of targeted therapies to restore or replace the lost function. Secondly, since LoF variants typically affect a specific gene, drugs can be designed to precisely target the gene or its product, enhancing efficacy while minimizing off-target effects or side effects. Additionally, gene therapy or other genetic interventions can be employed to repair or replace the faulty gene, offering a potential long-term solution to the problem. Therefore, the association of *FSCN3* with BP-related traits through LoF variants provides a clear functional link, target specificity, potential for genetic rescue, predictive value, and personalized medicine opportunities, making it a promising drug target for therapeutic development.

So far, little is known about the functions of *FSCN3*. *FSCN3* encodes the fascin actin-bundling protein 3. This gene is predicted to enable actin filament binding activity and to be involved in actin filament bundle assembly, cell migration, and establishment or maintenance of cell polarity. It is predicted to be located in the cytoskeleton and to be

active in several cellular components, including lamellipodium, microvillus, and ruffle (provided by the Alliance of Genome Resources, April 2022 (www.alliancegenome.org, accessed on 15 December 2022)). A related protein, Fascin 1, was shown to contribute to neuropathic pain by promoting inflammation in rat spinal cord [37]. It is highly expressed specifically in microglia after spinal cord injury and regulates microglial migration in mice [38]. FSCN3 is also involved in neurogenic pain. It showed decreased expression in the model of neuropathic pain 'spare nerve injury' in rats (p-value = 0.007, (https://maayanlab.cloud/Harmonizome/gene_set/Neurological+pain+disorder_Dorsal+Root+ Ganglia_GSE15041/GEO+Signatures+of+Differentially+Expressed+Genes+for+Diseases, accessed on 15 December 2022) [39]). It is also possible that FSCN3 is involved in other biological processes because the FSCN3 protein interacts with several other proteins, including FLAD1, GLMN, PPP5C, and PRAME (www.alliancegenome.org, accessed on 15 December 2022). One of them, GLMN, is involved in the inflammation process via the regulation of C-reactive protein [40]. We can speculate that the loss of FSCN3 function changes the protein complex and is involved in the inflammation process. Inflammation is often accompanied by pain, and due to it, the loss of FSCN3 function could theoretically explain back pain.

We did not have the ability to replicate *FSCN3* because its LoF variants are ultra-rare (MAF ~ 10^{-6} – 10^{-4}); they are absent in the list of GWAS variants and may be replicated only on exome sequencing data. Nevertheless, there are three suggestively significant (*p*-value < 2.5×10^{-7}) GWAS signals found in the FinnGen study for pain (limb, back, neck, head abdominally) within ~250 Kb from *FSCN3* borders (rs138017935, rs185233392, and rs147079677, (https://r8.finngen.fi/, accessed on 10 December 2022). The latter two SNVs also have *p*-value = 5×10^{-6} for dorsalgia. This supports the hypothesis of *FSCN3* involvement in pain development. The replication of more frequent LoF variants is also very difficult. For example, the LoF variants with MAF ~ 10^{-3} in the known *SLC13A3* gene can not be replicated using FinnGen because their frequencies in the FinnGen GWAS. Even without being replicated, the *FSCN3* gene deserves attention because it is only the second gene shown to have an effect on back pain due to LoF variants.

We also detected several genes associated with BP-related traits with a suggestive level of significance. The *IL1R5* gene was associated with SGIT (*p*-value = 1.14×10^{-5} for the LoF Burden test). For this gene, an association with osteoporosis has been shown in the Chinese population [41]. Osteoporosis is a common metabolic bone disease characterized by progressive bone mass loss and the degeneration of bone microarchitecture [42]. *IL1R2* has also been shown to be associated with ankylosing spondylitis [43]. This disease is characterized by inflammation of the spine and sacroiliac joints, causing pain and stiffness and, ultimately, new bone formation and progressive joint ankylosis.

The *KIF5B* gene was associated with CBP (p-value = 1.94×10^{-5} for LoF + protein coding variants). Three pathogenic variants in this gene have been described as associated with kyphomelic dysplasia (MIM: 211350), which belongs to a group of heterogeneous skeletal dysplasias [44]. Therefore, both genes *IL1R2* and *KIF5B* are good candidates for explaining some cases of back pain.

One more gene associated with CBP is *P4HTM*. A minimum *p*-value of 2.9×10^{-7} was obtained for the "LoF + protein coding" variant annotation. The protein coded by *P4HTM* belongs to the prolyl hydroxylase domain (PHD). The PHD subfamily has four members: PHD1, PHD2, PHD3, and PHD4, or P4HTM [45]. The PHD proteins play a critical role in the regulation of hypoxia-inducible transcription factors (HIF) under normoxia. HIF plays a role in adaptation to hypoxia and may be related to cellular oxygen sensing. Increasing evidence suggests that the HIF pathway plays an essential role in bone regeneration and repair [46] because the initial factor of bone regeneration is the upregulation of HIF-1 α induced by hypoxia [47]. We can speculate that the abnormal structure of the P4HTM protein caused by rare mutations impairs bone regeneration and thus causes back pain. It is also known that pathogenic variants of *P4HTM* in homozygous or compound heterozygous states are the cause of the HIDEA syndrome (OMIM # 618493) [48]. Among other symptoms,

the phenotype is characterized by muscular hypotonia [49]. We can speculate that one more possible mechanism of *P4HTM* involvement in back pain is muscular hypotonia resulting from mutations in *P4HTM*. However, this gene did not pass the conditional analysis. The conditional SNV that affected the *P4HTM* signal was not genome-wide significant or suggestive (*p*-value = 0.0009) and was neither present among the known CBP variants nor in LD with them. It is located in the intron of *MST1R*, a macrophage-stimulating 1 receptor gene, yet it showed a *p*-value lower than the exonic *P4HTM* SNV in LD with it. Given this, we cannot fully claim *P4HTM* to be associated with CBP, nor can we reject that it may have a potential role in CBP pathogenesis. The region is known to be associated with pain phenotypes, such as IDD, dorsalgia, pain (limb, back, neck, head abdominally), lower back pain or/and sciatica, hernia, and other dorsopathies, with 16 genome-wide significant SNVs already found within 1 Mb from *P4HTM* borders [8,33] (https://r8.finngen.fi/, accessed on 10 December 2022). The region is also dense in genes: 16 SNVs are located in 10 genes. Further studies are needed to prioritize the gene(s) in this location that actually affect the back pain phenotype.

5. Conclusions

We conducted the exome-wide gene-based association analysis of BP-related phenotypes using exome sequencing data from a UK biobank project. We identified two genes significantly associated with BP-related phenotypes. The effect of the known *SLC13A1* gene on SBP was explained not only by LoF but also by rare missense variants. A newly detected gene, *FSCN3*, is a promising potential drug target because its effect on BP is explained by LoF variants. The role of this gene should be clarified in further studies.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/genes14101962/s1, Supplementary Information: Supplementary Methods, Supplementary Tables S1–S2, Supplementary Figures S1–S3. References [22,25,50] are cited in the Supplementary Materials.

Author Contributions: Conceptualization, Y.A.T. and T.I.A.; Data curation, A.V.K., M.B.F., F.M.K.W. and P.S.; Funding acquisition, Y.A.T.; Investigation, I.V.Z., E.E.E. and N.M.B.; Methodology, G.R.S.; Supervision, T.I.A.; Visualization, I.V.Z.; Writing—original draft, T.I.A.; Writing—review and editing, I.V.Z., E.E.E., N.M.B., G.R.S., P.S. and Y.A.T. All co-authors discussed the results and contributed to preparing the final version of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: The work of I.V.Z. and A.V.K. was supported by the budget project of the Institute of Cytology and Genetics FWNR-2022-0020. The work of Y.A.T., T.I.A., E.E., and N.M.B. was supported by the Russian Science Foundation (RSF) grant No. 22-15-20037 and the Government of the Novosibirsk region. The work of G.R.S. was supported by the Russian Science Foundation (RSF) grant No. 23-25-00209. P.S. is a Staff Physician at the VA Puget Sound Health Care System and Co-Director of the Resource Core of the University of Washington Clinical Learning, Evidence and Research (CLEAR) Center for Musculoskeletal Disorders, which is funded by NIH/NIAMS P30AR072572. The contents of this work do not represent the views of the U.S. Department of Veterans Affairs, the National Institutes of Health, or the United States Government.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data are freely available in the ZENODO database https://doi.org/ 10.5281/zenodo.8118630 (accessed on 31 March 2023) and at https://mga.icgbio.ru/BP_phens_ewas/ (accessed on 13 October 2023).

Acknowledgments: This research has been conducted using the UK Biobank Resource under Applications #18219 and #59345.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Hoy, D.; March, L.; Brooks, P.; Blyth, F.; Woolf, A.; Bain, C.; Williams, G.; Smith, E.; Vos, T.; Barendregt, J.; et al. The global burden of low back pain: Estimates from the Global Burden of Disease 2010 study. *Ann. Rheum. Dis.* 2014, *73*, 968–974. [CrossRef]
- Manchikanti, L.; Singh, V.; Falco, F.J.; Benyamin, R.M.; Hirsch, J.A. Epidemiology of low back pain in adults. *Neuromodulation* 2014, 17 (Suppl. 2), 3–10. [CrossRef] [PubMed]
- Bortsov, A.V.; Parisien, M.; Khoury, S.; Martinsen, A.E.; Lie, M.U.; Heuch, I.; Hveem, K.; Zwart, J.A.; Winsvold, B.S.; Diatchenko, L. Brain-specific genes contribute to chronic but not to acute back pain. *Pain. Rep.* 2022, 7, e1018. [CrossRef] [PubMed]
- Gao, S.; Zhou, H.; Luo, S.; Cai, X.; Ye, F.; He, Q.; Huang, C.; Zheng, X.; Li, Y.; Du, Z.; et al. Investigating the Causal Relationship Between Physical Activity and Chronic Back Pain: A Bidirectional Two-Sample Mendelian Randomization Study. *Front. Genet.* 2021, 12, 758639. [CrossRef] [PubMed]
- Johnston, K.J.A.; Adams, M.J.; Nicholl, B.I.; Ward, J.; Strawbridge, R.J.; Ferguson, A.; McIntosh, A.M.; Bailey, M.E.S.; Smith, D.J. Genome-wide association study of multisite chronic pain in UK Biobank. *PLoS Genet.* 2019, 15, e1008164. [CrossRef] [PubMed]
- Tsepilov, Y.A.; Freidin, M.B.; Shadrina, A.S.; Sharapov, S.Z.; Elgaeva, E.E.; Zundert, J.V.; Karssen Lcapital Es, C.; Suri, P.; Williams, F.M.K.; Aulchenko, Y.S. Analysis of genetically independent phenotypes identifies shared genetic factors associated with chronic musculoskeletal pain conditions. *Commun. Biol.* 2020, *3*, 329. [CrossRef] [PubMed]
- Bjornsdottir, G.; Benonisdottir, S.; Sveinbjornsson, G.; Styrkarsdottir, U.; Thorleifsson, G.; Walters, G.B.; Bjornsson, A.; Olafsson, I.H.; Ulfarsson, E.; Vikingsson, A.; et al. Sequence variant at 8q24.21 associates with sciatica caused by lumbar disc herniation. *Nat. Commun.* 2017, *8*, 14265. [CrossRef] [PubMed]
- Bjornsdottir, G.; Stefansdottir, L.; Thorleifsson, G.; Sulem, P.; Norland, K.; Ferkingstad, E.; Oddsson, A.; Zink, F.; Lund, S.H.; Nawaz, M.S.; et al. Rare SLC13A1 variants associate with intervertebral disc disorder highlighting role of sulfate in disc pathology. *Nat. Commun.* 2022, *13*, 634. [CrossRef] [PubMed]
- 9. Belonogova, N.M.; Kirichenko, A.V.; Freidin, M.B.; Williams, F.M.K.; Suri, P.; Aulchenko, Y.S.; Axenovich, T.I.; Tsepilov, Y.A. Noncoding rare variants in PANX3 are associated with chronic back pain. *Pain* **2023**, *164*, 864–869. [CrossRef]
- Ou-Yang, D.C.; Kleck, C.J.; Ackert-Bicknell, C.L. Genetics of Intervertebral Disc Degeneration. Curr. Osteoporos. Rep. 2023, 21, 56–64. [CrossRef]
- 11. Suri, P.; Palmer, M.R.; Tsepilov, Y.A.; Freidin, M.B.; Boer, C.G.; Yau, M.S.; Evans, D.S.; Gelemanovic, A.; Bartz, T.M.; Nethander, M.; 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. [CrossRef] [PubMed]
- Battie, M.C.; Videman, T.; Levalahti, E.; Gill, K.; Kaprio, J. Heritability of low back pain and the role of disc degeneration. *Pain* 2007, 131, 272–280. [CrossRef] [PubMed]
- Hartvigsen, J.; Nielsen, J.; Kyvik, K.O.; Fejer, R.; Vach, W.; Iachine, I.; Leboeuf-Yde, C. Heritability of spinal pain and consequences of spinal pain: A comprehensive genetic epidemiologic analysis using a population-based sample of 15,328 twins ages 20–71 years. *Arthritis Rheum.* 2009, *61*, 1343–1351. [CrossRef] [PubMed]
- 14. Mather, L.; Karkkainen, S.; Narusyte, J.; Ropponen, A.; Mittendorfer-Rutz, E.; Svedberg, P. Sick leave due to back pain, common mental disorders and disability pension: Common genetic liability. *Eur. J. Pain.* **2020**, *24*, 1892–1901. [CrossRef] [PubMed]
- 15. Nyman, T.; Mulder, M.; Iliadou, A.; Svartengren, M.; Wiktorin, C. High heritability for concurrent low back and neck-shoulder pain: A study of twins. *Spine* **2011**, *36*, E1469–E1476. [CrossRef] [PubMed]
- Freidin, M.B.; Tsepilov, Y.A.; Palmer, M.; Karssen, L.C.; Suri, P.; Aulchenko, Y.S.; Williams, F.M.K.; Group, C.M.W. Insight into the genetic architecture of back pain and its risk factors from a study of 509,000 individuals. *Pain* 2019, 160, 1361–1373. [CrossRef] [PubMed]
- 17. Wainschtein, P.; Jain, D.; Zheng, Z.; Group, T.O.A.W.; Consortium, N.T.-O.f.P.M.; Cupples, L.A.; Shadyab, A.H.; McKnight, B.; Shoemaker, B.M.; Mitchell, B.D.; et al. Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. *Nat. Genet.* **2022**, *54*, 263–273. [CrossRef] [PubMed]
- 18. Ao, X.; Parisien, M.; Zidan, M.; Grant, A.V.; Martinsen, A.E.; Winsvold, B.S.; Diatchenko, L. Rare variant analyses in large-scale cohorts identified SLC13A1 associated with chronic pain. *Pain* **2023**, *164*, 1841–1851. [CrossRef]
- 19. Svishcheva, G.R.; Tiys, E.S.; Elgaeva, E.E.; Feoktistova, S.G.; Timmers, P.; Sharapov, S.Z.; Axenovich, T.I.; Tsepilov, Y.A. A Novel Framework for Analysis of the Shared Genetic Background of Correlated Traits. *Genes* **2022**, *13*, 1694. [CrossRef]
- Jiang, L.; Zheng, Z.; Fang, H.; Yang, J. A generalized linear mixed model association tool for biobank-scale data. *Nat. Genet.* 2021, 53, 1616–1621. [CrossRef]
- 21. Stephens, M. A unified framework for association analysis with multiple related phenotypes. *PLoS ONE* 2013, *8*, e65245. [CrossRef]
- Bulik-Sullivan, B.K.; Loh, P.R.; Finucane, H.K.; Ripke, S.; Yang, J.; Schizophrenia Working Group of the Psychiatric Genomics Consortium; Patterson, N.; Daly, M.J.; Price, A.L.; Neale, B.M. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* 2015, 47, 291–295. [CrossRef]
- McLaren, W.; Gil, L.; Hunt, S.E.; Riat, H.S.; Ritchie, G.R.; Thormann, A.; Flicek, P.; Cunningham, F. The Ensembl Variant Effect Predictor. *Genom. Biol.* 2016, 17, 122. [CrossRef] [PubMed]
- Van Hout, C.V.; Tachmazidou, I.; Backman, J.D.; Hoffman, J.D.; Liu, D.; Pandey, A.K.; Gonzaga-Jauregui, C.; Khalid, S.; Ye, B.; Banerjee, N.; et al. Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature* 2020, 586, 749–756. [CrossRef] [PubMed]

- Zhou, W.; Bi, W.; Zhao, Z.; Dey, K.K.; Jagadeesh, K.A.; Karczewski, K.J.; Daly, M.J.; Neale, B.M.; Lee, S. SAIGE-GENE+ improves the efficiency and accuracy of set-based rare variant association tests. *Nat. Genet.* 2022, 54, 1466–1469. [CrossRef]
- Belonogova, N.M.; Svishcheva, G.R.; Kirichenko, A.V.; Zorkoltseva, I.V.; Tsepilov, Y.A.; Axenovich, T.I. sumSTAAR: A flexible framework for gene-based association studies using GWAS summary statistics. *PLoS Comput. Biol.* 2022, 18, e1010172. [CrossRef] [PubMed]
- Lee, S.; Wu, M.C.; Lin, X. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics* 2012, 13, 762–775. [CrossRef] [PubMed]
- Wang, K.; Abbott, D. A principal components regression approach to multilocus genetic association studies. *Genet. Epidemiol.* 2008, 32, 108–118. [CrossRef] [PubMed]
- Svishcheva, G.R.; Belonogova, N.M.; Zorkoltseva, I.V.; Kirichenko, A.V.; Axenovich, T.I. Gene-based association tests using GWAS summary statistics. *Bioinformatics* 2019, 35, 3701–3708. [CrossRef] [PubMed]
- Liu, Y.; Chen, S.; Li, Z.; Morrison, A.C.; Boerwinkle, E.; Lin, X. ACAT: A Fast and Powerful p Value Combination Method for Rare-Variant Analysis in Sequencing Studies. *Am. J. Hum. Genet.* 2019, *104*, 410–421. [CrossRef]
- Yang, J.; Ferreira, T.; Morris, A.P.; Medland, S.E.; Genetic Investigation of ANthropometric Traits (GIANT) Consortium, DIAbetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium; Madden, P.A.; Heath, A.C.; Martin, N.G.; Montgomery, G.W.; Weedon, M.N.; et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* 2012, 44, 369–375. [CrossRef] [PubMed]
- 32. Yang, J.; Lee, S.H.; Goddard, M.E.; Visscher, P.M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 2011, 88, 76–82. [CrossRef]
- Rahman, M.S.; Winsvold, B.S.; Chavez Chavez, S.O.; Borte, S.; Tsepilov, Y.A.; Sharapov, S.Z.; Pain, H.A.-I.; Aulchenko, Y.S.; Hagen, K.; Fors, E.A.; et al. Genome-wide association study identifies RNF123 locus as associated with chronic widespread musculoskeletal pain. *Ann. Rheum. Dis.* 2021, *80*, 1227–1235. [CrossRef] [PubMed]
- Langford, R.; Hurrion, E.; Dawson, P.A. Genetics and pathophysiology of mammalian sulfate biology. J. Genet. Genom. 2017, 44, 7–20. [CrossRef] [PubMed]
- Gatchel, R.J.; McGeary, D.D.; McGeary, C.A.; Lippe, B. Interdisciplinary chronic pain management: Past, present, and future. *Am. Psychol.* 2014, 69, 119–130. [CrossRef] [PubMed]
- 36. Gatchel, R.J.; Peng, Y.B.; Peters, M.L.; Fuchs, P.N.; Turk, D.C. The biopsychosocial approach to chronic pain: Scientific advances and future directions. *Psychol. Bull.* 2007, 133, 581–624. [CrossRef] [PubMed]
- 37. Wang, B.; Fan, B.; Dai, Q.; Xu, X.; Jiang, P.; Zhu, L.; Dai, H.; Yao, Z.; Xu, Z.; Liu, X. Fascin-1 Contributes to Neuropathic Pain by Promoting Inflammation in Rat Spinal Cord. *Neurochem. Res.* 2018, 43, 287–296. [CrossRef] [PubMed]
- Yu, S.; Cheng, L.; Tian, D.; Li, Z.; Yao, F.; Luo, Y.; Liu, Y.; Zhu, Z.; Zheng, M.; Jing, J. Fascin-1 is Highly Expressed Specifically in Microglia After Spinal Cord Injury and Regulates Microglial Migration. *Front. Pharmacol.* 2021, 12, 729524. [CrossRef]
- Rouillard, A.D.; Gundersen, G.W.; Fernandez, N.F.; Wang, Z.; Monteiro, C.D.; McDermott, M.G.; Ma'ayan, A. The harmonizome: A collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database* 2016, 2016, baw100. [CrossRef]
- 40. Koskeridis, F.; Evangelou, E.; Said, S.; Boyle, J.J.; Elliott, P.; Dehghan, A.; Tzoulaki, I. Pleiotropic genetic architecture and novel loci for C-reactive protein levels. *Nat. Commun.* **2022**, *13*, 6939. [CrossRef]
- 41. Rong, K.; Liang, Z.; Xiang, W.; Wang, Z.; Wen, F.; Lu, L. IL1R2 polymorphisms and their interaction are associated with osteoporosis susceptibility in the Chinese Han population. *Int. J. Immunogenet.* **2021**, *48*, 510–525. [CrossRef]
- 42. Hendrickx, G.; Boudin, E.; Van Hul, W. A look behind the scenes: The risk and pathogenesis of primary osteoporosis. *Nat. Rev. Rheumatol.* **2015**, *11*, 462–474. [CrossRef] [PubMed]
- Australo-Anglo-American Spondyloarthritis, C.; Reveille, J.D.; Sims, A.M.; Danoy, P.; Evans, D.M.; Leo, P.; Pointon, J.J.; Jin, R.; Zhou, X.; Bradbury, L.A.; et al. Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci. *Nat. Genet.* 2010, *42*, 123–127. [CrossRef] [PubMed]
- 44. Itai, T.; Wang, Z.; Nishimura, G.; Ohashi, H.; Guo, L.; Wakano, Y.; Sugiura, T.; Hayakawa, H.; Okada, M.; Saisu, T.; et al. De novo heterozygous variants in KIF5B cause kyphomelic dysplasia. *Clin. Genet.* **2022**, *102*, 3–11. [CrossRef]
- Koivunen, P.; Tiainen, P.; Hyvarinen, J.; Williams, K.E.; Sormunen, R.; Klaus, S.J.; Kivirikko, K.I.; Myllyharju, J. An endoplasmic reticulum transmembrane prolyl 4-hydroxylase is induced by hypoxia and acts on hypoxia-inducible factor α. *J. Biol. Chem.* 2007, 282, 30544–30552. [CrossRef] [PubMed]
- 46. Fan, L.; Li, J.; Yu, Z.; Dang, X.; Wang, K. The hypoxia-inducible factor pathway, prolyl hydroxylase domain protein inhibitors, and their roles in bone repair and regeneration. *Biomed. Res. Int.* **2014**, 2014, 239356. [CrossRef] [PubMed]
- 47. Danis, A. Mechanism of bone lengthening by the Ilizarov technique. Bull. Mem. Acad. R Med. Belg. 2001, 156, 107–112. [PubMed]
- Kraatari-Tiri, M.; Soikkonen, L.; Myllykoski, M.; Jamshidi, Y.; Karimiani, E.G.; Komulainen-Ebrahim, J.; Kallankari, H.; Mignot, C.; Depienne, C.; Keren, B.; et al. HIDEA syndrome is caused by biallelic, pathogenic, rare or founder P4HTM variants impacting the active site or the overall stability of the P4H-TM protein. *Clin. Genet.* 2022, 102, 444–450. [CrossRef] [PubMed]

- Rahikkala, E.; Myllykoski, M.; Hinttala, R.; Vieira, P.; Nayebzadeh, N.; Weiss, S.; Plomp, A.S.; Bittner, R.E.; Kurki, M.I.; Kuismin, O.; et al. Biallelic loss-of-function P4HTM gene variants cause hypotonia, hypoventilation, intellectual disability, dysautonomia, epilepsy, and eye abnormalities (HIDEA syndrome). *Genet. Med.* 2019, 21, 2355–2363. [CrossRef]
- Grotzinger, A.D.; Rhemtulla, M.; de Vlaming, R.; Ritchie, S.J.; Mallard, T.T.; Hill, W.D.; Marioni, R.E.; McIntosh, A.M.; Koellinger, P.D.; Harden, K.P. Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat. Hum. Behav.* 2019, *3*, 513–525. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.